MÉTODOS FUNDAMENTALES de economía matemática

CUARTA EDICIÓN

ALPHA C. CHIANG

Métodos fundamentales de economía matemática

Métodos fundamentales de economía matemática

Cuarta edición

Alpha C. Chiang
Profesor emérito
University of Connecticut

Kevin Wainwright
British Columbia Institute of Technology and
Simon Fraser University

Traducción
Francisco Sánchez Fragoso
Raúl Arrioja Juárez
Traductores profesionales

Revisión técnica
Andrés González Nucamendi
Instituto Tecnológico y de Estudios Superiores de Monterrey
Campus Ciudad de México

Filadelfo León Cázares
Universidad de Guadalajara
Centro Universitario de Ciencias Económico-Administrativas

Mc Graw Hill

MÉXICO • BOGOTÁ • BUENOS AIRES • CARACAS • GUATEMALA • LISBOA
MADRID • NUEVA YORK • SAN JUAN • SANTIAGO
AUCKLAND • LONDRES • MILÁN • MONTREAL • NUEVA DELHI
SAN FRANCISCO • SINGAPUR • ST. LOUIS • SIDNEY • TORONTO

Director Higher Education: Miguel Ángel Toledo Castellanos
Director editorial: Ricardo A. del Bosque Alayón
Editor sponsor: Jesús Mares Chacón
Editor de desarrollo: Marcela I. Rocha Martínez
Supervisor de producción: Zeferino García García

MÉTODOS FUNDAMENTALES DE ECONOMÍA MATEMÁTICA
Cuarta edición

McGraw-Hill Interamericana

*A Subsidiary of **The McGraw-Hill** Companies, Inc.*
Corporativo Punta Santa Fe
Prolongación Paseo de la Reforma 1015, Torre A
Piso 17, Colonia Desarrollo Santa Fe,
Delegación Álvaro Obregón
C.P. 01376, México, D.F.
Miembro de la Cámara Nacional de la Industria Editorial Mexicana, Reg. Núm. 736

ISBN- 10: 970-10-5614-0
ISBN- 13: 978-970-10-5614-1

Traducido de la cuarta edición de *Fundamental Methods of Mathematical Economics*.

0-07-010910-9

Diseño de portada: Kamy Carter

Sobre la portada: La gráfica de la figura 20.1 de la página 635 muestra que la distancia más corta entre dos puntos es una línea recta. Escogimos este concepto como base del diseño de la portada porque una verdad tan simple necesita demostrarse con una de las técnicas más avanzadas de este libro.

2345678901 09875432106

Impreso en México Printed in Mexico

Esta obra se terminó de imprimir en mayo
de 2007 en Litográfica Ingramex S.A. de C.V.,
Centeno 162, Col. Granjas Esmeralda
Delegación Iztapalapa, México, D.F.

The McGraw-Hill Companies

A Emily, Darryl y Tracey

—Alpha C. Chiang

A Skippy y Myrtle

—Kevin Wainwright

Acerca de los autores

Alpha C. Chiang obtuvo su doctorado en la Columbia University en 1954, después de completar la licenciatura en 1946 en la Saint John's University (Shanghai, China) y la maestría en 1948 en la University of Colorado. En 1954 se incorporó a la facultad de la Universidad de Denison en Ohio, donde asumió la dirección del Departamento de Economía en 1961. De 1964 en adelante dio clases en la Universidad de Connecticut, donde, después de 28 años, se convirtió en profesor emérito de economía en 1992. Asimismo, impartió cátedras como profesor invitado en el Colegio New Asia de la Universidad de China en Hong Kong, la Universidad de Cornell, la Universidad de Lingnan en Hong Kong y la Escuela de Economía y Administración de Negocios de Helsinki. Sus publicaciones incluyen otro libro de economía matemática: *Elements of Dynamic Optimization*, Waveland Press, Inc., 1992. Ha recibido los premios de la Fundación Ford y las becas de la Fundación Nacional para las Ciencias, fue presidente electo de la Asociación de Economistas y Científicos Políticos de Ohio, 1963-1964 y se le menciona en *Who's Who in Economics: A Biographical Dictionary of Major Economists 1900-1994*, MIT Press.

Kevin Wainwright es miembro del profesorado del British Columbia Institute of Technology en Burnaby, B.C., Canadá. Desde 2001 se ha desempeñado como presidente de la asociación de profesores y jefe del programa de Administración de Negocios. Realizó sus estudios de licenciatura en la Universidad Simon Fraser en Burnaby, B.C., Canadá, y continúa enseñando en el Departamento de Economía de esta institución. Es especialista en teoría microeconómica y economía matemática.

Prefacio

Este libro está escrito para los estudiantes de economía decididos a aprender los métodos matemáticos básicos indispensables para entender las publicaciones de economía actuales. Por desgracia, estudiar matemáticas es para muchos algo parecido a tomar una medicina amarga, absolutamente necesaria, pero desagradable en extremo. Tal actitud, conocida como "ansiedad matemática", al parecer tiene sus raíces en la manera poco propicia con que se presentan las matemáticas a los estudiantes. Con la creencia de que lo conciso es elegante, las explicaciones ofrecidas suelen ser tan breves que no ofrecen claridad, de modo que los estudiantes se confunden y les queda una sensación injusta de inadecuación intelectual. Un estilo de presentación demasiado formal, cuando no va acompañado de ilustraciones o demostraciones intuitivas pertinentes, puede perjudicar la motivación. Un avance accidentado en el nivel del material puede hacer que ciertos temas matemáticos parezcan más difíciles de lo que en realidad son. Por último, los problemas de los ejercicios excesivamente complejos tienden a destrozar la confianza de los estudiantes, en vez de estimularles el pensamiento.

Con esto en mente, hemos hecho un gran esfuerzo para minimizar los aspectos que causan preocupación. En la medida de lo posible, ofrecemos explicaciones detalladas, más que crípticas. El estilo es deliberadamente informal y "amigable con el lector". Por lo general, intentamos prever y contestar preguntas que es probable que surjan en las mentes de los alumnos a medida que estudian. Con el fin de subrayar la importancia que la matemática tiene para la economía, dejamos que las necesidades analíticas de los economistas motiven el estudio de las técnicas matemáticas relacionadas, e inmediatamente después ilustramos estas últimas con modelos económicos apropiados. Por lo tanto, el juego de herramientas matemáticas se fortalece con un programa cuidadosamente clasificado, donde las herramientas elementales sirven como peldaños para las más avanzadas que se analizan después. Siempre que sea apropiado, las ilustraciones gráficas ofrecen un refuerzo visual a los resultados algebraicos. Además, hemos diseñado los problemas de los ejercicios como medios para ayudar a consolidar la comprensión y reforzar la confianza, y no como desafíos exactos que podrían frustrar e intimidar de manera inconsciente al estudiante inexperto.

En este libro se tratan los siguientes tipos principales de análisis económico: estática (análisis de equilibrio), estática comparativa, problemas de optimización (como un tipo especial de estática), dinámica y optimización dinámica. Para enfrentarlos, se introducen a su debido tiempo los siguientes métodos matemáticos: álgebra de matrices, cálculo diferencial e integral, ecuaciones diferenciales, ecuaciones en diferencias y teoría de control óptimo. Gracias al número considerable de modelos económicos ilustrativos, tanto de macro como de microeconomía, que aparecen aquí, este libro es útil también para quienes ya cuentan con capacitación matemática, pero que aún necesitan una guía para llevarlos del reino de las matemáticas al territorio de la economía. Por esta misma razón, el libro no sólo debe servir como texto para un curso sobre métodos matemáticos, sino también como lectura complementaria en cursos de teoría microeconómica, teoría macroeconómica y crecimiento y desarrollo económicos.

Se ha intentado conservar los objetivos principales y el estilo de las ediciones anteriores; sin embargo, esta edición contiene algunos cambios importantes. El material sobre programación matemática se presenta ahora en un nuevo capítulo, el 13, titulado "Temas adicionales de optimización". Este capítulo contiene dos temas muy importantes: optimización con restricciones de desigualdad y el teorema de la envolvente. Bajo el primer tema, las condiciones de Kuhn-Tucker se desarrollan casi de la misma manera que en la edición anterior; sin embargo, el tema se mejoró con algunas nuevas aplicaciones económicas, entre otras la fijación de precios de carga máxima y el racionamiento del consumidor. El segundo tema se relaciona con el desarrollo del teorema de la envolvente, la función de valor máximo y el concepto de dualidad. Al aplicar el teorema de la envolvente a varios modelos económicos se obtienen resultados importantes, como la identidad de Roy, el lema de Shephard y el lema de Hotelling.

La segunda adición importante a esta edición es un nuevo capítulo, el 20, sobre teoría de control óptimo. Su propósito es introducir al lector en los fundamentos del control óptimo y demostrar cómo se puede aplicar en la economía con ejemplos de economía de recursos naturales y teoría de crecimiento óptimo. El material de este capítulo se deriva en gran medida del análisis de la teoría de control óptimo que se encuentra en *Elements of Dynamic Optimization,* de Alpha C. Chiang (McGraw-Hill 1992, ahora publicado por Waveland Press, Inc.), que presenta un tratamiento completo del control óptimo y su precursor, el cálculo de variaciones.

Además de los dos nuevos capítulos, hay algunas adiciones importantes y mejoras a esta edición. En el capítulo 3 se amplió la explicación de cómo resolver ecuaciones polinomiales de orden superior por factorización (sección 3.3). En el capítulo 4 se agregó una nueva sección sobre cadenas de Markov (sección 4.7). Y en el capítulo 5 se introdujo la comprobación del rango de una matriz vía una matriz escalonada (sección 5.1) y la condición de Hawkins-Simon en relación con el modelo de Leontief de insumo-producto (sección 5.7). Respecto a las aplicaciones económicas, se añadieron muchos ejemplos nuevos y se mejoraron algunas de las aplicaciones existentes. En la sección 5.6 se incluyó una versión lineal del modelo IS-LM, y se amplió una forma más general del modelo en la sección 8.6 para abarcar tanto una economía cerrada como abierta, y demostrar así una aplicación mucho más abundante de estática comparativa para modelos de función general. Otros materiales que se añadieron son la explicación de la utilidad esperada y las preferencias de riesgo (sección 9.3), un modelo de maximización de ganancia que incorpora la función de producción de Cobb-Douglas (sección 11.6) y un problema de elección intertemporal de dos periodos (sección 12.3). Por último, los problemas de los ejercicios se han revisado y aumentado, lo cual ofrece a los estudiantes la oportunidad de perfeccionar sus habilidades.

Agradecimientos

Estamos en deuda con muchas de las personas que participaron en la escritura de este libro. Ante todo, se debe a los matemáticos y economistas cuyas ideas originales forman la base de este volumen. En segundo término, hay muchos estudiantes cuyos esfuerzos y preguntas durante años han ayudado a conformar los principios y enfoque de este libro.

Las tres ediciones anteriores se beneficiaron con los comentarios y sugerencias de (en orden alfabético): Nancy S. Barrett, Thomas Birnberg, E.J.R. Booth, Charles E. Butler, Roberta Grower Carey, Emily Chiang, Lloyd R. Cohen, Gary Cornell, Harald Dickson, John C.H. Fei, Warren L. Fisher, Roger N. Folsom, Dennis R. Heffley, Jack Hirshleifer, James C. Hsiao, KiJun Jeong, George Kondor, William F. Lott, Paul B. Manchester, Peter Morgan, Mark Nerlove, J. Frank Sharp, Alan G. Sleeman, Dennis Starleaf, Henry Y. Wan, Jr., y Chiou-Nan Yeh.

Para la presente edición, se agradece con sinceridad las sugerencias e ideas de Curt L. Anderson, David Andolfatto, James Bathgate, C.R. Birchenhall, Michael Bowe, John Carson, Kimoon Cheong, Youngsub Chun, Kamran M. Dadkhah, Robert Delorme, Patrick Emerson, Roger Nils Folsom, Paul Gomme, Terry Heaps, Suzanne Helburn, Melvin Iyogu, Ki-Jun Jeong, Robbie Jones, John Kane, Heon-Goo Kim, George Kondor, Hui-wen Koo, Stephen Layson, Boon T. Lim, Anthony M. Marino, Richard Miles, Peter Morgan, Rafael Hernández Núñez, Alex Panayides, Xinghe Wang y Hans-Olaf Wiesemann.

Nuestro aprecio más profundo para Sarah Dunn, quien se desempeñó como hábil mecanógrafa, lectora de pruebas y asistente de investigación. Se agradece especialmente a Denise Potten por sus esfuerzos y habilidades logísticas en la etapa de producción. Por último, se extiende nuestro sincero aprecio a Lucille Sutton, Bruce Gin y Lucy Mullins, de McGraw-Hill, por su paciencia y esfuerzos en la producción de este manuscrito. El producto final y errores que pudiera haber son sólo nuestra responsabilidad.

Sugerencias para aprovechar este libro

Como resultado del incremento gradual del juego de herramientas matemáticas en la organización de este libro, la forma ideal de estudio es seguir el orden específico en que se presentan los temas. Sin embargo, se pueden hacer algunas modificaciones en la secuencia de lectura: después de completar las ecuaciones diferenciales de primer orden (capítulo 15), se puede proceder directamente a la teoría de control óptimo (capítulo 20). No obstante, si se pasa del capítulo 15 al 20, quizás el lector desee revisar la sección 19.5, que trata de los diagramas de fase de dos variables.

Si la estática comparativa no es un área de interés principal, se puede omitir el análisis estático comparativo de los modelos de función general (capítulo 8) y saltar del capítulo 7 al 9. En ese caso sería necesario omitir también la sección 11.7, la parte de estática comparativa de la sección 12.5 y el estudio de dualidad del capítulo 13.

Alpha C. Chiang
Kevin Wainwright

Contenido breve

Contenido

Capítulo 19
Ecuaciones diferenciales y ecuaciones en diferencias simultáneas 592

Capítulo 20
Teoría de control óptimo 631

Parte 1

Introducción

Capítulo 1

Naturaleza de la economía matemática

La *economía matemática* no es otra rama de la economía en el sentido en que lo son las finanzas públicas o el comercio internacional. Más bien es un *método* utilizado en el análisis económico, en el cual el economista emplea símbolos matemáticos para enunciar los problemas y se basa en teoremas matemáticos para auxiliarse en el razonamiento. En cuanto al tema de análisis, éste puede ser teoría micro o macroeconómica, finanzas públicas, economía urbana u otra cosa.

Si el término *economía matemática* se utiliza en un sentido más amplio, se podría decir que, en la actualidad, todo libro elemental de economía ejemplifica la economía matemática en la medida en que se utilizan con frecuencia métodos geométricos para deducir resultados teóricos. Sin embargo, es común que la economía matemática se reserve para describir casos en los que se emplean técnicas matemáticas más complejas que la geometría simple, como el álgebra de matrices, el cálculo diferencial e integral, las ecuaciones diferenciales, las ecuaciones de diferencias, etc. El objetivo de este libro es introducir al lector en los aspectos más fundamentales de estos métodos matemáticos, que podemos encontrar a diario en las publicaciones actuales de economía.

1.1 Economía matemática *versus* economía no matemática

Puesto que la economía matemática es sólo un método del análisis económico, no debe diferir en modo fundamental, y de hecho no lo hace, del método *no* matemático del análisis económico. El objetivo de cualquier análisis teórico, sin importar el método, siempre es obtener un conjunto de conclusiones o teoremas a partir de un conjunto determinado de hipótesis o postulados, mediante un proceso de razonamiento. La diferencia principal entre "economía matemática" y "economía literaria" es dual: para comenzar, en la primera las suposiciones y conclusiones se expresan con símbolos matemáticos en vez de palabras, y en ecuaciones en vez de enunciados. Segundo, en lugar de lógica "literaria", se hace uso de teoremas matemáticos —los cuales abundan— en el proceso de razonamiento. En vista de que los símbolos y las palabras son en realidad equivalentes (tenga en cuenta que los símbolos por lo común se definen en palabras), poco importa qué se elija. Pero, sin duda, los símbolos son más convenientes en el razonamiento deductivo y, de hecho, contribuyen más a alcanzar la concisión y precisión del enunciado.

De nuevo, la elección entre lógica literaria y lógica matemática es un asunto de poca importancia; pero, la matemática tiene la ventaja de forzar a los analistas a hacer explícitas sus suposiciones en cada etapa del razonamiento. Esto se debe a que los teoremas matemáticos se expresan normalmente en la forma "si-entonces", de manera que para poder llegar a la parte "entonces" (la conclusión) del teorema, es necesario tener la seguridad de que la parte "si" (la condición) funciona conforme a las suposiciones explícitas adoptadas.

A pesar de estar de acuerdo con estos puntos, se podría preguntar: ¿por qué es necesario ir más allá de los métodos geométricos? La respuesta es que, si bien el análisis geométrico tiene la ventaja importante de ser visual, también adolece de una grave limitación dimensional. En el análisis gráfico usual de curvas de inferencia, por ejemplo, la suposición estándar es que sólo *dos* artículos estén disponibles para el consumidor. Esta suposición simplificadora no se adopta por gusto, sino que se impone debido a que la tarea de dibujar una gráfica tridimensional es sumamente difícil y la construcción de una gráfica de cuatro (o más) dimensiones es, en realidad, una imposibilidad física. Para tratar con el caso más general de 3, 4 o n artículos se debe recurrir, en cambio, a la herramienta más flexible de las ecuaciones. Esta única razón debe ser una motivación suficiente para estudiar los métodos matemáticos más allá de la geometría.

En resumen, se observa que el método matemático tiene las siguientes ventajas: (1) el "lenguaje" usado es más conciso y preciso; (2) existe una gran cantidad de teoremas matemáticos a nuestro servicio; (3) al obligarnos a expresar de forma explícita todas las suposiciones como un prerrequisito para el uso de teoremas matemáticos, se evita la adopción no intencional de suposiciones implícitas indeseables, y (4) permite tratar el caso general de n variables.

Contra estas ventajas, a veces se escucha la crítica de que una teoría deducida en forma matemática es, sin duda, *irreal*. Sin embargo, la crítica no es válida. De hecho, el epíteto "irreal" no se puede usar para criticar la teoría económica en general, ya sea que el método sea matemático o no. La teoría es por sí misma una abstracción del mundo real. Es un dispositivo para distinguir sólo los factores y relaciones más esenciales, de modo que se pueda estudiar el punto crucial del problema en cuestión, libre de muchas complicaciones que existen en la realidad. Así, el enunciado "la teoría carece de realismo" es sólo una obviedad que no puede ser aceptada como crítica válida de la teoría. De la misma manera, es bastante intrascendente calificar cualquier enfoque de la teoría como "irreal"; por ejemplo, la teoría de la empresa bajo la competencia pura es "irreal", al igual que la teoría de la empresa bajo la competencia imperfecta. Pero, si estas teorías se deducen matemáticamente o no, es irrelevante e inmaterial.

Para aprovechar la gran cantidad de herramientas matemáticas, por supuesto, es necesario adquirirlas primero. Por desgracia, las herramientas que son de interés para el economista se encuentran dispersas entre muchos cursos matemáticos; demasiados para ajustarse de modo confortable al plan de estudio de un estudiante de economía común. El servicio que realiza el presente volumen es reunir en un lugar los métodos matemáticos más importantes en relación con las publicaciones de economía, organizarlos en un orden lógico de progresión, explicar por completo cada método e ilustrar de inmediato cómo se aplica el método en el análisis económico. Al relacionar los métodos con sus aplicaciones, la importancia de las matemáticas en relación con la economía se hace más transparente que en los cursos regulares de matemáticas, en los que las aplicaciones ilustradas se asocian sobre todo con la física e ingeniería. La familiaridad con el contenido de este libro (y si es posible también con el volumen que le sigue: Alpha C. Chiang, *Elements of Dynamic Optimization*, McGraw-Hill, 1992; ahora publicado por Waveland Press, Inc.) debe permitirle comprender la mayor parte de los artículos profesionales encontrados en publicaciones periódicas como *American Economic Review*, *Quarterly Journal of Economics*, *Journal of Political Economy*, *Review of Economics and Statistics* y *Economic Journal*. Quienes, por medio de esta exposición, desarrollen un interés serio en la economía matemática pueden proceder a un estudio más riguroso y avanzado de las matemáticas.

1.2 Economía matemática *versus* econometría

El término *economía matemática* a veces se confunde con un término relacionado, *econometría*. Como indica la parte "métrica" del último término, la econometría se relaciona principalmente con la medición de datos económicos. Por lo tanto, ésta trata del estudio de observaciones *empíricas* por medio de métodos estadísticos de estimación y prueba de hipótesis. Por otro lado, la economía matemática se refiere a la aplicación de las matemáticas a los aspectos puramente *teóricos* del análisis económico, con poco o ningún interés en cuanto a problemas estadísticos como los errores de medición de las variables bajo estudio.

En el presente libro, el estudio se limita a la economía matemática; es decir, se centra la atención en la aplicación de las matemáticas al razonamiento deductivo y no al estudio inductivo y, como resultado, se estará tratando sobre todo con material teórico en vez de empírico. Esto es, por supuesto, sólo un asunto de elección del ámbito de análisis, y de ningún modo significa que la econometría sea menos importante.

De hecho, los estudios empíricos y el análisis teórico son complementarios y se refuerzan mutuamente. Por un lado, las teorías deben probarse contra los datos empíricos a fin de validarlas antes de aplicarlas con confianza. Por el otro, el trabajo estadístico necesita guiarse de la teoría económica, a fin de determinar la dirección de investigación más pertinente y fructífera.

Sin embargo, en un sentido, la economía matemática podría ser considerada como la más básica de ambas, porque, para tener un estudio estadístico y econométrico significativo, es indispensable un buen marco teórico, de preferencia en la formulación matemática. Por lo tanto, el contenido del presente libro debe ser útil no sólo para quienes estén interesados en la economía teórica, sino también para aquellos que buscan un fundamento que les sirva en la consecución de sus estudios econométricos.

Capítulo 2

Modelos económicos

Como ya se mencionó, cualquier teoría económica es necesariamente una abstracción del mundo real. Entre otras cosas, porque la inmensa complejidad de la economía real imposibilita comprender a la vez todas las interrelaciones; y tampoco, para el caso, todas estas interrelaciones son de igual importancia para la comprensión del fenómeno económico particular de estudio. Así, el procedimiento más razonable es elegir lo que, según nuestro criterio, son los factores y relaciones principales pertinentes del problema, y enfocar la atención sólo en éstos. Esta clase de marco analítico simplificado de forma deliberada se llama *modelo económico*, puesto que sólo es una estructura o representación aproximada de la economía real.

2.1 Elementos de un modelo matemático

Un modelo económico es simplemente un marco teórico, y no hay razón inherente de por qué debe ser matemático. Sin embargo, si el modelo *es* matemático, por lo general consistirá en un conjunto de *ecuaciones* diseñadas para describir la estructura del modelo. Al relacionar cierta cantidad de *variables* entre sí en ciertas maneras, estas ecuaciones dan forma matemática al conjunto de suposiciones analíticas adoptadas. Entonces, mediante la aplicación de las operaciones matemáticas destacadas en estas ecuaciones, se puede obtener un conjunto de conclusiones que se deduzcan de manera lógica de esas suposiciones.

Variables, constantes y parámetros

Una *variable* es algo cuya magnitud puede cambiar, es decir, algo que puede tomar valores diferentes. Las variables de uso común en economía son precio, ganancia, ingreso, costo, ingreso nacional, consumo, inversión, importaciones y exportaciones. Puesto que cada variable puede tomar varios valores, se debe representar mediante un símbolo en vez de un número específico. Por ejemplo, se podría representar al precio mediante la letra P, a la ganancia con π, al ingreso mediante I, al costo por medio de C, al ingreso nacional con Y, y así sucesivamente. Sin embargo, cuando se escribe $P = 3$ o $C = 18$, se "congelan" estas variables en valores específicos (en unidades elegidas de modo apropiado).

Un modelo económico, construido de manera apropiada, se puede resolver para obtener los *valores solución* de cierto conjunto de variables, como por ejemplo el nivel de precios de depuración del mercado o el nivel de producción maximización-ganancias. Esta clase de variables, cuyos valores solución se buscan desde el modelo, se conocen como *variables endógenas* (que se originan desde dentro). No obstante, el modelo también podría contener variables que se supone están determinadas por fuerzas externas al modelo y cuyas magnitudes se aceptan sólo como datos; este tipo de variables se llaman *variables exógenas* (que se origina desde fuera). Se debe observar que una variable que es endógena en un modelo podría muy bien ser exógena en otro. En un análisis de la determinación de mercado del precio del trigo (P), por ejemplo, la variable P debe ser en definitiva endógena; pero en el marco de una teoría de gasto del consumidor, P se convertiría en un dato para el consumidor individual y, por lo tanto, se debe considerar exógena.

Las variables suelen aparecer en combinación con números fijos o constantes, como por ejemplo en las expresiones $7P$ o $0.5R$. Una *constante* es una magnitud que no cambia y, por lo tanto, es la antítesis de una variable. Sin embargo, un coeficiente puede ser simbólico en vez de numérico. Se puede, por ejemplo, permitir que el símbolo a represente una determinada constante y usar la expresión aP en lugar de $7P$ en un modelo, a fin de obtener un mayor nivel de generalidad (véase la sección 2.7). Este símbolo a es un caso bastante peculiar, se supone que representa una constante dada y, sin embargo, puesto que todavía no se le asigna un número específico, puede tomar casi cualquier valor. En resumen, ¡es una *constante* que es *variable*! Para identificar su estado especial, se le da el nombre distintivo de *constante paramétrica* (o simplemente *parámetro*).

Se debe subrayar que, si bien es posible asignar valores distintos a un parámetro, no obstante se considera como un dato en el modelo. Es por esta razón que las personas en ocasiones simplemente dicen "constante", aun cuando la constante es paramétrica. En este sentido, los parámetros se asemejan mucho a las variables exógenas, porque ambos van a ser tratados como "presunciones" en un modelo. Esto explica por qué muchos escritores, por simplicidad, se refieren a ambos en conjunto con la designación única de "parámetros".

Por convención, las constantes paramétricas se representan por lo común mediante los símbolos a, b, c o sus contrapartes en el alfabeto griego: α, β y γ. Aunque, por supuesto, otros símbolos también son permisibles. En cuanto a las variables exógenas, a fin de distinguirlas visualmente de sus primas endógenas, se seguirá la práctica de anexar un subíndice 0 al símbolo elegido. Por ejemplo, si P simboliza precio, entonces P_0 indica un precio determinado de forma exógena.

Ecuaciones e identidades

Las variables podrían existir de forma independiente, pero en realidad no se vuelven interesantes hasta que se relacionan entre sí mediante ecuaciones o desigualdades. En este momento se analizarán sólo ecuaciones.

En las aplicaciones económicas se podría distinguir entre tres tipos de ecuación: ecuaciones definicionales, de comportamiento y condicionales.

Una *ecuación definicional* establece una identidad entre dos expresiones alternas que tienen el mismo significado. Para tal ecuación, suele usarse el signo de igualdad idéntica $\equiv$ (léase: "es idénticamente igual a") en lugar del signo igual $=$, aunque este último también es aceptable. Como ejemplo, la ganancia total se define como el exceso de ingreso total sobre el costo total; así, se puede escribir

$$\pi \equiv R - C$$

Una *ecuación de comportamiento*, por otro lado, especifica la manera en la cual se comporta una variable en respuesta a cambios en otras variables. Es posible que esto tenga que ver con el comportamiento humano (como el patrón de consumo agregado en relación con el ingreso nacional) o el comportamiento no humano (por ejemplo, cómo reacciona el costo total de una empresa a cambios en el producto). En términos generales, las ecuaciones de comportamiento se pueden usar para describir el entorno institucional general de un modelo, incluso los aspectos tecnológico (por ejemplo, función de producción) y legales (como estructura fiscal). Sin embargo, antes de escribir una ecuación de comportamiento, siempre es necesario adoptar suposiciones definidas respecto al patrón de conducta de la variable en cuestión. Considere dos funciones de costo

$$C = 75 + 10Q \tag{2.1}$$

$$C = 110 + Q^2 \tag{2.2}$$

donde Q denota la producción. Puesto que las dos ecuaciones tienen formas diferentes, resulta evidente que la condición de producción supuesta en cada una es diferente de la otra. En (2.1), el costo fijo (el valor de C cuando $Q = 0$) es 75, mientras que en (2.2) es 110. La variación de costo también es diferente. En (2.1), para cada incremento unitario en Q, hay un incremento constante de 10 en C. Pero en (2.2), cuando Q aumenta unidad tras unidad, C aumentará mediante cantidades progresivamente más grandes. Resulta claro que, ante todo, es por la especificación de la forma de las ecuaciones de comportamiento que se da expresión matemática a las suposiciones adoptadas en un modelo.

Como el tercer tipo, una *ecuación condicional* expresa que se debe satisfacer un requerimiento. Por ejemplo, en un modelo en el que interviene el concepto de equilibrio, se debe establecer una *condición de equilibrio*, que describe el prerrequisito para la consecución de equilibrio. Dos de las condiciones de equilibrio más conocidas en economía son

$$Q_d = Q_s \qquad \text{[cantidad demandada = cantidad suministrada]}$$

y

$$S = I \qquad \text{[ahorro previsto = inversión prevista]}$$

que pertenecen, respectivamente, al equilibrio de un modelo de mercado y al equilibrio del modelo de ingreso nacional en su forma más simple. De modo similar, un modelo de optimización deriva o aplica una o más *condiciones de optimización*. Una condición de este tipo que viene fácilmente a la mente es

$$\text{CM} = \text{IM} \qquad \text{[costo marginal = ingreso marginal]}$$

en la teoría de la empresa. Debido a que las ecuaciones de este tipo no son definicionales ni de comportamiento, por sí mismas constituyen una clase.

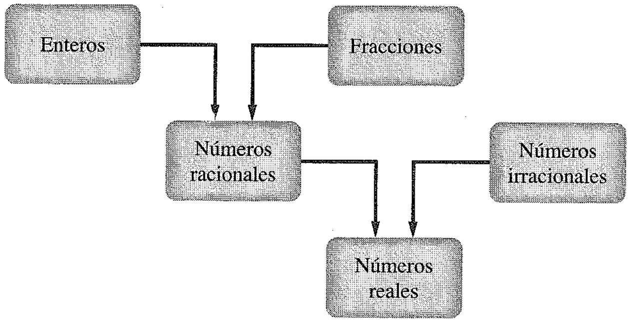
2.2 Sistema de números reales

Las ecuaciones y variables son los ingredientes esenciales de un modelo matemático. Sin embargo, puesto que los valores que toma una variable económica son por lo común numéricos, son pertinentes algunas palabras acerca del sistema numérico. Aquí se tratará sólo con los llamados números reales.

Los números enteros como 1, 2, 3, ... se llaman *enteros positivos*, los cuales se usan con mayor frecuencia en el conteo. Sus contrapartes negativas $-1, -2, -3, \ldots$ se llaman *enteros negativos*; éstos se pueden emplear, por ejemplo, para indicar temperaturas bajo cero (en grados). El número 0 (cero), por otro lado, no es positivo ni negativo, y en ese sentido es único. Se agrupan los enteros positivos y negativos y el número cero en una sola categoría, y se hace referencia a ellos en forma colectiva como el *conjunto de los enteros*.

Por supuesto, los enteros no agotan todos los números posibles, porque se tienen *fracciones*, como por ejemplo $\frac{2}{3}$, $\frac{5}{4}$ y $\frac{7}{3}$, que —si se colocan sobre una regla— caerían entre los enteros. Asimismo, se tienen fracciones negativas, como $-\frac{1}{2}$ y $-\frac{2}{5}$. Las fracciones positivas y negativas constituyen el *conjunto de las fracciones*.

FIGURA 2.1



La propiedad común de los números fraccionarios es que cada uno se puede expresar como una razón de dos enteros. Cualquier número que se puede expresar como una razón de dos enteros se llama *número racional*. Pero los enteros por sí mismos son también racionales, porque cualquier entero n se puede considerar como la razón $n/1$. El conjunto de los enteros y el conjunto de las fracciones forman el *conjunto de los números racionales*. Una característica alternativa definitoria de un número racional es que se puede expresar como un decimal finito (por ejemplo, $\frac{1}{4} = 0.25$) o un decimal periódico ($\frac{1}{3} = 0.3333\ldots$), donde algún número o serie de números a la derecha del punto decimal se repite de forma indefinida.

Una vez que se usa el concepto de números racionales, surge de manera natural el concepto de *números irracionales*, números que *no se pueden* expresar como razones de un par de enteros. Un ejemplo es el número $\sqrt{2} = 1.4142\ldots$, que es un decimal no periódico, no finito. Otro es la constante especial $\pi = 3.1415\ldots$ (que representa la razón entre la circunferencia de cualquier círculo y su diámetro), el cual de nuevo es un decimal no periódico, no finito —característica de los números irracionales—.

Cada número irracional, si se coloca en una regla, caería entre dos números racionales, así que, al igual que las fracciones llenan los espacios entre los enteros en una regla, los números irracionales llenan los espacios entre los números racionales. El resultado de este proceso de llenado es un continuo de números, los cuales se llaman números reales. Este continuo constituye el *conjunto de los números reales*, que por lo común se denota con el símbolo R. Cuando el conjunto R se muestra en una recta (una regla extendida), se hace referencia a la línea como *recta real*.

En la figura 2.1 se listan (en el orden analizado) los conjuntos de números, dispuestos en relación entre sí. Sin embargo, si se leen de abajo hacia arriba, se encuentra en efecto un esquema clasificatorio en el que el conjunto de números reales se descompone en sus conjuntos numéricos componentes y subcomponentes. Por lo tanto, esta figura es un resumen de la estructura del sistema de números reales.

Los números reales son todo lo que se necesita para los primeros 15 capítulos de este libro, pero ellos no son los únicos números usados en matemáticas. De hecho, la razón para el término *real* es que hay también números "imaginarios", los cuales tienen que ver con las raíces cuadradas de números negativos. Ese concepto se analizará más adelante en el capítulo 16.

2.3 Concepto de conjuntos

Ya se ha empleado la palabra *conjunto* varias veces. En vista de que el concepto de conjuntos sostiene cada rama de la matemática moderna, es recomendable familiarizarse con por lo menos sus conceptos más básicos.

Notación de conjuntos

Un *conjunto* es simplemente una colección de objetos distintos, los cuales pueden ser un grupo de números, personas, artículos alimentarios o alguna otra cosa (distintos). Así, los estudiantes inscritos en un curso de economía pueden ser considerados como un conjunto, igual que los enteros 2, 3 y 4 pueden formar un conjunto. Los objetos de un conjunto se llaman *elementos* del conjunto.

Hay dos formas de escribir un conjunto: por *enumeración* y por *descripción*. Si S representa el conjunto de tres números: 2, 3 y 4, se puede escribir, por enumeración de los elementos,

$$S = \{2, 3, 4\}$$

Pero si I denota el conjunto de *todos* los enteros positivos, la enumeración se vuelve difícil y, en cambio, simplemente se podrían describir los elementos y escribir

$$I = \{x \mid x \text{ es un entero positivo}\}$$

que se lee como sigue: "I es el conjunto de todos los (números) x, tales que x es un entero positivo". Observe que en cualquier caso se emplea un par de llaves para encerrar el conjunto. En el método descriptivo, se inserta siempre una barra vertical (o dos puntos) para separar el símbolo de designación de los elementos y la descripción de éstos. Como otro ejemplo, el conjunto de los números reales mayores que 2 pero menores que 5 (llamado J) se puede expresar de forma simbólica como

$$J = \{x \mid 2 < x < 5\}$$

Aquí, incluso el símbolo descriptivo se expresa de forma simbólica.

Un conjunto con una cantidad finita de elementos, ejemplificado mediante el conjunto S mencionado antes, se llama *conjunto finito*. El conjunto I y el conjunto J, cada uno con una cantidad infinita de elementos, son, por otro lado, ejemplos de un *conjunto infinito*. Los conjuntos finitos siempre son *numerables* (o *contables*), es decir, sus elementos se pueden contar uno por uno en la secuencia 1, 2, 3, . . . No obstante, los conjuntos infinitos pueden ser numerables (conjunto I) o *no numerables* (conjunto J). En el último caso, no hay manera de asociar los elementos del conjunto con los números de conteo naturales 1, 2, 3, . . . , y, por lo tanto, el conjunto es incontable.

La pertenencia en un conjunto se indica mediante el símbolo $\in$ (una variante de la letra griega épsilon ϵ para "elemento"), que se lee como sigue: "es un elemento de". Así, para los dos conjuntos S e I definidos antes, se puede escribir

$$2 \in S \qquad 3 \in S \qquad 8 \in I \qquad 9 \in I \qquad \text{(etc.)}$$

pero es evidente que $8 \notin S$ (léase: "8 no es un elemento del conjunto S"). Si se usa el símbolo R para denotar el conjunto de los números reales, entonces la expresión "x es algún número real" se puede expresar simplemente como

$$x \in R$$

Relaciones entre conjuntos

Cuando dos conjuntos se comparan entre sí, se puede observar varios tipos de relación. Si dos conjuntos S_1 y S_2 contienen elementos idénticos,

$$S_1 = \{2, 7, a, f\} \qquad \text{y} \qquad S_2 = \{2, a, 7, f\}$$

se dice entonces que S_1 y S_2 son *iguales* ($S_1 = S_2$). Note que el orden de aparición de los elementos en un conjunto es de poca importancia. Sin embargo, siempre que se encuentre incluso un elemento diferente en dos conjuntos cualesquiera, esos dos conjuntos no son iguales.

Otra clase de relación de conjuntos es que un conjunto puede ser un *subconjunto* del otro conjunto. Si se tienen dos conjuntos

$$S = \{1, 3, 5, 7, 9\} \qquad \text{y} \qquad T = \{3, 7\}$$

entonces T es un subconjunto de S, porque todo elemento de T es también un elemento de S. Un enunciado más formal de esto es: T es un subconjunto de S si y sólo si $x \in T$ implica que $x \in S$. Si se emplean los símbolos de inclusión de conjuntos $\subset$ (está contenido en) y $\supset$ (incluye a), entonces se puede escribir

$$T \subset S \qquad \text{o} \qquad S \supset T$$

Es posible que dos conjuntos sean subconjuntos entre sí. Cuando esto ocurre se puede estar seguro de que ambos conjuntos son iguales. Para expresar esto de modo formal, se puede tener $S_1 \subset S_2$ y $S_2 \subset S_1$ si y sólo si $S_1 = S_2$.

Note que mientras el símbolo $\in$ relaciona un *elemento* individual con un *conjunto*, el símbolo $\subset$ relaciona un *subconjunto* con un *conjunto*. Como una aplicación de esta idea, se puede expresar con base en la figura 2.1 que el conjunto de los números enteros es un subconjunto del conjunto de todos los números racionales. Asimismo, el conjunto de los números racionales es un subconjunto del conjunto de los números reales.

¿Cuántos subconjuntos se pueden formar a partir de los cinco elementos del conjunto $S = \{1, 3, 5, 7, 9\}$? Primero, cada elemento de S puede contar como un subconjunto distinto de S, tal como $\{1\}$ y $\{3\}$. Pero sucede lo mismo con cada par, tercia o cuádrupla de estos elementos, por ejemplo $\{1, 3\}$, $\{1, 5\}$ y $\{3, 7, 9\}$. Cualquier subconjunto que *no* contiene *a todos* los elementos de S se llama *subconjunto propio* de S. Pero el conjunto S por sí mismo (con sus cinco elementos) también puede ser considerado como uno de sus subconjuntos propios: todo elemento de S es un elemento de S y, por lo tanto, el conjunto S en sí satisface la definición de un subconjunto. Éste es, por supuesto, un caso límite, del cual se obtiene el subconjunto más grande posible de S, a saber, S mismo.

En el otro extremo, el subconjunto más pequeño posible de S es un conjunto que no contiene ningún elemento. A este conjunto se le conoce como *conjunto nulo* o *conjunto vacío*, y se denota mediante el símbolo $\varnothing$ o $\{\}$. La razón para considerar el conjunto nulo como un subconjunto de S es bastante interesante: si el conjunto nulo no es un subconjunto de S ($\varnothing \not\subset S$), entonces $\varnothing$ debe contener por lo menos un elemento x tal que $x \notin S$. Pero, puesto que por definición el conjunto nulo no tiene ningún elemento en absoluto, no se puede decir que $\varnothing \not\subset S$; por consiguiente, el conjunto nulo es un subconjunto de S.

Es extremadamente importante distinguir con claridad el símbolo $\varnothing$ o $\{\}$ de la notación $\{0\}$; el primero carece de elementos, pero el último contiene un elemento, cero. El conjunto nulo es único; en todo el mundo sólo hay un conjunto tal, y se considera un subconjunto de *cualquier* conjunto que pueda ser concebido.

Al contar los subconjuntos de S, incluyendo los dos casos límites S y $\varnothing$, se encuentra un total de $2^5 = 32$ subconjuntos. En general, si un conjunto tiene n elementos, se pueden formar un total de 2^n subconjuntos de estos elementos.[1]

[1] Dado un conjunto con n elementos $\{a, b, c, \ldots, n\}$ se puede clasificar primero sus subconjuntos en dos categorías: uno con el elemento a en él, y uno sin él. Además, cada uno de estos dos subconjuntos se puede clasificar en dos subcategorías: una con el elemento b en ella, y una sin el elemento. Note que al considerar el segundo elemento b, se duplica el número de categorías en la clasificación de 2 a 4 ($= 2^2$). De la misma manera, la consideración del elemento c incrementa el número total de categorías a 8 ($= 2^3$). Cuando se consideran los n elementos, el número total de categorías se convierte en el número de subconjuntos, y ese número es 2^n.

Como un tercer tipo posible de relación de conjuntos, dos conjuntos podrían no tener elementos en común en absoluto. En ese caso, se dice que los dos conjuntos son *disjuntos*. Por ejemplo, el conjunto de todos los enteros positivos y el conjunto de los enteros negativos son mutuamente excluyentes; así que son conjuntos disjuntos.

Un cuarto tipo de relación ocurre cuando dos conjuntos tienen algunos elementos en común, pero algunos elementos propios de cada uno. En ese caso, los dos conjuntos no son iguales ni disjuntos; asimismo, ningún conjunto es un subconjunto del otro.

Operaciones con conjuntos

Cuando se suman, restan, multiplican, dividen o se toma la raíz cuadrada de algunos números, se realizan operaciones matemáticas. Aunque los conjuntos son diferentes de los números, también se pueden llevar a cabo ciertas operaciones matemáticas en ellos. Las tres operaciones principales a analizar aquí tienen que ver con la unión, intersección y complemento de conjuntos.

Tomar la *unión* de dos conjuntos A y B significa formar un nuevo conjunto que contiene los elementos (y sólo esos elementos) que pertenecen a A, o a B, o a ambos A y B. El conjunto unión se simboliza mediante $A \cup B$ (léase "A unión B").

Ejemplo 1 Si $A = \{3, 5, 7\}$ y $B = \{2, 3, 4, 8\}$, entonces

$$A \cup B = \{2, 3, 4, 5, 7, 8\}$$

Este ejemplo, de forma incidental, ilustra el caso en el que dos conjuntos A y B no son iguales ni disjuntos y en el que ninguno es subconjunto del otro.

Ejemplo 2 De nuevo, en relación con la figura 2.1, se ve que la unión del conjunto de los enteros con el conjunto de las fracciones es el conjunto de todos los números racionales. De manera similar, la unión del conjunto de los números racionales y del conjunto de los números irracionales produce el conjunto de los números reales.

Por otro lado, la *intersección* de dos conjuntos A y B es un nuevo conjunto que contiene a los elementos (y sólo esos elementos) que pertenecen *tanto a A como a B*. El conjunto intersección se simboliza mediante $A \cap B$ (léase: "A intersección B").

Ejemplo 3 De los conjuntos A y B del ejemplo 1, se puede escribir

$$A \cap B = \{3\}$$

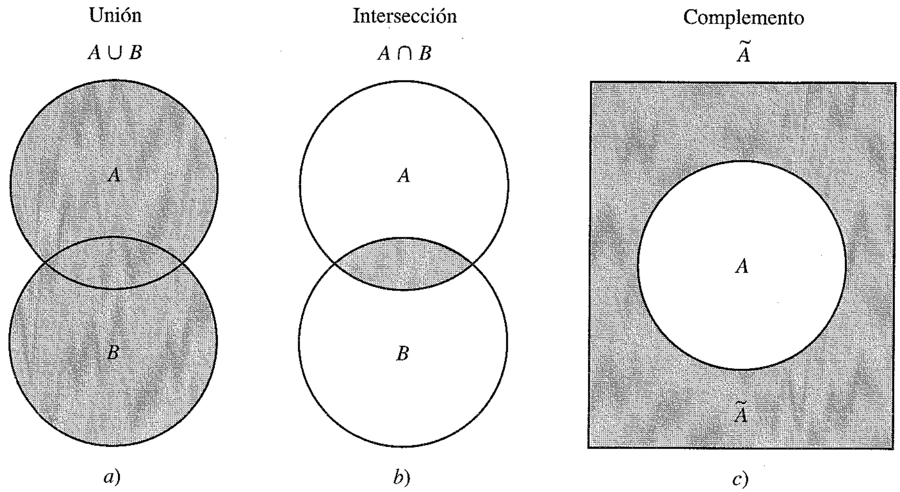
Ejemplo 4 Si $A = \{-3, 6, 10\}$ y $B = \{9, 2, 7, 4\}$, entonces $A \cap B = \varnothing$. El conjunto A y el conjunto B son disjuntos; por lo tanto, su intersección es el conjunto vacío: ningún elemento es común a A y B.

Resulta obvio que la intersección es un concepto más restrictivo que la unión. En la primera, sólo los elementos *comunes a A y B* son aceptables, mientras que en la última, la pertenencia en uno u otro (*ya sea A o B*) es suficiente para establecer la pertenencia en el conjunto unión. Así, los símbolos operadores $\cap$ y $\cup$ —que, por cierto, tienen el mismo tipo de estatus general que los símbolos $\sqrt{\ }$, $+$, $\div$, etc.— tienen las connotaciones "y" y "o", respectivamente. Este punto se puede apreciar mejor si se comparan las siguientes definiciones formales de intersección y unión.

$$\text{Intersección:} \qquad A \cap B = \{x \mid x \in A \quad \text{y} \quad x \in B\}$$

$$\text{Unión:} \qquad A \cup B = \{x \mid x \in A \quad \text{o} \quad x \in B\}$$

FIGURA 2.2



¿Qué hay acerca del *complemento* de un conjunto? Para explicar esto, se introduce primero el concepto de *conjunto universal*. En un contexto de análisis particular, si los únicos números que se emplean son el conjunto de los primeros siete enteros positivos, se puede hacer referencia a éste como el conjunto universal U. Entonces, con un conjunto específico, por ejemplo $A = \{3, 6, 7\}$, se puede definir otro conjunto $\tilde{A}$ (léase: "el complemento de A") como el conjunto que contiene los números del conjunto universal U que no están en el conjunto A. Es decir,

$$\tilde{A} = \{x \mid x \in U \quad \text{y} \quad x \notin A\} = \{1, 2, 4, 5\}$$

Observe que mientras el símbolo $\cup$ tiene la connotación "o" y el símbolo $\cap$ significa "y", el símbolo de complemento $\sim$ lleva la implicación de "no".

Ejemplo 5

Si $U = \{5, 6, 7, 8, 9\}$ y $A = \{5, 6\}$, entonces $\tilde{A} = \{7, 8, 9\}$.

Ejemplo 6

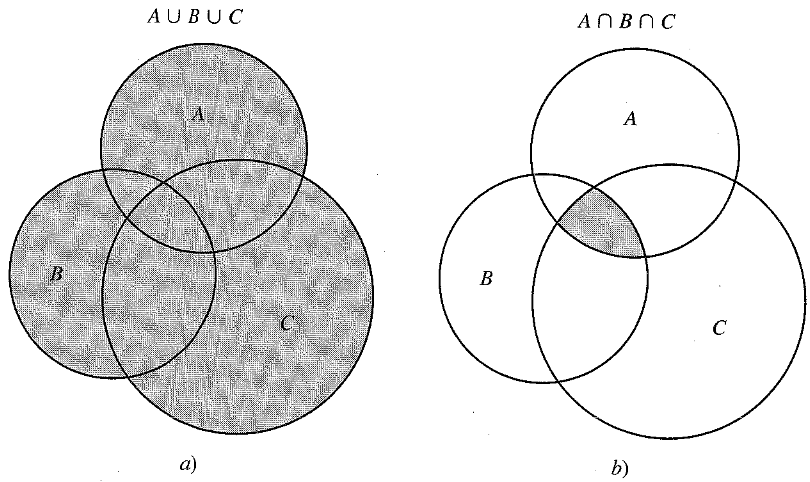
¿Cuál es el complemento de U? Puesto que todo objeto (número) en consideración se incluye en el conjunto universal, el complemento de U debe ser vacío. Así, $\tilde{U} = \varnothing$.

Los tres tipos de operación con conjuntos se representan en los tres diagramas de la figura 2.2, conocidos como *diagramas de Venn*. En el diagrama *a*, los puntos en el círculo superior forman un conjunto A y los puntos en el círculo inferior forman un conjunto B. La unión de A y B consiste entonces en el área sombreada que abarca ambos círculos. En el diagrama *b* se muestran los mismos dos conjuntos (círculos). Puesto que su intersección debe comprender sólo los puntos comunes a ambos conjuntos, sólo la porción de traslape (sombreada) de los dos círculos satisfacen la definición. En el diagrama *c*, sean los puntos del rectángulo el conjunto universal y sea A el conjunto de puntos en el círculo; entonces el conjunto complemento $\tilde{A}$ será el área (sombreada) fuera del círculo.

Leyes de operaciones con conjuntos

En la figura 2.2 es posible observar que el área sombreada del diagrama *a* representa no sólo $A \cup B$, sino que también $B \cup A$. De manera análoga, en el diagrama *b* la pequeña área som-

FIGURA 2.3



breada es la representación visual no sólo de $A \cap B$, sino que también de $B \cap A$. Cuando se formaliza, este resultado se conoce como *ley conmutativa* (de uniones e intersecciones):

$$A \cup B = B \cup A \qquad A \cap B = B \cap A$$

Estas relaciones son muy similares a las leyes algebraicas $a + b = b + a$ y $a \times b = b \times a$.

Para tomar la unión de tres conjuntos A, B y C, primero se toma la unión de dos conjuntos cualesquiera y luego la "unión" del conjunto resultante con el tercero; un procedimiento similar se aplica a la operación de intersección. Los resultados de esta clase de operaciones se ilustran en la figura 2.3. Es interesante que el orden en el que se eligen los conjuntos sea intrascendente. Este hecho da lugar a la *ley asociativa* (de uniones e intersecciones):

$$A \cup (B \cup C) = (A \cup B) \cup C$$

$$A \cap (B \cap C) = (A \cap B) \cap C$$

Estas ecuaciones recuerdan en forma importante las leyes algebraicas $a + (b + c) = (a + b) + c$ y $a \times (b \times c) = (a \times b) \times c$.

Hay también una ley de operación que se aplica cuando las uniones e intersecciones se usan en combinación. Ésta es la *ley distributiva* (de uniones e intersecciones):

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

Éstas se asemejan a la ley algebraica $a \times (b + c) = (a \times b) + (a \times c)$.

Ejemplo 7

Compruebe la ley distributiva, dados $A = \{4, 5\}$, $B = \{3, 6, 7\}$ y $C = \{2, 3\}$. Para comprobar la primera parte de la ley, se determinan por separado las expresiones del lado izquierdo y del lado derecho.

Izquierdo: $A \cup (B \cap C) = \{4, 5\} \cup \{3\} = \{3, 4, 5\}$

Derecho: $(A \cup B) \cap (A \cup C) = \{3, 4, 5, 6, 7\} \cap \{2, 3, 4, 5\} = \{3, 4, 5\}$

Puesto que los dos lados producen el mismo resultado, se verifica la ley. Repitiendo el procedimiento para la segunda parte de la ley, se tiene

$$\text{Izquierda:} \qquad A \cap (B \cup C) = \{4, 5\} \cap \{2, 3, 6, 7\} = \varnothing$$

$$\text{Derecha:} \qquad (A \cap B) \cup (A \cap C) = \varnothing \cup \varnothing = \varnothing$$

Así, de nuevo se verifica la ley.

Verificar una ley significa examinar mediante un ejemplo específico si ésta funciona en realidad. Si la ley es válida, entonces cualquier ejemplo específico debería, por ende, funcionar. Esto indica que si la ley no se cumple en por lo menos un ejemplo, entonces se invalida. Por otro lado, la verificación exitosa mediante ejemplos específicos (no obstante muchos) por sí misma no demuestra la ley. Para *probar* una ley, es necesario demostrar que la ley es válida para todos los casos posibles. El procedimiento que interviene en tal demostración se ilustra más adelante (véase, por ejemplo, la sección 2.5).

EJERCICIO 2.3

1. Escriba lo siguiente en notación de conjuntos:
   (*a*) El conjunto de los números reales mayores que 34.
   (*b*) El conjunto de los números reales mayores que 8 pero menores que 65.
2. Dados los conjuntos $S_1 = \{2, 4, 6\}$, $S_2 = \{7, 2, 6\}$, $S_3 = \{4, 2, 6\}$ y $S_4 = \{2, 4\}$, ¿cuáles de los siguientes enunciados son verdaderos?
   (*a*) $S_1 = S_3$
   (*b*) $S_1 = R$ (conjunto de los números reales)
   (*c*) $8 \in S_2$
   (*d*) $3 \notin S_2$
   (*e*) $4 \notin S_3$
   (*f*) $S_4 \subset R$
   (*g*) $S_1 \supset S_4$
   (*h*) $\varnothing \subset S_2$
   (*i*) $S_3 \supset \{1, 2\}$
3. En relación con los cuatro conjuntos dados en el problema 2, determine:
   (*a*) $S_1 \cup S_2$
   (*b*) $S_1 \cup S_3$
   (*c*) $S_2 \cap S_3$
   (*d*) $S_2 \cap S_4$
   (*e*) $S_4 \cap S_2 \cap S_1$
   (*f*) $S_3 \cup S_1 \cup S_4$
4. ¿Cuáles de los siguientes enunciados son válidos?
   (*a*) $A \cup A = A$
   (*b*) $A \cap A = A$
   (*c*) $A \cup \varnothing = A$
   (*d*) $A \cup U = U$
   (*e*) $A \cap \varnothing = \varnothing$
   (*f*) $A \cap U = A$
   (*g*) El complemento de $\tilde{A}$ es A.
5. Dados $A = \{4, 5, 6\}$, $B = \{3, 4, 6, 7\}$ y $C = \{2, 3, 6\}$, compruebe la ley distributiva.
6. Compruebe la ley distributiva por medio de diagramas de Venn, con diferentes órdenes de sombreado sucesivo.
7. Enumere los subconjuntos del conjunto $\{5, 6, 7\}$.
8. Enumere los subconjuntos del conjunto $S = \{a, b, c, d\}$. ¿Cuántos subconjuntos hay?
9. En el ejemplo 6 se muestra que $\varnothing$ es el complemento de U. Pero puesto que el conjunto vacío es un subconjunto de *cualquier* conjunto, $\varnothing$ debe ser un subconjunto de U. En vista de que el término "complemento de U" indica la idea de *no estar en U*, mientras que el término "subconjunto de U" indica la idea de estar *en U*, parece paradójico que $\varnothing$ sea ambas cosas. ¿Cómo resuelve esta paradoja?

2.4 Relaciones y funciones

El análisis de conjuntos se suscitó por el uso de ese término en relación con varias clases de números en el sistema numérico. Sin embargo, los conjuntos también se pueden referir a objetos distintos a números. En particular, se puede hablar de conjuntos de "pares ordenados", que se definen en seguida, los cuales conducen a conceptos importantes de relaciones y funciones.

Pares ordenados

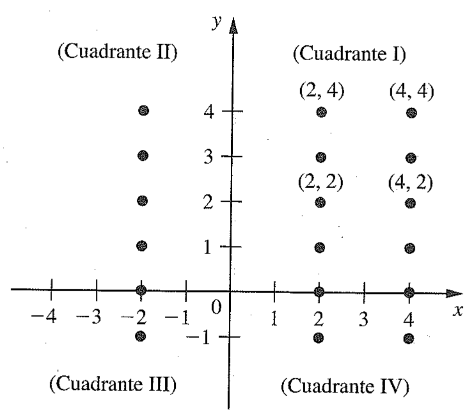
Al escribir un conjunto $\{a, b\}$, no importa el orden en el que aparecen los elementos a y b, porque por definición $\{a, b\} = \{b, a\}$. El par de elementos a y b, en este caso, no es un *par ordenado*. Sin embargo, cuando el orden de a y b tiene importancia, se pueden escribir dos *pares ordenados* diferentes denotados por (a, b) y (b, a), que tienen la propiedad de que $(a, b) \neq (b, a)$ a menos que $a = b$. Conceptos similares se aplican a un conjunto con más de dos elementos, en cuyo caso se puede distinguir entre ternas, cuádruplas, quíntuplas, etc., ordenadas y no ordenadas. Los pares, ternas, etc., ordenados se pueden llamar en forma colectiva *conjuntos ordenados*; éstos se encierran entre paréntesis y no entre llaves.

Ejemplo 1 Para mostrar la edad y el peso de cada estudiante en una clase, se pueden formar pares ordenados (a, w), en los que el primer elemento indica la edad (en años) y el segundo elemento indica el peso (en libras). Entonces (19, 127) y (127, 19) obviamente indican cosas diferentes. Además, el último par ordenado difícilmente se ajusta a algún estudiante en cualquier parte.

Ejemplo 2 Cuando se habla del conjunto de los competidores de un juego olímpico, el orden en el que se listan no tiene importancia y se tiene un conjunto no ordenado. Pero el conjunto {medallista de oro, medallista de plata, medallista de bronce} es una terna ordenada.

Los pares ordenados, como otros objetos, pueden ser elementos de un conjunto. Considere el plano coordenado rectangular (cartesiano) de la figura 2.4, donde un eje x y un eje y se cruzan entre sí en un ángulo recto, dividiendo el plano en cuatro cuadrantes. Este plano xy es un conjunto infinito de puntos, cada uno de los cuales representa un par ordenado cuyo primer elemento es un valor x y el segundo elemento, un valor y. Resulta claro que el punto marcado con (4, 2) es diferente del punto (2, 4); así, en este caso es importante el orden.

FIGURA 2.4



Con esta comprensión visual, se está listo para considerar el proceso de generación de pares ordenados. Suponga, a partir de dos conjuntos específicos, $x = \{1, 2\}$ y $y = \{3, 4\}$, que se desea formar los pares ordenados posibles con el primer elemento tomado del conjunto x y el segundo elemento tomado del conjunto y. El resultado, por supuesto, será el conjunto de cuatro pares ordenados $(1, 3)$, $(1, 4)$, $(2, 3)$ y $(2, 4)$. Este conjunto se llama *producto cartesiano* (en honor a Descartes), o *producto directo*, de los conjuntos x y y, y se denota mediante $x \times y$ (léase: "x cruz y"). Es importante recordar que mientras que x y y son conjuntos de números, el producto cartesiano resulta ser un conjunto de pares ordenados. Por enumeración, o por descripción, se puede de otro modo expresar este producto como

$$x \times y = \{(1, 3), (1, 4), (2, 3), (2, 4)\}$$

o bien

$$x \times y = \{(a, b) \mid a \in x \quad \text{y} \quad b \in y\}$$

La última expresión de hecho se podría tomar como la definición general del producto cartesiano para cualesquiera conjuntos dados x y y.

Para ampliar el horizonte, ahora se permite que x y y incluyan a los números reales. Entonces el producto cartesiano resultante

$$x \times y = \{(a, b) \mid a \in R \quad \text{y} \quad b \in R\} \tag{2.3}$$

representa el conjunto de pares ordenados con elementos reales. Además, cada par ordenado corresponde a un punto *único* en el plano coordenado cartesiano de la figura 2.4 y, recíprocamente, cada punto del plano coordenado también corresponde a un par ordenado *único* en el conjunto $x \times y$. En vista de esta doble unicidad, se dice que existe una *correspondencia uno a uno* entre el conjunto de pares ordenados en el producto cartesiano (2.3) y el conjunto de puntos en el plano coordenado rectangular. Ahora es fácil percibir la lógica de la notación $x \times y$; esto se podría asociar con el cruce del eje x y el eje y en la figura 2.4. Una forma más simple de expresar el conjunto $x \times y$ en (2.3) es escribirlo directamente como $R \times R$; esto por lo común se denota también como R^2.

Al ampliar esta idea, se podría definir también el producto cartesiano de tres conjuntos x, y y z, como sigue:

$$x \times y \times z = \{(a, b, c) \mid a \in x, b \in y, c \in z\}$$

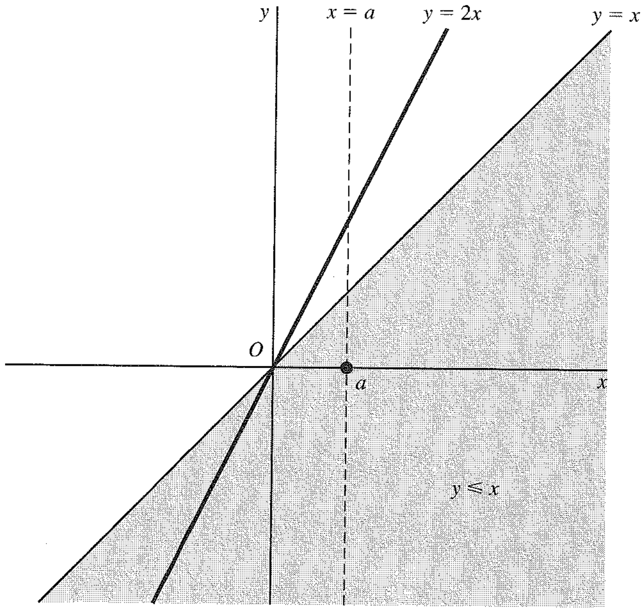
el cual es un conjunto de ternas ordenadas. Además, si los conjuntos x, y y z constan de los números reales, el producto cartesiano corresponderá al conjunto de los puntos en un espacio tridimensional. Esto se puede denotar mediante $R \times R \times R$, o de forma concisa como R^3. En el presente análisis, las variables se toman como valores reales; así, por lo común, el marco es $R^2, R^3, \ldots,$ o R^n.

Relaciones y funciones

Puesto que cualquier par ordenado asocia un valor y con un valor x, cualquier colección de pares ordenados —cualquier subconjunto del producto cartesiano (2.3)— constituirá una *relación* entre y y x. Dado un valor x, uno o más valores y se especificarán mediante esta relación. Por conveniencia, en general los elementos de $x \times y$ se escribirán ahora como (x, y) —en vez de (a, b)—, como se hizo en (2.3), donde tanto x como y son variables.

Ejemplo 3 El conjunto $\{(x, y) \mid y = 2x\}$ es un conjunto de pares ordenados que incluye, por ejemplo, $(1, 2)$, $(0, 0)$ y $(-1, -2)$. Esto constituye una relación, y su contraparte gráfica es el conjunto de puntos que yacen en la recta $y = 2x$, como se ve en la figura 2.5.

FIGURA 2.5



Ejemplo 4

El conjunto $\{(x, y) \mid y \leq x\}$, que consiste en los pares ordenados tales como (1, 0), (1, 1) y (1, –4), constituye otra relación. En la figura 2.5, este conjunto corresponde al conjunto de los puntos en el área sombreada que satisface la desigualdad $y \leq x$.

Observe que, cuando se da el valor x, no siempre es posible determinar un valor *único* y de una relación. En el ejemplo 4, los tres pares ordenados muestran que si $x = 1$, y puede tomar varios valores, como 0, 1 o –4, y aún satisfacer en cada caso la relación expresada. Desde el punto de vista gráfico, dos o más puntos de una relación podrían caer en una sola recta vertical en el plano xy. Esto se ejemplifica en la figura 2.5, donde muchos puntos del área sombreada (que representan la relación $y \leq x$) caen en la recta vertical discontinua identificada con $x = a$.

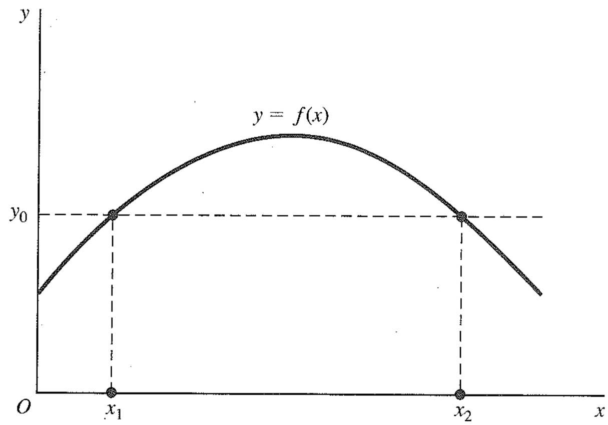
Sin embargo, como un caso especial, una relación puede ser tal que para cada valor x existe sólo *un* valor y correspondiente. La relación del ejemplo 3 es un ejemplo claro. En ese caso, se dice que y es una *función* de x, y esto se denota mediante $y = f(x)$, que se lee "y es igual a f de x". [*Nota:* $f(x)$ *no* significa f multiplicada por x.] Por lo tanto, una función es un conjunto de pares ordenados con la propiedad de que cualquier valor x determina un valor y *único*.[2] Debe resultar claro que una función debe ser una relación, pero es posible que una relación no sea una función.

Aunque la definición de función estipula una y única para cada x, no se requiere lo contrario. En otras palabras, se podría asociar más de un valor x al mismo valor y. Esta posibilidad se ilustra en la figura 2.6, donde los valores x_1 y x_2 en el conjunto x se relacionan con el mismo valor (y_0) en el conjunto y mediante la función $y = f(x)$.

Una función es también un *mapeo* o *transformación*; ambas palabras indican la acción de asociar una cosa con otra. En la expresión $y = f(x)$, la notación funcional f podría ser

[2] Esta definición de *función* corresponde a lo que se podría llamar *función de un solo valor* en la terminología antigua. Lo que antes se llamaba *función de varios valores* ahora se conoce como *relación* o correspondencia.

FIGURA 2.6



interpretada como una regla mediante la cual el conjunto x se "mapea" ("transforma") en el conjunto y. Entonces se puede escribir

$$f: x \rightarrow y$$

donde la flecha indica el mapeo y la letra f especifica de forma simbólica una regla de mapeo. Puesto que f representa una regla de mapeo *particular*, se debe emplear una notación funcional distinta para denotar otra función que pudiera aparecer en el mismo modelo. Los símbolos usuales (además de f) usados para este propósito son g, F, G, las letras griegas ϕ (fi) y ψ (psi) y sus mayúsculas Φ y Ψ. Por ejemplo, dos variables y y z pueden ser funciones de x, pero si una función se escribe como $y = f(x)$, la otra se debe escribir como $z = g(x)$, o bien, $z = \phi(x)$. No obstante, también se permite escribir $y = y(x)$ y $z = z(x)$, y de este modo deshacerse de los símbolos f y g.

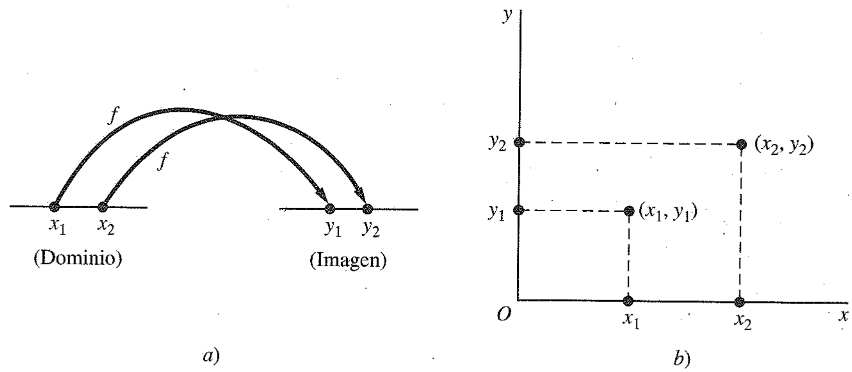
En la función $y = f(x)$, x se denomina *argumento* de la función y la letra y se llama *valor* de la función. También se hará referencia a x como la *variable independiente* y a y como la *variable dependiente*. El conjunto de los valores permisibles que x puede tomar en un contexto dado se conoce como el *dominio* de la función, que puede ser un subconjunto del conjunto de los números reales. El valor y hacia el cual se envía un valor x se llama la *imagen* de ese valor x. El conjunto de las imágenes se llama *imagen* de la función (o codominio), que es el conjunto de los valores que puede tomar la variable y. Así, el dominio pertenece a la variable independiente x, y la imagen tiene que ver con la variable dependiente y.

Como se ilustra en la figura 2.7*a*, se podría considerar la función f como una regla para el mapeo de cada punto en algún segmento de recta (el dominio) hacia algún punto en otro segmento de recta (la imagen). Al colocar el dominio en el eje x y la imagen en el eje y, como en la figura 2.7*b*, de inmediato se obtiene la gráfica familiar de dos dimensiones, en la cual la asociación entre valores x y valores y se especifica mediante un conjunto de pares ordenados, como (x_1, y_1) y (x_2, y_2).

En los modelos económicos, las ecuaciones de comportamiento normalmente entran como funciones. Puesto que la mayor parte de las variables económicas, por su naturaleza, están restringidas a ser números reales no negativos,[3] sus dominios también están restringidos. Ésta es

[3] Se dice "no negativo" en vez de "positivo" cuando son permisibles valores cero.

FIGURA 2.7



la razón de por qué la mayor parte de las representaciones geométricas en economía se trazan sólo en el primer cuadrante. En general, no se especifica el dominio de toda función en cada modelo económico. Cuando no se da especificación alguna, se entiende que el dominio (y la imagen) sólo incluyen números para los que una función tiene sentido económico.

Ejemplo 5 El costo total C de una empresa por día es una función de su producción diaria Q: $C = 150 + 7Q$. La empresa tiene una capacidad límite de 100 unidades de producto por día. ¿Cuáles son el dominio y la imagen de la función de costo? En vista de que Q puede variar sólo entre 0 y 100, el dominio es el conjunto de valores $0 \leq Q \leq 100$; o de manera más formal,

$$\text{Dominio} = \{Q \mid 0 \leq Q \leq 100\}$$

En cuanto a la imagen, puesto que la función tiene la forma de una recta, con el valor mínimo C en 150 (cuando $Q = 0$) y el valor máximo C en 850 (cuando $Q = 100$), se tiene

$$\text{Imagen} = \{C \mid 150 \leq C \leq 850\}$$

No obstante, tenga en cuenta la posibilidad de que los valores extremos de la imagen no ocurran siempre donde se obtienen los valores extremos del dominio.

EJERCICIO 2.4

1. Dados $S_1 = \{3, 6, 9\}$, $S_2 = \{a, b\}$ y $S_3 = \{m, n\}$, determine los productos cartesianos:
   (*a*) $S_1 \times S_2$ (*b*) $S_2 \times S_3$ (*c*) $S_3 \times S_1$
2. De la información del problema 1, encuentre el producto cartesiano $S_1 \times S_2 \times S_3$.
3. En general, ¿se cumple que $S_1 \times S_2 = S_2 \times S_1$? ¿En qué condiciones estos dos productos cartesianos son iguales?
4. ¿Alguna de las siguientes representaciones, dibujadas en un plano coordenado rectangular, representa una función?
   (*a*) Un círculo (*c*) Un rectángulo
   (*b*) Un triángulo (*d*) Una recta con pendiente descendente
5. Si el dominio de la función $y = 5 + 3x$ es el conjunto $\{x \mid 1 \leq x \leq 9\}$, determine la imagen de la función y exprésela como un conjunto.

6. Para la función $y = -x^2$, si el dominio es el conjunto de los números reales no negativos, ¿cuál es la imagen?
7. En la teoría de la empresa, los economistas consideran el costo total C como una función del nivel de producción Q: $C = f(Q)$.
   (*a*) De acuerdo con la definición de una función, ¿se debe relacionar cada cifra de costo con un nivel de producción único?
   (*b*) ¿Cada nivel de producción debe determinar una cifra de costo única?
8. Si un nivel de producción Q_1 se puede producir a un costo de C_1, entonces también debe ser posible (por ser menos eficaz) producir Q_1 a un costo de $C_1 + \$1$, o $C_1 + \$2$, etc. Así, parecería que el producto Q no sólo determina el costo total C. Si así fuera, escribir $C = f(Q)$ violaría la definición de una función. ¿Cómo, a pesar de este razonamiento, justificaría el uso de la función $C = f(Q)$?

2.5 Tipos de función

La expresión $y = f(x)$ es una expresión general para el efecto de que un mapeo sea posible, pero la regla real de mapeo no se hace por consiguiente explícita. Considérense ahora varios tipos específicos de función, donde cada uno representa una regla de mapeo distinta.

Funciones constantes

Una función cuya imagen consiste sólo en un elemento se llama *función constante*. Como ejemplo, se cita la función

$$y = f(x) = 7$$

que se puede expresar de otro modo como $y = 7$ o $f(x) = 7$, cuyo valor no cambia sin importar el valor de x. En el plano coordenado, esta función aparecerá como una recta horizontal. En los modelos de ingreso nacional, cuando la inversión I se determina de manera exógena, se podría tener una función de inversión de la forma $I = \$100$ millones, o bien, $I = I_0$, que ejemplifica la función constante.

Funciones polinomiales

La función constante es en realidad un caso "degenerado" de los que se conocen como *funciones polinomiales*. La palabra *polinomio* significa "varios términos", y una función polinomial de una sola variable x tiene la forma general

$$y = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \tag{2.4}$$

en la cual cada término contiene un coeficiente así como una potencia entera no negativa de la variable x. (Como se explica después en esta sección, se puede escribir $x^1 = x$ y $x^0 = 1$ en general; así, los dos primeros términos se pueden tomar como a_0x^0 y a_1x^1, respectivamente.) Note que en lugar de los símbolos $a, b, c, \ldots,$ se han empleado los símbolos $a_0, a_1, \ldots, a_n$ para los coeficientes. Esto sucede por dos consideraciones: (1) se economizan símbolos, ya que de esta manera sólo se emplea la letra a; y (2) el subíndice ayuda a precisar la ubicación de un coeficiente particular en la ecuación completa. Por ejemplo, en (2.4), a_2 es el coeficiente de x^2, y así sucesivamente.

Dependiendo del valor del entero n (el cual especifica la mayor potencia de x), se tienen varias subclases de función polinomial:

Caso de $n = 0$:	$y = a_0$	[función *constante*]
Caso de $n = 1$:	$y = a_0 + a_1x$	[función *lineal*]
Caso de $n = 2$:	$y = a_0 + a_1x + a_2x^2$	[función *cuadrática*]
Caso de $n = 3$:	$y = a_0 + a_1x + a_2x^2 + a_3x^3$	[función *cúbica*]

etc. Los superíndices de las potencias de x se llaman *exponentes*. La mayor potencia que interviene, es decir, el valor de n, se denomina *grado* de la función polinomial; una función cuadrática, por ejemplo, es un polinomio de segundo grado, y una función cúbica es un polinomio de tercer grado.[4] El orden en el que aparecen los términos a la derecha del signo igual es irrelevante: podrían estar ordenados de la mayor a la menor potencia. Además, aunque se colocó el símbolo y a la izquierda, también es aceptable escribir $f(x)$ en su lugar.

Cuando se grafica en el plano coordenado, una función lineal aparece como una recta, según se ilustra en la figura 2.8*a*. Cuando $x = 0$, la función lineal produce $y = a_0$; así, el par ordenado $(0, a_0)$ está sobre la recta. Esto da la ordenada al origen y, porque es en este punto que el eje vertical se cruza con la recta. El otro coeficiente, a_1, mide la *pendiente* (el grado de inclinación) de la recta. Esto significa que un incremento unitario en x da como resultado un incremento en y en la cantidad de a_1. En la figura 2.8*a* se ilustra el caso de $a_1 > 0$, lo cual tiene que ver con una pendiente positiva y, por lo tanto, una recta de pendiente ascendente; si $a_1 < 0$, la recta tiene pendiente descendente.

Por otro lado, una función cuadrática traza una *parábola*, en términos generales, una curva con una sola protuberancia o serpenteante. La ilustración de la figura 2.8*b* indica una a_2 negativa; en el caso de $a_2 > 0$, la curva se invierte y muestra un valle en vez de una colina. La gráfica de una función cúbica, en general, muestra dos ondulaciones, como se ilustra en la figura 2.8*c*. Estas funciones se usan con mucha frecuencia en los modelos económicos analizados más adelante.

Funciones racionales

Una función, como por ejemplo,

$$y = \frac{x - 1}{x^2 + 2x + 4}$$

en la cual y se expresa como una razón de dos polinomios en la variable x, se conoce como *función racional*. Según esta definición, cualquier función polinomial debe ser por sí misma una función racional, porque siempre se puede expresar como una razón respecto a 1, y 1 es una función constante.

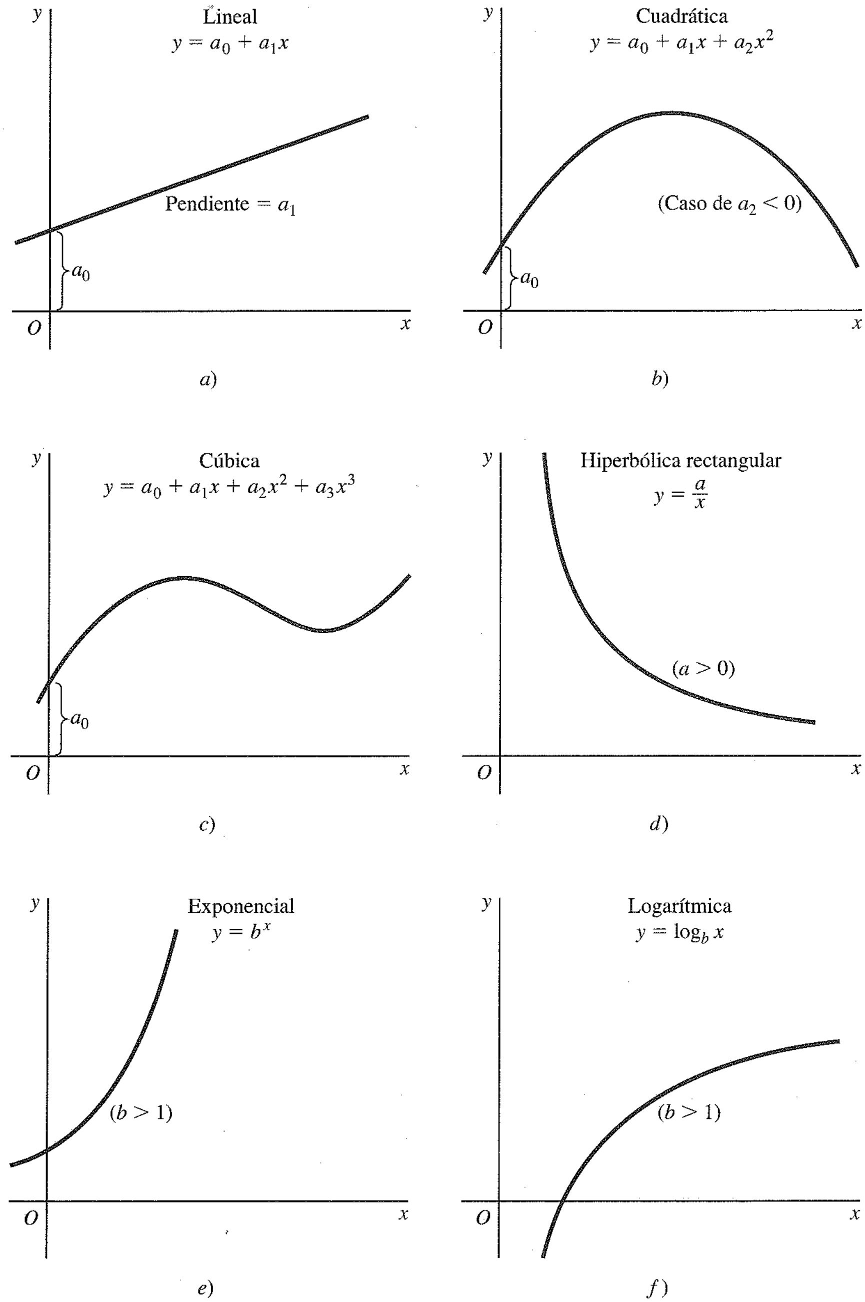
Una función racional especial que tiene aplicaciones interesantes en economía es la función

$$y = \frac{a}{x} \qquad \text{o} \qquad xy = a$$

que describe una *hipérbola rectangular*, como en la figura 2.8*d*. Puesto que el producto de las dos variables es siempre una constante fija en este caso, esta función se puede usar para representar esa curva de demanda especial, con precio P y cantidad Q en los dos ejes, para la

[4] En las distintas ecuaciones recién citadas, se supone siempre que el último coeficiente (a_n) no es igual a cero; de otro modo, la función degeneraría en un polinomio de grado menor.

FIGURA 2.8



cual el gasto total PQ es constante en todos los niveles de precio. (Esta curva de demanda es aquella con una elasticidad unitaria en cada punto sobre la curva.) Otra aplicación es para la curva del costo fijo promedio (CFP) (*average fixed cost*, AFC). Con CFP en un eje y el producto Q en el otro, la curva CFP debe ser hiperbólica rectangular porque CFP $\times$ Q(= costo fijo total) es una constante fija.

La hipérbola rectangular trazada a partir de $xy = a$ nunca toca los ejes, incluso si se extiende de forma indefinida hacia arriba y a la derecha. En cambio, la curva se aproxima *asintóticamente* a los ejes: cuando y crece, la curva se aproxima cada vez más al eje y pero en realidad nunca lo toca, y ocurre algo similar para el eje x. Los ejes constituyen las *asíntotas* de esta función.

Funciones no algebraicas

Cualquier función expresada en términos de polinomios o raíces, o ambas cosas (por ejemplo, la raíz cuadrada) de polinomios es una *función algebraica*. En consecuencia, las funciones analizadas hasta aquí son algebraicas.

Sin embargo, las *funciones exponenciales* como $y = b^x$, en la cual la variable independiente aparece en el exponente, son *no algebraicas*. Las estrechamente relacionadas *funciones logarítmicas*, como $y = \log_b x$, son también no algebraicas. Estos dos tipos de funciones tienen una función especial en ciertos tipos de aplicaciones económicas y, desde el punto de vista pedagógico, es deseable posponer su estudio hasta el capítulo 10. Aquí, simplemente se prevén sus formas gráficas generales en la figura 2.8*e* y *f*. Otras clases de funciones no algebraicas son las *funciones trigonométricas* (o *circulares*), las cuales se estudian en el capítulo 16 junto con el análisis dinámico. Se debe añadir aquí que las funciones no algebraicas también se conocen con el nombre más esotérico de *funciones trascendentes*.

Digresión acerca de exponentes

En el estudio de las funciones polinomiales se introdujo el término *exponente* como indicador de la potencia a la que está elevada una variable (o número). La expresión 6^2 significa que 6 está elevado a la segunda potencia; es decir, el 6 se va a multiplicar por sí mismo, o $6^2 \equiv 6 \times 6 = 36$. En general, se define, para un entero positivo n,

$$x^n \equiv \underbrace{x \times x \times \cdots \times x}_{n \text{ términos}}$$

como un caso especial, se nota que $x^1 = x$. De la definición general, se deduce que para enteros positivos m y n, los exponentes obedecen las reglas siguientes:

Regla I $\qquad x^m \times x^n = x^{m+n} \qquad$ (por ejemplo, $x^3 \times x^4 = x^7$)

PRUEBA

$$x^m \times x^n = (\underbrace{x \times x \times \cdots \times x}_{m \text{ términos}})(\underbrace{x \times x \times \cdots \times x}_{n \text{ términos}})$$
$$= \underbrace{x \times x \times \cdots \times x}_{m+n \text{ términos}} = x^{m+n}$$

Observe que en esta prueba no se asignó ningún valor específico al número x, o a los exponentes m y n. Así, el resultado obtenido es *generalmente* cierto. Es por esta razón que la demostración dada constituye una prueba, en oposición a una mera comprobación. Lo mismo se puede decir acerca de la prueba de la regla II siguiente.

Regla II $\qquad \dfrac{x^m}{x^n} = x^{m-n} \qquad (x \neq 0) \qquad \left(\text{por ejemplo, } \dfrac{x^4}{x^3} = x\right)$

PRUEBA

$$\frac{x^m}{x^n} = \frac{\overbrace{x \times x \times \cdots \times x}^{m \text{ términos}}}{\underbrace{x \times x \times \cdots \times x}_{n \text{ términos}}} = \underbrace{x \times x \times \cdots \times x}_{m-n \text{ términos}} = x^{m-n}$$

porque los n términos del denominador cancelan n de los m términos del numerador. Note que el caso de $x = 0$ se descarta en la enunciación de esta regla. Esto se debe a que cuando $x = 0$, la expresión x^m/x^n implicaría la división entre cero, la cual no está definida.

¿Qué pasa si $m < n$, por ejemplo, $m = 2$ y $n = 5$? En ese caso se obtiene, de acuerdo con la regla II, $x^{m-n} = x^{-3}$, una *potencia negativa* de x. ¿Qué significa esto? La respuesta la proporciona la regla II: cuando $m = 2$ y $n = 5$, se tiene

$$\frac{x^2}{x^5} = \frac{x \times x}{x \times x \times x \times x \times x} = \frac{1}{x \times x \times x} = \frac{1}{x^3}$$

Así, $x^{-3} = 1/x^3$, y esto se puede generalizar en otra regla:

Regla III $$x^{-n} = \frac{1}{x^n} \qquad (x \neq 0)$$

Elevar un número (no cero) a una potencia de *n negativa* es tomar el *recíproco* de su n-ésima potencia.

Otro caso especial en la aplicación de la regla II se da cuando $m = n$, que produce la expresión $x^{m-n} = x^{m-m} = x^0$. Para interpretar el significado de elevar un número x a la potencia cero, se puede escribir el término x^{m-m} de acuerdo con la regla II, con el resultado de que $x^m/x^m = 1$. Así, se puede concluir que cualquier número (no cero) elevado a la potencia cero es igual a 1. (La expresión 0^0 no está definida.) Esto se puede expresar como otra regla:

Regla IV $$x^0 = 1 \qquad (x \neq 0)$$

Mientras que se tenga que ver sólo con funciones polinomiales, únicamente se requieren potencias enteras (no negativas). Sin embargo, en las funciones exponenciales, el exponente es una variable que puede tomar valores no enteros también. A fin de interpretar un número, como $x^{1/2}$, considérese el hecho de que, por la regla I, se tiene

$$x^{1/2} \times x^{1/2} = x^1 = x$$

Puesto que $x^{1/2}$ multiplicada por sí misma es x, $x^{1/2}$ debe ser la raíz cuadrada de x. De manera similar, se puede demostrar que $x^{1/3}$ es la raíz cúbica de x. En general, por lo tanto, se puede expresar la regla siguiente:

Regla V $$x^{1/n} = \sqrt[n]{x}$$

Otras dos reglas que siguen los exponentes son

Regla VI $$(x^m)^n = x^{mn}$$

Regla VII $$x^m \times y^m = (xy)^m$$

EJERCICIO 2.5

1. Grafique las funciones
   (*a*) $y = 16 + 2x$ (*b*) $y = 8 - 2x$ (*c*) $y = 2x + 12$
   (En cada caso considere que el dominio consiste sólo en los números reales no negativos.)
2. ¿Cuál es la diferencia principal entre (*a*) y (*b*) en el problema 1? ¿Cómo se refleja esta diferencia en las gráficas? ¿Cuál es la diferencia principal entre (*a*) y (*c*)? ¿Cómo la reflejan sus gráficas?

3. Gráficas de funciones

   (*a*) $y = -x^2 + 5x - 2$ (*b*) $y = x^2 + 5x - 2$

   con el conjunto de valores $-5 \leq x \leq 5$ que constituye el dominio. Es bien sabido que el signo del coeficiente del término x^2 determina si la gráfica de la función cuadrática tendrá una "colina" o un "valle". Con base en el presente problema, ¿qué signo se relaciona con la colina? Proporcione una explicación intuitiva para esto.
4. Grafique la función $y = 36/x$, suponiendo que x y y pueden tomar valores positivos solamente. A continuación, suponga que ambas variables pueden tomar valores negativos también; ¿cómo se debe modificar la gráfica para reflejar este cambio?
5. Condense las expresiones siguientes:

   (*a*) $x^4 \times x^{15}$ (*b*) $x^a \times x^b \times x^c$ (*c*) $x^3 \times y^3 \times z^3$
6. Determine: (*a*) x^3/x^{-3} (*b*) $(x^{1/2} \times x^{1/3})/x^{2/3}$
7. Demuestre que $x^{m/n} = \sqrt[n]{x^m} = (\sqrt[n]{x})^m$. Especifique las reglas aplicadas en cada paso.
8. Pruebe la regla VI y la regla VII.

2.6 Funciones de dos o más variables independientes

Hasta aquí se han considerado sólo funciones de una sola variable independiente, $y = f(x)$. Pero el concepto de función se puede ampliar fácilmente al caso de dos o más variables independientes. Dada una función

$$z = g(x, y)$$

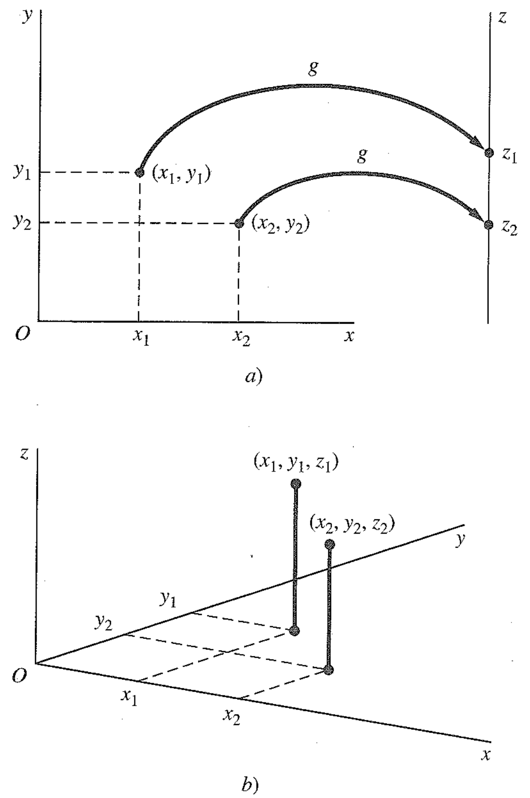
un determinado par de valores x y y determinan de manera única un valor de la variable dependiente z. Esta función se ejemplifica mediante

$$z = ax + by \qquad \text{o} \qquad z = a_0 + a_1x + a_2x^2 + b_1y + b_2y^2$$

Así como la función $y = f(x)$ mapea un punto en el dominio hacia un punto en la imagen, la función g hará precisamente lo mismo. Sin embargo, el dominio en este caso ya no es un conjunto de números, sino un conjunto de pares ordenados (x, y), porque se puede determinar z sólo cuando se especifican *tanto* x como y. La función g es, por lo tanto, un mapeo de un punto en un espacio bidimensional hacia un punto en un segmento de recta (es decir, un punto en un espacio unidimensional), por ejemplo del punto (x_1, y_1) hacia el punto z_1 o de (x_2, y_2) hacia z_2 en la figura 2.9*a*.

Sin embargo, si se coloca un eje z vertical perpendicular al plano xy, como se hizo en el diagrama *b*, se obtiene un espacio tridimensional en el cual se puede dar a la función g una representación gráfica como sigue. El dominio de la función será algún subconjunto de los puntos en el plano xy, y el valor de la función (valor z) para un determinado punto en el dominio, por ejemplo, (x_1, y_1), se puede indicar mediante la altura de una recta vertical plantada en ese punto. La relación entre las tres variables se resume entonces por la terna ordenada (x_1, y_1, z_1), que es un punto específico en el espacio tridimensional. El lugar de tales ternas ordenadas, que tomará la forma de una *superficie*, constituye entonces la gráfica de la función g. Mientras que la función $y = f(x)$ es un conjunto de *pares* ordenados, la función $z = g(x, y)$ será un conjunto de *ternas* ordenadas. Habrá muchas ocasiones para usar funcio-

FIGURA 2.9



nes de este tipo en modelos económicos. Una aplicación inmediata es en el área de funciones de producción. Suponga que la producción se determina mediante cantidades de capital (K) y mano de obra (L) empleada; entonces se puede escribir una función de producción de la forma general $Q = Q(K, L)$.

La posibilidad de abarcar casos de tres o más variables independientes ahora es evidente por sí misma. Con la función $y = h(u, v, w)$, por ejemplo, se puede mapear un punto del espacio tridimensional (u_1, v_1, w_1), hacia un punto en un espacio unidimensional (y_1). Esta función se podría usar para indicar que la utilidad de un consumidor es una función de su consumo de tres artículos diferentes, y el mapeo es de un espacio de artículos tridimensional hacia un espacio de utilidad unidimensional. Pero esta vez será físicamente imposible graficar la función, porque para esa tarea se requiere un diagrama de cuatro dimensiones a fin de dibujar las cuádruplas ordenadas, pero el mundo en que vivimos sólo tiene tres dimensiones. No obstante, en vista del atractivo intuitivo de la analogía geométrica, se pueden seguir denotando las cuádruplas ordenadas (u_1, v_1, w_1, y_1) como un "punto" en el espacio de cuatro dimensiones. El lugar geométrico de estos puntos produce la "gráfica" (que no es posible trazar) de la función $y = h(u, v, w)$, que se llama *hipersuperficie*. Estos términos, a saber, punto e hipersuperficie, también son llevados al caso general del espacio de n dimensiones.

Las funciones de más de una variable se clasifican en varios tipos. Por ejemplo, una función de la forma

$$y = a_1x_1 + a_2x_2 + \cdots + a_nx_n$$

es una función *lineal*, cuya característica es que cada variable sea elevada sólo a la primera potencia. Por otro lado, una función *cuadrática*, tiene que ver con primeras y segundas potencias de una o más variables independientes, pero la suma de los exponentes de las variables que aparecen en cualquier término simple no debe pasar de 2.

Note que en lugar de denotar las variables independientes mediante x, u, v, w, etc., se han intercambiado por los símbolos $x_1, x_2, \ldots, x_n$. La última notación, como el sistema de coeficientes con subíndices, tiene el mérito de economía de alfabeto, así como de un conteo más fácil del número de variables que intervienen en la función.

2.7 Niveles de generalidad

Al estudiar los distintos tipos de función, se han introducido, sin advertencia explícita, ejemplos de funciones que pertenecen a diversos niveles de generalidad. En ciertos casos, se han escrito funciones de la forma

$$y = 7 \qquad y = 6x + 4 \qquad y = x^2 - 3x + 1 \qquad \text{(etcétera)}$$

Éstas no sólo se expresan en términos de coeficientes numéricos, sino que también indican de modo específico si cada función es constante, lineal o cuadrática. En términos de gráficas, cada función de este tipo da lugar a una curva única bien definida. En vista de la naturaleza numérica de estas funciones, las soluciones del modelo basadas en ellas emergerán también como valores numéricos. La desventaja es que si se desea saber cómo cambiará la conclusión analítica cuando entra en vigor un conjunto diferente de coeficientes numéricos, se debe examinar de nuevo el proceso de razonamiento cada vez. Así, los resultados obtenidos de funciones específicas tienen muy poca generalidad.

En un nivel más general de descripción y análisis, hay funciones de la forma

$$y = a \qquad y = a + bx \qquad y = a + bx + cx^2 \qquad \text{(etcétera)}$$

Puesto que se emplean parámetros, cada función representa no sólo una curva sino una familia completa de curvas. La función $y = a$, por ejemplo, comprende no sólo los casos específicos $y = 0$, $y = 1$ y $y = 2$, sino también $y = \frac{1}{3}$, $y = -5, \ldots,$ hasta el infinito. Con funciones paramétricas, el resultado de las operaciones matemáticas también será en términos de parámetros. Estos resultados son más generales en el sentido de que, al asignar varios valores a los parámetros que aparecen en la solución del modelo, se puede obtener una familia completa de respuestas específicas sin tener que repetir otra vez el proceso de razonamiento.

A fin de obtener un nivel de generalidad incluso mayor, se puede recurrir a la expresión general de función $y = f(x)$ o bien, $z = g(x, y)$. Cuando se expresa en esta forma, la función no está restringida a ser lineal, cuadrática, exponencial o trigonométrica, las cuales están incluidas en la notación. El resultado analítico basado en tal formulación general tendrá por consiguiente la aplicabilidad más general. No obstante, como se encontrará a continuación, a fin de obtener resultados con significado económico, suele ser necesario imponer ciertas restricciones cualitativas en las funciones generales integradas en un modelo, tal como la restricción de que una función de demanda tiene una gráfica con pendiente negativa o de que una función de consumo tiene una gráfica con una pendiente positiva menor que la unidad.

Para resumir el presente capítulo, ahora es clara la estructura de un modelo económico matemático. En general, éste consistirá en un sistema de ecuaciones, que podría ser definicional, de comportamiento o de la naturaleza de las condiciones de equilibrio.[5] Las ecuaciones de comportamiento están por lo común en la forma de funciones, que podrían ser lineales o no lineales, numéricas o paramétricas y con una variable independiente o muchas. Es mediante éstas que las suposiciones analíticas adoptadas en el modelo reciben una expresión matemática.

Por lo tanto, al atacar un problema analítico, el primer paso es seleccionar las variables apropiadas, tanto exógenas como endógenas, para su inclusión en el modelo. A continuación, se debe traducir en ecuaciones el conjunto de suposiciones analíticas elegidas en relación con los aspectos humano, institucional, tecnológico, legal y otros aspectos de comportamiento del ambiente que afecta el funcionamiento de las variables. Sólo entonces se puede intentar deducir un conjunto de conclusiones a través de operaciones y manejos matemáticos pertinentes y darles interpretaciones económicas apropiadas.

[5] Las desigualdades también podrían entrar como un elemento importante de un modelo, pero por el momento no se tomarán en cuenta.

Parte 2

Análisis estático (o de equilibrio)

Capítulo 3

Análisis de equilibrio en economía

El procedimiento analítico descrito en el capítulo 2 se aplicará primero a lo que se conoce como *análisis estático*, o *análisis de equilibrio*. Para este propósito, es imperativo tener primero una comprensión clara de lo que significa *equilibrio*.

3.1 El significado de equilibrio

Al igual que cualquier término económico, *equilibrio* se puede definir de varias maneras. Según una definición, un equilibrio es "un conjunto de variables seleccionadas e interrelacionadas, tan ajustadas entre sí que ninguna tendencia inherente a cambiar prevalece en el modelo que constituyen".[1] Varias palabras de esta definición merecen atención especial. Primero, la palabra *seleccionadas* subraya el hecho de que existen variables que, por elección del analista, no han sido incluidas en el modelo. Puesto que el equilibrio en estudio puede tener relevancia sólo en el contexto particular del conjunto de variables elegido, si el modelo se extiende para incluir más variables, entonces ya no se aplica el estado de equilibrio que pertenece al modelo más pequeño.

Segundo, la palabra *interrelacionadas* sugiere que, a fin de que ocurra el equilibrio, las variables del modelo deben estar al mismo tiempo en un estado de reposo. Además, el estado de reposo de cada variable debe ser compatible con el de todas las demás variables; de otro modo, alguna(s) variable(s) cambiaría(n), causando que las otras, por lo tanto, cambiasen también en una reacción en cadena, y se podría decir que no existe ningún equilibrio.

Tercero, la palabra *inherente* significa que, al definir un equilibrio, el estado de reposo en cuestión se basa sólo en el balance de las fuerzas internas del modelo, mientras se suponen fijos los factores externos. Desde el punto de vista operacional, esto significa que los parámetros y las variables exógenas se tratan como constantes. Cuando en realidad cambian los factores externos, habrá un nuevo equilibrio definido con base en los nuevos valores de parámetro; pero, al definir el nuevo equilibrio, se supone que los nuevos valores de parámetro persisten y permanecen sin cambio.

[1] Fritz Machlup, "Equilibrium and Disequilibrium: Misplaced Concreteness and Disguised Politics", en *Economic Journal*, marzo de 1958, p. 9 (reimpreso en F. Machlup, *Essays on Economic Semantics*, Prentice Hall, Inc., Englewood Cliffs, NJ, 1963).

En esencia, un equilibrio de un modelo específico es una situación caracterizada por la falta de una tendencia a cambiar. Es por esta razón que el análisis de equilibrio (o, de modo más específico, el estudio de lo que se parece al estado de equilibrio) se denomina *estático*.

El hecho de que un equilibrio signifique la ausencia de una tendencia a cambiar podría llevarnos a concluir, con base en que sólo en el estado ideal habría una falta de motivación para cambiar, que un equilibrio constituye necesariamente un estado deseable o ideal de asuntos. Esta clase de conclusiones no está garantizada. Aun cuando cierta posición de equilibrio podría representar un estado deseable y algo por qué luchar —como por ejemplo una situación de maximización de ganancia desde el punto de vista de la empresa—, otra posición de equilibrio puede ser bastante indeseable y, por lo tanto, algo que se debe evitar, como un nivel de equilibrio en el subempleo de ingreso nacional. La única interpretación garantizada es que un equilibrio es una situación que, de lograrse, tendería a perpetuarse, a menos de que cambien las fuerzas externas.

La variedad de equilibrio deseable, a la que se denominará *equilibrio final*, se trata después en la parte 4 como problemas de optimización. En el presente capítulo, el estudio se confina al tipo de equilibrio *intermedio*, que no resulta de alguna finalidad consciente en un objetivo particular, sino de un proceso impersonal, o suprapersonal, de interacción y ajuste de fuerzas económicas. Ejemplos de esto son el equilibrio que obtiene el mercado bajo ciertas condiciones de la oferta y la demanda y el equilibrio del ingreso nacional bajo determinadas condiciones de consumo y patrones de inversión.

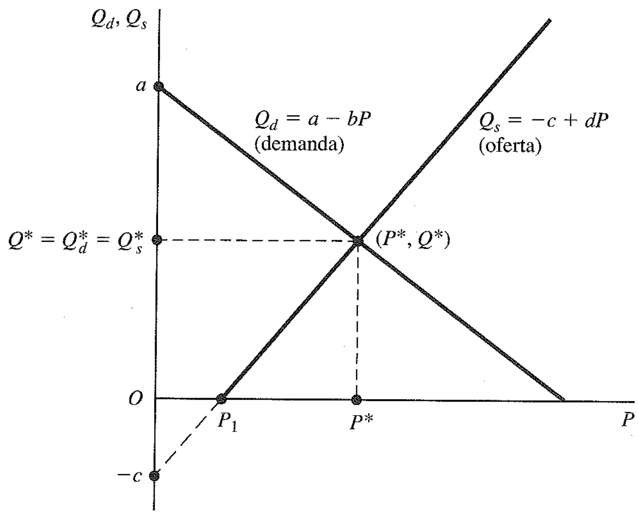
3.2 Equilibrio de mercado parcial: un modelo lineal

En un modelo de equilibrio estático, el problema estándar es hallar el conjunto de valores de las variables endógenas que satisfagan la condición de equilibrio del modelo. Esto se debe a que una vez que se han identificado esos valores, en efecto se ha identificado la condición de equilibrio. Considere como ilustración un modelo de equilibrio parcial de mercado, es decir, un modelo de determinación del precio en un mercado aislado.

Construcción del modelo

Dado que sólo se considera un artículo, es necesario incluir sólo tres variables en el modelo: la cantidad demandada del artículo (Q_d), la cantidad ofrecida del artículo (Q_s) y su precio (P). La cantidad se mide, por ejemplo, en libras por semana y el precio, en dólares. Una vez elegidas las variables, lo siguiente es hacer ciertas suposiciones en relación con el desempeño del mercado. Primero, se debe especificar la condición de equilibrio, algo indispensable en un modelo de equilibrio. La suposición estándar es que el equilibrio ocurra en el mercado si y sólo si la demanda excedente es cero ($Q_d - Q_s = 0$), es decir, si y sólo si el mercado está limpio. Pero esto hace surgir de inmediato la pregunta de cómo se determinan por sí mismas Q_d y Q_s. Para responder esto, se puede suponer que Q_d es una función lineal decreciente de P (cuando P aumenta, Q_d decrece). Por otro lado, Q_s se puede postular como una función lineal creciente de P (si P aumenta, Q_s también), con la condición de que no se ofrece ninguna cantidad a menos que el precio rebase un nivel positivo particular. Entonces, en total el modelo contendrá una condición de equilibrio más dos ecuaciones de comportamiento que gobiernan la demanda y proporcionan aspectos del mercado, respectivamente.

FIGURA 3.1



Traducido en expresiones matemáticas, el modelo se puede escribir como

$$\begin{aligned} Q_d &= Q_s \\ Q_d &= a - bP \qquad (a, b > 0) \\ Q_s &= -c + dP \qquad (c, d > 0) \end{aligned} \tag{3.1}$$

Aparecen en las dos funciones lineales cuatro parámetros, a, b, c y d, y todos se especifican como positivos. Cuando se grafica la función de demanda, como en la figura 3.1, su intersección vertical está en a y su pendiente es $-b$, que, como se requiere, es negativa. La función de oferta también tiene el tipo de pendiente requerida, d positiva, pero su intersección vertical es negativa, en $-c$. ¿Por qué se deseó especificar tal intersección vertical negativa? La respuesta es que de esta manera se fuerza a la curva de oferta a tener una intersección horizontal positiva en P_1, y, por lo tanto, se satisface la condición de que el suministro no llegará a menos que el precio sea positivo y suficientemente alto.

El lector debe observar que, en contra de la práctica usual, la cantidad y no el precio se graficó en la figura 3.1 sobre la vertical. Sin embargo, esto concuerda con la convención matemática de colocar la variable *dependiente* en el eje vertical. En un contexto diferente, en el que la curva de demanda se considera desde el punto de vista de una empresa comercial como descripción de la curva de ingreso promedio, IP $\equiv P = f(Q_d)$, se invertirán los ejes y se graficará a P sobre el eje vertical.

Con el modelo así construido, el siguiente paso es resolverlo, es decir, obtener los valores solución de las tres variables endógenas, Q_d, Q_s y P. Los valores solución son los que satisfacen de forma simultánea las tres ecuaciones en (3.1); es decir, son los valores que, cuando se sustituyen en las tres ecuaciones, las convierten en un conjunto de expresiones verdaderas. En el contexto de un modelo de equilibrio, esos valores también se pueden llamar *valores de equilibrio* de las variables mencionadas.

Muchos autores no emplean símbolos especiales para denotar los valores solución de las variables endógenas. Así, Q_d se usa para representar ya sea la variable de cantidad demandada (con una gama completa de valores) o su valor solución (un valor específico); sucede de manera similar con los símbolos Q_s y P. Desafortunadamente, esta práctica puede dar lugar a posibles confusiones; en particular, en el contexto del análisis estático comparativo (por ejemplo, sección 7.5). Para evitar tal fuente de confusión, se denota el valor solución de una variable endógena con un asterisco. Así, los valores solución de Q_d, Q_s y P, se denotan mediante Q_d^*, Q_s^* y P^*, respectivamente. Sin embargo, puesto que $Q_d^* = Q_s^*$, se pueden reemplazar incluso mediante un solo símbolo Q^*. Por consiguiente, una solución de equilibrio del modelo se puede denotar simplemente mediante un par ordenado (P^*, Q^*). En caso de que la solución no sea única, varios pares ordenados pueden satisfacer el sistema de ecuaciones simultáneas; entonces habrá un conjunto solución con más de un elemento en él. No obstante, la situación de equilibrio múltiple no surge en un modelo lineal como el presente.

Solución mediante eliminación de variables

Una forma de hallar una solución para un sistema de ecuaciones es mediante la eliminación sucesiva de variables y ecuaciones por sustitución. En (3.1), el modelo contiene tres ecuaciones con tres variables. Sin embargo, en vista de la igualdad de Q_d y Q_s mediante la condición de equilibrio, se puede hacer que $Q = Q_d = Q_s$ y reescribir el modelo, en forma equivalente, como sigue:

$$\begin{aligned} Q &= a - bP \\ Q &= -c + dP \end{aligned} \qquad \textbf{(3.2)}$$

De este modo se reduce el modelo a dos ecuaciones con dos variables. Además, al sustituir la primera ecuación en la segunda (3.2), el modelo se reduce a una sola ecuación con una variable:

$$a - bP = -c + dP$$

O bien, después de restar $(a + dP)$ de ambos lados de la ecuación y multiplica por -1,

$$(b + d)P = a + c \qquad \textbf{(3.3)}$$

Este resultado también se obtiene directamente de (3.1) al sustituir las ecuaciones segunda y tercera en la primera.

Puesto que $b + d \neq 0$, es válido dividir ambos lados de (3.3) entre $(b + d)$. El resultado es el valor solución de P:

$$P^* = \frac{a + c}{b + d} \qquad \textbf{(3.4)}$$

Observe que P^* se expresa —como todos los valores solución— en términos de los parámetros que representan los datos del modelo. Así, P^* es un valor definido, como debe ser. Asimismo, note que P^* es positivo —como debe ser un precio— debido a que los cuatro parámetros son positivos por especificación de modelo.

Para hallar la cantidad de equilibrio Q^* $(= Q_d^* = Q_s^*)$ que corresponde al valor P^*, se sustituye (3.4) en *cualquier* ecuación de (3.2), y luego se soluciona la ecuación resultante. Al sustituir (3.4) en la función de demanda, por ejemplo, se obtiene

$$Q^* = a - \frac{b(a + c)}{b + d} = \frac{a(b + d) - b(a + c)}{b + d} = \frac{ad - bc}{b + d} \qquad \textbf{(3.5)}$$

que de nuevo es una expresión sólo en términos de parámetros. Puesto que el denominador $(b + d)$ es positivo, la positividad de Q^* requiere que el número $(ad - bc)$ sea positivo también. Por consiguiente, a fin de tener un significado económico, el presente modelo debe contener la restricción adicional de que $ad > bc$.

El significado de esta restricción se puede ver en la figura 3.1. Es bien sabido que la P^* y la Q^* de un modelo de mercado se puede determinar de manera gráfica en la intersección de la curva de demanda y la curva de oferta. Para tener $Q^* > 0$ se requiere que el punto de intersección esté localizado arriba del eje horizontal en la figura 3.1, lo cual a su vez requiere que las pendientes e intersecciones verticales de las dos curvas satisfagan cierta restricción en sus magnitudes relativas. Esa restricción, según (3.5), es $ad > bc$, dado que tanto b como d son positivas.

Por cierto, la intersección de las curvas de la oferta y la demanda en la figura 3.1 no es diferente en concepto de la intersección mostrada en el diagrama de Venn de la figura 2.2*b*. Sólo hay una diferencia: en vez de que los puntos estén dentro de dos círculos, el presente caso tiene que ver con los puntos que están sobre dos líneas. Denótese el conjunto de puntos sobre las curvas de la oferta y la demanda por D y S, respectivamente. Entonces, al utilizar el símbolo $Q\ (= Q_d = Q_s)$, los dos conjuntos y su intersección se escriben

$$D = \{(P, Q) \mid Q = a - bP\}$$

$$S = \{(P, Q) \mid Q = -c + dP\}$$

y

$$D \cap S = (P^*, Q^*)$$

El conjunto intersección contiene en este caso sólo un elemento, el par ordenado (P^*, Q^*). El equilibrio de mercado es único.

EJERCICIO 3.2

1. Dado el modelo de mercado

   $Q_d = Q_s$

   $Q_d = 21 - 3P$

   $Q_s = -4 + 8P$

   obtenga P^* y Q^* por (*a*) eliminación de variables y (*b*) por medio de las fórmulas (3.4) y (3.5). (Use fracciones en vez de decimales.)
2. Sean las funciones de la oferta y la demanda como sigue:

   (*a*) $Q_d = 51 - 3P$ $\quad$ (*b*) $Q_d = 30 - 2P$

   $\quad\ Q_s = 6P - 10$ $\quad\quad\ Q_s = -6 + 5P$

   determine P^* y Q^* mediante eliminación de variables. (Use fracciones en vez de decimales.)
3. Según la ecuación (3.5), para que Q^* sea positiva, es necesario que la expresión $(ad - bc)$ tenga el mismo signo algebraico que $(b + d)$. Compruebe que esta condición se satisface en realidad en los modelos de los problemas 1 y 2.
4. Si $(b + d) = 0$ en el modelo de mercado lineal, ¿se puede encontrar una solución de equilibrio al usar (3.4) y (3.5)? ¿Por qué?
5. Si $(b + d) = 0$ en el modelo de mercado lineal, ¿qué se puede concluir en relación con las posiciones de las curvas de demanda y equilibrio en la figura 3.1? ¿Qué concluye entonces con respecto a la solución de equilibrio?

3.3 Equilibrio de mercado parcial: un modelo no lineal

Supóngase que la demanda lineal en el modelo de mercado aislado se sustituye por una función de demanda cuadrática, mientras que la función de oferta sigue siendo lineal. También, utilícense coeficientes numéricos en vez de parámetros. Entonces podría surgir un modelo como el siguiente:

$$\begin{aligned} Q_d &= Q_s \\ Q_d &= 4 - P^2 \\ Q_s &= 4P - 1 \end{aligned} \tag{3.6}$$

Como antes, este sistema de tres ecuaciones de puede reducir a una sola ecuación mediante eliminación de variables (por sustitución):

$$4 - P^2 = 4P - 1$$

o bien,

$$P^2 + 4P - 5 = 0 \tag{3.7}$$

Ésta es una ecuación cuadrática porque la expresión de la izquierda es una función cuadrática de la variable P. Una diferencia importante entre una ecuación cuadrática y una lineal es que, en general, la primera produce dos valores solución.

Ecuación cuadrática *versus* función cuadrática

Antes de analizar el método de solución, se debe hacer una distinción clara entre los dos términos *ecuación cuadrática* y *función cuadrática*. De acuerdo con la explicación anterior, la expresión $P^2 + 4P - 5$ constituye una *función* cuadrática, por ejemplo, $f(P)$. Por lo tanto, podemos escribir

$$f(P) = P^2 + 4P - 5 \tag{3.8}$$

Lo que hace (3.8) es especificar una regla de asignación de P a $f(P)$, tal como

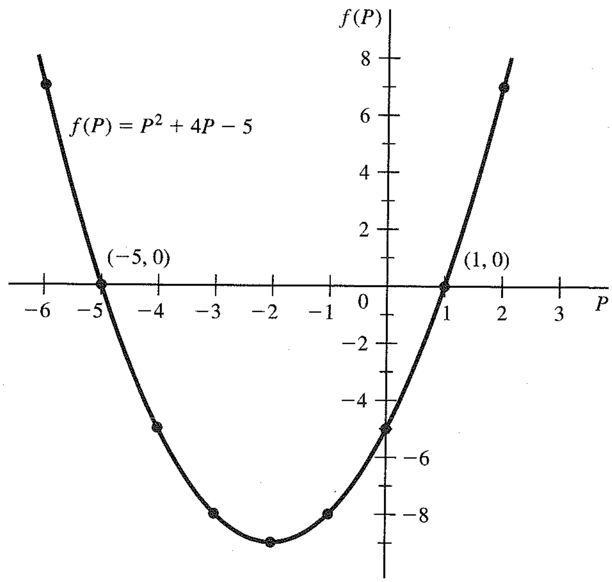
P	...	−6	−5	−4	−3	−2	−1	0	1	2	...
$f(P)$	...	7	0	−5	−8	−9	−8	−5	0	7	...

Aunque se han listado sólo nueve valores de P en esta tabla, en realidad *todos* los valores de P en el dominio de la función son elegibles para listar. Es quizá por esta razón que rara vez se habla de "resolver" la ecuación $f(P) = P^2 + 4P - 5$, porque normalmente se espera que los "valores solución" sean pocos, pero aquí pueden intervenir todos los valores. No obstante, se podría considerar de modo legítimo cada par ordenado de la tabla, por ejemplo $(-6, 7)$ y $(-5, 0)$, como solución de (3.8), puesto que cada par ordenado satisface esa ecuación. En vista de que se puede escribir una cantidad infinita de estos pares ordenados, uno para cada valor de P, hay un número infinito de soluciones para (3.8). Cuando se grafican como una curva, estos pares ordenados producen la parábola de la figura 3.2.

En (3.7), donde se igualó a cero la función cuadrática $f(P)$, la situación cambió de forma fundamental. Puesto que ahora desaparece la variable $f(P)$ (por habérsele asignado el valor cero), el resultado es una *ecuación* cuadrática con la única variable P.[2] Ahora que $f(P)$ está

[2] La distinción entre la función cuadrática y la ecuación cuadrática recién analizada también se puede extender a casos de polinomiales distintos de los cuadráticos. Así, una ecuación cúbica resulta cuando una función cúbica es distinta de cero.

FIGURA 3.2



restringida al valor cero, sólo una cantidad selecta de valores P puede satisfacer (3.7) y calificar como valores solución; a saber, aquellos valores de P en los que la parábola de la figura 3.2 cruza el eje horizontal, en los que $f(P)$ es cero. Note que esta vez los valores solución son sólo valores de P, no pares ordenados. Los valores solución P se denominan comúnmente las *raíces* de la ecuación *cuadrática* $f(P) = 0$, o bien, los *ceros* de la *función* cuadrática $f(P)$.

Hay dos puntos de intersección en la figura 3.2, a saber, $(1, 0)$ y $(-5, 0)$. Como se requiere, el segundo elemento de cada uno de estos pares ordenados (la *ordenada* del punto correspondiente) resulta $f(P) = 0$ en ambos casos. Por otro lado, el primer elemento de cada par ordenado (la *abscisa* del punto) da el valor solución de P. Aquí se obtienen dos soluciones,

$$P_1^* = 1 \qquad \text{y} \qquad P_2^* = -5$$

pero, ya que se descartan precios negativos, sólo el primer elemento es admisible desde el punto de vista económico.

Fórmula cuadrática

La ecuación (3.7) se ha resuelto en forma gráfica, pero también se puede resolver mediante un método algebraico. En general, dada una ecuación cuadrática de la forma

$$ax^2 + bx + c = 0 \qquad (a \neq 0) \tag{3.9}$$

hay dos raíces que se pueden obtener de la *fórmula cuadrática*:

$$x_1^*, x_2^* = \frac{-b \pm (b^2 - 4ac)^{1/2}}{2a} \tag{3.10}$$

donde la parte + del signo ± produce x_1^* y la parte − produce x_2^*.

Observe también que mientras que $b^2 - 4ac > 0$, diferirían los valores de x_1^* y x_2^*, de modo que se obtienen dos números reales distintos como raíces. Pero en el caso especial donde

$b^2 - 4ac = 0$, se encontraría que $x_1^* = x_2^* = -b/2a$. En este caso, las dos raíces comparten el mismo valor; éstas se conocen como *raíces repetidas*. En el otro caso especial, donde $b^2 - 4ac < 0$, se tendría la tarea de sacar la raíz cuadrada de un número negativo, lo cual no es posible en el sistema de números reales. En este último caso, no existen raíces de valores reales. Este asunto se analiza después en la sección 16.1.

Esta fórmula utilizada ampliamente se deduce por medio de un proceso conocido como "completar el cuadrado". Primero, al dividir cada término de (3.9) entre a se obtiene la ecuación

$$x^2 + \frac{b}{a}x + \frac{c}{a} = 0$$

Al restar c/a y sumar $b^2/4a^2$ a ambos lados de la ecuación, se obtiene

$$x^2 + \frac{b}{a}x + \frac{b^2}{4a^2} = \frac{b^2}{4a^2} - \frac{c}{a}$$

El lado izquierdo ahora es un "cuadrado perfecto" y, por lo tanto, la ecuación se puede expresar como

$$\left(x + \frac{b}{2a}\right)^2 = \frac{b^2 - 4ac}{4a^2}$$

o bien, después de sacar la raíz cuadrada en ambos lados,

$$x + \frac{b}{2a} = \pm\frac{(b^2 - 4ac)^{1/2}}{2a}$$

Por último, al restar $b/2a$ de ambos lados, se obtiene el resultado en (3.10).

La aplicación de la fórmula a (3.7), donde $a = 1$, $b = 4$, $c = -5$ y $x = P$, produce las raíces

$$P_1^*, P_2^* = \frac{-4 \pm (16 + 20)^{1/2}}{2} = \frac{-4 \pm 6}{2} = 1, -5$$

que coinciden con las soluciones gráficas de la figura 3.2. De nuevo, se rechaza $P_2^* = -5$ sobre bases económicas y, después de omitir el subíndice 1, se escribe simplemente $P^* = 1$.

Con esta información a la mano, la cantidad de equilibrio Q^* se determina sin dificultad a partir de la ecuación segunda o tercera de (3.6) como $Q^* = 3$.

Otra solución gráfica

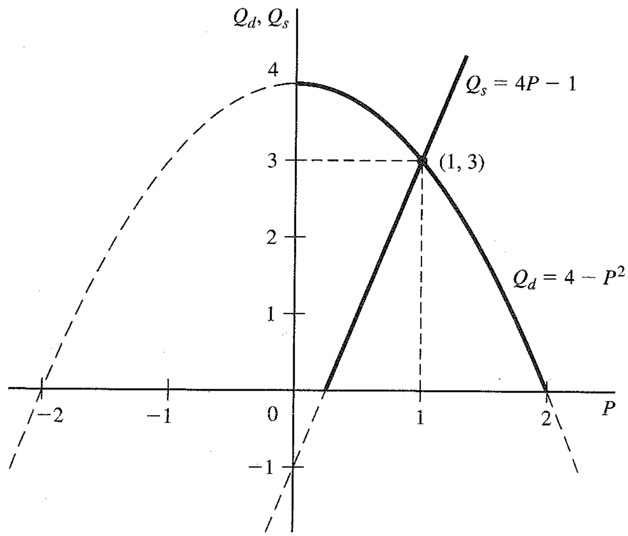
En la figura 3.2 se presenta un método de solución gráfica del presente modelo. Sin embargo, puesto que la cantidad variable se eliminó al deducir la ecuación cuadrática, sólo se puede encontrar P^* en esa figura. Si se tiene interés en determinar al mismo tiempo P^* y Q^* de la gráfica, se debe usar un diagrama con Q en un eje y P en el otro, similar en construcción a la figura 3.1. Esto se ilustra en la figura 3.3. Por supuesto que de nuevo el problema es encontrar la intersección de dos conjuntos de puntos, a saber,

$$D = \{(P, Q) \mid Q = 4 - P^2\}$$

y

$$S = \{(P, Q) \mid Q = 4P - 1\}$$

FIGURA 3.3



Si no se pone restricción en el dominio y la imagen, el conjunto de intersección contendrá dos elementos, a saber,

$$D \cap S = \{(1, 3), (-5, -21)\}$$

El primero se localiza en el cuadrante I y el último en el cuadrante III (no dibujado). Si el dominio y la imagen se restringen a valores positivos, sólo se puede aceptar el primer par ordenado (1, 3). Entonces el equilibrio de nuevo es único.

Ecuaciones polinomiales de grado superior

Si un sistema de ecuaciones simultáneas no se reduce a una ecuación lineal como (3.3)[3] o a una ecuación cuadrática como (3.7), sino a una ecuación cúbica (polinomio de tercer grado) o a una cuártica (polinomio de cuarto grado), será más difícil hallar las raíces. Un método útil que podría funcionar es el de *factorizar* la función.

Ejemplo 1

La expresión $x^3 - x^2 - 4x + 4$ se puede escribir como el producto de tres factores $(x - 1)$, $(x + 2)$ y $(x - 2)$. Así, la ecuación cúbica

$$x^3 - x^2 - 4x + 4 = 0$$

se puede escribir después de factorizar como

$$(x - 1)(x + 2)(x - 2) = 0$$

Pag. 173 Algebra

A fin de que el producto del lado izquierdo sea cero, al menos uno de los tres términos del producto debe ser cero. Al igualar a cero cada término, se obtiene

$$x - 1 = 0, \qquad \text{o bien,} \qquad x + 2 = 0, \qquad \text{o bien,} \qquad x - 2 = 0$$

Estas tres ecuaciones suministrarán las tres raíces de la ecuación cúbica, a saber,

$$x_1^* = 1 \qquad x_2^* = -2 \qquad \text{y} \qquad x_3^* = 2$$

[3] La ecuación (3.3) se puede considerar como el resultado de igualar a cero la función $(b + d)P - (a + c)$.

En el ejemplo 1 se ilustran dos hechos interesantes y útiles acerca de la factorización. Primero, dada una ecuación polinomial de tercer grado, la factorización da como resultado tres términos de la forma (x − raíz), y de este modo se obtienen tres raíces. Por lo común, una ecuación polinomial de grado n debe producir un total de n raíces. Segundo, y más importante para el propósito de la investigación de raíces, se nota la siguiente relación entre las tres raíces $(1, -2, 2)$ y el término constante 4: puesto que el término constante debe ser el producto de las tres raíces, cada raíz debe ser un divisor del término constante. Esta relación se puede formalizar en el siguiente teorema:

Teorema I Dada la ecuación polinomial

$$x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0$$

donde los coeficientes son enteros y el coeficiente de x^n es la unidad, si existen raíces enteras, entonces cada una de ellas debe ser un divisor de a_0.

Sin embargo, a veces se encuentran fracciones como coeficientes de la ecuación polinomial, como en

$$x^4 + \tfrac{5}{2}x^3 - \tfrac{11}{2}x^2 - 10x + 6 = 0$$

y éstas no forman parte de la hipótesis del teorema I. Aun cuando se multiplique por 2 para eliminar las fracciones (terminando en la forma mostrada en el ejemplo 2 que sigue), todavía no se puede aplicar el teorema I, porque el coeficiente del término de mayor grado no es la unidad. En tales casos, se puede recurrir a un teorema más general:

Teorema II Dada la ecuación polinomial con coeficientes enteros

$$a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0$$

si existe una raíz racional r/s, donde r y s son enteros sin un divisor común diferente de la unidad, entonces r es un divisor de a_0 y s es un divisor de a_n.

Ejemplo 2

¿Tiene la ecuación cuártica

$$2x^4 + 5x^3 - 11x^2 - 20x + 12 = 0$$

raíces racionales? Con $a_0 = 12$, los únicos valores posibles para el numerador r en r/s son el conjunto de divisores $\{1, -1, 2, -2, 3, -3, 4, -4, 6, -6, 12, -12\}$. Y, con $a_n = 2$, los únicos valores posibles para s son el conjunto de divisores $\{1, -1, 2, -2\}$. Tomando a su vez cada elemento del conjunto r, y dividiéndolo entre cada elemento del conjunto s, respectivamente, se encuentra que r/s sólo puede tomar los valores

$$1, -1, \frac{1}{2}, -\frac{1}{2}, 2, -2, 3, -3, \frac{3}{2}, -\frac{3}{2}, 4, -4, 6, -6, 12, -12$$

Entre estos candidatos para las raíces, muchos no satisfacen la ecuación dada. Por ejemplo, al fijar $x = 1$ en la ecuación cuártica se obtiene el resultado absurdo $-12 = 0$. De hecho, puesto que se está resolviendo una ecuación cuártica, se puede esperar que a lo sumo cuatro de los valores r/s listados califiquen como raíces. Los cuatro candidatos exitosos resultan ser $\frac{1}{2}$, 2, −2 y −3. Según el principio de factorización, la ecuación cuártica dada se puede escribir en forma equivalente como

$$(x - \tfrac{1}{2})(x - 2)(x + 2)(x + 3) = 0$$

donde el primer factor también se puede escribir como $(2x - 1)$.

En el ejemplo 2 se rechaza la raíz posible 1 porque $x = 1$ no satisface la ecuación dada; es decir, la sustitución de $x = 1$ en la ecuación no produce la identidad $0 = 0$ como se requiere. Ahora considere el caso donde $x = 1$ es una raíz de alguna ecuación polinomial. En ese caso, puesto que $x^n = x^{n-1} = \cdots = x = 1$, la ecuación polinomial se reduciría a la forma simple $a_n + a_{n-1} + \cdots + a_1 + a_0 = 0$. Este hecho proporciona el fundamento para el siguiente teorema:

Teorema III Dada la ecuación polinomial

$$a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0$$

si la suma de los coeficientes $a_n, a_{n-1}, \ldots, a_0$ es cero, entonces $x = 1$ es una raíz de la ecuación.

EJERCICIO 3.3

1. Determine en forma gráfica los ceros de las siguientes funciones:
   (*a*) $f(x) = x^2 - 8x + 15$ (*b*) $g(x) = 2x^2 - 4x - 16$
2. Resuelva el problema 1 mediante la fórmula cuadrática.
3. (*a*) Encuentre una ecuación cúbica con raíces 6, −1 y 3.
   (*b*) Obtenga una ecuación cuártica con raíces 1, 2, 3 y 5.
4. Para cada una de las siguientes ecuaciones polinomiales, determine si $x = 1$ es una raíz.
   (*a*) $x^3 - 2x^2 + 3x - 2 = 0$ (*c*) $3x^4 - x^2 + 2x - 4 = 0$
   (*b*) $2x^3 - \frac{1}{2}x^2 + x - 2 = 0$
5. Halle las raíces racionales, si existen, de las siguientes ecuaciones.
   (*a*) $x^3 - 4x^2 + x + 6 = 0$ (*c*) $x^3 + \frac{3}{4}x^2 - \frac{3}{8}x - \frac{1}{8} = 0$
   (*b*) $8x^3 + 6x^2 - 3x - 1 = 0$ (*d*) $x^4 - 6x^3 + 7\frac{3}{4}x^2 - \frac{3}{2}x - 2 = 0$
6. Obtenga la solución de equilibrio para cada uno de los siguientes modelos:
   (*a*) $Q_d = Q_s$ (*b*) $Q_d = Q_s$
   $Q_d = 3 - P^2$ $Q_d = 8 - P^2$
   $Q_s = 6P - 4$ $Q_s = P^2 - 2$
7. La condición de equilibrio de mercado, $Q_d = Q_s$, suele expresarse en una forma alternativa equivalente, $Q_d - Q_s = 0$, que tiene la interpretación económica "la demanda excedente es cero". ¿Representa la ecuación (3.7) esta última versión de la condición de equilibrio? Si no, provea una interpretación económica apropiada para (3.7).

3.4 Equilibrio general de mercado

Las dos últimas secciones tratan con modelos de un mercado aislado, en donde Q_d y Q_s son funciones del precio de un producto solamente. Sin embargo, en realidad ningún producto goza (o experimenta) nunca de tan solitaria existencia; para cualquier producto, normalmente existen muchos sustitutos y bienes complementarios. Así, una ilustración más real de la función de demanda de un producto también debe tomar en cuenta el efecto, no sólo del precio del producto, sino de los precios de productos relacionados. Lo mismo se cumple para la función de oferta. Sin embargo, una vez que se consideran los precios de los otros artículos o productos, la estructura del modelo en sí se debe ampliar para que pueda producir también los valores de equilibrio de estos otros precios. Como resultado, las variables de precio y cantidad de múltiples productos deben entrar juntos de forma endógena en el modelo.

En un modelo de mercado aislado, la condición de equilibrio consiste sólo en una ecuación, $Q_d = Q_s$, o bien, $E \equiv Q_d - Q_s = 0$, donde E representa la demanda excedente. Cuando se consideran al mismo tiempo varios artículos interdependientes, el equilibrio requeriría la ausencia de demanda excedente para cada artículo incluido en el modelo porque, si por lo menos *un* artículo se enfrenta a un exceso de demanda, el ajuste de precio de ese artículo afectará necesariamente las cantidades demandadas y ofrecidas de los artículos relacionados, lo que causaría cambios en general. En consecuencia, la condición de equilibrio de un modelo de mercado de n artículos requerirá n ecuaciones, una para cada artículo, en la forma

$$E_i \equiv Q_{di} - Q_{si} = 0 \qquad (i = 1, 2, \ldots, n) \tag{3.11}$$

Si existe una solución, habrá un conjunto de precios P_i^* y cantidades correspondientes Q_i^* de modo que las n ecuaciones en la condición de equilibrio se satisfagan simultáneamente.

Modelo de mercado de dos artículos

Para ilustrar el problema, procedamos a analizar un modelo simple en el que sólo dos artículos se relacionan entre sí. Para simplificar, se supone que las funciones de la oferta y la demanda de ambos artículos son lineales. En términos paramétricos, este modelo se puede escribir como

$$\begin{aligned}
Q_{d1} - Q_{s1} &= 0 \\
Q_{d1} &= a_0 + a_1 P_1 + a_2 P_2 \\
Q_{s1} &= b_0 + b_1 P_1 + b_2 P_2 \\
Q_{d2} - Q_{s2} &= 0 \\
Q_{d2} &= \alpha_0 + \alpha_1 P_1 + \alpha_2 P_2 \\
Q_{s2} &= \beta_0 + \beta_1 P_1 + \beta_2 P_2
\end{aligned} \tag{3.12}$$

donde los coeficientes a y b pertenecen a las funciones de la oferta y la demanda del primer artículo, y los coeficientes α y β a los del segundo. Se ha dejado de lado la especificación de los signos de los coeficientes, pero, en el curso del análisis, ciertas restricciones surgirán como un prerrequisito para obtener resultados económicamente razonables. Asimismo, en un ejemplo numérico posterior se harán algunos comentarios acerca de los signos específicos que se darán a los coeficientes.

Como un primer paso hacia la solución de este modelo, se puede recurrir de nuevo a la eliminación de variables. Al sustituir las ecuaciones segunda y tercera en la primera (para el primer artículo) y la quinta y sexta ecuaciones en la cuarta (para el segundo artículo), el modelo se reduce a dos ecuaciones con dos variables:

$$\begin{aligned}
(a_0 - b_0) + (a_1 - b_1)P_1 + (a_2 - b_2)P_2 &= 0 \\
(\alpha_0 - \beta_0) + (\alpha_1 - \beta_1)P_1 + (\alpha_2 - \beta_2)P_2 &= 0
\end{aligned} \tag{3.13}$$

Éstas representan la versión de dos artículos de (3.11), después que las funciones de la oferta y la demanda se sustituyen en las dos condiciones de equilibrio.

Aunque éste es un sistema simple de sólo dos ecuaciones, se requieren 12 parámetros y las manipulaciones algebraicas resultarán difíciles de manejar a menos que se introduzca algún tipo de abreviatura. Así, se definen los símbolos abreviados:

$$\left.\begin{aligned} c_i &\equiv a_i - b_i \\ \gamma_i &\equiv \alpha_i - \beta_i \end{aligned}\right. \qquad (i = 0, 1, 2)$$

Entonces, después de pasar los términos c_0 y γ_0 al lado derecho, se obtiene

$$\begin{aligned} c_1 P_1 + c_2 P_2 &= -c_0 \\ \gamma_1 P_1 + \gamma_2 P_2 &= -\gamma_0 \end{aligned} \tag{3.13'}$$

que se puede resolver mediante eliminación de variables. De la primera ecuación, se encuentra que $P_2 = -(c_0 + c_1 P_1)/c_2$. Al sustituir esto en la segunda ecuación y resolver, se obtiene

$$P_1^* = \frac{c_2\gamma_0 - c_0\gamma_2}{c_1\gamma_2 - c_2\gamma_1} \tag{3.14}$$

Note que P_1^* se expresa por completo —como debe ser un valor solución— en términos de los datos (parámetros) del modelo. Mediante un proceso similar, se encuentra que el precio de equilibrio del segundo artículo es

$$P_2^* = \frac{c_0\gamma_1 - c_1\gamma_0}{c_1\gamma_2 - c_2\gamma_1} \tag{3.15}$$

Sin embargo, para que estos dos valores tengan sentido es necesario imponer ciertas restricciones al modelo. Primero, puesto que la división entre cero no está definida, se requiere que el denominador común de (3.14) y (3.15) sea distinto de cero, es decir, $c_1\gamma_2 \neq c_2\gamma_1$. Segundo, para asegurar que la solución sea positiva, el numerador debe tener el mismo signo que el denominador.

Una vez obtenidos los precios de equilibrio, las cantidades de equilibrio Q_1^* y Q_2^* se calculan fácilmente al sustituir (3.14) y (3.15) en la segunda (o tercera) ecuación y la quinta (o sexta) ecuación de (3.12). Estos valores de solución también son expresados de forma natural en términos de los parámetros. (Su cálculo se deja como ejercicio.)

Ejemplo numérico

Supóngase que las funciones de la oferta y la demanda son numéricamente como sigue:

$$\begin{aligned} Q_{d1} &= 10 - 2P_1 + P_2 \\ Q_{s1} &= -2 + 3P_1 \\ Q_{d2} &= 15 + P_1 - P_2 \\ Q_{s2} &= -1 \qquad\quad + 2P_2 \end{aligned} \tag{3.16}$$

¿Cuál es la solución de equilibrio?

Antes de contestar la pregunta, se consideran los coeficientes numéricos. Para cada artículo, se ve que Q_{si} depende solamente de P_i, pero Q_{di} se muestra como una función de ambos precios. Nótese que mientras P_1 tiene un coeficiente negativo en Q_{d1}, como se esperaría, el coeficiente de P_2 es positivo. El hecho de que un aumento en P_2 tienda a aumentar a Q_{d1} hace pensar que los dos artículos son sustitutos entre sí. El papel de P_1 en la función Q_{d2} tiene una interpretación similar.

Con estos coeficientes, los símbolos abreviados c_i y γ_i tomarán los siguientes valores:

$$\begin{array}{lll} c_0 = 10 - (-2) = 12 & c_1 = -2 - 3 = -5 & c_2 = 1 - 0 = 1 \\ \gamma_0 = 15 - (-1) = 16 & \gamma_1 = 1 - 0 = 1 & \gamma_2 = -1 - 2 = -3 \end{array}$$

Por sustitución directa de éstos en (3.14) y (3.15), se obtiene

$$P_1^* = \tfrac{52}{14} = 3\tfrac{5}{7} \qquad \text{y} \qquad P_2^* = \tfrac{92}{14} = 6\tfrac{4}{7}$$

Y la sustitución subsecuente de P_1^* y P_2^* en (3.16) produce

$$Q_1^* = \tfrac{64}{7} = 9\tfrac{1}{7} \qquad \text{y} \qquad Q_2^* = \tfrac{85}{7} = 12\tfrac{1}{7}$$

Así, los valores de equilibrio resultan positivos, como se requiere. A fin de conservar los valores exactos de P_1^* y P_2^* en el siguiente cálculo de Q_1^* y Q_2^*, se recomienda expresarlos como fracciones en vez de decimales.

¿Podemos obtener la gráfica del precio de equilibrio? Sí, de (3.13), es claro que un modelo de dos artículos se puede resumir mediante dos ecuaciones con dos variables P_1 y P_2. Ambas ecuaciones se pueden graficar en el plano coordenado $P_1 P_2$, si se conocen los coeficientes numéricos y, entonces, la intersección de las dos curvas se puede ubicar con precisión en P_1^* y P_2^*.

Caso de *n* artículos

El análisis anterior del mercado de múltiples artículos ha sido limitado al caso de dos artículos, pero debe ser evidente que se avanza del análisis de *equilibrio parcial* en la dirección del análisis de *equilibrio general*. A medida que entran más artículos al modelo, habrá más variables y más ecuaciones, y las ecuaciones se volverán más grandes y complicadas. Si todos los artículos de una economía se incluyen en un modelo de mercado integral, el resultado será un modelo de equilibrio general de tipo walrasiano, en el cual el exceso de demanda para cada artículo se considera una función de los precios de todos los artículos en la economía.

Por supuesto, algunos de los precios pueden llevar coeficientes cero cuando no tienen que ver en la determinación del exceso de demanda de un determinado artículo; por ejemplo, en la función de exceso de demanda de pianos, el precio de las palomitas de maíz podría tener un coeficiente cero. Sin embargo, en general, con n artículos en total, se pueden expresar las funciones de la oferta y la demanda como sigue (con Q_{di} y Q_{si} como símbolos de función, en lugar de f y g):

$$\begin{aligned} Q_{di} &= Q_{di}(P_1, P_2, \ldots, P_n) \\ Q_{si} &= Q_{si}(P_1, P_2, \ldots, P_n) \end{aligned} \qquad (i = 1, 2, \ldots, n) \tag{3.17}$$

En vista de los subíndices, estas dos ecuaciones representan la totalidad de las $2n$ funciones que contiene el modelo. (Estas funciones no necesariamente son lineales.) Además, la condición de equilibrio está compuesta de un conjunto de n ecuaciones,

$$Q_{di} - Q_{si} = 0 \qquad (i = 1, 2, \ldots, n) \tag{3.18}$$

Cuando se agrega la ecuación (3.18) a (3.17), se completa el modelo. Así, se debe contar un total de $3n$ ecuaciones.

Al sustituir (3.17) en (3.18), el modelo se puede reducir a un solo conjunto de n ecuaciones simultáneas:

$$Q_{di}(P_1, P_2, \ldots, P_n) - Q_{si}(P_1, P_2, \ldots, P_n) = 0 \qquad (i = 1, 2, \ldots, n)$$

Además, dado que $E_i \equiv Q_{di} - Q_{si}$, donde E_i es necesariamente también una función de los n precios, el último conjunto de ecuaciones de puede escribir como

$$E_i(P_1, P_2, \ldots, P_n) = 0 \qquad (i = 1, 2, \ldots, n)$$

Si en realidad existe una solución, al resolver de forma simultánea estas n ecuaciones se determinan los n precios de equilibrio P_i^*. Luego, Q_i^* se puede deducir de las funciones de la oferta y la demanda.

Solución de un sistema general de ecuaciones

Si un modelo está provisto de coeficientes numéricos, como en (3.16), los valores de equilibrio de las variables serán también en términos numéricos. En un nivel más general, si un modelo se expresa en términos de constantes paramétricas, como en (3.12), los valores de equilibrio también tendrán que ver con parámetros y, por lo tanto, aparecerán como "fórmulas", como se ejemplifica en (3.14) y (3.15). Sin embargo, si incluso las formas de función se dejan sin especificar en un modelo para generalizar, como en (3.17), la manera de expresar los valores solución será también por necesidad sumamente general.

Con base en la experiencia adquirida en modelos paramétricos, se sabe que un valor solución siempre es una expresión en términos de los parámetros. Para un modelo de función general que contiene, por ejemplo, un total de m parámetros $(a_1, a_2, \ldots, a_m)$, donde m no necesariamente es igual a n, se puede esperar que los n precios de equilibrio tomen la forma analítica general de

$$P_i^* = P_i^*(a_1, a_2, \ldots, a_m) \qquad (i = 1, 2, \ldots, n) \tag{3.19}$$

Ésta es una expresión simbólica para que el valor solución de *cada* variable (en este caso el precio) sea una función del conjunto de todos los parámetros de modelo. Como ésta es una expresión muy general, en realidad no da mucha información detallada acerca de la solución. Pero, como se verá en el capítulo 8, en el tratamiento analítico general de algunos tipos de problemas, incluso esta forma en apariencia poco informativa de expresar una solución, probará ser útil.

Escribir tal solución es una tarea fácil. Pero existe un problema importante: la expresión en (3.19) se puede justificar si y sólo si en realidad existe una solución *única*, porque entonces, y sólo entonces, se puede asignar la m-tupla ordenada $(a_1, a_2, \ldots, a_m)$ a un valor determinado para cada precio P_i^*. Sin embargo, una desventaja es que no hay una razón *a priori* para presumir que todo modelo producirá de forma automática una solución única. En este sentido, es necesario remarcar que el proceso de "contar ecuaciones e incógnitas" no es suficiente como argumento. Algunos ejemplos muy simples deben servir para tener la certeza de que un número igual de ecuaciones e incógnitas (variables endógenas) no garantiza necesariamente la existencia de una solución única.

Considere los tres sistemas de ecuaciones simultáneas

$$\begin{aligned} x + y &= 8 \\ x + y &= 9 \end{aligned} \tag{3.20}$$

$$\begin{aligned} 2x + y &= 12 \\ 4x + 2y &= 24 \end{aligned} \tag{3.21}$$

$$\begin{aligned} 2x + 3y &= 58 \\ y &= 18 \\ x + y &= 20 \end{aligned} \tag{3.22}$$

En (3.20), a pesar de que dos incógnitas están relacionadas por exactamente dos ecuaciones, no hay solución. Sucede que estas dos ecuaciones son *inconsistentes*, porque si la suma de x y y es 8, no puede ser al mismo tiempo 9. En (3.21), otro caso de dos ecuaciones con dos variables, las dos ecuaciones son *funcionalmente dependientes*, lo que significa que una se obtiene (es consecuencia) de la otra. (Aquí, la segunda ecuación es igual a dos veces la primera.) Por lo tanto, una ecuación es redundante y se podría eliminar del sistema, dejando vigente sólo una ecuación con dos incógnitas. La solución será entonces la ecuación $y = 12 - 2x$, que no produce un par ordenado (x^*, y^*) sino un número infinito de ellos, entre los que están (0, 12), (1, 10), (2, 8), etc., todos los cuales satisfacen dicha ecuación. Por último, el caso de (3.22) tiene más ecuaciones que incógnitas, sin embargo, el par ordenado (2, 18) constituye una solución única. La razón es que, en vista de la existencia de una dependencia funcional entre las ecuaciones (la primera es igual a la segunda más dos veces la tercera), se tienen de hecho sólo dos ecuaciones consistentes e independientes en dos variables.

Estos ejemplos simples deben ser suficientes para comunicar la importancia de la *consistencia* y la *independencia funcional* como los dos prerrequisitos para aplicación del proceso de contar ecuaciones e incógnitas. En general, a fin de aplicar ese proceso, asegúrese de que (1) la satisfacción de cualquier ecuación del modelo no imposibilita la satisfacción de otra y (2) que ninguna ecuación es redundante. Por ejemplo, en (3.17) se podría suponer con seguridad que las n funciones de demanda y las n de oferta son independientes entre sí, cada una obtenida de una fuente distinta: cada demanda surge de las decisiones de un grupo de consumidores y cada oferta surge de las decisiones de un grupo de empresas. Así, cada función sirve para describir un aspecto de la situación de mercado, y ninguna es redundante. Quizá se puede suponer también consistencia mutua. Además, las ecuaciones de condición de equilibrio en (3.18) son también independientes y es de suponer que consistentes. Por consiguiente, la solución analítica como se escribe en (3.19) puede considerarse en general justificable.[4]

Para modelos de ecuaciones simultáneas, existen métodos sistemáticos para probar la existencia de una solución única (o determinada). Para modelos lineales, éstos requieren de una aplicación del concepto de *determinante* que se introduce en el capítulo 5. En el caso de modelos no lineales, se requiere también de conocer las llamadas derivadas parciales y un tipo especial de determinante, llamado *determinante jacobiano*, que se analiza en los capítulos 7 y 8.

EJERCICIO 3.4

1. Desarrolle la solución de (3.13′), paso a paso, y de este modo compruebe los resultados en (3.14) y (3.15).
2. Vuelva a escribir (3.14) y (3.15) en términos de los parámetros originales del modelo en (3.12).
3. Las funciones de la oferta y la demanda de un modelo de mercado de dos artículos son como sigue:

$$Q_{d1} = 18 - 3P_1 + P_2 \qquad Q_{d2} = 12 + P_1 - 2P_2$$
$$Q_{s1} = -2 + 4P_1 \qquad Q_{s2} = -2 \qquad + 3P_2$$

Determine P_i^* y Q_i^* $(i = 1, 2)$. (Use fracciones en vez de decimales.)

[4] En esencia ésta es la forma en la que Léon Walras abordó el problema de la existencia de un equilibrio de mercado general. En las publicaciones modernas, se pueden encontrar varias pruebas matemáticas complejas de la existencia de un equilibrio de mercado competitivo bajo ciertas condiciones económicas postuladas. Pero la matemática utilizada es avanzada. La más fácil de entender es quizá la prueba provista en Robert Dorfman, Paul A. Samuelson y Robert M. Solow, *Linear Programming and Economic Analysis*, McGraw-Hill Book Company, Nueva York, 1958, capítulo 13.

3.5 Equilibrio en el análisis de ingreso nacional

Aun cuando hasta ahora el estudio de análisis estático se ha restringido a *modelos de mercado* de diversas maneras —lineales y no lineales, de un artículo y muchos artículos, específicos y generales—, éste tiene también, por supuesto, aplicaciones en otras áreas de la economía. Como un ejemplo, se puede citar el modelo de ingreso nacional keynesiano más simple,

$$\begin{aligned} Y &= C + I_0 + G_0 \\ C &= a + bY \end{aligned} \qquad (a > 0, \quad 0 < b < 1) \tag{3.23}$$

donde Y y C representan a las variables endógenas de ingreso nacional y gasto de consumo (planificado), respectivamente, e I_0 y G_0 representan la inversión y los gastos de gobierno determinados de manera exógena. La primera ecuación es una condición de equilibrio (ingreso nacional = gasto total planeado). La segunda, la función de consumo, es de comportamiento. Los dos parámetros en la función de consumo, a y b, representan el gasto de consumo autónomo y la propensión marginal al consumo, respectivamente.

Es suficientemente claro que estas dos ecuaciones, con dos variables endógenas, no son funcionalmente dependientes ni inconsistentes entre sí. Así, se podrían hallar valores de equilibrio de ingreso y gasto de consumo, Y^* y C^*, en términos de los parámetros a y b y de las variables exógenas I_0 y G_0.

La sustitución de la segunda ecuación en la primera reduce la ecuación (3.23) a una sola ecuación en la variable Y:

$$Y = a + bY + I_0 + G_0$$

o bien, $\quad (1 - b)Y = a + I_0 + G_0 \qquad$ (reuniendo los términos relacionados con Y)

Para hallar el valor solución de Y (ingreso nacional de equilibrio), sólo se tiene que dividir entre $(1 - b)$:

$$Y^* = \frac{a + I_0 + G_0}{1 - b} \tag{3.24}$$

Nótese de nuevo que el valor solución se expresa por completo en términos de los parámetros y variables exógenas, los datos del modelo. Al colocar (3.24) en la segunda ecuación de (3.23) se produce entonces el nivel de equilibrio de gasto de consumo:

$$\begin{aligned} C^* &= a + bY^* = a + \frac{b(a + I_0 + G_0)}{1 - b} \\ &= \frac{a(1 - b) + b(a + I_0 + G_0)}{1 - b} = \frac{a + b(I_0 + G_0)}{1 - b} \end{aligned} \tag{3.25}$$

Esto se expresa de nuevo por completo en términos de los datos proporcionados.

Tanto Y^* como C^* tienen la expresión $(1 - b)$ en el denominador; así que es necesaria la restricción $b \neq 1$ para evitar la división entre cero. Puesto que se supuso que b, la propensión marginal al consumo, es una fracción positiva, esta restricción se satisface de manera automática. Además, para que Y^* y C^* sean positivas, los numeradores de (3.24) y (3.25) deben ser positivos. Puesto que los gastos exógenos I_0 y G_0 son positivos normalmente, como lo es el parámetro a (la interseccíon vertical de la función de consumo), el signo de las expresiones del numerador también funcionará.

Como comprobación del cálculo, se puede sumar la expresión C^* en (3.25) a $(I_0 + G_0)$ y comprobar que la suma es igual a la expresión de Y^* en (3.24).

Resulta claro que este modelo es de extrema simplicidad y crudeza, pero se pueden construir también otros modelos de determinación de ingreso nacional con distintos grados de complejidad. No obstante, en cada caso los principios requeridos en la construcción y análisis del modelo son idénticos a los ya analizados. Por esta razón, ya no se dan más ejemplos aquí. Un modelo de ingreso nacional más completo, en el que interviene el equilibrio simultáneo del mercado de dinero y el mercado de bienes, se analiza en la sección 8.6.

EJERCICIO 3.5

1. Dado el siguiente modelo:

$$Y = C + I_0 + G_0$$
$$C = a + b(Y - T) \qquad (a > 0, \quad 0 < b < 1) \qquad [T\text{: impuestos}]$$
$$T = d + tY \qquad (d > 0, \quad 0 < t < 1) \qquad [t\text{: tasa de impuesto sobre la renta}]$$

   (*a*) ¿Cuántas variables endógenas hay?

   (*b*) Determine Y^*, T^* y C^*.

2. Sea el modelo de ingreso nacional:

$$Y = C + I_0 + G$$
$$C = a + b(Y - T_0) \qquad (a > 0, \quad 0 < b < 1)$$
$$G = gY \qquad (0 < g < 1)$$

   (*a*) Identifique las variables endógenas.

   (*b*) Dé el significado económico del parámetro g.

   (*c*) Determine el ingreso nacional de equilibrio.

   (*d*) ¿Qué restricción se requiere en los parámetros para que exista una solución?

3. Determine Y^* y C^* a partir de lo siguiente:

$$Y = C + I_0 + G_0$$
$$C = 25 + 6Y^{1/2}$$
$$I_0 = 16$$
$$G_0 = 14$$

Capítulo 4

Modelos lineales y álgebra de matrices

Para el modelo de un artículo (3.1), las soluciones P^* y Q^* como se expresan en (3.4) y (3.5), respectivamente, son simples, aun cuando intervienen varios parámetros. A medida que más y más artículos se incorporan al modelo, las fórmulas solución muy pronto se vuelven difíciles de manejar. Ésa es la razón de haber recurrido a una abreviatura, incluso para el caso de dos artículos, a fin de que las soluciones (3.14) y (3.15) aún se puedan escribir de manera más o menos concisa. No se intentó atacar ningún modelo de tres o cuatro artículos, aun en la versión lineal, sobre todo porque no se disponía de un método adecuado para manejar un sistema grande de ecuaciones simultáneas. Esta clase de método se encuentra en el *álgebra de matrices*, el tema de este capítulo y el siguiente.

El álgebra de matrices permite hacer muchas cosas. En primer lugar, proporciona una forma compacta de escribir un sistema de ecuaciones, incluso uno muy grande. Segundo, conduce a una forma de probar la existencia de una solución mediante la evaluación de un *determinante*, un concepto que tiene relación estrecha con el de una matriz. Tercero, proporciona un método para hallar la solución (si existe). Puesto que los sistemas de ecuaciones se encuentran no sólo en el análisis estático sino también en los análisis estático comparativo y dinámico y en problemas de optimización, se encontrará amplia aplicación del álgebra de matrices en casi todos los capítulos que siguen; por esta razón es deseable introducirla a tiempo.

Sin embargo, surge un ligero problema ya que el álgebra de matrices es aplicable sólo a sistemas de ecuaciones *lineales*. Cómo en realidad las ecuaciones lineales pueden describir relaciones económicas reales depende, por supuesto, de la naturaleza de las relaciones en cuestión. En muchos casos, aun cuando la suposición de linealidad conlleva sacrificar cierto realismo, una relación lineal supuesta puede producir una aproximación suficientemente estrecha a una relación no lineal para garantizar su uso.

En otros casos, mientras se conserva la no linealidad del modelo, se puede llevar a cabo una transformación de variables para obtener una relación lineal con la cual trabajar. Por ejemplo, la función no lineal

$$y = ax^b$$

se puede transformar de manera fácil al aplicar el logaritmo en ambos lados de la función

$$\log y = \log a + b \log x$$

que es lineal en las dos variables $(\log y)$ y $(\log x)$. (Los logaritmos se estudian con más detalle en el capítulo 10.) Lo que es más importante, en muchas aplicaciones como el análisis estático comparativo y los problemas de optimización, que se estudian más tarde, aunque la formulación del modelo económico es de naturaleza no lineal, los sistemas de ecuaciones lineales surgirán en el curso del análisis. Así, la restricción de linealidad no es tan restrictiva como parecía al principio.

4.1 Matrices y vectores

El modelo de mercado de dos artículos (3.12) se puede escribir, después de eliminar las variables de cantidad, como un sistema de dos ecuaciones lineales, como en (3.13′),

$$\begin{aligned} c_1 P_1 + c_2 P_2 &= -c_0 \\ \gamma_1 P_1 + \gamma_2 P_2 &= -\gamma_0 \end{aligned}$$

donde los parámetros c_0 y γ_0 aparecen a la derecha del signo igual. En general, un sistema de m ecuaciones lineales en n variables $(x_1, x_2, \ldots, x_n)$ se puede disponer en este tipo de formato:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= d_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= d_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= d_m \end{aligned} \tag{4.1}$$

En (4.1), la variable x_1 aparece sólo dentro de la columna del extremo izquierdo y, en general, la variable x_j aparece sólo en la j-ésima columna en el lado izquierdo del símbolo igual. El símbolo de parámetro con doble subíndice a_{ij} representa al coeficiente que aparece en la i-ésima ecuación, que multiplica a la j-ésima variable. Por ejemplo, a_{21} es el coeficiente en la segunda ecuación, que multiplica a la variable x_1. Por otro lado, el parámetro d_i que no está unido a ninguna variable, representa el término constante en la i-ésima ecuación. Por ejemplo, d_i es el término constante en la primera ecuación. Por lo tanto, los subíndices se adaptan a los lugares específicos de las variables y parámetros en (4.1).

Matrices como arreglos

Hay en esencia tres tipos de componentes en el sistema de ecuaciones (4.1). El primero es el conjunto de coeficientes a_{ij}; el segundo es el conjunto de variables $x_1, \ldots, x_n$, y el último es el conjunto de términos constantes $d_1, \ldots, d_m$. Si se arreglan los tres conjuntos como tres configuraciones rectangulares y se etiquetan, respectivamente, como A, x y d (sin subíndices), entonces se tiene

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \qquad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad d = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_m \end{bmatrix} \tag{4.2}$$

Como un ejemplo simple, dado el sistema de ecuaciones lineales

$$\begin{aligned} 6x_1 + 3x_2 + x_3 &= 22 \\ x_1 + 4x_2 - 2x_3 &= 12 \\ 4x_1 - x_2 + 5x_3 &= 10 \end{aligned} \tag{4.3}$$

se puede escribir

$$A = \begin{bmatrix} 6 & 3 & 1 \\ 1 & 4 & -2 \\ 4 & -1 & 5 \end{bmatrix} \qquad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad d = \begin{bmatrix} 22 \\ 12 \\ 10 \end{bmatrix} \tag{4.4}$$

Cada uno de los tres arreglos en (4.2) o (4.4) constituye una *matriz*.

Una matriz se define como un arreglo rectangular de números, parámetros o variables. Los miembros del arreglo, conocidos como *elementos* de la matriz, normalmente se encierran entre corchetes, como en (4.2), o a veces entre paréntesis o con líneas verticales dobles: || ||. Note que en la matriz A (la *matriz de coeficientes* del sistema de ecuaciones), los elementos se separan no por comas, sino sólo por espacios en blanco. Como un mecanismo de abreviación, el arreglo en la matriz A se puede escribir de forma más simple como

$$A = [a_{ij}] \qquad \begin{pmatrix} i = 1, 2, \ldots, m \\ j = 1, 2, \ldots, n \end{pmatrix}$$

En vista de que la ubicación de cada elemento de la matriz se fija sin ambigüedad mediante el subíndice, toda matriz es un conjunto ordenado.

Vectores como matrices especiales

El número de renglones y el número de columnas en una matriz juntos definen la *dimensión* de la matriz. Puesto que la matriz A en (4.2) contiene m renglones y n columnas, se dice que es de dimensión $m \times n$ (léase "m por n"). Es importante recordar que el número de renglones siempre precede al número de columnas; esto concuerda con la forma en la que están ordenados los dos subíndices en a_{ij}. En el caso especial en el que $m = n$, la matriz se llama *matriz cuadrada*; así, la matriz A en (4.4) es una matriz cuadrada de 3×3.

Algunas matrices pueden contener sólo una columna, como x y d en (4.2) o (4.4). Esta clase de matrices reciben el nombre especial de *vectores columna*. En (4.2), la dimensión de x es $n \times 1$ y la de d es $m \times 1$; en (4.4) tanto x como d son de 3×1. Si se disponen las variables x_j en un arreglo horizontal, resultaría una matriz de $1 \times n$, que se llama *vector renglón*. Para propósitos de notación, un vector renglón suele distinguirse de un vector columna por el uso de un símbolo al que se le agrega un apóstrofo.

$$x' = [x_1 \quad x_2 \quad \cdots \quad x_n]$$

Es posible observar que el vector (ya sea renglón o columna) es solamente una n-tupla ordenada y, como tal, a veces se puede interpretar como un punto en un espacio de dimensión n. A su vez, la matriz A de $m \times n$ se interpreta como un conjunto ordenado de m vectores renglón o como un conjunto ordenado de n vectores columna. En el capítulo 5 se prosigue con estas ideas.

Un asunto de interés más inmediato es cómo la notación de matrices posibilita, como se prometió, expresar un sistema de ecuaciones en una forma compacta. Con las matrices definidas en (4.4), el sistema de ecuaciones (4.3) se expresa simplemente como

$$Ax = d$$

De hecho, si A, x y d reciben los significados de (4.2), entonces incluso el sistema general de ecuaciones en (4.1) se puede escribir como $Ax = d$. La compactibilidad de esta notación es, por lo tanto, inconfundible.

Sin embargo, la ecuación $Ax = d$ lleva a por lo menos dos preguntas. ¿Cómo se multiplican dos matrices A y x? ¿Qué denota la igualdad de Ax y d? Puesto que las matrices implican bloques completos de números, las operaciones algebraicas familiares definidas para números simples no son directamente aplicables, y es necesario un nuevo conjunto de reglas operacionales.

EJERCICIO 4.1

1. Reescriba el modelo de mercado (3.1) en el formato de (4.1), y muestre que, si se disponen las tres variables en el orden Q_d, Q_s y P, la matriz de coeficientes será

$$\begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & b \\ 0 & 1 & -d \end{bmatrix}$$

¿Cómo escribiría el vector de constantes?

2. Reescriba el modelo de mercado (3.12) en el formato de (4.1) con las variables dispuestas en el siguiente orden: Q_{d1}, Q_{s1}, Q_{d2}, Q_{s2}, P_1, P_2. Escriba la matriz de coeficientes, el vector de variables y el vector de constantes.
3. ¿Se puede reescribir el modelo de mercado (3.6) en el formato de (4.1)? ¿Por qué?
4. Reescriba el modelo de ingreso nacional (3.23) en el formato de (4.1), con Y como la primera variable. Escriba la matriz de coeficientes y el vector de constantes.
5. Reescriba el modelo de ingreso nacional del ejercicio 3.5-1 en el formato de (4.1), con las variables en el orden Y, T y C. [*Sugerencia:* tenga cuidado con la expresión multiplicativa $b(Y - T)$ en la función de consumo.]

4.2 Operaciones con matrices

Como un requisito, se define primero la palabra *igualdad*. Se dice que dos matrices $A = [a_{ij}]$ y $B = [b_{ij}]$ son *iguales* si y sólo si tienen la misma dimensión y tienen elementos idénticos en las ubicaciones correspondientes del arreglo. En otras palabras, $A = B$ si y sólo si $a_{ij} = b_{ij}$ para todos los valores de i y de j. Así, por ejemplo, se encuentra

$$\begin{bmatrix} 4 & 3 \\ 2 & 0 \end{bmatrix} = \begin{bmatrix} 4 & 3 \\ 2 & 0 \end{bmatrix} \neq \begin{bmatrix} 2 & 0 \\ 4 & 3 \end{bmatrix}$$

Como otro ejemplo, si $\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 7 \\ 4 \end{bmatrix}$, esto significa que $x = 7$ y $y = 4$.

Suma y resta de matrices

Dos matrices se pueden sumar si y sólo si tienen la misma dimensión. Cuando se satisface este requerimiento dimensional, se dice que las matrices son *conformables para suma*. En ese caso, la suma de $A = [a_{ij}]$ y $B = [b_{ij}]$ se define como la suma de cada par de elementos correspondientes.

Ejemplo 1

$$\begin{bmatrix} 4 & 9 \\ 2 & 1 \end{bmatrix} + \begin{bmatrix} 2 & 0 \\ 0 & 7 \end{bmatrix} = \begin{bmatrix} 4+2 & 9+0 \\ 2+0 & 1+7 \end{bmatrix} = \begin{bmatrix} 6 & 9 \\ 2 & 8 \end{bmatrix}$$

Ejemplo 2

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} \\ a_{21}+b_{21} & a_{22}+b_{22} \\ a_{31}+b_{31} & a_{32}+b_{32} \end{bmatrix}$$

En general, se puede expresar esta regla así:

$$[a_{ij}] + [b_{ij}] = [c_{ij}] \qquad \text{donde } c_{ij} = a_{ij} + b_{ij}$$

Note que la matriz suma $[c_{ij}]$ debe tener la misma dimensión que las matrices de componentes $[a_{ij}]$ y $[b_{ij}]$.

La operación de resta $A - B$ se puede definir de manera similar si y sólo si A y B tienen la misma dimensión. La operación conlleva el resultado

$$[a_{ij}] - [b_{ij}] = [d_{ij}] \qquad \text{donde } d_{ij} = a_{ij} - b_{ij}$$

Ejemplo 3

$$\begin{bmatrix} 19 & 3 \\ 2 & 0 \end{bmatrix} - \begin{bmatrix} 6 & 8 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 19-6 & 3-8 \\ 2-1 & 0-3 \end{bmatrix} = \begin{bmatrix} 13 & -5 \\ 1 & -3 \end{bmatrix}$$

La operación de resta $A - B$ se puede considerar alternativamente como una operación de suma en la que participan una matriz A y otra matriz $(-1)B$. Sin embargo, esto origina la pregunta de qué denota la multiplicación de una matriz por un solo número (en este caso, -1).

Multiplicación escalar

Multiplicar una matriz por un número, o en terminología de álgebra de matrices, por un *escalar*, es multiplicar *cada* elemento de esa matriz por el escalar dado.

Ejemplo 4

$$7\begin{bmatrix} 3 & -1 \\ 0 & 5 \end{bmatrix} = \begin{bmatrix} 21 & -7 \\ 0 & 35 \end{bmatrix}$$

Ejemplo 5

$$\frac{1}{2}\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} \frac{1}{2}a_{11} & \frac{1}{2}a_{12} \\ \frac{1}{2}a_{21} & \frac{1}{2}a_{22} \end{bmatrix}$$

De estos ejemplos, se debe aclarar la razón fundamental del nombre escalar: porque "incrementa (o disminuye)" la matriz por un cierto múltiplo. Por supuesto, el escalar puede ser también un número negativo.

Ejemplo 6

$$-1\begin{bmatrix} a_{11} & a_{12} & d_1 \\ a_{21} & a_{22} & d_2 \end{bmatrix} = \begin{bmatrix} -a_{11} & -a_{12} & -d_1 \\ -a_{21} & -a_{22} & -d_2 \end{bmatrix}$$

Observe que si la matriz de la izquierda representa los coeficientes y los términos constantes de las ecuaciones simultáneas

$$a_{11}x_1 + a_{12}x_2 = d_1$$

$$a_{21}x_1 + a_{22}x_2 = d_2$$

entonces la multiplicación por el escalar -1 equivale a multiplicar ambos lados de ambas ecuaciones por -1, y de este modo cambia el signo de todo término en el sistema.

Multiplicación de matrices

Mientras que un escalar se puede usar para multiplicar una matriz de cualquier dimensión, la multiplicación de dos matrices depende de la satisfacción de un requerimiento dimensional diferente.

Supóngase que, dadas dos matrices A y B, se desea encontrar el producto AB, la condición de conformabilidad para la multiplicación es que la dimensión *columna* de A (la matriz "primaria" en la expresión AB) debe ser igual a la dimensión renglón de B (la matriz "secundaria"). Por ejemplo, si

$$\underset{(1\times 2)}{A} = [a_{11} \quad a_{12}] \qquad \underset{(2\times 3)}{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix} \tag{4.5}$$

entonces está definido el producto AB, puesto que A tiene *dos columnas* y B tiene *dos renglones*, exactamente el mismo número.[1] Esto se puede comprobar de un vistazo al comparar el *segundo* número en el indicador de dimensión de A, que es (1×2), con el *primer* número en el indicador de dimensión de B, (2×3). Por otro lado, el producto inverso BA *no* está definido en este caso, porque B (ahora la matriz primaria) tiene *tres* columnas mientras que A (la matriz secundaria) tiene sólo *un* renglón; por lo tanto, se viola la condición de conformabilidad.

En general, si A es de dimensión $m \times n$ y B es de dimensión $p \times q$, la matriz producto AB estará definida si y sólo si $n = p$. Por otra parte, si se define la matriz producto AB tendrá la dimensión $m \times q$ —el mismo número de *renglones* que la matriz primaria A y el mismo número de *columnas* que la matriz secundaria B—. Para las matrices expresadas en (4.5), AB será 1×3.

Queda por definir el procedimiento exacto de multiplicación. Para este propósito, se toman las matrices A y B de (4.5) para ilustración. Puesto que el producto AB está definido y se espera que sea de dimensión 1×3, se puede escribir en general (usando el símbolo C y no c' para el vector renglón) que

$$AB = C = [c_{11} \quad c_{12} \quad c_{13}]$$

Cada elemento de la matriz producto C, que se denota mediante c_{ij}, se define como una suma de productos, que se calculan de los elementos del i-ésimo renglón de la matriz primaria A, y los de la j-ésima columna de la matriz secundaria B. Por ejemplo, para determinar c_{11}, se debe tomar el *primer renglón* en A (puesto que $i = 1$) y la *primera columna* en B (puesto que $j = 1$), como se ilustra en el panel superior de la figura 4.1, y luego juntar los elementos en secuencia, multiplicar cada par y tomar la suma de los productos resultantes, para obtener

$$c_{11} = a_{11}b_{11} + a_{12}b_{21} \tag{4.6}$$

De modo similar, para c_{12}, se toma el *primer renglón* en A (puesto que $i = 1$) y la *segunda columna* en B (puesto que $j = 2$) y se calcula la suma indicada de productos, de acuerdo con el panel inferior de la figura 4.1, como sigue:

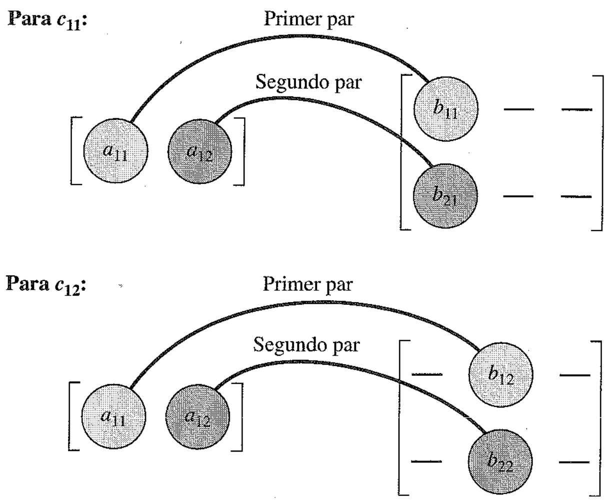
$$c_{12} = a_{11}b_{12} + a_{12}b_{22} \tag{4.6'}$$

De la misma manera, también se debe tener

$$c_{13} = a_{11}b_{13} + a_{12}b_{23} \tag{4.6''}$$

[1] La matriz A, un vector renglón, normalmente se denotaría mediante a'. Aquí se emplea el símbolo A para remarcar el hecho de que la regla de multiplicación que está siendo explicada se aplica en general a las matrices, no sólo al producto de un vector y una matriz.

FIGURA 4.1



Es el requerimiento particular de formar parejas en este proceso el que requiere la comparación de la dimensión columna de la matriz primaria y la dimensión renglón de la matriz secundaria antes de llevar a cabo la multiplicación.

El procedimiento de multiplicación ilustrado en la figura 4.1 también se puede describir por medio del concepto del *producto interior* de dos vectores. Dados dos vectores u y v con n componentes cada uno, por ejemplo, $(u_1, u_2, \ldots, u_n)$ y $(v_1, v_2, \ldots, v_n)$, ordenados *ya sea* como dos renglones *o* como dos columnas *o* como un renglón y una columna, su producto interior, escrito como $u \cdot v$ (con un punto en medio), se define como

$$u \cdot v = u_1v_1 + u_2v_2 + \cdots + u_nv_n$$

Ésta es una suma de productos de componentes correspondientes y, por lo tanto, el producto interior de dos vectores es un escalar.

Ejemplo 7

Si, después de un viaje de compras, se ordenan las cantidades compradas de n bienes como un vector renglón $Q' = [Q_1 \quad Q_2 \quad \cdots \quad Q_n]$, y la lista de precios de esos bienes en un vector precio $P' = [P_1 \quad P_2 \quad \cdots \quad P_n]$, entonces el producto interior de esos dos vectores es

$$Q' \cdot P' = Q_1P_1 + Q_2P_2 + \cdots + Q_nP_n = \text{costo de compra total}$$

Con este concepto, se puede describir el elemento c_{ij} en la matriz producto $C = AB$ simplemente como el producto interior del i-ésimo renglón de la matriz primaria A y la j-ésima columna de la matriz secundaria B. Al examinar la figura 4.1, se comprueba la validez de esta descripción.

La regla de multiplicación recién descrita se aplica con igual validez cuando las dimensiones de A y B son distintas a las que se ilustran en la figura 4.1; el único prerrequisito es que se satisfaga la condición de conformabilidad.

Ejemplo 8

Dadas

$$\underset{(3\times2)}{A} = \begin{bmatrix} 1 & 3 \\ 2 & 8 \\ 4 & 0 \end{bmatrix} \qquad \text{y} \qquad \underset{(2\times1)}{B} = \begin{bmatrix} 5 \\ 9 \end{bmatrix}$$

determine AB. El producto AB está definido porque A tiene dos columnas y B tiene dos renglones. Su matriz producto debe ser 3×1, un vector columna:

$$AB = \begin{bmatrix} 1(5)+3(9) \\ 2(5)+8(9) \\ 4(5)+0(9) \end{bmatrix} = \begin{bmatrix} 32 \\ 82 \\ 20 \end{bmatrix}$$

Ejemplo 9 Dadas

$$\underset{(3\times3)}{A} = \begin{bmatrix} 3 & -1 & 2 \\ 1 & 0 & 3 \\ 4 & 0 & 2 \end{bmatrix} \quad \text{y} \quad \underset{(3\times3)}{B} = \begin{bmatrix} 0 & -\frac{1}{5} & \frac{3}{10} \\ -1 & \frac{1}{5} & \frac{7}{10} \\ 0 & \frac{2}{5} & -\frac{1}{10} \end{bmatrix}$$

obtenga AB. La misma regla de multiplicación ahora produce una matriz producto muy especial:

$$AB = \begin{bmatrix} 0+1+0 & -\frac{3}{5}-\frac{1}{5}+\frac{4}{5} & \frac{9}{10}-\frac{7}{10}-\frac{2}{10} \\ 0+0+0 & -\frac{1}{5}+0+\frac{6}{5} & \frac{3}{10}+0-\frac{3}{10} \\ 0+0+0 & -\frac{4}{5}+0+\frac{4}{5} & \frac{12}{10}+0-\frac{2}{10} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Esta última matriz —una matriz cuadrada con unos en su *diagonal principal* (la diagonal que va del noroeste al sudeste) y cero en los demás lugares— ejemplifica el tipo importante de matriz conocida como *matriz identidad*. La explicación de esto se amplía en la sección 4.5.

Ejemplo 10 Tómese la matriz A y el vector x como se define en (4.4) y determine Ax. La matriz producto es un vector columna de 3×1:

$$Ax = \underset{(3\times3)}{\begin{bmatrix} 6 & 3 & 1 \\ 1 & 4 & -2 \\ 4 & -1 & 5 \end{bmatrix}} \underset{(3\times1)}{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}} = \underset{(3\times1)}{\begin{bmatrix} 6x_1+3x_2+\ \ x_3 \\ x_1+4x_2-2x_3 \\ 4x_1-\ \ x_2+5x_3 \end{bmatrix}}$$

Nota: el producto de la derecha es un vector *columna*, ¡a pesar de su apariencia corpulenta! Por lo tanto, cuando se escribe $Ax = d$, se tiene

$$\begin{bmatrix} 6x_1+3x_2+\ \ x_3 \\ x_1+4x_2-2x_3 \\ 4x_1-\ \ x_2+5x_3 \end{bmatrix} = \begin{bmatrix} 22 \\ 12 \\ 10 \end{bmatrix}$$

que, de acuerdo con la definición de igualdad de matrices, es equivalente a la expresión del sistema de ecuaciones completo en (4.3).

Observe que, para usar la notación matricial $Ax = d$, es necesario, como resultado de la condición de conformabilidad, disponer las variables x_j en un vector *columna*, aun cuando estas variables se listan en un orden horizontal en el sistema original de ecuaciones.

Ejemplo 11 El modelo de ingreso nacional simple de dos variables endógenas Y y C,

$$Y = C + I_0 + G_0$$
$$C = a + bY$$

se pueden disponer en el formato estándar de (4.1) como sigue:

$$Y - C = I_0 + G_0$$
$$-bY + C = a$$

Por consiguiente, la matriz de coeficientes A, el vector de variables x y el vector de constantes d son

$$\underset{(2\times2)}{A} = \begin{bmatrix} 1 & -1 \\ -b & 1 \end{bmatrix} \qquad \underset{(2\times1)}{x} = \begin{bmatrix} Y \\ C \end{bmatrix} \qquad \underset{(2\times1)}{d} = \begin{bmatrix} I_0 + G_0 \\ a \end{bmatrix}$$

Compruebe que este sistema se puede expresar mediante la ecuación $Ax = d$.

Por la regla de multiplicación de matrices, se tiene

$$Ax = \begin{bmatrix} 1 & -1 \\ -b & 1 \end{bmatrix}\begin{bmatrix} Y \\ C \end{bmatrix} = \begin{bmatrix} 1(Y) + (-1)(C) \\ -b(Y) + 1(C) \end{bmatrix} = \begin{bmatrix} Y - C \\ -bY + C \end{bmatrix}$$

Así, la ecuación matricial $Ax = d$ produciría

$$\begin{bmatrix} Y - C \\ -bY + C \end{bmatrix} = \begin{bmatrix} I_0 + G_0 \\ a \end{bmatrix}$$

Puesto que la igualdad de matrices indica la igualdad entre elementos correspondientes, es claro que la ecuación $Ax = d$ representa con precisión el sistema de ecuaciones original, como se expresa en el formato (4.1).

El asunto de la división

Si bien las matrices, al igual que los números, pueden experimentar las operaciones de suma, resta y multiplicación, sujetas a las condiciones de conformabilidad, no es posible dividir una matriz entre otra. Es decir, no se puede escribir A/B.

Para dos números a y b, el cociente a/b (con $b \neq 0$) se puede escribir de forma alternativa como ab^{-1} o $b^{-1}a$, donde b^{-1} representa el *inverso* o *recíproco* de b. Puesto que $ab^{-1} = b^{-1}a$, la expresión del cociente a/b se puede usar para representar tanto ab^{-1} como $b^{-1}a$. En el caso de las matrices es diferente. En ciertas ocasiones, al aplicar el concepto de inversos para matrices, se puede definir (como se estudió en la sección 4.6) una matriz B^{-1}, que es la inversa de la matriz B. Pero, de la explicación de la condición de conformabilidad, se deduce que si AB^{-1} está definido, no se puede estar seguro de que $B^{-1}A$ también esté definido. Aun cuando tanto AB^{-1} como $B^{-1}A$ estén definidos, podrían no representar el mismo producto. Por consiguiente, la expresión A/B no se puede usar sin ambigüedad, por lo que se debe evitar. En cambio, se debe especificar si se hace referencia a AB^{-1} o bien a $B^{-1}A$, siempre y cuando exista el inverso B^{-1} y que esté definido el producto de las matrices en cuestión. Las matrices inversas se analizan con más detalle en la sección 4.6.

Notación $\sum$

El uso de símbolos con subíndice no sólo ayuda a designar los lugares de parámetros y variables, sino que también se presta a una abreviatura flexible para denotar sumas de términos, como los que surgen durante el proceso de multiplicación de matrices.

La abreviatura de suma hace uso de la letra griega Σ (sigma, para "suma"). Para expresar la suma de x_1, x_2 y x_3, por ejemplo, se puede escribir

$$x_1 + x_2 + x_3 = \sum_{j=1}^{3} x_j$$

que se lee como "la suma de x_j cuando j varía de 1 a 3". El símbolo j, conocido como *índice de sumatoria*, toma sólo valores enteros. La expresión x_j representa al *sumando* (aquel que se va a sumar) y es de hecho una función de j. Aparte de la letra j, los índices de sumatoria también se denotan por lo común mediante i o k, como

$$\sum_{i=3}^{7} x_i = x_3 + x_4 + x_5 + x_6 + x_7$$

$$\sum_{k=0}^{n} x_k = x_0 + x_1 + \cdots + x_n$$

La aplicación de la notación $\sum$ se puede extender fácilmente a casos en los que el término x lleva como prefijo un coeficiente o en los que cada término de la suma se eleva a alguna potencia entera. Por ejemplo, se puede escribir:

$$\sum_{j=1}^{3} ax_j = ax_1 + ax_2 + ax_3 = a(x_1 + x_2 + x_3) = a\sum_{j=1}^{3} x_j$$

$$\sum_{j=1}^{3} a_j x_j = a_1x_1 + a_2x_2 + a_3x_3$$

$$\sum_{i=0}^{n} a_i x^i = a_0x^0 + a_1x^1 + a_2x^2 + \cdots + a_nx^n$$

$$= a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

En particular, el último ejemplo muestra que de hecho la expresión $\sum_{i=0}^{n} a_i x^i$ se puede usar como una forma abreviada de la función polinomial general de (2.4).

Se puede mencionar de paso que, siempre y cuando el contexto de la explicación no deje ambigüedad en cuanto al alcance de la suma, el símbolo $\sum$ se puede usar solo, sin un índice (como $\sum x_i$), o sólo con la letra índice debajo (como $\sum_i x_i$).

A continuación se aplica la abreviatura $\sum$ a la multiplicación de matrices. En (4.6), (4.6′) y (4.6″), cada elemento de la matriz producto $C = AB$ se define como una suma de términos, que ahora se puede reescribir como sigue:

$$c_{11} = a_{11}b_{11} + a_{12}b_{21} = \sum_{k=1}^{2} a_{1k}b_{k1}$$

$$c_{12} = a_{11}b_{12} + a_{12}b_{22} = \sum_{k=1}^{2} a_{1k}b_{k2}$$

$$c_{13} = a_{11}b_{13} + a_{12}b_{23} = \sum_{k=1}^{2} a_{1k}b_{k3}$$

En cada caso, el primer subíndice de c_{1j} se refleja en el primer subíndice de a_{1k}, y el segundo subíndice de c_{1j} se refleja en el segundo subíndice de b_{kj} en la expresión de $\sum$. Por otro lado, el índice k es un subíndice "simulado"; sirve para indicar cuál par particular de elementos está siendo multiplicado, pero no aparece en el símbolo c_{1j}.

Al ampliar esto a la multiplicación de una matriz $A = [a_{ik}]$ de $m \times n$ y de una matriz $B = [b_{kj}]$ de $n \times p$, se pueden escribir ahora los elementos de la matriz producto $AB = C = [c_{ij}]$ de $m \times p$ como

$$c_{11} = \sum_{k=1}^{n} a_{1k}b_{k1} \qquad c_{12} = \sum_{k=1}^{n} a_{1k}b_{k2} \qquad \cdots$$

o de modo más general,

$$c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj} \qquad \begin{pmatrix} i = 1, 2, \ldots, m \\ j = 1, 2, \ldots, p \end{pmatrix}$$

Esta última ecuación representa otra forma de expresar la regla de multiplicación para las matrices definidas antes.

EJERCICIO 4.2

1. Dadas $A = \begin{bmatrix} 7 & -1 \\ 6 & 9 \end{bmatrix}$, $B = \begin{bmatrix} 0 & 4 \\ 3 & -2 \end{bmatrix}$ y $C = \begin{bmatrix} 8 & 3 \\ 6 & 1 \end{bmatrix}$, obtenga:

   (*a*) $A + B$ (*b*) $C - A$ (*c*) $3A$ (*d*) $4B + 2C$

2. Dadas $A = \begin{bmatrix} 2 & 8 \\ 3 & 0 \\ 5 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 2 & 0 \\ 3 & 8 \end{bmatrix}$ y $C = \begin{bmatrix} 7 & 2 \\ 6 & 3 \end{bmatrix}$:

   (*a*) ¿Está definido AB? Calcule AB. ¿Puede calcular BA? ¿Por qué?

   (*b*) ¿Está definido BC? Calcule BC. ¿Está definido CB? En caso afirmativo, calcule CB. ¿Es cierto que $BC = CB$?

3. Con base en las matrices dadas en el ejemplo 9, ¿está definido el producto BA? Si es así, calcule el producto. ¿En este caso se tiene $AB = BA$?

4. Obtenga las matrices producto de los siguientes casos (en cada uno, anexe debajo de cada matriz un indicador de dimensión):

   (*a*) $\begin{bmatrix} 0 & 2 & 0 \\ 3 & 0 & 4 \\ 2 & 3 & 0 \end{bmatrix} \begin{bmatrix} 8 & 0 \\ 0 & 1 \\ 3 & 5 \end{bmatrix}$ (*c*) $\begin{bmatrix} 3 & 5 & 0 \\ 4 & 2 & -7 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$

   (*b*) $\begin{bmatrix} 6 & 5 & -1 \\ 1 & 0 & 4 \end{bmatrix} \begin{bmatrix} 4 & -1 \\ 5 & 2 \\ 0 & 1 \end{bmatrix}$ (*d*) $\begin{bmatrix} a & b & c \end{bmatrix} \begin{bmatrix} 7 & 0 \\ 0 & 2 \\ 1 & 4 \end{bmatrix}$

5. En el ejemplo 7, si se disponen las cantidades y precios como vectores columna en lugar de vectores renglón, ¿está definido $Q \cdot P$? ¿Es posible expresar el costo total de compra como $Q \cdot P$, $Q' \cdot P$, $Q \cdot P'$?

6. Desarrolle las siguientes expresiones de suma:

   (*a*) $\sum_{i=2}^{5} x_i$ (*d*) $\sum_{i=1}^{n} a_i x^{i-1}$

   (*b*) $\sum_{i=5}^{8} a_i x_i$ (*e*) $\sum_{i=0}^{3} (x + i)^2$

   (*c*) $\sum_{i=1}^{4} bx_i$

7. Reescriba lo siguiente en notación de $\sum$:
   (*a*) $x_1(x_1 - 1) + 2x_2(x_2 - 1) + 3x_3(x_3 - 1)$
   (*b*) $a_2(x_3 + 2) + a_3(x_4 + 3) + a_4(x_5 + 4)$
   (*c*) $\frac{1}{x} + \frac{1}{x^2} + \cdots + \frac{1}{x^n} \quad (x \neq 0)$
   (*d*) $1 + \frac{1}{x} + \frac{1}{x^2} + \cdots + \frac{1}{x^n} \quad (x \neq 0)$
8. Muestre que las siguientes expresiones son ciertas:
   (*a*) $\left(\sum_{i=0}^{n} x_i\right) + x_{n+1} = \sum_{i=0}^{n+1} x_i$
   (*b*) $\sum_{j=1}^{n} ab_j y_j = a \sum_{j=1}^{n} b_j y_j$
   (*c*) $\sum_{j=1}^{n} (x_j + y_j) = \sum_{j=1}^{n} x_j + \sum_{j=1}^{n} y_j$

4.3 Notas sobre operaciones con vectores

En las secciones 4.1 y 4.2, los vectores se consideran como un tipo especial de matriz. Como tales, califican para la aplicación de todas las operaciones algebraicas analizadas. Sin embargo, debido a sus peculiaridades dimensionales, vienen al caso algunos comentarios adicionales sobre operaciones con vectores.

Multiplicación de vectores

Un vector columna u de $m \times 1$ y un vector renglón v' de $1 \times n$ producen una matriz producto uv' de dimensión $m \times n$.

Ejemplo 1

Dada $u = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$ y $v' = [1 \quad 4 \quad 5]$, se obtiene

$$uv' = \begin{bmatrix} 3(1) & 3(4) & 3(5) \\ 2(1) & 2(4) & 2(5) \end{bmatrix} = \begin{bmatrix} 3 & 12 & 15 \\ 2 & 8 & 10 \end{bmatrix}$$

Puesto que cada renglón en u consta sólo de un elemento, al igual que cada columna en v', cada elemento de uv' resulta ser un solo producto en vez de una suma de productos. El producto uv' es una matriz de 2×3, aun cuando con lo que se empezó son un par de vectores.

Por otro lado, dados un vector renglón u' de $1 \times n$ y un vector columna v, de $n \times 1$, el producto $u'v$ será de dimensión 1×1.

Ejemplo 2

Dado $u' = [3 \quad 4]$ y $v = \begin{bmatrix} 9 \\ 7 \end{bmatrix}$, se tiene

$$u'v = [3(9) + 4(7)] = [55]$$

Como está escrita, $u'v$ es una matriz, a pesar del hecho de que sólo está presente un solo elemento. Sin embargo, las matrices de 1×1 se comportan exactamente como los escalares con respecto a la suma y la multiplicación: $[4] + [8] = [12]$, así como $4 + 8 = 12$, y

$[3][7] = [21]$, de igual modo que $3(7) = 21$. Además, las matrices de 1×1 no poseen propiedades importantes que no tengan los escalares. De hecho, hay una correspondencia uno a uno entre el conjunto de los escalares y el conjunto de las matrices de 1×1 cuyos elementos son escalares. Por esta razón, se puede redefinir $u'v$ como el *escalar* que corresponde a la matriz producto de 1×1. Para el ejemplo 2, se puede escribir en consecuencia $u'v = 55$. Este producto se llama *producto escalar.*[2] Sin embargo, recuerde que si bien una matriz de 1×1 se puede tratar como un escalar, éste no se puede reemplazar por una matriz de 1×1 a voluntad si se van a realizar más cálculos, porque podrían surgir complicaciones en relación con las condiciones de conformabilidad.

Ejemplo 3

Dado un vector renglón $u' = [3 \quad 6 \quad 9]$, obtenga $u'u$. Puesto que u es solamente el vector columna con los elementos de u' ordenados verticalmente, se tiene

$$u'u = [3 \quad 6 \quad 9]\begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix} = (3)^2 + (6)^2 + (9)^2$$

donde se han omitido los corchetes de la matriz producto de 1×1 a la derecha. Note que el producto $u'u$ da la suma de cuadrados de los elementos de u.

En general, si $u' = [u_1 \quad u_2 \quad \cdots \quad u_n]$, entonces $u'u$ será la suma de cuadrados (un escalar) de los elementos u_j:

$$u'u = u_1^2 + u_2^2 + \cdots + u_n^2 = \sum_{j=1}^{n} u_j^2$$

Si se hubiera calculado el producto interior $u \cdot u$ (o $u' \cdot u'$), se habría obtenido el mismo resultado.

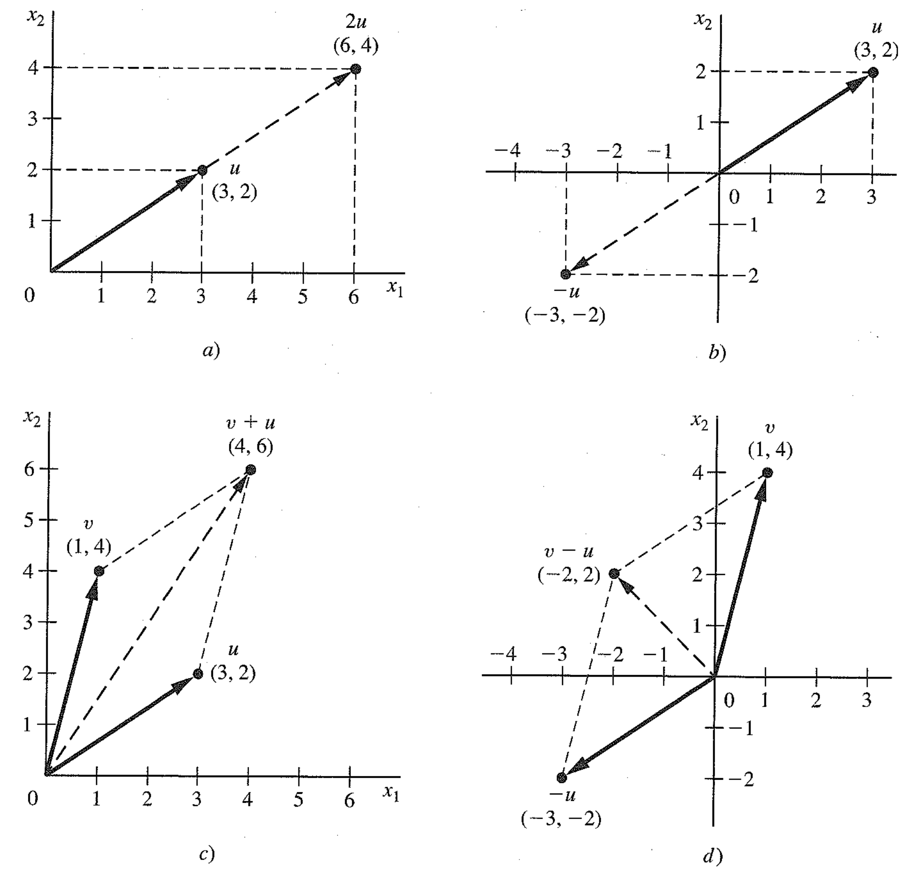
Para concluir, es importante distinguir entre los significados de uv' (una matriz más grande que 1×1) y $u'v$ (una matriz de 1×1, o un escalar). Observe, en particular, que un producto escalar debe tener un vector *renglón* como la matriz primaria y un vector *columna* como la matriz secundaria, de lo contrario el producto no puede ser 1×1.

Interpretación geométrica de operaciones con vectores

Ya se mencionó que un vector columna o renglón con n componentes (al cual de aquí en adelante se le llama *vector n-dimensional*) se puede considerar como una n-tupla y, por lo tanto, como un punto en un espacio n-dimensional (al cual en lo sucesivo se denomina espacio n-dimensional). Enseguida se dan más explicaciones acerca de esta idea. En la figura 4.2*a*, se grafica un punto (3, 2) en un espacio bidimensional y se identifica con una u. Ésta es la contraparte geométrica del vector $u = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$ o el vector $u' = [3 \quad 2]$, ambos indican en este contexto el mismo par ordenado. Si se dibuja una flecha (un segmento de recta dirigido) del punto de origen (0, 0) al punto u, ésta especifica la única ruta recta mediante la cual se llega al punto de destino u desde el punto de origen. Puesto que existe una flecha única para cada punto, se puede considerar que el vector u está representado de forma gráfica *ya sea* por el punto (3, 2) *o bien* por la flecha correspondiente. Esta flecha —que parte del origen (0, 0) como la manecilla de un reloj, con una longitud y dirección definidas— se llama *radio vector*.

[2] El concepto de producto escalar es, por lo tanto, similar al concepto de producto interior de dos vectores con el mismo número de elementos en cada uno, que también produce un escalar. Sin embargo, recuerde que el producto interior está exento de la condición de conformabilidad para la multiplicación, así que es posible escribirlo como $u \cdot v$. Por otro lado, en el caso del producto escalar (denotado sin un punto entre los dos símbolos de vector), se puede expresar sólo como un vector renglón multiplicado por un vector columna, con el vector renglón en primer lugar.

FIGURA 4.2



Siguiendo esta nueva interpretación de un vector, es posible dar significados geométricos a (1) la multiplicación escalar de un vector, (2) la suma y resta de vectores y, de modo más general, (3) la denominada combinación lineal de vectores.

Primero, si se grafica el vector $\begin{bmatrix} 6 \\ 4 \end{bmatrix} = 2u$ en la figura 4.2*a*, la flecha resultante se traslapa con la anterior pero tiene el doble de longitud. De hecho, la multiplicación de un vector u por algún escalar k produce una flecha superpuesta, pero la cabeza tendrá otra ubicación a menos que $k = 1$. Si el multiplicador escalar es $k > 1$, la flecha se extiende (aumenta); si $0 < k < 1$, la flecha se acorta (disminuye); si $k = 0$, la flecha se reduce al punto de origen, lo que representa un *vector nulo*, $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$. Un multiplicador escalar negativo incluso invierte la dirección de la flecha. Si el vector u se multiplica por -1, por ejemplo, se obtiene $-u = \begin{bmatrix} -3 \\ -2 \end{bmatrix}$. Esto se traza en la figura 4.2*b* como una flecha de la misma longitud que u pero en dirección contraria.

A continuación, considere la suma de dos vectores, $v = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$ y $u = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$. La suma de $v + u = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$ se puede graficar de manera directa como la flecha discontinua de la figura 4.2*c*. Si se construye un paralelogramo con los dos vectores u y v (flechas continuas) como dos de sus lados, la diagonal del paralelogramo resulta ser exactamente la flecha que representa al vector suma $v + u$. En general, un vector suma se obtiene de forma geométrica a partir de un paralelogramo. Además, este método también da la *diferencia de vectores* $v - u$, puesto que el último es equivalente a la *suma* de v y $(-1)u$. En la figura 4.2*d*, se reproduce primero el vector v y el vector negativo $-u$ de los diagramas *c* y *b*, respectivamente, y luego se construye un paralelogramo. La diagonal resultante representa la diferencia de vectores $v - u$.

La interpretación geométrica de una combinación lineal (es decir, una suma o diferencia lineal) de vectores ocupa sólo una pequeña parte de estos resultados. Considere el caso sencillo de

$$3v + 2u = 3\begin{bmatrix} 1 \\ 4 \end{bmatrix} + 2\begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 9 \\ 16 \end{bmatrix}$$

La parte de multiplicación escalar de esta operación requiere reubicar las flechas respectivas de los dos vectores v y u, y la parte de suma requiere la construcción de un paralelogramo. Más allá de estas dos operaciones gráficas básicas, no hay nada nuevo en una combinación lineal de vectores. Esto se cumple incluso si hay más términos en la combinación lineal, como en

$$\sum_{i=1}^{n} k_i v_i = k_1 v_1 + k_2 v_2 + \cdots + k_n v_n$$

donde las k_i son un conjunto de escalares; pero, los símbolos con subíndice v_i ahora denotan un conjunto de vectores. Para formar esta suma, se podrían sumar los dos primeros términos y después agregar la suma resultante a la tercera, y así sucesivamente, hasta incluir todos los términos.

Dependencia lineal

Un conjunto de vectores $v_1, \ldots, v_n$ es *linealmente dependiente* si (y sólo si) cualquiera de ellos se puede expresar como una combinación lineal de los demás vectores; de lo contrario, son *linealmente independientes*.

Ejemplo 4

Los vectores $v_1 = \begin{bmatrix} 2 \\ 7 \end{bmatrix}$, $v_2 = \begin{bmatrix} 1 \\ 8 \end{bmatrix}$ y $v_3 = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$ son linealmente dependientes porque v_3 es una combinación lineal de v_1 y v_2:

$$3v_1 - 2v_2 = \begin{bmatrix} 6 \\ 21 \end{bmatrix} - \begin{bmatrix} 2 \\ 16 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \end{bmatrix} = v_3$$

Note que esta última ecuación se puede expresar de otra forma como

$$3v_1 - 2v_2 - v_3 = 0$$

donde $0 \equiv \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ representa un vector nulo (conocido también como *vector cero*).

Ejemplo 5

Los dos vectores renglón $v'_1 = [5 \quad 12]$ y $v'_2 = [10 \quad 24]$ son linealmente dependientes porque

$$2v'_1 = 2[5 \quad 12] = [10 \quad 24] = v'_2$$

El hecho de que un vector es un múltiplo de otro vector ilustra el caso más simple de combinación lineal. Note de nuevo que esta última ecuación se puede escribir en forma equivalente como

$$2v'_1 - v'_2 = 0'$$

donde $0'$ representa el vector renglón nulo $[0 \quad 0]$.

Con la introducción de vectores nulos, la dependencia lineal se puede redefinir como sigue. Un conjunto de m vectores $v_1, \ldots, v_n$ es *linealmente dependiente* si y sólo si existe un conjunto de escalares $k_1, \ldots, k_n$ (no todos cero) tal que

$$\sum_{i=1}^{n} k_i v_i = \underset{(m\times 1)}{0}$$

Por otro lado, si esta ecuación *sólo se satisface cuando* $k_i = 0$ para toda i, estos vectores son linealmente independientes.

El concepto de dependencia lineal admite también una interpretación geométrica fácil. Dos vectores u y $2u$ (uno un múltiplo del otro) son obviamente dependientes. Desde el punto de vista geométrico, en la figura 4.2*a*, sus flechas yacen en una sola recta. Lo mismo se cumple para los dos vectores dependientes u y $-u$ de la figura 4.2*b*. En contraste, los dos vectores u y v de la figura 4.2*c* son linealmente *independientes*, porque es imposible expresar uno como un múltiplo del otro. Geométricamente, sus flechas no yacen en una sola recta.

Cuando se consideran más de dos vectores en el espacio bidimensional, surge esta conclusión importante: una vez hallados dos vectores linealmente *independientes* en el espacio bidimensional (por ejemplo, u y v), los demás vectores en ese espacio se pueden expresar como una combinación lineal de éstos (u y v). En la figura 4.2*c* y *d*, ya se ilustró cómo se determinan las dos combinaciones lineales simples $v + u$ y $v - u$. Además, al ampliar, acortar e invertir los vectores u y v, y combinarlos después en varios paralelogramos, se genera un número infinito de nuevos vectores, que agotarán el conjunto de los vectores en dos dimensiones. Como resultado de esto, cualquier conjunto de tres o más vectores en dos dimensiones (tres o más vectores en un espacio bidimensional) deben ser linealmente dependientes. Dos de ellos pueden ser independientes, pero entonces el tercero debe ser una combinación lineal de los dos primeros.

Espacio vectorial

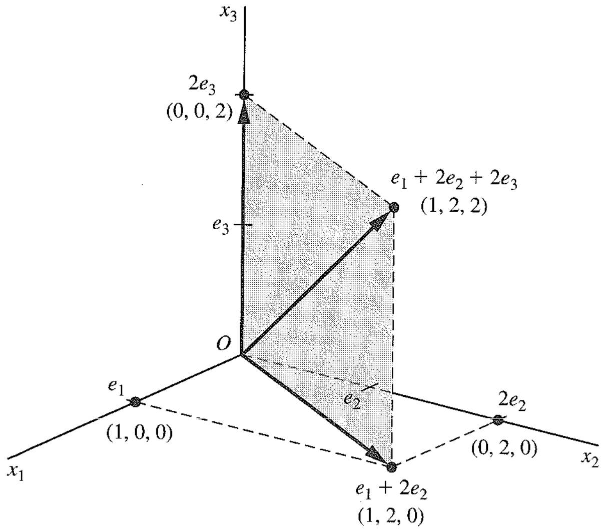
La totalidad de los vectores en dos dimensiones generados mediante las distintas combinaciones lineales de dos vectores independientes u y v constituye el *espacio vectorial* bidimensional. Puesto que se trata sólo con vectores que tienen componentes de valores reales, este espacio vectorial no es otro que R^2, el espacio bidimensional al que se ha hecho referencia todo el tiempo. El espacio bidimensional no se puede generar mediante un solo vector bidimensional, porque sus combinaciones lineales sólo dan lugar al conjunto de vectores que yacen en una sola recta. Tampoco la generación del espacio bidimensional requiere más de dos vectores bidimensionales linealmente independientes; de todos modos, sería imposible hallar más de dos.

Se dice que los vectores linealmente independientes u y v *generan* el espacio bidimensional. Asimismo, se dice que constituyen una *base* para el espacio bidimensional. Note que se dijo *una* base, no *la* base, porque cualquier par de vectores en dos dimensiones puede servir siempre y cuando sean linealmente independientes. En particular, considere los dos vectores $[1 \quad 0]$ y $[0 \quad 1]$, que se llaman *vectores unitarios*. El primero traza una flecha que yace a lo largo del eje horizontal, y el segundo, una flecha que yace a lo largo del eje vertical. Debido a que son linealmente independientes, pueden servir como base del espacio bidimensional y, de hecho, se considera por lo común que el espacio bidimensional se genera por sus dos ejes, los cuales no son sino las versiones extendidas de los dos vectores unitarios.

Por analogía, el espacio vectorial de tres dimensiones es la totalidad de vectores en tres dimensiones, y debe ser generado exactamente por tres vectores tridimensionales linealmente independientes. Como ilustración, considere el conjunto de tres vectores unitarios

$$e_1 \equiv \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad e_2 \equiv \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \qquad e_3 \equiv \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{4.7}$$

FIGURA 4.3



donde cada e_i es un vector con 1 como su i-ésimo elemento y con ceros en los demás lugares.[3] Resulta obvio que estos tres vectores son linealmente independientes; de hecho, sus flechas yacen en los tres ejes del espacio tridimensional de la figura 4.3. Así que ellos abarcan el espacio tridimensional, lo que significa que todo el espacio tridimensional (R^3) se puede generar a partir de estos vectores unitarios. Por ejemplo, el vector $\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ se puede considerar una combinación lineal $e_1 + 2e_2 + 2e_3$. Desde el punto de vista geométrico, se puede sumar primero los vectores e_1 y $2e_2$ de la figura 4.3 por el método del paralelogramo, con el fin de obtener el vector representado por el punto $(1, 2, 0)$ en el plano x_1x_2, y luego sumar el último vector a $2e_3$, vía el paralelogramo construido en el plano sombreado vertical, para obtener el resultado final deseado: el punto $(1, 2, 2)$.

La extensión ulterior al espacio n-dimensional debe ser evidente. El espacio n-dimensional se puede definir como la totalidad de los vectores n-dimensionales. Aunque no es posible graficarlo, se puede considerar que un total de n (n componentes) vectores unitarios que son linealmente independientes cruzan el espacio n-dimensional. Cada vector n-dimensional, una n-tupla ordenada, representa un *punto* en el espacio n-dimensional, o una flecha que va del punto de origen (es decir, el vector nulo de n componentes) al punto antes mencionado. Y cualquier conjunto dado de n vectores n-dimensionales linealmente independientes es, de hecho, capaz de generar todo el espacio n-dimensional. Puesto que en el análisis cada componente del vector n-dimensional está restringido a ser un número real, este espacio n es de hecho R^n.

El espacio n-dimensional al que se ha hecho referencia a veces recibe el nombre más específico de *espacio euclidiano* de dimensión n (en honor a Euclides). Para explicar este último concepto, primero se debe comentar brevemente acerca del concepto de *distancia* entre dos puntos vectoriales. Para cualquier par de puntos u y v en un determinado espacio, la distancia de u a v es alguna función de valores reales

$$d = d(u, v)$$

[3] El símbolo *e* se puede relacionar con la palabra de origen alemán *eins*, para "uno".

con las siguientes propiedades: (1) cuando u y v coinciden, la distancia es cero; (2) cuando los dos puntos son distintos, la distancia de u a v y la distancia de v a u están representadas por el mismo número real positivo, y (3) la distancia entre u y v nunca es mayor que la distancia de u a w (un punto distinto de u y v) más la distancia de w a v. Expresado con símbolos,

$$
\begin{aligned}
&d(u, v) = 0 && (\text{para } u = v)\\
&d(u, v) = d(v, u) > 0 && (\text{para } u \neq v)\\
&d(u, v) \leq d(u, w) + d(w, v) && (\text{para } w \neq u, v)
\end{aligned}
$$

La última propiedad se conoce como la *desigualdad triangular*, porque los tres puntos u, v y w por lo común definen un triángulo cuando están juntos.

Cuando un espacio vectorial tiene una función de distancia definida que satisface las tres propiedades previas, se llama *espacio métrico*. Sin embargo, note que la distancia $d(u, v)$ ha sido analizada sólo en términos generales. Dependiendo de la forma específica asignada a la función d, podría resultar una diversidad de espacios métricos. El denominado espacio euclidiano es un tipo específico de espacio métrico, con una función de distancia definida como sigue: sea el punto u la n-tupla $(a_1, a_2, \ldots, a_n)$ y el punto v la n-tupla $(b_1, b_2, \ldots, b_n)$; entonces, la función de distancia euclidiana es

$$d(u, v) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \cdots + (a_n - b_n)^2}$$

donde la raíz cuadrada se toma como positiva. Como se puede comprobar fácilmente, esta función de distancia específica satisface las tres propiedades enumeradas antes. Aplicada al espacio bidimensional de la figura 4.2a, se encuentra que la distancia entre los dos puntos (6, 4) y (3, 2) es

$$\sqrt{(6-3)^2 + (4-2)^2} = \sqrt{3^2 + 2^2} = \sqrt{13}$$

Este resultado es congruente con el *teorema de Pitágoras*, el cual establece que la longitud de la hipotenusa de un triángulo rectángulo es igual a la raíz cuadrada (positiva) de la suma de los cuadrados de las longitudes de sus otros dos lados. Porque si se toma (6, 4) y (3, 2) como u y v, y se grafica un nuevo punto w en (6, 2), se tendrá de hecho un triángulo rectángulo con las longitudes de sus lados horizontal y vertical iguales a 3 y 2, respectivamente, y la longitud de la hipotenusa (la distancia entre u y v) igual a $\sqrt{3^2 + 2^2} = \sqrt{13}$.

La función de distancia euclidiana también se puede expresar en términos de la raíz cuadrada de un producto escalar de dos vectores. Puesto que u y v denotan las dos n-tuplas $(a_1, \ldots, a_n)$ y $(b_1, \ldots, b_n)$, se puede escribir un vector columna $u - v$, con elementos $a_1 - b_1, a_2 - b_2, \ldots, a_n - b_n$. Lo que va debajo del signo de raíz cuadrada en la función de distancia euclidiana es, por supuesto, simplemente la suma de cuadrados de estos n elementos, que, en vista del ejemplo 3 de esta sección, se puede escribir como el producto escalar $(u - v)'(u - v)$. Por consiguiente, se tiene

$$d(u, v) = \sqrt{(u - v)'(u - v)}$$

EJERCICIO 4.3

1. Dado $u' = [5 \quad 1 \quad 3]$, $v' = [3 \quad 1 \quad -1]$, $w' = [7 \quad 5 \quad 8]$ y $x' = [x_1 \quad x_2 \quad x_3]$, escriba los vectores columna, u, v, w y x, y encuentre

   (*a*) uv' (*c*) xx' (*e*) $u'v$ (*g*) $u'u$

   (*b*) uw' (*d*) $v'u$ (*f*) $w'x$ (*h*) $x'x$

2. Dados $w = \begin{bmatrix} 3 \\ 2 \\ 16 \end{bmatrix}$, $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ y $z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$:

   (*a*) ¿Cuáles de los siguientes están definidos: $w'x$, $x'y'$, xy', $y'y$, zz', yw', $x \cdot y$?

   (*b*) Encuentre los productos que están definidos.

3. Habiendo vendido n artículos de mercancía en cantidades $Q_1, \ldots, Q_n$ y precios $P_1, \ldots, P_n$, ¿cómo expresaría el ingreso total en (*a*) notación de sumatoria y (*b*) notación vectorial?

4. Dados dos vectores no nulos w_1 y w_2, el ángulo θ ($0° \leq \theta \leq 180°$) que forman, se relaciona con el producto escalar $w_1'w_2$ ($= w_2'w_1$) como sigue:

$$\theta \text{ es un ángulo } \begin{Bmatrix} \text{agudo} \\ \text{recto} \\ \text{obtuso} \end{Bmatrix} \text{ si y sólo si } \quad w_1'w_2 \begin{Bmatrix} > \\ = \\ < \end{Bmatrix} 0$$

   Compruebe esto calculando el producto escalar de cada uno de los pares de vectores siguientes (véanse las figuras 4.2 y 4.3):

   (*a*) $w_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$, $w_2 = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$

   (*b*) $w_1 = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$, $w_2 = \begin{bmatrix} -3 \\ -2 \end{bmatrix}$

   (*c*) $w_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$, $w_2 = \begin{bmatrix} -3 \\ -2 \end{bmatrix}$

   (*d*) $w_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $w_2 = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}$

   (*e*) $w_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$, $w_2 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$

5. Dado $u = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$ y $v = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$, obtenga en forma gráfica lo siguiente:

   (*a*) $2v$ (*b*) $u + v$ (*c*) $u - v$ (*d*) $v - u$ (*e*) $2u + 3v$ (*f*) $4u - 2v$

6. Puesto que los tres vectores unitarios definidos en (4.7) generan el espacio tridimensional, cualquier otro vector de tres dimensiones se debe poder expresar como una combinación lineal de e_1, e_2 y e_3. Muestre que los vectores de tres dimensiones siguientes se pueden expresar de este modo:

   (*a*) $\begin{bmatrix} 4 \\ 7 \\ 0 \end{bmatrix}$ (*b*) $\begin{bmatrix} 25 \\ -2 \\ 1 \end{bmatrix}$ (*c*) $\begin{bmatrix} -1 \\ 6 \\ 9 \end{bmatrix}$ (*d*) $\begin{bmatrix} 2 \\ 0 \\ 8 \end{bmatrix}$

7. En el espacio euclidiano tridimensional, ¿cuál es la distancia entre los siguientes puntos?

   (*a*) (3, 2, 8) y (0, −1, 5) (*b*) (9, 0, 4) y (2, 0, −4)

8. La desigualdad triangular se escribe con el signo de desigualdad *débil* $\leq$, y no con el signo de desigualdad estricto $<$. ¿En qué circunstancias aplicaría la parte "$=$" de la desigualdad?

9. Exprese la longitud del radio vector v en el espacio euclidiano n-dimensional (es decir, la distancia del origen al punto v) al usar cada uno de los siguientes conceptos:

   (*a*) escalares (*b*) un producto escalar (*c*) un producto interior

4.4 Leyes conmutativa, asociativa y distributiva

En el álgebra escalar ordinaria, las operaciones aditiva y multiplicativa obedecen las leyes conmutativa, asociativa y distributiva como sigue:

Ley conmutativa de adición:	$a + b = b + a$
Ley conmutativa de multiplicación:	$ab = ba$
Ley asociativa de adición:	$(a + b) + c = a + (b + c)$
Ley asociativa de multiplicación:	$(ab)c = a(bc)$
Ley distributiva:	$a(b + c) = ab + ac$

Éstas han sido mencionadas durante la explicación de las leyes con nombre similar aplicables a la unión e intersección de conjuntos. La mayoría de estas leyes, pero no todas, se aplica también a operaciones con matrices, la excepción importante es la ley conmutativa de la multiplicación.

Suma de matrices

La suma de matrices es conmutativa y asociativa. Esto se deduce del hecho de que la suma de matrices sólo requiere la suma de los elementos correspondientes de dos matrices y de que poco importa el orden en el que se suma cada par de elementos correspondientes. En este contexto, por cierto, la operación de resta $A - B$ se puede considerar simplemente como la operación de suma $A + (-B)$ y, por lo tanto, es innecesaria cualquier otra explicación.

Las leyes conmutativa y asociativa se expresan como sigue:

Ley conmutativa $\quad A + B = B + A$

PRUEBA $\quad A + B = [a_{ij}] + [b_{ij}] = [a_{ij} + b_{ij}] = [b_{ij} + a_{ij}] = B + A$

Ejemplo 1 Dadas $A = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$ y $B = \begin{bmatrix} 6 & 2 \\ 3 & 4 \end{bmatrix}$, se encuentra que

$$A + B = B + A = \begin{bmatrix} 9 & 3 \\ 3 & 6 \end{bmatrix}$$

Ley asociativa $\quad (A + B) + C = A + (B + C)$

PRUEBA
$$\begin{aligned}(A + B) + C &= [a_{ij} + b_{ij}] + [c_{ij}] = [a_{ij} + b_{ij} + c_{ij}] \\ &= [a_{ij}] + [b_{ij} + c_{ij}] = A + (B + C)\end{aligned}$$

Ejemplo 2 Dadas $v_1 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$, $v_2 = \begin{bmatrix} 9 \\ 1 \end{bmatrix}$ y $v_3 = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$, se encuentra que

$$(v_1 + v_2) - v_3 = \begin{bmatrix} 12 \\ 5 \end{bmatrix} - \begin{bmatrix} 2 \\ 5 \end{bmatrix} = \begin{bmatrix} 10 \\ 0 \end{bmatrix}$$

que es igual a

$$v_1 + (v_2 - v_3) = \begin{bmatrix} 3 \\ 4 \end{bmatrix} + \begin{bmatrix} 7 \\ -4 \end{bmatrix} = \begin{bmatrix} 10 \\ 0 \end{bmatrix}$$

Aplicada a la combinación lineal de vectores $k_1v_1 + \cdots + k_nv_n$, la ley asociativa permite seleccionar primero cualquier par de términos para suma (o resta), en vez de tener que seguir la secuencia en la cual se listan los n términos.

Multiplicación de matrices

La multiplicación de matrices *no* es conmutativa, es decir,

$$AB \neq BA$$

Como ya se explicó, aun cuando está definido AB, es posible que BA no lo esté; pero aun cuando ambos productos estén definidos, la regla general todavía es $AB \neq BA$.

Ejemplo 3 Sea $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ y $B = \begin{bmatrix} 0 & -1 \\ 6 & 7 \end{bmatrix}$, entonces

$$AB = \begin{bmatrix} 1(0)+2(6) & 1(-1)+2(7) \\ 3(0)+4(6) & 3(-1)+4(7) \end{bmatrix} = \begin{bmatrix} 12 & 13 \\ 24 & 25 \end{bmatrix}$$

pero

$$BA = \begin{bmatrix} 0(1)-1(3) & 0(2)-1(4) \\ 6(1)+7(3) & 6(2)+7(4) \end{bmatrix} = \begin{bmatrix} -3 & -4 \\ 27 & 40 \end{bmatrix}$$

Ejemplo 4 Sea u' un vector renglón de 1×3; entonces el vector columna correspondiente u debe ser 3×1. El producto $u'u$ será 1×1, pero el producto uu' será 3×3. Así, es evidente que $u'u \neq uu'$.

En vista de la regla general $AB \neq BA$, los términos *premultiplicar* y *posmultiplicar* se utilizan a menudo para especificar el orden de multiplicación. En el producto AB, se dice que la matriz B está *pre*multiplicada por A, y A está *pos*multiplicada por B.

Sin embargo, existen excepciones interesantes a la regla $AB \neq BA$. Un caso es cuando A es una matriz cuadrada y B es una matriz identidad. Otro es cuando A es la inversa de B, es decir, cuando $A = B^{-1}$. Ambas se consideran de nuevo más adelante. También se debe remarcar aquí que la multiplicación de una matriz por un escalar *obedece* la ley conmutativa; así, si k es un escalar, entonces

$$kA = Ak$$

Aunque en general no es conmutativa, la multiplicación de matrices *es* asociativa.

Ley asociativa $$(AB)C = A(BC) = ABC$$

Al formar el producto ABC, cada par *adyacente* de matrices debe satisfacer de modo natural la condición de conformabilidad. Si A es $m \times n$ y si C es $p \times q$, entonces la conformabilidad requiere que B sea $n \times p$:

$$\underset{(m\times n)}{A} \quad \underset{(n\times p)}{B} \quad \underset{(p\times q)}{C}$$

Note la apariencia dual de n y p en los indicadores de dimensión. Si se satisface la condición de conformabilidad, la ley asociativa establece que cualquier par *adyacente* de matrices se puede multiplicar primero, siempre que el producto sea insertado debidamente en el lugar exacto del par original.

Ejemplo 5 Si $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ y $A = \begin{bmatrix} a_{11} & 0 \\ 0 & a_{22} \end{bmatrix}$, entonces

$$x'Ax = x'(Ax) = [x_1 \quad x_2] \begin{bmatrix} a_{11}x_1 \\ a_{22}x_2 \end{bmatrix} = a_{11}x_1^2 + a_{22}x_2^2$$

El mismo resultado se obtiene de

$$(x'A)x = [a_{11}x_1 \quad a_{22}x_2] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = a_{11}x_1^2 + a_{22}x_2^2$$

En el ejemplo 5, la matriz cuadrada A tiene elementos no nulos a_{11} y a_{22} en la diagonal principal y ceros en cualquier otra parte. Esta clase de matriz se llama *matriz diagonal*. Cuando una matriz diagonal A aparece en el producto $x'Ax$, el producto resultante da una suma de cuadrados "ponderada"; los elementos de la diagonal de A suministran las ponderaciones para los términos x_1^2 y x_2^2. Este resultado contrasta con el producto escalar $x'x$, que produce una suma de cuadrados (no ponderada) simple.

Ejemplo 6

Sea una economía ideal definida por un nivel de ingreso nacional Y^0, además de una tasa de inflación p^0. Y suponga que cualquier desviación positiva del ingreso real Y respecto a Y^0 se considera igualmente indeseable que una desviación negativa de la misma magnitud, y de manera similar para desviaciones de la tasa de inflación real p respecto a p^0. Entonces se puede escribir una *función de pérdida social* como

$$\Lambda = \alpha(Y - Y^0)^2 + \beta(p - p^0)^2$$

donde α y β son las ponderaciones asignadas a las dos fuentes de pérdida social. Si se considera que las desviaciones de Y son el tipo de pérdida más grave, entonces α debe ser mayor que β. Note que el cuadrado de las desviaciones produce dos efectos. Primero, al elevar al cuadrado, una desviación positiva recibirá el mismo valor de pérdida que una desviación negativa de la misma magnitud numérica. Segundo, elevar al cuadrado causa que las desviaciones más grandes sean mucho más significativas en la medida de pérdida social que las desviaciones menores. Esta función de pérdida social se puede expresar, si se desea, mediante el producto de matrices

$$[Y - Y^0 \quad p - p^0]\begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix}\begin{bmatrix} Y - Y^0 \\ p - p^0 \end{bmatrix}$$

La multiplicación de matrices también es distributiva.

Ley distributiva $\quad A(B + C) = AB + AC \quad$ [premultiplicación por A]

$\quad (B + C)A = BA + CA \quad$ [posmultiplicación por A]

Por supuesto que en cada caso se deben observar las condiciones de conformabilidad para la suma, así como para la multiplicación.

EJERCICIO 4.4

1. Dadas $A = \begin{bmatrix} 3 & 6 \\ 2 & 4 \end{bmatrix}$, $B = \begin{bmatrix} -1 & 7 \\ 8 & 4 \end{bmatrix}$ y $C = \begin{bmatrix} 3 & 4 \\ 1 & 9 \end{bmatrix}$, compruebe que
   (*a*) $(A + B) + C = A + (B + C)$
   (*b*) $(A + B) - C = A + (B - C)$
2. La resta de una matriz B se puede considerar como la suma de la matriz $(-1)B$. ¿La ley conmutativa de suma permite expresar que $A - B = B - A$? Si no, ¿cómo corregiría el enunciado?
3. Pruebe la ley asociativa de multiplicación con las siguientes matrices:

$$A = \begin{bmatrix} 5 & 3 \\ 0 & 5 \end{bmatrix} \qquad B = \begin{bmatrix} -8 & 0 & 7 \\ 1 & 3 & 2 \end{bmatrix} \qquad C = \begin{bmatrix} 1 & 0 \\ 0 & 3 \\ 7 & 1 \end{bmatrix}$$

4. Pruebe que para dos escalares cualesquiera g y k
   (*a*) $k(A+B) = kA + kB$
   (*b*) $(g+k)A = gA + kA$
   (*Nota:* para *probar* un resultado, no se pueden usar ejemplos específicos.)
5. Para (*a*) a (*d*) obtenga $C = AB$.
   (*a*) $A = \begin{bmatrix} 12 & 14 \\ 20 & 5 \end{bmatrix} \qquad B = \begin{bmatrix} 3 & 9 \\ 0 & 2 \end{bmatrix}$
   (*b*) $A = \begin{bmatrix} 4 & 7 \\ 9 & 1 \end{bmatrix} \qquad B = \begin{bmatrix} 3 & 8 & 5 \\ 2 & 6 & 7 \end{bmatrix}$
   (*c*) $A = \begin{bmatrix} 7 & 11 \\ 2 & 9 \\ 10 & 6 \end{bmatrix} \qquad B = \begin{bmatrix} 12 & 4 & 5 \\ 3 & 6 & 1 \end{bmatrix}$
   (*d*) $A = \begin{bmatrix} 6 & 2 & 5 \\ 7 & 9 & 4 \end{bmatrix} \qquad B = \begin{bmatrix} 10 & 1 \\ 11 & 3 \\ 2 & 9 \end{bmatrix}$
   (*e*) Determine (i) $C = AB$ e (ii) $D = BA$, si
   $$A = \begin{bmatrix} -2 \\ 4 \\ 7 \end{bmatrix} \qquad B = \begin{bmatrix} 3 & 6 & -2 \end{bmatrix}$$
6. Pruebe que $(A+B)(C+D) = AC + AD + BC + BD$.
7. Si en la matriz A del ejemplo 5 todos sus elementos fueran cero, ¿$x'Ax$ todavía daría una suma de cuadrados ponderada? ¿Aún se aplicaría la ley asociativa?
8. Mencione algunas situaciones o contextos donde pudiera ser relevante el concepto de suma de cuadrados ponderada o no ponderada.

4.5 Matrices identidad y matrices nulas

Matrices identidad

Ya se ha hecho referencia al término *matriz identidad*, que se define como una matriz *cuadrada* con unos en su diagonal principal y ceros en cualquier otra parte. Se denota mediante el símbolo I, o I_n, en la que el subíndice n sirve para indicar su dimensión de renglón (así como de columna). Así,

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Ambas se pueden denotar también mediante I.

La importancia de este tipo especial de matriz radica en el hecho de que desempeña un papel similar al del número 1 en el álgebra escalar. Para cualquier número a, se tiene $1(a) = a(1) = a$. De manera similar, para cualquier matriz A, se tiene

$$IA = AI = A \tag{4.8}$$

Ejemplo 1 Sea $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & 3 \end{bmatrix}$, entonces

$$IA = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & 3 \end{bmatrix} = A$$

$$AI = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & 3 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & 3 \end{bmatrix} = A$$

Debido a que A es 2×3, la premultiplicación y posmultiplicación de A por I requerirían matrices identidad de dimensiones distintas, a saber, I_2 e I_3, respectivamente. Pero en caso de que A sea $n \times n$, entonces se puede usar la misma matriz identidad I_n, así que (4.8) se convierte en $I_nA = AI_n$, y esto ilustra una excepción a la regla de que la multiplicación de matrices no es conmutativa.

La naturaleza especial de las matrices identidad hace posible, durante el proceso de multiplicación, *insertar* o *eliminar* una matriz identidad sin afectar el producto de matrices. Esto se deduce directamente de (4.8). Recordando la ley asociativa, se tiene, por ejemplo,

$$\underset{(m\times n)}{A}\;\underset{(n\times n)}{I}\;\underset{(n\times p)}{B} = (AI)B = \underset{(m\times n)}{A}\;\underset{(n\times p)}{B}$$

la cual muestra que la presencia o ausencia de I no afecta el producto. Observe que la conformabilidad de dimensión se conserva ya sea que I aparezca en el producto o no.

Un caso interesante de (4.8) ocurre cuando $A = I_n$, porque entonces se tiene

$$AI_n = (I_n)^2 = I_n$$

que establece que una matriz identidad cuadrada es igual a sí misma. Una generalización de este resultado es que

$$(I_n)^k = I_n \qquad (k = 1, 2, \ldots)$$

Una matriz identidad permanece sin cambio cuando se multiplica por sí misma cualquier número de veces. Cualquier matriz con tal propiedad (a saber, $AA = A$) se denomina *matriz idempotente*.

Matrices nulas

Así como una matriz identidad I desempeña la función del número 1, una *matriz nula*, o *matriz cero*, denotada por 0, juega el papel del número cero. Una matriz nula es simplemente una matriz cuyos elementos son cero. A diferencia de I, la matriz cero no está restringida a ser cuadrada. Así, es posible escribir

$$\underset{(2\times 2)}{0} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \qquad \text{y} \qquad \underset{(2\times 3)}{0} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

etc. Una matriz nula cuadrada es idempotente, pero una rectangular no lo es. (¿Por qué?)

Como contraparte del número cero, las matrices nulas obedecen las siguientes reglas de operación (sujetas a conformabilidad) con respecto a la suma y la multiplicación:

$$\underset{(m\times n)}{A} + \underset{(m\times n)}{0} = \underset{(m\times n)}{0} + \underset{(m\times n)}{A} = \underset{(m\times n)}{A}$$

$$\underset{(m\times n)}{A}\;\underset{(n\times p)}{0} = \underset{(m\times p)}{0} \qquad \text{y} \qquad \underset{(q\times m)}{0}\;\underset{(m\times n)}{A} = \underset{(q\times n)}{0}$$

Note que, en la multiplicación, la matriz nula a la izquierda del signo igual y la que está a la derecha pueden tener dimensiones distintas.

Ejemplo 2

$$A+0=\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}+\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}=\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}=A$$

Ejemplo 3

$$\underset{(2\times 3)}{A}\ \underset{(3\times 1)}{0}=\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \end{bmatrix}=\underset{(2\times 1)}{0}$$

A la izquierda, la matriz nula es un vector nulo de 3×1; a la derecha, es un vector nulo de 2×1.

Características del álgebra de matrices

A pesar de las aparentes similitudes entre el álgebra de matrices y el álgebra escalar, el caso de las matrices presenta ciertas características que sirven como advertencia para no confiar ciegamente en el álgebra escalar. Ya se ha visto que, en general, $AB \neq BA$ en el álgebra de matrices. A continuación se consideran dos características más del álgebra de matrices.

Entre otras cosas, en el caso de escalares, la ecuación $ab = 0$ siempre implica que a o b es cero, pero esto no es así en la multiplicación de matrices. Así, se tiene

$$AB=\begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} -2 & 4 \\ 1 & -2 \end{bmatrix}=\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}=0$$

aunque ni A ni B es una matriz cero.

Como otra ilustración, para escalares, la ecuación $cd = ce$ (con $c \neq 0$) implica que $d = e$. Esto no se cumple para matrices. Así, dadas

$$C=\begin{bmatrix} 2 & 3 \\ 6 & 9 \end{bmatrix} \qquad D=\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \qquad E=\begin{bmatrix} -2 & 1 \\ 3 & 2 \end{bmatrix}$$

se encuentra que

$$CD=CE=\begin{bmatrix} 5 & 8 \\ 15 & 24 \end{bmatrix}$$

aun cuando $D \neq E$.

Estos resultados extraños pertenecen en realidad sólo a la clase especial de matrices conocidas como *matrices singulares*, de las cuales las matrices A, B y C son ejemplos. (Aproximadamente, estas matrices contienen un renglón que es un múltiplo de otro renglón.) No obstante, estos ejemplos revelan las dificultades de la extensión injustificada de teoremas algebraicos a operaciones con matrices.

EJERCICIO 4.5

Dadas $A=\begin{bmatrix} -1 & 5 & 7 \\ 0 & -2 & 4 \end{bmatrix}$, $b=\begin{bmatrix} 9 \\ 6 \\ 0 \end{bmatrix}$ y $x=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$:

1. Calcule: (*a*) AI (*b*) IA (*c*) Ix (*d*) $x'I$

Indique la dimensión de la matriz identidad utilizada en cada caso.

2. Calcule: (*a*) Ab (*b*) AIb (*c*) $x'IA$ (*d*) $x'A$
¿La inserción de I en (*b*) afecta el resultado en (*a*)? ¿La eliminación de I en (*d*) afecta el resultado en (*c*)?
3. ¿Cuál es la dimensión de la matriz nula que resulta de cada una de las siguientes operaciones?
(*a*) Premultiplicar A por una matriz nula de 5×2.
(*b*) Posmultiplicar A por una matriz nula de 3×6.
(*c*) Premultiplicar b por una matriz nula de 2×3.
(*d*) Posmultiplicar x por una matriz nula de 1×5.
4. Muestre que la matriz diagonal

$$\begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}$$

puede ser idempotente sólo si cada elemento diagonal es 1 o 0. ¿Cuántas matrices diagonales, idempotentes, numéricas, distintas de dimensión $n \times n$ se pueden construir en total de esta matriz?

4.6 Transpuestas e inversas

Cuando se intercambian los renglones y las columnas de una matriz A —de modo que su primer renglón se convierta en la primer columna, y viceversa—, se obtiene la *transpuesta* de A, que se denota por A' o A^T. La prima de ningún modo resulta una novedad; se usó antes para distinguir un vector renglón de un vector columna. En la terminología recién introducida, un vector renglón x' constituye el transpuesto del vector columna x. Es evidente que el superíndice T en el símbolo alternativo es la abreviatura para la palabra transpuesta.

Ejemplo 1

Dadas $\underset{(2\times3)}{A} = \begin{bmatrix} 3 & 8 & -9 \\ 1 & 0 & 4 \end{bmatrix}$ y $\underset{(2\times2)}{B} = \begin{bmatrix} 3 & 4 \\ 1 & 7 \end{bmatrix}$, se pueden intercambiar los renglones y las columnas y escribir

$$\underset{(3\times2)}{A'} = \begin{bmatrix} 3 & 1 \\ 8 & 0 \\ -9 & 4 \end{bmatrix} \qquad \text{y} \qquad \underset{(2\times2)}{B'} = \begin{bmatrix} 3 & 1 \\ 4 & 7 \end{bmatrix}$$

Por definición, si una matriz A es $m \times n$, entonces su transpuesta A' debe ser $n \times m$. Sin embargo, una matriz cuadrada $n \times n$ posee una transpuesta con la misma dimensión.

Ejemplo 2

Si $C = \begin{bmatrix} 9 & -1 \\ 2 & 0 \end{bmatrix}$ y $D = \begin{bmatrix} 1 & 0 & 4 \\ 0 & 3 & 7 \\ 4 & 7 & 2 \end{bmatrix}$, entonces

$$C' = \begin{bmatrix} 9 & 2 \\ -1 & 0 \end{bmatrix} \qquad \text{y} \qquad D' = \begin{bmatrix} 1 & 0 & 4 \\ 0 & 3 & 7 \\ 4 & 7 & 2 \end{bmatrix}$$

Aquí, la dimensión de cada transpuesta es idéntica a la de la matriz original.

¡En D', se observa también el resultado notable de que D' hereda no sólo la dimensión de D, sino también la disposición original de los elementos! El hecho de que $D' = D$ es el resultado de la simetría de los elementos con referencia a la diagonal principal. Si se considera a la diagonal principal en D como un espejo, los elementos localizados en el noreste son imágenes exactas de los elementos en su sudoeste; por lo tanto, el primer renglón es idéntico a la primera columna, y así sucesivamente. La matriz D ejemplifica la clase especial de matrices cuadradas conocidas como *matrices simétricas*. Otro ejemplo de esta clase de matriz es la matriz identidad I, la cual, como una matriz simétrica, tiene la transpuesta $I' = I$.

Propiedades de las transpuestas

Las transpuestas se caracterizan por las siguientes propiedades:

$$(A')' = A \tag{4.9}$$

$$(A+B)' = A' + B' \tag{4.10}$$

$$(AB)' = B'A' \tag{4.11}$$

La primera establece que la transpuesta de la transpuesta es la matriz original, una conclusión bastante evidente por sí misma.

La segunda propiedad se puede expresar de forma verbal así: la transpuesta de una suma es la suma de las transpuestas.

Ejemplo 3 Si $A = \begin{bmatrix} 4 & 1 \\ 9 & 0 \end{bmatrix}$ y $B = \begin{bmatrix} 2 & 0 \\ 7 & 1 \end{bmatrix}$, entonces

$$(A+B)' = \begin{bmatrix} 6 & 1 \\ 16 & 1 \end{bmatrix}' = \begin{bmatrix} 6 & 16 \\ 1 & 1 \end{bmatrix}$$

y

$$A' + B' = \begin{bmatrix} 4 & 9 \\ 1 & 0 \end{bmatrix} + \begin{bmatrix} 2 & 7 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 16 \\ 1 & 1 \end{bmatrix}$$

La tercera propiedad es que la transpuesta de un producto es el producto de las transpuestas *en orden inverso*. Para apreciar por qué se debe invertir el orden, se examina la conformabilidad de dimensión de dos productos en ambos lados de (4.11). Si se establece que A sea de $m \times n$ y B, de $n \times p$, entonces AB será de $m \times p$ y $(AB)'$ será de $p \times m$. Para que se cumpla la igualdad, es necesario que la expresión de la derecha $B'A'$ sea de la misma dimensión. Puesto que B' es de $p \times n$ y A' es de $n \times m$, el producto $B'A'$ es de hecho de $p \times m$, como se requiere. Así, se resuelve la dimensión de $B'A'$. Por otro lado, note que el producto $A'B'$ aún no está definido a menos que $m = p$.

Ejemplo 4 Dadas $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ y $B = \begin{bmatrix} 0 & -1 \\ 6 & 7 \end{bmatrix}$, se tiene

$$(AB)' = \begin{bmatrix} 12 & 13 \\ 24 & 25 \end{bmatrix}' = \begin{bmatrix} 12 & 24 \\ 13 & 25 \end{bmatrix}$$

y

$$B'A' = \begin{bmatrix} 0 & 6 \\ -1 & 7 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 12 & 24 \\ 13 & 25 \end{bmatrix}$$

Esto comprueba la propiedad.

Inversas y sus propiedades

Para una matriz dada A, siempre se puede deducir la transpuesta A'. Por otro lado, su matriz *inversa*, otro tipo de matriz "deducida", podría existir o no. La inversa de la matriz A, denotada por A^{-1}, se define sólo si A es una matriz cuadrada, en cuyo caso el inverso es la matriz que satisface la condición

$$AA^{-1} = A^{-1}A = I \tag{4.12}$$

Es decir, si A se premultiplica (o posmultiplica) por A^{-1}, el producto será la misma matriz identidad. Ésta es otra excepción a la regla de que la multiplicación de matrices no es conmutativa.

Es importante mencionar los siguientes puntos:

1. No toda matriz cuadrada tiene un inverso, la forma cuadrada es una condición *necesaria*, pero *no* una condición *suficiente*, para la existencia de un inverso. Si una matriz cuadrada A tiene un inverso, se dice que A es no singular; si A no posee inverso, se llama matriz *singular*.
2. Si A^{-1} existe, entonces se considera a la matriz A como el inverso de A^{-1}, así como A^{-1} es el inverso de A. En resumen, A y A^{-1} son inversas entre sí.
3. Si A es de $n \times n$, entonces A^{-1} también debe ser de $n \times n$; de lo contrario, no puede ser conformable para la premultiplicación ni para la posmultiplicación. La matriz identidad producida por la multiplicación también será de $n \times n$.
4. Si existe un inverso, entonces es único. Para probar su unicidad, supóngase que se encontró que B es un inverso para A, así que

$$AB = BA = I$$

Ahora suponga que hay otra matriz C tal que $AC = CA = I$. Al premultiplicar ambos lados de $AB = I$ por C, se encuentra que

$$CAB = CI(= C) \qquad \text{[por (4.8)]}$$

Puesto que por suposición $CA = I$, la ecuación anterior se reduce a

$$IB = C, \quad \text{o bien,} \quad B = C$$

Es decir, B y C debe ser una y la misma matriz inversa. Por esta razón, se puede hablar de *la* (a diferencia de *una*) inversa de A.

5. Las dos partes de la condición (4.12) —a saber, $AA^{-1} = I$ y $A^{-1}A = I$—, en realidad se implican mutuamente, así que para satisfacer cualquier ecuación es suficiente con establecer la relación inversa entre A y A^{-1}. Para probar esto, se debe mostrar que si $AA^{-1} = I$, y si hay una matriz B tal que $BA = I$, entonces $B = A^{-1}$ (así que $BA = I$ debe ser en efecto la ecuación $A^{-1}A = I$). A continuación se posmultiplican ambos lados de la ecuación dada $BA = I$ por A^{-1}; entonces

$$\begin{aligned} (BA)A^{-1} &= IA^{-1} \\ B(AA^{-1}) &= IA^{-1} && \text{[ley asociativa]} \\ BI &= IA^{-1} && [AA^{-1} = I \text{ por suposición}] \end{aligned}$$

Por lo tanto, como se requería,

$$B = A^{-1} \qquad \text{[por (4.8)]}$$

De manera análoga, se puede demostrar que, si $A^{-1}A = I$, entonces la única matriz C que produce $CA^{-1} = I$ es $C = A$.

Ejemplo 5 Sea $A = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$ y $B = \frac{1}{6}\begin{bmatrix} 2 & -1 \\ 0 & 3 \end{bmatrix}$; entonces, puesto que el multiplicador escalar $(\frac{1}{6})$ en B se puede mover a la parte posterior (ley conmutativa), se puede escribir

$$AB = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 2 & -1 \\ 0 & 3 \end{bmatrix}\frac{1}{6} = \begin{bmatrix} 6 & 0 \\ 0 & 6 \end{bmatrix}\frac{1}{6} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Esto establece a B como la inversa de A, y viceversa. La multiplicación inversa, como se esperaba, también produce la misma matriz identidad:

$$BA = \frac{1}{6}\begin{bmatrix} 2 & -1 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} = \frac{1}{6}\begin{bmatrix} 6 & 0 \\ 0 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Las siguientes tres propiedades de matrices inversas son de interés. Si A y B son matrices no singulares con dimensión $n \times n$, entonces

$$(A^{-1})^{-1} = A \tag{4.13}$$

$$(AB)^{-1} = B^{-1}A^{-1} \tag{4.14}$$

$$(A')^{-1} = (A^{-1})' \tag{4.15}$$

La primera establece que la inversa de una inversa es la matriz original. La segunda establece que la inversa de un producto es el producto de las inversas *en orden inverso*. Y la última significa que la inversa de la transpuesta es la transpuesta de la inversa. Note que en estas expresiones se presupone la existencia de las inversas y la satisfacción de la condición de conformabilidad.

La validez de (4.13) es bastante obvia, pero se procede a probar (4.14) y (4.15). Dado el producto AB, determínese su inverso, llamado C. De (4.12) se sabe que $CAB = I$; así, la posmultiplicación de ambos lados por $B^{-1}A^{-1}$ produce

$$CABB^{-1}A^{-1} = IB^{-1}A^{-1}\ (= B^{-1}A^{-1}) \tag{4.16}$$

Pero, el lado izquierdo se puede reducir a

$$\begin{aligned} CA(BB^{-1})A^{-1} &= CAIA^{-1} && \text{[por (4.12)]} \\ &= CAA^{-1} = CI = C && \text{[por (4.12) y (4.8)]} \end{aligned}$$

La sustitución de esto en (4.16) pone de manifiesto entonces que $C = B^{-1}A^{-1}$ o, en otras palabras, que el inverso de AB es igual a $B^{-1}A^{-1}$, como se supuso. En esta prueba, la ecuación $AA^{-1} = A^{-1}A = I$ se utilizó dos veces. Note que la aplicación de esta ecuación es permisible si y sólo si una matriz y su inversa son estrictamente adyacentes entre sí en un producto. Se puede escribir $AA^{-1}B = IB = B$, pero *nunca* $ABA^{-1} = B$.

La prueba de (4.15) es como sigue. Dada A', se procede a encontrar su inversa, llámela D. Por definición, se tiene entonces $DA' = I$. Pero se sabe que

$$(AA^{-1})' = I' = I$$

produce la misma identidad. Así, se puede escribir

$$\begin{aligned} DA' &= (AA^{-1})' \\ &= (A^{-1})'A' \qquad \text{[por (4.11)]} \end{aligned}$$

Al posmultiplicar ambos lados por $(A')^{-1}$, se obtiene

$$DA'(A')^{-1} = (A^{-1})'A'(A')^{-1}$$

o bien,
$$D = (A^{-1})' \qquad \text{[por (4.12)]}$$

Así, el inverso de A' es igual a $(A^{-1})'$, como se supuso.

En las pruebas recién presentadas, las operaciones matemáticas se llevaron a cabo en bloques completos de números. Si esos bloques de números no se hubieran tratado como entidades matemáticas (matrices), las mismas operaciones habrían sido mucho más largas y complicadas. La belleza del álgebra de matrices radica precisamente en la simplificación de tales operaciones.

Matriz inversa y solución de sistemas de ecuaciones lineales

La aplicación del concepto de matriz inversa a la solución de un sistema de ecuaciones simultáneas es inmediata y directa. En relación con el sistema de ecuaciones en (4.3), se señaló antes que se puede escribir en notación matricial como

$$\underset{(3\times3)}{A}\ \underset{(3\times1)}{x} = \underset{(3\times1)}{d} \tag{4.17}$$

donde A, x y d son como se definió en (4.4). Ahora bien, si existe la matriz inversa A^{-1}, la premultiplicación de ambos lados de la ecuación (4.17) por A^{-1} producirá

$$A^{-1}Ax = A^{-1}d$$

o bien,
$$\underset{(3\times1)}{x} = \underset{(3\times3)}{A^{-1}}\ \underset{(3\times1)}{d} \tag{4.18}$$

El lado izquierdo de (4.18) es un vector columna de variables, mientras que el producto del lado derecho es un vector columna de ciertos números conocidos. Así, por la definición de la igualdad de matrices o vectores, (4.18) muestra el conjunto de valores de las variables que satisfacen el sistema de ecuaciones, es decir, los valores solución. Además, puesto que A^{-1} es única si existe, $A^{-1}d$ debe ser un vector único de valores solución. Por lo tanto, el vector x de (4.18) se escribirá como x^*, para indicar su estado como una solución (única).

Los métodos para probar la existencia de la inversa y de su cálculo se analizan en el capítulo 5. Sin embargo, se puede expresar aquí que la inversa de la matriz A en (4.4) es

$$A^{-1} = \frac{1}{52}\begin{bmatrix} 18 & -16 & -10 \\ -13 & 26 & 13 \\ -17 & 18 & 21 \end{bmatrix}$$

Así (4.18) resulta ser

$$\begin{bmatrix} x_1^* \\ x_2^* \\ x_3^* \end{bmatrix} = \frac{1}{52}\begin{bmatrix} 18 & -16 & -10 \\ -13 & 26 & 13 \\ -17 & 18 & 21 \end{bmatrix}\begin{bmatrix} 22 \\ 12 \\ 10 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$$

que da la solución: $x_1^* = 2$, $x_2^* = 3$ y $x_3^* = 1$.

El resultado es que, como una manera de hallar la solución de un sistema de ecuaciones lineales $Ax = d$, donde la matriz de coeficientes A es no singular, primero se tiene que determinar la inversa A^{-1}, y luego posmultiplicar A^{-1} por el vector constante d. Así, el producto $A^{-1}d$ da los valores solución de las variables.

Ejemplo 6 Como se muestra en el ejemplo 11 de la sección 4.2, el modelo de ingreso nacional simple

$$Y = C + I_0 + G_0$$
$$C = a + bY$$

se puede escribir en notación de matrices como $Ax = d$, donde

$$A = \begin{bmatrix} 1 & -1 \\ -b & 1 \end{bmatrix} \qquad x = \begin{bmatrix} Y \\ C \end{bmatrix} \qquad \text{y} \qquad d = \begin{bmatrix} I_0 + G_0 \\ a \end{bmatrix}$$

La inversa de la matriz A es (véase la explicación en la sección 5.6)

$$A^{-1} = \frac{1}{1-b} \begin{bmatrix} 1 & 1 \\ b & 1 \end{bmatrix}$$

Por consiguiente, la solución del modelo es $x^* = A^{-1}d$, o bien,

$$\begin{bmatrix} Y^* \\ C^* \end{bmatrix} = \frac{1}{1-b} \begin{bmatrix} 1 & 1 \\ b & 1 \end{bmatrix} \begin{bmatrix} I_0 + G_0 \\ a \end{bmatrix} = \frac{1}{1-b} \begin{bmatrix} I_0 + G_0 + a \\ b(I_0 + G_0) + a \end{bmatrix}$$

EJERCICIO 4.6

1. Dada $A = \begin{bmatrix} 0 & 4 \\ -1 & 3 \end{bmatrix}$, $B = \begin{bmatrix} 3 & -8 \\ 0 & 1 \end{bmatrix}$ y $C = \begin{bmatrix} 1 & 0 & 9 \\ 6 & 1 & 1 \end{bmatrix}$, obtenga A', B' y C'.
2. Por medio de las matrices del problema 1, compruebe que
   (*a*) $(A + B)' = A' + B'$ (*b*) $(AC)' = C'A'$
3. Generalice el resultado (4.11) al caso de un producto de tres matrices al probar que, para matrices conformables cualesquiera A, B y C, se cumple la ecuación $(ABC)' = C'B'A'$.
4. Dadas las siguientes cuatro matrices, pruebe si alguna de ellas es la inversa de la otra:
$$D = \begin{bmatrix} 1 & 12 \\ 0 & 3 \end{bmatrix} \qquad E = \begin{bmatrix} 1 & 1 \\ 6 & 8 \end{bmatrix} \qquad F = \begin{bmatrix} 1 & -4 \\ 0 & \frac{1}{3} \end{bmatrix} \qquad G = \begin{bmatrix} 4 & -\frac{1}{2} \\ -3 & \frac{1}{2} \end{bmatrix}$$
5. Generalice el resultado (4.14) al probar que, para matrices no singulares conformables A, B y C, se cumple la ecuación $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$.
6. Sea $A = I - X(X'X)^{-1}X'$.
   (*a*) ¿A debe ser cuadrada? ¿$(X'X)$ debe ser cuadrado? ¿X debe ser cuadrada?
   (*b*) Muestre que la matriz A es idempotente. [*Nota:* si X' y X no son cuadradas, es inapropiado que se aplique (4.14).]

4.7 Cadenas de Markov finitas

Una aplicación común del álgebra de matrices se encuentra en lo que se conoce como procesos de Markov o cadenas de Markov. Los *procesos de Markov* se emplean para medir o estimar movimientos en el tiempo. Esto requiere el uso de una matriz de transición de Markov, donde cada valor en la matriz de transición es una probabilidad de pasar de un estado (ubicación, trabajo, etc.) a otro estado. También hay un vector que contiene la distribución inicial en los distintos estados. Al multiplicar repetidamente tal vector por la matriz de transición, se pueden estimar los cambios en los estados con el tiempo.

Considere el problema del movimiento interno de empleados dentro de una compañía que tiene muchas filiales o tiendas de distribución.[4] Una ilustración simple con dos filiales, como Abbotsford y Burnaby, ayudará a demostrar los conceptos básicos del proceso de Markov. Para determinar el número de empleados en Abbotsford mañana, se toma la probabilidad de que los empleados permanecerán en Abbotsford multiplicada por el número total de empleados que actualmente están en Abbotsford, lo cual da el total de empleados actuales en Abbotsford que seguirán en la tienda mañana. Aunado a esta cantidad está el número de empleados de Burnaby que son transferidos a Abbotsford. Esta cantidad se encuentra al multiplicar el número total de empleados que actualmente están en Burnaby por la probabilidad de que un empleado de Burnaby sea transferido a Abbotsford. De manera similar, el proceso sería el mismo para determinar el número de empleados que habrá mañana en la región de Burnaby, constituido por los empleados de Burnaby que optaron por permanecer allí y los empleados de Abbotsford que son transferidos hoy a la región de Burnaby. El proceso descrito tiene que ver con cuatro probabilidades. Estas cuatro probabilidades se pueden acomodar en una matriz. Esto se conoce como una matriz de transición de Markov (o simplemente, una "Markov").

Sean A_t y B_t las poblaciones de Abbotsford y Burnaby, respectivamente, en algún instante, t. Además, defina las probabilidades transicionales como sigue

$$P_{AA} \equiv \text{probabilidad de que una } A \text{ actual siga siendo una } A$$

$$P_{AB} \equiv \text{probabilidad de que una } A \text{ actual se mueva a } B$$

$$P_{BB} \equiv \text{probabilidad de que una } B \text{ actual siga siendo } B$$

$$P_{BA} \equiv \text{probabilidad de que una } B \text{ actual se mueva a } A.$$

Si la distribución de empleados en los lugares en el tiempo t se denota como un vector

$$x'_t = [A_t \quad B_t]$$

y las probabilidades de transición en forma de matriz

$$M = \begin{bmatrix} P_{AA} & P_{AB} \\ P_{BA} & P_{BB} \end{bmatrix}$$

entonces la distribución de empleados en estos lugares en el siguiente periodo $(t+1)$ es

$$\underset{(1\times 2)}{x'_t} \ \underset{(2\times 2)}{M} = \underset{(1\times 2)}{x'_{t+1}}$$

$$\begin{aligned}[A_t \quad B_t]\begin{bmatrix} P_{AA} & P_{AB} \\ P_{BA} & P_{BB} \end{bmatrix} &= [(A_tP_{AA} + B_tP_{BA})\,(A_tP_{AB} + B_tP_{BB})] \\ &= [A_{t+1} \quad B_{t+1}]\end{aligned}$$

[4] Nos gustaría agradecer a Sarah Dunn por este ejemplo. Este trabajo resulta de su proyecto final mientras era estudiante en el British Columbia Institute of Techonology, Burnaby, BC, Canadá (junio de 2003).

Para hallar la distribución de empleados después de dos periodos

$$[A_{t+1} \quad B_{t+1}]\begin{bmatrix} P_{AA} & P_{AB} \\ P_{BA} & P_{BB} \end{bmatrix} = [A_{t+2} \quad B_{t+2}]$$

$$[A_t \quad B_t]\begin{bmatrix} P_{AA} & P_{AB} \\ P_{BA} & P_{BB} \end{bmatrix}\begin{bmatrix} P_{AA} & P_{AB} \\ P_{BA} & P_{BB} \end{bmatrix} = [A_{t+2} \quad B_{t+2}]$$

$$[A_t \quad B_t]\begin{bmatrix} P_{AA} & P_{AB} \\ P_{BA} & P_{BB} \end{bmatrix}^2 = [A_{t+2} \quad B_{t+2}]$$

En general, para n periodos

$$[A_t \quad B_t]\begin{bmatrix} P_{AA} & P_{AB} \\ P_{BA} & P_{BB} \end{bmatrix}^n = [A_{t+n} \quad B_{t+n}]$$

La matriz de probabilidad M de 2×2 se conoce como *matriz de transición de Markov*. Para el caso donde n es exógena, el proceso se conoce como una cadena finita de Markov.

Ejemplo 1

Supóngase que la distribución inicial de empleados en los dos lugares en el tiempo $t = 0$ es

$$x_0' = [A_0 \quad B_0] = [100 \quad 100]$$

En otras palabras, al inicio hay cantidades iguales en cada lugar. Además, sean las probabilidades de transición en forma matricial como sigue:

$$M = \begin{bmatrix} P_{AA} & P_{AB} \\ P_{BA} & P_{BB} \end{bmatrix} = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}$$

Entonces la distribución de empleados en los lugares en el siguiente periodo ($t = 1$) es

$$[100 \quad 100]\begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} = [110 \quad 90] = [A_1 \quad B_1]$$

La distribución después de dos periodos se determina mediante

$$[100 \quad 100]\begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}^2 = [100 \quad 100]\begin{bmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{bmatrix}$$
$$= [113 \quad 87] = [A_2 \quad B_2]$$

La distribución después de 10 periodos ($t = 10$) está dada por

$$[100 \quad 100]\begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}^{10} = [100 \quad 100]\begin{bmatrix} 0.5174 & 0.4286 \\ 0.5174 & 0.4286 \end{bmatrix}$$
$$= [114.3 \quad 85.7] = [A_{10} \quad B_{10}]$$

Observe lo que sucede cuando la matriz de transición de Markov se eleva a potencias cada vez más altas. La nueva matriz de transición encontrada al elevar la matriz original a potencias cada vez más grandes converge en una matriz donde los renglones son idénticos. Ésta se conoce como el *estado estable*. ¿A qué esperaría que se pareciera el decimoprimer periodo o periodos superiores de la distribución?

Caso especial: cadenas absorbentes de Markov

Ahora, se extenderá el modelo al agregar una tercera opción: los empleados pueden salir de la compañía, con

$$P_{AE} \equiv \text{probabilidad de que una } A \text{ actual elija salir } (E)$$

$$P_{BE} \equiv \text{probabilidad de que una } B \text{ actual elija salir } (E)$$

En este punto, se añaden las suposiciones siguientes:

$$P_{EA} = 0 \qquad P_{EB} = 0 \qquad P_{EE} = 1$$

donde P_{EA}, P_{EB} y P_{EE} son las probabilidades de que un empleado que en la actualidad está en E opte por A, B o E, respectivamente. En otras palabras, nadie que sale de la compañía vuelve a entrar. Asimismo, estas restricciones implican que la compañía nunca reemplaza a los empleados que salen (no hay nuevas contrataciones).

Comenzando en el tiempo $t = 0$, la cadena de Markov ahora se convierte en

$$[A_0 \quad B_0 \quad E_0] \begin{bmatrix} P_{AA} & P_{AB} & P_{AE} \\ P_{BA} & P_{BB} & P_{BE} \\ P_{EA} & P_{EB} & P_{EE} \end{bmatrix}^n = [A_n \quad B_n \quad E_n]$$

$$[A_0 \quad B_0 \quad E_0] \begin{bmatrix} P_{AA} & P_{AB} & P_{AE} \\ P_{BA} & P_{BB} & P_{BE} \\ 0 & 0 & 1 \end{bmatrix}^n = [A_n \quad B_n \quad E_n]$$

(Suponga que $E_0 = 0$.)

Este tipo de proceso de Markov se conoce como *cadena de Markov absorbente*. Como resultado de los valores de las probabilidades de transición encontrados en el tercer renglón, se ve que una vez que un empleado se convierte en una E en un estado (tiempo), ese empleado permanece allí en los futuros estados (tiempos). A medida que n va al infinito, A_n y B_n se aproximan a cero y E_n se aproxima al valor del número total de trabajadores en el tiempo cero (es decir, $A_0 + B_0 + E_0$).

EJERCICIO 4.7

1. Considere la situación de un despido masivo (es decir, el cierre de una fábrica) donde laboraban 1 200 empleados y ahora comienzan a buscar trabajo. En este caso hay dos estados: empleado (E) y desempleado (U) con un vector inicial

   $$x_0' = [E \quad U] = [0 \quad 1\,200]$$

   Suponga que en un periodo determinado una persona desempleada encontrará trabajo con probabilidad 0.7 y, por lo tanto, permanecerá desempleado con una probabilidad de 0.3. Además, las personas que tienen empleo podrían perderlo en algún periodo con una probabilidad de 0.1 (y tendrán una probabilidad 0.9 de permanecer empleados).

   (*a*) Plantee una matriz de transición de Markov para este problema.

   (*b*) ¿Cuál será el número de personas desempleadas después de (i) 2 periodos; (ii) 3 periodos; (iii) 5 periodos; (iv) 10 periodos?

   (*c*) ¿Cuál es el nivel de estado estable de desempleo?

Capítulo 5

Modelos lineales y álgebra de matrices (continuación)

En el capítulo 4 se mostró que un sistema de ecuaciones lineales, por grande que sea, se puede escribir en una notación matricial compacta. Además, esta clase de sistema de ecuaciones se puede resolver al hallar la inversa de la matriz de coeficientes, siempre que exista la inversa. Ahora centraremos la atención en las preguntas de cómo probar la existencia de la inversa y cómo hallarla. Sólo después de que hayamos contestado estas preguntas será posible aplicar de forma significativa el álgebra matricial a modelos económicos.

5.1 Condiciones de la no singularidad de una matriz

Una determinada matriz de coeficientes A puede tener inversa (es decir, puede ser "no singular") sólo si es cuadrada. Sin embargo, como ya señalamos, la condición de ser cuadrada es necesaria pero no suficiente para que la inversa A^{-1} exista. Una matriz puede ser cuadrada y a la vez singular (sin una inversa).

Condiciones necesarias *versus* suficientes

Los conceptos "condición necesaria" y "condición suficiente" se emplean con frecuencia en economía, Por ello es importante conocer su significado preciso antes de seguir adelante.

Una condición necesaria tiene el carácter de prerrequisito: suponga que una afirmación p es cierta *sólo si* otra afirmación q es verdadera; entonces q constituye una condición necesaria de p. De forma simbólica esto se expresa como sigue:

$$p \Rightarrow q \tag{5.1}$$

que se lee como "*p sólo si q*", o de otro modo, "si p, entonces q". También es correcto interpretar que (5.1) significa "*p implica q*". Por supuesto, puede suceder que también se tenga $p \Rightarrow w$ al mismo tiempo. Entonces tanto q como w son condiciones necesarias para p.

Ejemplo 1 Si p es la afirmación "una persona es un padre" y q es la frase "una persona es varón", entonces se aplica el enunciado lógico $p \Rightarrow q$. Una persona es un padre *sólo si* es varón, y ser varón es una condición necesaria para la paternidad. No obstante, observe que no se cumple lo contrario: la paternidad no es una condición necesaria para la masculinidad.

Un tipo diferente de situación es una en la que una afirmación p es verdadera si q lo es, pero p puede ser verdadera también cuando q no lo es. En este caso, se dice que q es una condición suficiente para p. La verdad de q es suficiente para establecer la verdad de p, pero no es una condición necesaria para p. Este caso se expresa en forma simbólica mediante

$$p \Leftarrow q \tag{5.2}$$

que se lee: "*p si q*" (sin la palabra *sólo*), o en forma alternativa, "si q, entonces p", como si (5.2) se leyera hacia atrás. También se puede interpretar que significa "q implica p".

Ejemplo 2

Si p simboliza la afirmación "se puede ir a Europa" y q es la expresión "se viaja en avión a Europa", entonces $p \Leftarrow q$. Volar puede servir para que uno llegue a Europa, pero puesto que el transporte por el océano también es factible, volar no es un prerrequisito. Se puede escribir $p \Leftarrow q$, pero no $p \Rightarrow q$.

En una tercera situación posible, q es necesaria y suficiente para p. En tal caso, se escribe

$$p \Leftrightarrow q \tag{5.3}$$

que se lee: "*p si y sólo si q*". La flecha de doble cabeza es en realidad una combinación de los dos tipos de flecha en (5.1) y (5.2); por consiguiente, la unión utiliza los dos términos "si" y "sólo si". Note que (5.3) establece no sólo que p implica q sino también que q implica p.

Ejemplo 3

Si la afirmación "hay menos de treinta días en el mes" se representa con p y la frase "es el mes de febrero" con q, entonces $p \Leftrightarrow q$. Para tener menos de treinta días en el mes, es necesario que éste sea febrero. A la inversa, la especificación de febrero es suficiente para establecer que hay menos de treinta días en el mes. Así, q es una condición necesaria y suficiente para p.

A fin de probar $p \Rightarrow q$, es necesario mostrar que q se deduce de manera lógica de p. Análogamente, para probar $p \Leftarrow q$, se requiere una demostración de que p se deduce lógicamente de q. Pero para probar $p \Leftrightarrow q$ se necesita una demostración de que p y q se deducen una de otra.

Las condiciones necesarias y suficientes son importantes como mecanismos de selección. Considere un grupo de candidatos para becas, o para posiciones de trabajo. Puesto que las condiciones necesarias se consideran prerrequisitos, sirven para separar los candidatos en dos grupos: los que no satisfacen las condiciones necesarias se descalifican automáticamente; los que satisfacen las condiciones necesarias permanecen como candidatos admisibles. Sin embargo, permanecer como un candidato admisible no garantiza que el candidato al final tenga éxito. Así, las condiciones necesarias son más concluyentes en seleccionar a los candidatos no exitosos que en identificar a los exitosos. En general, debemos tener en mente que las condiciones necesarias *no son suficientes* por sí mismas.

A diferencia de las condiciones necesarias, las condiciones suficientes sirven directamente para identificar a candidatos exitosos. Un candidato que satisface una condición suficiente es de forma automática uno exitoso. Así como las condiciones necesarias no son suficientes por sí mismas, las condiciones suficientes no son necesarias por sí mismas. Esto es porque, junto con alguna condición suficiente dada, podrían existir otras condiciones suficientes menos estrictas, y el candidato que no satisface la condición suficiente establecida aún puede calificar con una condición suficiente menos rígida. Por ejemplo, una calificación A es suficiente para aprobar un curso, pero no es una condición necesaria puesto que una calificación B también es suficiente.

El mecanismo de selección más eficaz se encuentra en las condiciones necesarias y suficientes. El hecho de no satisfacer esta clase de condición significa que el candidato en definitiva está fuera, y la satisfacción de esta condición significa que el candidato está dentro. Se puede hallar una aplicación inmediata de esto en el presente análisis de no singularidad de una matriz.

Condiciones de no singularidad

Después de que se satisface la condición de ser cuadrada (una condición necesaria), una condición suficiente de la no singularidad de una matriz es que sus renglones sean linealmente independientes (o, lo que es lo mismo, que sus *columnas* sean linealmente independientes). Cuando se toman juntas las dos condiciones: ser cuadrada e independencia lineal, constituyen la condición necesaria y suficiente de no singularidad (no singularidad $\Leftrightarrow$ cuadratura *e* independencia lineal).

Una matriz A de coeficientes de $n \times n$ se puede considerar como un conjunto ordenado de vectores renglón, es decir, como un vector columna cuyos elementos son por sí mismos vectores renglón:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} v'_1 \\ v'_2 \\ \vdots \\ v'_n \end{bmatrix}$$

donde $v'_i = [a_{i1} \quad a_{i2} \quad \cdots \quad a_{in}]$, $i = 1, 2, \ldots, n$. Para que los renglones (vectores renglón) sean linealmente independientes, ninguno debe ser una combinación lineal del resto. De manera más formal, como se mencionó en la sección 4.3, la independencia lineal de los renglones requiere que el único conjunto de escalares k_i que puede satisfacer la ecuación vectorial

$$\sum_{i=1}^{n} k_i v'_i = \underset{(1\times n)}{0} \qquad \textbf{(5.4)}$$

sea $k_i = 0$ para toda i.

Ejemplo 4

Si la matriz de coeficientes es

$$A = \begin{bmatrix} 3 & 4 & 5 \\ 0 & 1 & 2 \\ 6 & 8 & 10 \end{bmatrix} = \begin{bmatrix} v'_1 \\ v'_2 \\ v'_3 \end{bmatrix}$$

entonces, puesto que $[6 \quad 8 \quad 10] = 2[3 \quad 4 \quad 5]$, se tiene $v'_3 = 2v'_1 = 2v'_1 + 0v'_2$. Así, el tercer renglón se puede expresar como una combinación lineal de los dos primeros, y los renglones *no* son linealmente independientes. Alternativamente, la ecuación previa se puede escribir como

$$2v'_1 + 0v'_2 - v'_3 = [6 \quad 8 \quad 10] + [0 \quad 0 \quad 0] - [6 \quad 8 \quad 10] = [0 \quad 0 \quad 0]$$

En vista de que el conjunto de escalares que condujo al vector cero de (5.4) no es $k_i = 0$ para toda i, se deduce que los renglones son linealmente dependientes.

A diferencia de la condición de cuadratura, la condición de independencia lineal normalmente no se puede determinar de un vistazo. Así que se debe contar con un método para probar la independencia lineal entre renglones (o columnas). Sin embargo, antes de que emprendamos la tarea, convendría primero fortalecer nuestra motivación para tener una idea intuitiva de por qué la condición de independencia lineal se agrega a la condición de cuadratura como un todo. Del procedimiento de contar ecuaciones e incógnitas en la sección 3.4, recordamos la conclusión general de que, para que un sistema de ecuaciones posea una solución única, no es suficiente tener el mismo número de ecuaciones e incógnitas. Además, las *ecuaciones* deben ser consistentes entre sí y funcionalmente independientes entre sí (lo cual significa independencia lineal, en el contexto presente de los sistemas lineales). Hay una relación bastante obvia entre el criterio "mismo número de ecuaciones que incógnitas" y la *cuadratura* (mismo número de renglones y columnas) de la matriz de coeficientes. Lo que hace el requerimiento "independencia lineal entre los renglones" es imposibilitar la inconsistencia así como la dependencia lineal *entre las ecuaciones*. Por lo tanto, en conjunto, el doble requerimiento de cuadratura e independencia lineal de renglones en la matriz de coeficientes es equivalente a las condiciones de existencia de una solución única enunciada en la sección 3.4.

A continuación se ilustra cómo la dependencia lineal *entre los renglones* de la matriz de coeficientes puede causar inconsistencia o dependencia lineal *entre las propias ecuaciones*. Sea que el sistema de ecuaciones $Ax = d$ toma la forma

$$\begin{bmatrix} 10 & 4 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$$

donde la matriz de coeficientes A contiene renglones linealmente dependientes: $v_1' = 2v_2'$. (Observe que sus columnas también son dependientes, la primera es $\frac{5}{2}$ de la segunda.) No se han especificado los valores de los términos constantes d_1 y d_2, pero sólo hay *dos* posibilidades distintas en relación con sus valores relativos: (1) $d_1 = 2d_2$ y (2) $d_1 \neq 2d_2$. En la primera —con $d_1 = 12$ y $d_2 = 6$, por ejemplo—, las dos ecuaciones son consistentes pero *linealmente dependientes* (al igual que los dos renglones de la matriz A), porque la primera ecuación es simplemente la segunda multiplicada por 2. Así, una ecuación es redundante, y el sistema se reduce en efecto a una sola ecuación, $5x_1 + 2x_2 = 6$, con un número infinito de soluciones. Para la segunda posibilidad —con $d_1 = 12$ pero $d_2 = 0$, por ejemplo—, las dos ecuaciones son *inconsistentes*, porque si la primera ecuación ($10x_1 + 4x_2 = 12$) es cierta, entonces, al dividir entre dos cada término, se deduce que $5x_1 + 2x_2 = 6$; en consecuencia, si esta primera ecuación es cierta, la segunda ecuación ($5x_1 + 2x_2 = 0$) no es posible que también sea cierta. Por lo que no existe solución.

La conclusión es que no habrá solución única (habrá infinidad de soluciones o no habrá solución) mientras los renglones de la matriz de coeficientes A sean linealmente dependientes. De hecho, la única forma de tener una solución única es que los renglones (o columnas) sean linealmente independientes en la matriz de coeficientes. En ese caso, la matriz A será no singular, lo que significa que la inversa A^{-1} existe, y se puede hallar una solución única $x^* = A^{-1}d$.

Rango de una matriz

Aunque el concepto de independencia en los renglones ha sido explicado sólo en relación con las matrices cuadradas, es igualmente aplicable a cualquier matriz rectangular de $m \times n$. Si el número máximo de renglones linealmente independientes que se puede hallar en esta matriz es r, se dice que la matriz es de *rango r*. (El rango también indica el número máximo de *columnas* linealmente independientes en dicha matriz.) El rango de una matriz de $m \times n$ puede ser a lo sumo m o n, cualquiera que sea el más pequeño.

Dada una matriz con sólo dos renglones (o dos columnas), la independencia de renglones (o columnas) se comprueba fácilmente mediante una inspección visual, sólo se tiene que comprobar si un renglón (columna) es el múltiplo exacto del otro. Pero, para una matriz de dimensión mayor, la inspección visual podría no ser factible, y se requiere un método más formal. Un método para hallar el rango de una matriz A (no necesariamente cuadrada), es decir, para determinar el número de renglones independientes en A, implica transformar A en una *matriz escalonada* mediante ciertas "operaciones elementales en los renglones". Una característica estructural particular de la matriz escalonada indica entonces el rango de la matriz A.

Hay sólo tres tipos de *operaciones elementales en los renglones* en una matriz:[1]

1. Intercambio de dos renglones cualesquiera en la matriz.
2. Multiplicación (o división) de un renglón por algún escalar $k \neq 0$.
3. Suma de "k veces cualquier renglón" a otro renglón.

Si bien cada una de estas operaciones convierte una determinada matriz A en una forma diferente, ninguna de ellas modifica el rango. Es esta característica de las operaciones elementales en los renglones la que permite leer el rango de A de su matriz escalonada. La forma más fácil de explicar el método de la matriz escalonada es mediante un ejemplo específico.

Ejemplo 5

Determine el rango de la matriz

$$A = \begin{bmatrix} 0 & -11 & -4 \\ 2 & 6 & 2 \\ 4 & 1 & 0 \end{bmatrix}$$

a partir de su forma escalonada. Primero, se comprueba en la primera columna de A la presencia de elementos cero. Si hay elementos cero en la columna 1, se mueven a la parte inferior de la matriz. En el caso de A, se desea mover el 0 (primer elemento de la columna 1) a la parte inferior de esa columna, lo cual se puede llevar a cabo al intercambiar el renglón 1 y el renglón 3 (por medio de la primera operación elemental). El resultado es

$$A_1 = \begin{bmatrix} 4 & 1 & 0 \\ 2 & 6 & 2 \\ 0 & -11 & -4 \end{bmatrix}$$

El siguiente objetivo es cambiar la primera columna de A_1 en un vector unitario e_1 como se define en (4.7). Para transformar el elemento 4 en la unidad, se divide el renglón 1 de A_1 entre el escalar 4 (aplicando la segunda operación elemental), lo cual produce

$$A_2 = \begin{bmatrix} 1 & \frac{1}{4} & 0 \\ 2 & 6 & 2 \\ 0 & -11 & -4 \end{bmatrix}$$

Entonces, para transformar el elemento 2 de la columna 1 de A_2 en 0, se multiplica el renglón 1 de A_2 por -2, y luego se suma el resultado al renglón 2 de A_2 (aplicando la tercera operación elemental). La matriz resultante,

$$A_3 = \begin{bmatrix} 1 & \frac{1}{4} & 0 \\ 0 & 5\frac{1}{2} & 2 \\ 0 & -11 & -4 \end{bmatrix}$$

[1] De manera similar a las operaciones elementales de los renglones, pueden definirse operaciones elementales en las columnas. Para los fines que aquí perseguimos, son suficientes las operaciones elementales en los renglones.

ahora tiene el vector unitario deseado e_1 como su primera columna. Una vez logrado esto, dejamos de lado el primer renglón de A_3 —el cual consideraremos más adelante— y continuamos sólo con los dos renglones restantes, donde deseamos crear un vector unitario de dos elementos en la segunda columna, al transformar el elemento $5\frac{1}{2}$ en 1, y el elemento -11 en 0. Para este fin, se requiere dividir el renglón 2 de A_3 entre $5\frac{1}{2}$, y de este modo se cambia el renglón en el vector $[0 \quad 1 \quad \frac{4}{11}]$; luego, se agrega 11 veces este vector al renglón 3 de A_3. El resultado final, en la forma de

$$A_4 = \begin{bmatrix} 1 & \frac{1}{4} & 0 \\ 0 & 1 & \frac{4}{11} \\ 0 & 0 & 0 \end{bmatrix}$$

ejemplifica la matriz escalonada, que, por definición, posee tres características estructurales. Primera, los renglones no cero (renglones con por lo menos un elemento no cero) aparecen arriba de los renglones cero (renglones que contienen solamente ceros). Segunda, en todo renglón no cero, el primer elemento no cero es la unidad. Tercera, el elemento unitario (el primer elemento no cero) en cualquier renglón debe aparecer a la izquierda del elemento unitario homólogo del renglón justo después. Por ahora debe quedar claro que todas las operaciones elementales en los renglones que hemos llevado a cabo están diseñadas para producir estas características en A_4.

Ahora bien, podemos leer simplemente el rango de A por el número de renglones no cero presentes en la matriz escalonada A_4. Puesto que A_4 contiene dos renglones no cero, podemos concluir que $r(A) = 2$. Por supuesto, éste es también el rango de las matrices A_1 a A_4, porque las operaciones elementales de los renglones no modifican el rango de una matriz.

El método de transformación de matriz escalonada se aplica tanto a matrices cuadradas como a no cuadradas. Para el ejemplo 5, hemos elegido una matriz cuadrada porque nuestro objetivo inmediato es atender la cuestión de no singularidad, que pertenece sólo a matrices cuadradas. Por definición, para que una matriz A de $n \times n$ sea no singular, debe tener n renglones linealmente independientes (o columnas); en consecuencia, debe ser de rango n y su matriz escalonada debe contener exactamente n renglones no cero, con ningún renglón cero en absoluto. Por el contrario, una matriz de $n \times n$ que tenga rango n debe ser no singular. Así, una matriz escalonada de $n \times n$ sin ningún renglón cero debe ser no singular, ya que es la matriz de la cual se obtiene la matriz escalonada mediante operaciones elementales en los renglones. En el ejemplo 5, la matriz A es de 3×3, pero $r(A) = 2$; por lo tanto, A es singular.

EJERCICIO 5.1

1. En las siguientes proposiciones por pares, sea p la primera y q la segunda proposición. Indique para cada caso si se aplica (5.1), (5.2) o (5.3).
   (*a*) Es un día de fiesta; es el día de acción de gracias.
   (*b*) Una figura geométrica tiene cuatro lados; es un rectángulo.
   (*c*) Dos pares ordenados (a, b) y (b, a) son iguales; a es igual a b.
   (*d*) Un número es racional; éste se puede expresar como un cociente de dos enteros.
   (*e*) Una matriz de 4×4 es no singular; el rango de la matriz de 4×4 es 4.
   (*f*) El tanque de gasolina de mi automóvil está vacío; no puedo encender mi automóvil.
   (*g*) La carta se devuelve al remitente con la leyenda "destinatario desconocido"; el remitente escribió mal la dirección en el sobre.

2. Sea p la proposición "una figura geométrica es un cuadrado" y sea q como sigue:
   (*a*) Tiene cuatro lados.
   (*b*) Tiene cuatro lados iguales.
   (*c*) Tiene cuatro lados iguales, cada uno perpendicular al adyacente.
   ¿Cuál es verdadero en cada caso: $p \Rightarrow q$, $p \Leftarrow q$ o $p \Leftrightarrow q$?
3. ¿Son linealmente independientes los renglones en cada una de las siguientes matrices?
   (*a*) $\begin{bmatrix} 24 & 8 \\ 9 & -3 \end{bmatrix}$ (*b*) $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ (*c*) $\begin{bmatrix} 0 & 4 \\ 3 & 2 \end{bmatrix}$ (*d*) $\begin{bmatrix} -1 & 5 \\ 2 & -10 \end{bmatrix}$
4. Compruebe si las columnas de cada matriz del problema 3 son también linealmente independientes. ¿Obtuvo la misma respuesta para la independencia de los renglones?
5. Determine el rango de cada una de las siguientes matrices a partir de su matriz escalonada, y comente acerca de la cuestión de no singularidad.
   (*a*) $A = \begin{bmatrix} 1 & 5 & 1 \\ 0 & 3 & 9 \\ -1 & 0 & 0 \end{bmatrix}$ (*c*) $C = \begin{bmatrix} 7 & 6 & 3 & 3 \\ 0 & 1 & 2 & 1 \\ 8 & 0 & 0 & 8 \end{bmatrix}$
   (*b*) $B = \begin{bmatrix} 0 & -1 & -4 \\ 3 & 1 & 2 \\ 6 & 1 & 0 \end{bmatrix}$ (*d*) $D = \begin{bmatrix} 2 & 7 & 9 & -1 \\ 1 & 1 & 0 & 1 \\ 0 & 5 & 9 & -3 \end{bmatrix}$
6. Por definición de la dependencia lineal entre los renglones de una matriz, uno o más renglones se pueden expresar como una combinación lineal de algunos otros renglones. En la matriz escalonada, la presencia de uno o más renglones ceros indica la dependencia lineal. ¿Qué proporciona el enlace entre la presencia de una combinación de renglones lineales en una determinada matriz y la presencia de renglones cero en la matriz escalonada?

5.2 Prueba de no singularidad mediante el uso del determinante

Para determinar si una matriz es no singular, se puede hacer uso también del determinante.

Determinantes y no singularidad

El determinante de una matriz cuadrada A, denotada por $|A|$, es un escalar (número) definido unívocamente relacionado con esa matriz. Los determinantes se definen sólo en matrices *cuadradas*. La matriz más pequeña posible es, por supuesto, la matriz $A = [a_{11}]$ de 1×1. Por definición, su determinante es igual al único elemento a_{11}: $|A| = |a_{11}| = a_{11}$. El símbolo $|a_{11}|$ no se debe confundir con el de valor absoluto de un número, que se escribe de manera similar. En el contexto de valor absoluto, se tiene, por ejemplo, no sólo $|5| = 5$, sino también $|-5| = 5$, porque el valor absoluto de un número es su valor numérico sin considerar el signo algebraico. En cambio, el símbolo de determinante conserva el signo del elemento, así que mientras $|8| = 8$ (un número positivo), tenemos $|-8| = -8$ (un número negativo). Esta distinción prueba ser crucial en la última explicación cuando se aplica a pruebas de determinante cuyos resultados dependen en forma crítica de los signos de determinantes de varias dimensiones, entre otros los de 1×1, como por ejemplo $|a_{11}| = a_{11}$.

Para una matriz $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ de 2×2, su determinante se define como la suma de dos términos como sigue:

$$|A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12} \qquad [= \text{un escalar}] \qquad \textbf{(5.5)}$$

que se obtiene al multiplicar los dos elementos de la diagonal principal de A y luego restar el producto de los dos elementos restantes. En vista de la dimensión de la matriz A, el determinante $|A|$ dado en (5.5) se llama *determinante de segundo orden.*

Ejemplo 1 Dadas $A = \begin{bmatrix} 10 & 4 \\ 8 & 5 \end{bmatrix}$ y $B = \begin{bmatrix} 3 & 5 \\ 0 & -1 \end{bmatrix}$, sus determinantes son

$$|A| = \begin{vmatrix} 10 & 4 \\ 8 & 5 \end{vmatrix} = 10(5) - 8(4) = 18$$

y

$$|B| = \begin{vmatrix} 3 & 5 \\ 0 & -1 \end{vmatrix} = 3(-1) - 0(5) = -3$$

Si bien un determinante (encerrado entre dos barras verticales en vez de corchetes) es por definición un escalar, una matriz como tal no tiene un valor numérico. En otras palabras, un determinante es reducible a un número, pero una matriz es, en cambio, un bloque completo de números. Se debe remarcar también que un determinante está definido sólo para una matriz cuadrada, mientras que una matriz no tiene que ser cuadrada.

Incluso en esta primera etapa de explicación, es posible tener un indicio de la relación entre la dependencia lineal de los renglones en una matriz A, por un lado, y su determinante $|A|$, por otro. Las dos matrices

$$C = \begin{bmatrix} c_1' \\ c_2' \end{bmatrix} = \begin{bmatrix} 3 & 8 \\ 3 & 8 \end{bmatrix} \qquad \text{y} \qquad D = \begin{bmatrix} d_1' \\ d_2' \end{bmatrix} = \begin{bmatrix} 2 & 6 \\ 8 & 24 \end{bmatrix}$$

tienen renglones linealmente dependientes, porque $c_1' = c_2'$ y $d_2' = 4d_1'$. Ambos determinantes también resultan ser igual a cero:

$$|C| = \begin{vmatrix} 3 & 8 \\ 3 & 8 \end{vmatrix} = 3(8) - 3(8) = 0$$

$$|D| = \begin{vmatrix} 2 & 6 \\ 8 & 24 \end{vmatrix} = 2(24) - 8(6) = 0$$

Este resultado indica de forma enfática que un determinante "nulo" (un determinante de valor cero) puede tener algo que ver con la dependencia lineal. Veremos que éste es de hecho el caso. Además, el valor de un determinante $|A|$ puede ser no sólo un criterio para probar la independencia lineal de los renglones (por consiguiente la no singularidad) de la matriz A, sino también como una entrada en el cálculo de la inversa A^{-1}, si existe.

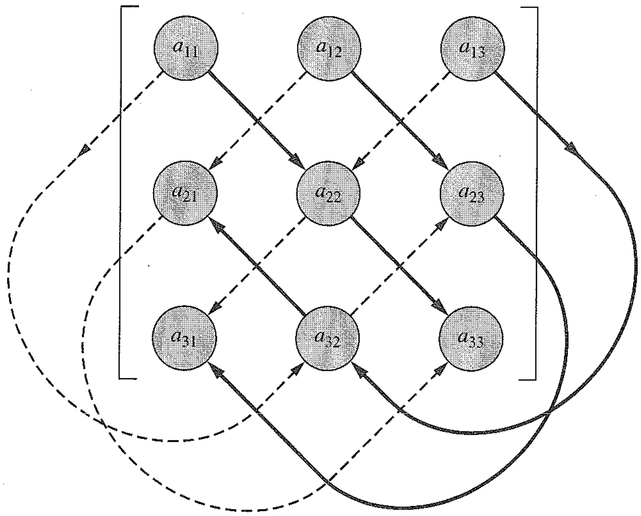
Sin embargo, primero se debe ampliar la perspectiva mediante un análisis de los determinantes de orden superior.

Evaluación de un determinante de tercer orden

Un determinante de orden 3 se relaciona con una matriz de 3×3. Dada

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

FIGURA 5.1



su determinante tiene el valor

$$
\begin{aligned}
|A| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} &= a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\
&= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} + a_{12}a_{23}a_{31} - a_{12}a_{21}a_{33} \\
&\quad + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} \qquad [= \text{un escalar}]
\end{aligned}
\tag{5.6}
$$

Al examinar la línea inferior de (5.6), observamos el valor de $|A|$ expresado como una suma de seis términos de producto, tres de los cuales van precedidos de un signo menos y tres de un signo más. Por complicada que parezca esta suma, hay una forma muy fácil de entender estos seis términos de un determinante de tercer orden. Esto se explica mejor mediante un diagrama (figura 5.1). En el determinante mostrado en la figura 5.1, cada elemento del renglón superior se relaciona con otros dos elementos mediante dos flechas *continuas* como sigue: $a_{11} \to a_{22} \to a_{33}$, $a_{12} \to a_{23} \to a_{31}$ y $a_{13} \to a_{32} \to a_{21}$. Cada terna de elementos enlazada de esta manera se puede multiplicar, y su producto se puede tomar como uno de los seis términos de (5.6). Los términos del producto de las flechas continuas van precedidos de un signo más.

Por otro lado, cada elemento del renglón superior ha sido conectado con otros dos elementos mediante dos flechas *discontinuas* como sigue: $a_{11} \to a_{32} \to a_{23}$, $a_{12} \to a_{21} \to a_{33}$ y $a_{13} \to a_{22} \to a_{31}$. Cada terna de elementos así conectados se multiplica también, y se toma su producto nuevamente como uno de los seis términos de (5.6). Cada uno de estos productos va precedido de un signo menos. La suma de los seis productos será entonces el valor del determinante.

Ejemplo 2

$$
\begin{vmatrix} 2 & 1 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix} = (2)(5)(9) + (1)(6)(7) + (3)(8)(4) - (2)(8)(6) - (1)(4)(9) - (3)(5)(7) = -9
$$

Ejemplo 3

$$
\begin{vmatrix} -7 & 0 & 3 \\ 9 & 1 & 4 \\ 0 & 6 & 5 \end{vmatrix} = (-7)(1)(5) + (0)(4)(0) + (3)(6)(9) - (-7)(6)(4) - (0)(9)(5) - (3)(1)(0)
$$

$$
= 295
$$

Este método de multiplicación cruzada proporciona una manera conveniente de evaluar un determinante de tercer orden, pero desafortunadamente *no* es aplicable a determinantes de orden superior a 3. Para estos últimos, se debe recurrir a la expansión de Laplace del determinante.

Evaluación de un determinante de *n*-ésimo orden mediante la expansión de Laplace

Primero explicaremos el proceso de desarrollo o *expansión de Laplace* para un determinante de tercer orden. Volviendo a la primera línea de (5.6), se ve que el valor de $|A|$ se puede considerar también como una suma de *tres términos*, cada uno de los cuales es un producto de un elemento del primero renglón y un determinante particular de *segundo* orden. Este último proceso de evaluar $|A|$, por medio de ciertos determinantes de orden menor, ilustra la expansión de Laplace del determinante.

Los tres determinantes de segundo orden en (5.6) no se determinan de modo arbitrario, sino se especifican por medio de una regla definida. El primero, $\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$, es un *sub*determinante de $|A|$ obtenido al eliminar el *primer* renglón y la *primera* columna de $|A|$. Éste se llama el *menor* del elemento a_{11} (el elemento en la intersección del renglón y la columna eliminados) y se denota mediante $|M_{11}|$. En general, el símbolo $|M_{ij}|$ se emplea para representar el menor obtenido al eliminar el i-ésimo renglón y la j-ésima columna de un determinante dado. Puesto que un menor es por sí mismo un determinante, tiene un valor. Como puede comprobar el lector, los otros dos determinantes de segundo orden en (5.6) son, respectivamente, los menores $|M_{12}|$ y $|M_{13}|$; es decir,

$$|M_{11}| \equiv \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} \qquad |M_{12}| \equiv \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} \qquad |M_{13}| \equiv \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

Un concepto que tiene relación estrecha con el menor es el del *cofactor*: un cofactor, denotado por $|C_{ij}|$, es un menor con un signo algebraico prescrito unido a él.[2] La regla de signo es como sigue: si la suma de los dos subíndices i y j del menor $|M_{ij}|$ es par, entonces el cofactor toma el mismo signo que el menor; es decir, $|C_{ij}| \equiv |M_{ij}|$. Si es impar, entonces el cofactor toma el signo opuesto al menor; es decir, $|C_{ij}| \equiv -|M_{ij}|$. En resumen, tenemos

$$|C_{ij}| \equiv (-1)^{i+j}|M_{ij}|$$

donde es obvio que la expresión $(-1)^{i+j}$ puede ser positiva si y sólo si $(i + j)$ es par. El hecho de que un cofactor tenga un signo específico es de extrema importancia y siempre se debe tener presente.

Ejemplo 4

En el determinante $\begin{vmatrix} 9 & 8 & 7 \\ 6 & 5 & 4 \\ 3 & 2 & 1 \end{vmatrix}$, el menor del elemento 8 es

$$|M_{12}| = \begin{vmatrix} 6 & 4 \\ 3 & 1 \end{vmatrix} = -6$$

En tanto que el cofactor del mismo elemento es

$$|C_{12}| = -|M_{12}| = 6$$

porque $i + j = 1 + 2 = 3$ es impar. De manera similar, el cofactor del elemento 4 es

$$|C_{23}| = -|M_{23}| = -\begin{vmatrix} 9 & 8 \\ 3 & 2 \end{vmatrix} = 6$$

[2] Muchos autores utilizan los símbolos M_{ij} y C_{ij} (sin las barras verticales) para menores y cofactores. Nosotros agregamos las barras verticales para dar énfasis visual al hecho de que los menores y cofactores están de la naturaleza de los determinantes y, como tales, tienen valores escalares.

Usando estos nuevos conceptos, un determinante de tercer orden se puede expresar como

$$\begin{aligned} |A| &= a_{11}|M_{11}| - a_{12}|M_{12}| + a_{13}|M_{13}| \\ &= a_{11}|C_{11}| + a_{12}|C_{12}| + a_{13}|C_{13}| = \sum_{j=1}^{3} a_{1j}|C_{1j}| \end{aligned} \tag{5.7}$$

es decir, como una suma de tres términos, cada uno de los cuales es el producto de un elemento del primer renglón y de su cofactor correspondiente. Note la diferencia en los signos de los términos $a_{12}|M_{12}|$ y $a_{12}|C_{12}|$ en (5.7). Esto es porque $1 + 2$ da un número impar.

La expansión de Laplace de un determinante de *tercer* orden reduce el problema de evaluación a evaluar sólo ciertos determinantes de *segundo* orden. Una reducción similar se obtiene con la expansión de Laplace de determinantes de orden superior. En un determinante de cuarto orden $|B|$, por ejemplo, el renglón superior contendrá cuatro elementos $b_{11} \ldots b_{14}$; así, de la misma manera que en (5.7), se puede escribir

$$|B| = \sum_{j=1}^{4} b_{1j}|C_{1j}|$$

donde los cofactores $|C_{1j}|$ son de orden 3. Cada cofactor de tercer orden se puede evaluar entonces como en (5.6). En general, la expansión de Laplace de un determinante de n-ésimo orden reducirá el problema a uno en el que se evalúan n cofactores, cada uno de los cuales es de orden $n - 1$. La aplicación repetida del proceso conducirá de forma metódica a órdenes cada vez menores de los determinantes para culminar finalmente en los determinantes básicos de segundo orden definidos en (5.5). Entonces el valor del determinante original se puede calcular con facilidad.

Aunque el proceso de expansión de Laplace se ha expresado en términos de los cofactores de los elementos del primer renglón, también es factible desarrollar un determinante mediante el cofactor de cualquier renglón o, de igual manera, de cualquier columna. Por ejemplo, si la primera columna de un determinante de tercer orden $|A|$ consta de los elementos a_{11}, a_{21} y a_{31}, el desarrollo por cofactores de estos elementos también produce el valor de $|A|$:

$$|A| = a_{11}|C_{11}| + a_{21}|C_{21}| + a_{31}|C_{31}| = \sum_{i=1}^{3} a_{i1}|C_{i1}|$$

Ejemplo 5

Dada $|A| = \begin{vmatrix} 5 & 6 & 1 \\ 2 & 3 & 0 \\ 7 & -3 & 0 \end{vmatrix}$, el desarrollo por el primer *renglón* produce el resultado

$$|A| = 5\begin{vmatrix} 3 & 0 \\ -3 & 0 \end{vmatrix} - 6\begin{vmatrix} 2 & 0 \\ 7 & 0 \end{vmatrix} + \begin{vmatrix} 2 & 3 \\ 7 & -3 \end{vmatrix} = 0 + 0 - 27 = -27$$

En tanto que el desarrollo por la primera *columna* produce la respuesta idéntica:

$$|A| = 5\begin{vmatrix} 3 & 0 \\ -3 & 0 \end{vmatrix} - 2\begin{vmatrix} 6 & 1 \\ -3 & 0 \end{vmatrix} + 7\begin{vmatrix} 6 & 1 \\ 3 & 0 \end{vmatrix} = 0 - 6 - 21 = -27$$

En la medida en que tenga que ver el cálculo numérico, este hecho nos ofrece una oportunidad de elegir algún renglón o columna "fácil" para el desarrollo. Un renglón o columna con mayor número de ceros o unos siempre es preferible para este propósito, porque un 0 multiplicado por su cofactor es simplemente 0, de modo que se eliminará el término, y un 1 multiplicado por su cofactor es el cofactor en sí, así que por lo menos se puede ahorrar un paso de multiplicación. En el ejemplo 5, la forma más fácil de desarrollar el determinante es mediante la tercera columna, que consta de los elementos 1, 0 y 0. Por lo tanto, se podría haber evaluado así:

$$|A| = 1\begin{vmatrix} 2 & 3 \\ 7 & -3 \end{vmatrix} = -6 - 21 = -27$$

En resumen, el valor de un determinante $|A|$ de orden n se determina mediante la expansión de Laplace de *cualquier renglón* o *columna* como sigue:

$$\begin{aligned} |A| &= \sum_{j=1}^{n} a_{ij}|C_{ij}| \qquad \text{[expansión mediante el } i\text{-ésimo renglón]} \\ &= \sum_{i=1}^{n} a_{ij}|C_{ij}| \qquad \text{[expansión mediante la } j\text{-ésima columna]} \end{aligned} \tag{5.8}$$

EJERCICIO 5.2

1. Evalúe los siguientes determinantes:

(*a*) $\begin{vmatrix} 8 & 1 & 3 \\ 4 & 0 & 1 \\ 6 & 0 & 3 \end{vmatrix}$ (*c*) $\begin{vmatrix} 4 & 0 & 2 \\ 6 & 0 & 3 \\ 8 & 2 & 3 \end{vmatrix}$ (*e*) $\begin{vmatrix} a & b & c \\ b & c & a \\ c & a & b \end{vmatrix}$

(*b*) $\begin{vmatrix} 1 & 2 & 3 \\ 4 & 7 & 5 \\ 3 & 6 & 9 \end{vmatrix}$ (*d*) $\begin{vmatrix} 1 & 1 & 4 \\ 8 & 11 & -2 \\ 0 & 4 & 7 \end{vmatrix}$ (*f*) $\begin{vmatrix} x & 5 & 0 \\ 3 & y & 2 \\ 9 & -1 & 8 \end{vmatrix}$

2. Determine los signos que se anexarán a los menores pertinentes a fin de obtener los siguientes cofactores de un determinante: $|C_{13}|$, $|C_{23}|$, $|C_{33}|$, $|C_{41}|$ y $|C_{34}|$.

3. Dada $\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix}$, obtenga los menores y cofactores de los elementos a, b y f.

4. Evalúe los siguientes determinantes:

(*a*) $\begin{vmatrix} 1 & 2 & 0 & 9 \\ 2 & 3 & 4 & 6 \\ 1 & 6 & 0 & -1 \\ 0 & -5 & 0 & 8 \end{vmatrix}$ (*b*) $\begin{vmatrix} 2 & 7 & 0 & 1 \\ 5 & 6 & 4 & 8 \\ 0 & 0 & 9 & 0 \\ 1 & -3 & 1 & 4 \end{vmatrix}$

5. En el primer determinante del problema 4, obtenga el valor del cofactor del elemento 9.

6. Determine los menores y cofactores del tercer renglón, a partir de

$$A = \begin{bmatrix} 9 & 11 & 4 \\ 3 & 2 & 7 \\ 6 & 10 & 4 \end{bmatrix}$$

7. Utilice la expansión de Laplace para hallar el determinante de

$$A = \begin{bmatrix} 15 & 7 & 9 \\ 2 & 5 & 6 \\ 9 & 0 & 12 \end{bmatrix}$$

5.3 Propiedades básicas de determinantes

Ahora es posible analizar algunas propiedades de determinantes que nos permitirán "descubrir" la relación entre dependencia lineal de los renglones de una matriz cuadrada y la anulación del determinante de esa matriz.

Aquí se analizan cinco propiedades básicas. Éstas son propiedades comunes a los determinantes de todos los órdenes, aunque deberíamos ilustrar la mayoría con determinantes de segundo orden:

Propiedad I El intercambio de renglones y columnas no afecta el valor de un determinante. En otras palabras, el determinante de una matriz A tiene el mismo valor que el de su transpuesta A', es decir, $|A| = |A'|$.

Ejemplo 1

$$\begin{vmatrix} 4 & 3 \\ 5 & 6 \end{vmatrix} = \begin{vmatrix} 4 & 5 \\ 3 & 6 \end{vmatrix} = 9$$

Ejemplo 2

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & c \\ b & d \end{vmatrix} = ad - bc$$

Propiedad II El intercambio de *dos* renglones cualesquiera (o *dos* columnas cualesquiera) modificará el signo, pero no el valor numérico del determinante. (Es obvio que esta propiedad se relaciona con la primera operación elemental del renglón en una matriz.)

Ejemplo 3 $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$, pero el intercambio de los dos renglones produce

$$\begin{vmatrix} c & d \\ a & b \end{vmatrix} = cb - ad = -(ad - bc)$$

Ejemplo 4 $\begin{vmatrix} 0 & 1 & 3 \\ 2 & 5 & 7 \\ 3 & 0 & 1 \end{vmatrix} = -26$, pero el intercambio de las columnas primera y tercera produce

$$\begin{vmatrix} 3 & 1 & 0 \\ 7 & 5 & 2 \\ 1 & 0 & 3 \end{vmatrix} = 26.$$

Propiedad III La multiplicación de *cualquier* renglón (o columna) por un escalar k cambiará el valor del determinante k veces. (Esta propiedad se relaciona con la segunda operación elemental de los renglones de una matriz.)

Ejemplo 5 Al multiplicar el renglón superior del determinante del ejemplo 3 por k, se obtiene

$$\begin{vmatrix} ka & kb \\ c & d \end{vmatrix} = kad - kbc = k(ad - bc) = k\begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

Es importante distinguir entre las dos expresiones kA y $k|A|$. Al multiplicar una *matriz* A por un escalar k, todos los elementos de A se multiplican por k. Pero, si leemos de derecha a izquierda la ecuación del ejemplo presente, debe quedar claro que, al multiplicar un *determinante* $|A|$ por k, sólo un renglón (o columna) se debe multiplicar por k. Por lo tanto, esta ecuación en efecto nos proporciona una regla para factorizar un determinante: siempre que un solo renglón o columna contenga un divisor común, éste se puede usar como factor determinante.

Ejemplo 6 Factorizando a su vez la primera columna y el segundo renglón, se tiene

$$\begin{vmatrix} 15a & 7b \\ 12c & 2d \end{vmatrix} = 3\begin{vmatrix} 5a & 7b \\ 4c & 2d \end{vmatrix} = 3(2)\begin{vmatrix} 5a & 7b \\ 2c & d \end{vmatrix} = 6(5ad - 14bc)$$

La evaluación directa del determinante original produce, por supuesto, la misma respuesta.

Por el contrario, la factorización de una *matriz* requiere la presencia de un divisor común para *todos* sus elementos, como en

$$\begin{bmatrix} ka & kb \\ kc & kd \end{bmatrix} = k\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Propiedad IV La suma (resta) de un múltiplo de cualquier renglón a (de) otro renglón dejará sin cambio al determinante. Lo mismo se cumple si se sustituye la palabra *renglón* por *columna* en el enunciado anterior. (Esta propiedad se relaciona con la tercera operación elemental de los renglones de una matriz.)

Ejemplo 7 Al sumar k veces el renglón superior del determinante del ejemplo 3 a su segundo renglón, terminamos con el determinante original:

$$\begin{vmatrix} a & b \\ c+ka & d+kb \end{vmatrix} = a(d+kb) - b(c+ka) = ad - bc = \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

Propiedad V Si un renglón (o columna) es un múltiplo de otro renglón (o columna), el valor del determinante será cero. Como un caso especial de esto, cuando dos renglones (o dos columnas) son *idénticos*, se anula el determinante.

Ejemplo 8 $\begin{vmatrix} 2a & 2b \\ a & b \end{vmatrix} = 2ab - 2ab = 0 \qquad \begin{vmatrix} c & c \\ d & d \end{vmatrix} = cd - cd = 0$

En el ejercicio 5.2-1 se encuentran más ejemplos de este tipo de determinante "nulo".

Esta propiedad importante es, de hecho, una consecuencia lógica de la propiedad IV. Para entender esto, aplicamos la propiedad IV a los dos determinantes del ejemplo 8 y observamos el resultado. Para el primero, intente restar dos veces el segundo renglón del renglón de la parte superior; para el segundo determinante, reste la segunda columna de la primera. Puesto que estas operaciones no modifican los valores de los determinantes, se puede escribir

$$\begin{vmatrix} 2a & 2b \\ a & b \end{vmatrix} = \begin{vmatrix} 0 & 0 \\ a & b \end{vmatrix} \qquad \begin{vmatrix} c & c \\ d & d \end{vmatrix} = \begin{vmatrix} 0 & c \\ 0 & d \end{vmatrix}$$

Los nuevos determinantes (reducidos) ahora contienen, respectivamente, un renglón y una columna de ceros; así que su expansión de Laplace produce un valor cero en ambos casos. En general, cuando un renglón (columna) es un múltiplo de otro renglón (columna), la aplicación de la propiedad IV siempre puede reducir los elementos de ese renglón (columna) a cero y, por lo tanto, se deduce la propiedad V.

Las propiedades básicas recién explicadas son útiles en varios sentidos. En primer lugar, pueden ser de gran ayuda para simplificar la tarea de evaluar determinantes. Al restar múltiplos de un renglón (o columna) de otro, por ejemplo, los elementos del determinante se pueden reducir a números mucho más pequeños y más simples. Con la factorización, si es factible, también se logra lo mismo. De hecho, si se pueden aplicar estas propiedades para transformar algún renglón o columna de manera que contengan principalmente ceros o unos, la expansión de Laplace del determinante se convierte en una tarea mucho más manejable.

Criterio del determinante en relación con la no singularidad

Nuestra preocupación actual es, ante todo, vincular la dependencia lineal de los renglones con la anulación de un determinante. Para este propósito, se puede invocar la propiedad V. Considere un sistema de ecuaciones $Ax = d$:

$$\begin{bmatrix} 3 & 4 & 2 \\ 15 & 20 & 10 \\ 4 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix}$$

Este sistema puede tener una solución única si y sólo si los renglones de la matriz de coeficientes A son linealmente independientes, de manera que A sea no singular. Pero el segundo renglón es cinco veces el primero; los renglones son de hecho *dependientes* y, por lo tanto, no existe solución única. La detección de esta dependencia de renglones se hizo mediante la inspección visual; pero, en virtud de la propiedad V, lo habríamos podido descubrir también por el hecho de que $|A| = 0$.

Por supuesto que la dependencia de renglones en una matriz puede asumir un patrón más intrincado y hermético. Por ejemplo, en la matriz

$$B = \begin{bmatrix} 4 & 1 & 2 \\ 5 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} v_1' \\ v_2' \\ v_3' \end{bmatrix}$$

existe dependencia de renglones porque $2v_1' - v_2' - 3v_3' = 0$; sin embargo, este hecho desafía la detección visual. No obstante, incluso en este caso la propiedad V nos proporcionará un determinante nulo, $|B| = 0$, puesto que al sumar tres veces v_3' a v_2' y restarle dos veces v_1', el segundo renglón se puede reducir a un vector cero. En general, *cualquier* patrón de dependencia lineal entre renglones se reflejará en el determinante nulo, ¡y en esto radica la belleza de la propiedad V! Por el contrario, si los renglones son linealmente independientes, el determinante debe tener un valor diferente a cero.

En los dos párrafos anteriores, la no singularidad de una matriz se ha relacionado principalmente con la independencia lineal entre *renglones*. Pero, de vez en cuando, se ha hecho la afirmación de que, para una matriz *cuadrada* A, los *renglones* son independientes $\Leftrightarrow$ las columnas son independientes. Ahora estamos preparados para probar esa afirmación:

> De acuerdo con la propiedad I, sabemos que $|A| = |A'|$. Puesto que los renglones en A son independientes $A \Leftrightarrow |A| \neq 0$, podemos establecer también que los *renglones* de A son independientes $\Leftrightarrow |A'| \neq 0$. Pero $|A'| \neq 0 \Leftrightarrow$ los renglones de la transpuesta A' son independientes $\Leftrightarrow$ las columnas de A son independientes (los renglones de A' son por definición las columnas de A). Por lo tanto, los *renglones* de A son independientes $\Leftrightarrow$ las *columnas* de A son independientes.

Ahora podemos resumir la explicación de la prueba de no singularidad. Dado un sistema de ecuaciones lineales $Ax = d$, donde A es una matriz de coeficientes $n \times n$,

$$\begin{aligned} |A| \neq 0 &\Leftrightarrow \text{hay independencia de renglones (columnas) en la matriz } A \\ &\Leftrightarrow A \text{ es no singular} \\ &\Leftrightarrow \text{existe } A^{-1} \\ &\Leftrightarrow \text{existe una solución única } x^* = A^{-1}d \end{aligned}$$

Así, el valor del determinante de la matriz de coeficientes, $|A|$, proporciona un criterio conveniente para probar la no singularidad de matriz A y la existencia de una solución única para el sistema de ecuaciones $Ax = d$. Sin embargo, note que el criterio de determinantes no dice nada acerca de los signos algebraicos de los valores solución; es decir, aun cuando estamos seguros de una solución única si $|A| \neq 0$, a veces podemos obtener valores solución negativos que son inadmisibles desde el punto de vista económico.

Ejemplo 9 ¿El sistema de ecuaciones

$$\begin{aligned} 7x_1 - 3x_2 - 3x_3 &= 7 \\ 2x_1 + 4x_2 + x_3 &= 0 \\ -2x_2 - x_3 &= 2 \end{aligned}$$

posee una solución única? El determinante $|A|$ es

$$\begin{vmatrix} 7 & -3 & -3 \\ 2 & 4 & 1 \\ 0 & -2 & -1 \end{vmatrix} = -8 \neq 0$$

Por lo tanto, existe una solución única.

Redefinición del rango de una matriz

El rango de una matriz A se definió antes como el número máximo de renglones linealmente independientes en A. En vista de la relación entre la independencia de renglones y la no anulación del determinante, se puede redefinir el rango de una matriz de $m \times n$ como el orden máximo de un determinante no nulo que se puede construir de los renglones y columnas de esa matriz. El rango de cualquier matriz es un número único.

Es evidente que el rango puede ser a lo sumo m o n, cualquiera que sea el más pequeño, porque un determinante se define sólo para una matriz cuadrada, y de una matriz de dimensión, por ejemplo 3×5, los determinantes más grandes posibles (nulos o no) serán de orden 3. Este hecho se puede expresar simbólicamente como sigue:

$$r(A) \leq \text{mín}\,\{m, n\}$$

que se lee: "el rango de A es menor que o igual al mínimo del conjunto de dos números m y n". El rango de una matriz no singular A de dimensión $n \times n$ debe ser n; es ese caso, podemos escribir $r(A) = n$.

Algunas veces, uno podría estar interesado en el rango del producto de dos matrices. En ese caso, puede ser útil la siguiente regla:

$$r(AB) \leq \text{mín}\,\{r(A), r(B)\} \qquad \textbf{(5.9)}$$

Aunque esta regla no produce un valor único de $r(AB)$, la aplicación de la regla puede, no obstante, conducir a resultados únicos. En particular, podemos usar la ecuación (5.9) para mostrar que si una matriz A, con $r(A) = j$, se multiplica por alguna matriz B no singular (conformable), el rango de la matriz producto AB (o BA, como podría ser el caso), debe ser j. Podríamos probar esto para el producto AB (el caso de BA es análogo). Primero, en el lado derecho de (5.9) observamos sólo tres casos posibles: (*i*) $r(A) < r(B)$, (*ii*) $r(A) = r(B)$ y (*iii*) $r(A) > r(B)$.

Para los casos (*i*) e (*ii*), (5.9) se reduce a $r(AB) \leq r(A) = j$. Para el caso (*iii*), se encuentra que $r(AB) \leq r(B) < r(A) = j$. Así, de cualquier manera, se obtiene

$$r(AB) \leq r(A) = j \tag{5.10}$$

Ahora considere la identidad $(AB)B^{-1} = A$, Por (5.9), se puede escribir

$$r[(AB)B^{-1}] \leq \text{mín}\{r(AB), r(B^{-1})\}$$

Al aplicar el mismo razonamiento con el que se llegó a (5.10), se concluye de esto que

$$r[(AB)B^{-1}] \leq r(AB)$$

Puesto que la expresión del lado izquierdo de esta desigualdad es igual a $r(A) = j$, se puede escribir

$$j \leq r(AB) \tag{5.11}$$

Pero (5.10) y (5.11) no se pueden satisfacer al mismo tiempo a menos que $r(AB) = j$. Así, el rango de la matriz producto AB debe ser j, como se afirma.

EJERCICIO 5.3

1. Use el determinante $\begin{vmatrix} 4 & 0 & -1 \\ 2 & 1 & -7 \\ 3 & 3 & 9 \end{vmatrix}$ para comprobar las primeras cuatro propiedades de los determinantes.
2. Muestre que, cuando los elementos de un determinante de n-ésimo orden $|A|$ se multiplican por un número k, el resultado será $k^n|A|$.
3. ¿Cuáles propiedades de los determinantes nos permiten escribir lo siguiente?

   (*a*) $\begin{vmatrix} 9 & 18 \\ 27 & 56 \end{vmatrix} = \begin{vmatrix} 9 & 18 \\ 0 & 2 \end{vmatrix}$ (*b*) $\begin{vmatrix} 9 & 27 \\ 4 & 2 \end{vmatrix} = 18\begin{vmatrix} 1 & 3 \\ 2 & 1 \end{vmatrix}$
4. Pruebe si las siguientes matrices son no singulares:

   (*a*) $\begin{bmatrix} 4 & 0 & 1 \\ 19 & 1 & -3 \\ 7 & 1 & 0 \end{bmatrix}$ (*c*) $\begin{bmatrix} 7 & -1 & 0 \\ 1 & 1 & 4 \\ 13 & -3 & -4 \end{bmatrix}$

   (*b*) $\begin{bmatrix} 4 & -2 & 1 \\ -5 & 6 & 0 \\ 7 & 0 & 3 \end{bmatrix}$ (*d*) $\begin{bmatrix} -4 & 9 & 5 \\ 3 & 0 & 1 \\ 10 & 8 & 6 \end{bmatrix}$
5. ¿Qué se puede concluir acerca del rango de cada matriz del problema 4?
6. ¿Alguno de los siguientes conjuntos de vectores de tres dimensiones generan el espacio tridimensional? ¿Por qué sí o por qué no?

   (*a*) $[1\ \ 2\ \ 1]$ $[2\ \ 3\ \ 1]$ $[3\ \ 4\ \ 2]$

   (*b*) $[8\ \ 1\ \ 3]$ $[1\ \ 2\ \ 8]$ $[-7\ \ 1\ \ 5]$
7. Reescriba el modelo de ingreso nacional simple (3.23) en la forma $Ax = d$ (con Y como la primera variable en el vector x), y luego pruebe si la matriz de coeficientes A es no singular.
8. Comente acerca de la validez de las siguientes afirmaciones:

   (*a*) "Dada cualquier matriz A, se puede obtener de ésta siempre una transpuesta y un determinante."

   (*b*) "Al multiplicar por 2 cada elemento de un determinante de $n \times n$ se duplica el valor de ese determinante."

   (*c*) "Si se anula una matriz cuadrada A, entonces se puede tener la seguridad de que el sistema de ecuaciones $Ax = d$ es no singular."

5.4 Obtención de la matriz inversa

Si la matriz A en el sistema de ecuaciones lineales $Ax = d$ es no singular, entonces existe A^{-1}, y la solución del sistema será $x^* = A^{-1}d$. Ya aprendimos a probar la no singularidad de A por medio del criterio $|A| \neq 0$. La siguiente pregunta es, ¿cómo se puede obtener la inversa A^{-1} si A pasa esa prueba?

Expansión de un determinante por cofactores ajenos

Antes de contestar esta pregunta, procedemos a analizar otra propiedad importante de los determinantes.

Propiedad VI La expansión de un determinante por *cofactores ajenos* (los cofactores de un renglón, o columna, "erróneo") siempre produce un valor de cero.

Ejemplo 1

Si desarrollamos el determinante $\begin{vmatrix} 4 & 1 & 2 \\ 5 & 2 & 1 \\ 1 & 0 & 3 \end{vmatrix}$ por medio de sus elementos de *primer* renglón, pero los cofactores de los elementos del *segundo* renglón

$$|C_{21}| = -\begin{vmatrix} 1 & 2 \\ 0 & 3 \end{vmatrix} = -3 \qquad |C_{22}| = \begin{vmatrix} 4 & 2 \\ 1 & 3 \end{vmatrix} = 10 \qquad |C_{23}| = -\begin{vmatrix} 4 & 1 \\ 1 & 0 \end{vmatrix} = 1$$

obtenemos $a_{11}|C_{21}| + a_{12}|C_{22}| + a_{13}|C_{23}| = 4(-3) + 1(10) + 2(1) = 0$.

En términos más generales, aplicar el mismo tipo de desarrollo mediante cofactores ajenos, como se describió en el ejemplo 1 del determinante $|A| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$, produce una suma cero de productos como sigue:

$$\begin{aligned}\sum_{i=1}^{3} a_{1j}|C_{2j}| &= a_{11}|C_{21}| + a_{12}|C_{22}| + a_{13}|C_{23}| \\ &= -a_{11}\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{12}\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} - a_{13}\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} \qquad \textbf{(5.12)} \\ &= -a_{11}a_{12}a_{33} + a_{11}a_{13}a_{32} + a_{11}a_{12}a_{33} - a_{12}a_{13}a_{31} \\ &\quad - a_{11}a_{13}a_{32} + a_{12}a_{13}a_{31} = 0\end{aligned}$$

La razón para este resultado radica en el hecho de que la suma de productos en (5.12) se puede considerar como el resultado de la expansión *regular*, mediante el segundo renglón de otro determinante $|A^*| \equiv \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$, que difiere de $|A|$ sólo en su segundo renglón y cuyos primeros dos renglones son idénticos. Como ejercicio, escriba los cofactores de los renglones segundos de $|A^*|$ y compruebe que éstos son precisamente los cofactores que aparecen en (5.12), y con los signos correctos.Puesto que $|A^*| = 0$, como resultado de sus dos renglones idénticos, la expansión por cofactores ajenos mostrada en (5.12) producirá necesariamente un valor de cero también.

La propiedad VI es válida para determinantes de todos los órdenes y se aplica cuando se desarrolla un determinante por cofactores ajenos de cualquier renglón o columna. Así, se puede expresar, en general, que para un determinante de orden n se cumple lo siguiente:

$$\sum_{j=1}^{n} a_{ij}|C_{i'j}| = 0 \qquad (i \neq i') \quad \text{[desarrollo por el } i\text{-ésimo renglón y los cofactores del } i'\text{-ésimo renglón]}$$

$$\sum_{i=1}^{n} a_{ij}|C_{ij'}| = 0 \qquad (j \neq j') \quad \text{[desarrollo por la } j\text{-ésima columna y los cofactores de la } j'\text{-ésima columna]} \tag{5.13}$$

Compare cuidadosamente (5.13) con (5.8). En el último (expansión de Laplace regular), los subíndices de a_{ij} y de $|C_{ij}|$ deben ser idénticos en cada término del producto en la suma. Por otro lado, en la expansión por cofactores ajenos, como en (5.13), uno de los dos subíndices (un valor elegido de i' o j') está de modo inevitable "fuera de lugar".

Inversión de matriz

La propiedad VI, como se resumió en (5.13), es de ayuda directa para desarrollar un método de inversión de matriz, es decir, hallar la inversa de una matriz.

Suponga que se tiene la matriz no singular A de $n \times n$:

$$\underset{(n\times n)}{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \qquad (|A| \neq 0) \tag{5.14}$$

Puesto que cada elemento de A tiene un cofactor $|C_{ij}|$, es posible formar una matriz de cofactores al sustituir cada elemento a_{ij} en (5.4) por su cofactor $|C_{ij}|$. Esta matriz de cofactores, denotada por $C = [|C_{ij}|]$, debe ser también de $n \times n$. Sin embargo, para los fines que perseguimos, la transpuesta de C tiene más interés. Esta transpuesta C' se conoce como la *adjunta* de A y se simboliza mediante adj A. Cuando la escribimos, la adjunta toma la forma

$$\underset{(n\times n)}{C'} \equiv \operatorname{adj} A \equiv \begin{bmatrix} |C_{11}| & |C_{21}| & \cdots & |C_{n1}| \\ |C_{12}| & |C_{22}| & \cdots & |C_{n2}| \\ \cdots & \cdots & \cdots & \cdots \\ |C_{1n}| & |C_{2n}| & \cdots & |C_{nn}| \end{bmatrix} \tag{5.15}$$

Las matrices A y C' son conformables en relación con la multiplicación, y su producto AC' es otra matriz de $n \times n$, en la cual cada elemento es una suma de productos. Al utilizar la fórmula para la expansión de Laplace así como la propiedad VI de los determinantes, el producto AC' se puede expresar como sigue:

$$\underset{(n\times n)}{AC'} = \begin{bmatrix} \sum_{j=1}^{n} a_{1j}|C_{1j}| & \sum_{j=1}^{n} a_{1j}|C_{2j}| & \cdots & \sum_{j=1}^{n} a_{1j}|C_{nj}| \\ \sum_{j=1}^{n} a_{2j}|C_{1j}| & \sum_{j=1}^{n} a_{2j}|C_{2j}| & \cdots & \sum_{j=1}^{n} a_{2j}|C_{nj}| \\ \vdots & \vdots & & \vdots \\ \sum_{j=1}^{n} a_{nj}|C_{1j}| & \sum_{j=1}^{n} a_{nj}|C_{2j}| & \cdots & \sum_{j=1}^{n} a_{nj}|C_{nj}| \end{bmatrix}$$

$$= \begin{bmatrix} |A| & 0 & \cdots & 0 \\ 0 & |A| & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & |A| \end{bmatrix} \qquad \text{[por (5.8) y (5.13)]}$$

$$= |A| \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = |A| I_n \qquad \text{[factorizando]}$$

Como el determinante $|A|$ es un escalar no nulo, es lícito dividir ambos lados de la ecuación $AC' = |A|I$ entre $|A|$. El resultado es

$$\frac{AC'}{|A|} = I, \quad \text{o bien,} \quad A\frac{C'}{|A|} = I$$

Si se premultiplican ambos lados de la última ecuación por A^{-1}, y se utiliza el resultado $A^{-1} A$, se obtiene $\dfrac{C'}{|A|} = A^{-1}$, o bien

$$A^{-1} = \frac{1}{|A|} \text{adj } A \qquad \text{[por (5.15)]} \tag{5.16}$$

¡Ahora hemos encontrado una forma de invertir la matriz A!

El procedimiento general para hallar la inversa de una matriz cuadrada A tiene que ver con los siguientes pasos: (1) halle $|A|$ [debemos proceder con los pasos posteriores si y sólo si $|A| \neq 0$, porque si $|A| = 0$, la inversa en (5.16) estará indefinida]; (2) determine los cofactores de los elementos de A y ordénelos como una matriz $C = [|C_{ij}|]$; (3) tome la transpuesta de C para obtener adj A, y (4) divida adj A entre el determinante $|A|$. El resultado será la inversa deseada A^{-1}.

Ejemplo 2

Obtenga la inversa de $A = \begin{bmatrix} 3 & 2 \\ 1 & 0 \end{bmatrix}$. Puesto que $|A| = -2 \neq 0$, existe la inversa A^{-1}. El cofactor de cada elemento es en este caso un determinante 1×1, que se define simplemente como el elemento escalar de ese mismo determinante (es decir, $|a_{ij}| \equiv a_{ij}$). Así, se tiene

$$C = \begin{bmatrix} |C_{11}| & |C_{12}| \\ |C_{21}| & |C_{22}| \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ -2 & 3 \end{bmatrix}$$

Observe los signos menos que acompañan a 1 y 2, según se necesite para los cofactores. La transposición de la matriz de cofactores produce

$$\text{adj } A = \begin{bmatrix} 0 & -2 \\ -1 & 3 \end{bmatrix}$$

de modo que la inversa A^{-1} se puede escribir como

$$A^{-1} = \frac{1}{|A|} \text{adj } A = -\frac{1}{2} \begin{bmatrix} 0 & -2 \\ -1 & 3 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & -\frac{3}{2} \end{bmatrix}$$

Ejemplo 3

Obtenga la inversa de $B = \begin{bmatrix} 4 & 1 & -1 \\ 0 & 3 & 2 \\ 3 & 0 & 7 \end{bmatrix}$. Puesto que $|B| = 99 \neq 0$, también existe la inversa B^{-1}. La matriz de cofactores es

$$\begin{bmatrix} \begin{vmatrix} 3 & 2 \\ 0 & 7 \end{vmatrix} & -\begin{vmatrix} 0 & 2 \\ 3 & 7 \end{vmatrix} & \begin{vmatrix} 0 & 3 \\ 3 & 0 \end{vmatrix} \\ -\begin{vmatrix} 1 & -1 \\ 0 & 7 \end{vmatrix} & \begin{vmatrix} 4 & -1 \\ 3 & 7 \end{vmatrix} & -\begin{vmatrix} 4 & 1 \\ 3 & 0 \end{vmatrix} \\ \begin{vmatrix} 1 & -1 \\ 3 & 2 \end{vmatrix} & -\begin{vmatrix} 4 & -1 \\ 0 & 2 \end{vmatrix} & \begin{vmatrix} 4 & 1 \\ 0 & 3 \end{vmatrix} \end{bmatrix} = \begin{bmatrix} 21 & 6 & -9 \\ -7 & 31 & 3 \\ 5 & -8 & 12 \end{bmatrix}$$

Por lo tanto,

$$\text{adj } B = \begin{bmatrix} 21 & -7 & 5 \\ 6 & 31 & -8 \\ -9 & 3 & 12 \end{bmatrix}$$

y la matriz inversa deseada es

$$B^{-1} = \frac{1}{|B|} \text{adj } B = \frac{1}{99} \begin{bmatrix} 21 & -7 & 5 \\ 6 & 31 & -8 \\ -9 & 3 & 12 \end{bmatrix}$$

Se puede comprobar que los resultados de los ejemplos 2 y 3 satisfacen $AA^{-1} = A^{-1}A = I$ y $BB^{-1} = B^{-1}B = I$, respectivamente.

EJERCICIO 5.4

1. Suponga que expandimos un determinante de cuarto orden por su *tercera columna* y los cofactores de los elementos de la *segunda columna*. ¿Cómo escribiría la suma resultante de productos en la notación de $\sum$? ¿Cuál será la suma de productos en la notación de $\sum$ si la expandimos por el *segundo renglón* y los cofactores de los elementos del *cuarto renglón*?
2. Obtenga la inversa de cada una de las siguientes matrices:

   (*a*) $A = \begin{bmatrix} 5 & 2 \\ 0 & 1 \end{bmatrix}$ (*b*) $B = \begin{bmatrix} -1 & 0 \\ 9 & 2 \end{bmatrix}$ (*c*) $C = \begin{bmatrix} 3 & 7 \\ 3 & -1 \end{bmatrix}$ (*d*) $D = \begin{bmatrix} 7 & 6 \\ 0 & 3 \end{bmatrix}$
3. (*a*) Con base en sus respuestas del problema 2, formule una regla de dos pasos para obtener la adjunta de una matriz A de 2×2: en el primer paso, indique lo que se les debe hacer a los dos elementos diagonales de A, con el fin de obtener los elementos diagonales de adj A; en el segundo paso, indique qué se debe hacer a los dos elementos fuera de la diagonal de A. (*Advertencia:* esta regla se aplica sólo a matrices de 2×2.)

   (*b*) Añada un tercer paso que, junto con los dos pasos previos, produzca la matriz inversa A^{-1} de 2×2.
4. Obtenga la inversa de cada una de las siguientes matrices:

   (*a*) $E = \begin{bmatrix} 4 & -2 & 1 \\ 7 & 3 & 0 \\ 2 & 0 & 1 \end{bmatrix}$ (*c*) $G = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$

   (*b*) $F = \begin{bmatrix} 1 & -1 & 2 \\ 1 & 0 & 3 \\ 4 & 0 & 2 \end{bmatrix}$ (*d*) $H = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

5. Determine la inversa de

$$A = \begin{bmatrix} 4 & 1 & -5 \\ -2 & 3 & 1 \\ 3 & -1 & 4 \end{bmatrix}$$

6. Resuelva la matriz $Ax = d$ por inversión de matriz, donde

(*a*) $4x + 3y = 28$
$2x + 5y = 42$

(*b*) $4x_1 + x_2 - 5x_3 = 8$
$-2x_1 + 3x_2 + x_3 = 12$
$3x_1 - x_2 + 4x_3 = 5$

7. ¿Es posible que una matriz sea su propia inversa?

5.5 Regla de Cramer

El método de inversión de matriz estudiado en la sección 5.4 nos permite deducir una forma práctica (acaso siempre eficaz) de resolver un sistema de ecuaciones lineales, conocida como *regla de Cramer.*

Deducción de la regla

Dado un sistema de ecuaciones $Ax = d$, donde A es de $n \times n$, la solución se puede escribir como

$$x^* = A^{-1}d = \frac{1}{|A|}(\text{adj } A)d \qquad [\text{por } (5.16)]$$

siempre que A sea no singular. De acuerdo con (5.15), esto significa que

$$\begin{bmatrix} x_1^* \\ x_2^* \\ \vdots \\ x_n^* \end{bmatrix} = \frac{1}{|A|} \begin{bmatrix} |C_{11}| & |C_{21}| & \cdots & |C_{n1}| \\ |C_{12}| & |C_{22}| & \cdots & |C_{n2}| \\ \cdots & \cdots & \cdots & \cdots \\ |C_{1n}| & |C_{2n}| & \cdots & |C_{nn}| \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix}$$

$$= \frac{1}{|A|} \begin{bmatrix} d_1|C_{11}| + d_2|C_{21}| + \cdots + d_n|C_{n1}| \\ d_1|C_{12}| + d_2|C_{22}| + \cdots + d_n|C_{n2}| \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ d_1|C_{1n}| + d_2|C_{2n}| + \cdots + d_n|C_{nn}| \end{bmatrix}$$

$$= \frac{1}{|A|} \begin{bmatrix} \sum_{i=1}^{n} d_i|C_{i1}| \\ \sum_{i=1}^{n} d_i|C_{i2}| \\ \vdots \\ \sum_{i=1}^{n} d_i|C_{in}| \end{bmatrix}$$

Al igualar los elementos correspondientes en ambos lados de la ecuación, obtenemos los valores solución

$$x_1^* = \frac{1}{|A|}\sum_{i=1}^{n} d_i|C_{i1}| \qquad x_2^* = \frac{1}{|A|}\sum_{i=1}^{n} d_i|C_{i2}| \qquad \text{(etc.)} \tag{5.17}$$

Los términos de $\sum$ en (5.17) parecen desconocidos. ¿Qué significan? De (5.8), observamos que la expansión de Laplace de un determinante $|A|$ por su primera columna se puede expresar en la forma $\sum_{i=1}^{n} a_{i1}|C_{i1}|$. Si se reemplaza la primera columna de $|A|$ por el vector columna d, pero se mantienen intactas las otras columnas, entonces resulta un nuevo determinante, al cual se le puede llamar $|A_1|$ —el subíndice 1 indica que la primera columna se sustituyó por d—. La expansión de $|A_1|$ por su primera columna (la columna d) produce la expresión $\sum_{i=1}^{n} d_i|C_{i1}|$, porque los elementos d_i ahora toman el lugar de los elementos a_{i1}. Volviendo a (5.17), vemos, por lo tanto, que

$$x_1^* = \frac{1}{|A|}|A_1|$$

De manera similar, si se reemplaza la segunda columna de $|A|$ por el vector columna d, mientras se retienen las otras columnas, el desarrollo del nuevo determinante $|A_2|$ por su segunda columna (la columna d) da como resultado la expresión $\sum_{i=1}^{n} d_i|C_{i2}|$. Cuando se divide entre $|A|$, esta última suma proporciona el valor solución x_2^*, y así sucesivamente.

Este procedimiento se puede generalizar. Para hallar el valor solución de la j-ésima variable x_j^*, sólo podemos reemplazar la j-ésima columna del determinante $|A|$ por los términos constantes $d_1 \cdots d_n$ para obtener un nuevo determinante $|A_j|$ y luego dividir $|A_j|$ entre el determinante original $|A|$. Así, la solución del sistema $Ax = d$ se puede expresar como

$$x_j^* = \frac{|A_j|}{|A|} = \frac{1}{|A|}\begin{vmatrix} a_{11} & a_{12} & \cdots & d_1 & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & d_2 & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & d_n & \cdots & a_{nn} \end{vmatrix} \tag{5.18}$$

(j-ésima columna reemplazada por d)

El resultado en (5.18) es el enunciado de la regla de Cramer. Note que, mientras el método de inversión de matriz produce los valores solución de *todas* las variables endógenas a la vez (x^* es un vector), la regla de Cramer sólo nos da el valor solución de una variable endógena a la vez (x_j^* es un escalar); por esta razón puede no ser eficaz.

Ejemplo 1 Obtenga la solución del sistema de ecuaciones

$$\begin{aligned} 5x_1 + 3x_2 &= 30 \\ 6x_1 - 2x_2 &= 8 \end{aligned}$$

Los coeficientes y los términos constantes proporcionan los siguientes determinantes:

$$|A| = \begin{vmatrix} 5 & 3 \\ 6 & -2 \end{vmatrix} = -28 \qquad |A_1| = \begin{vmatrix} 30 & 3 \\ 8 & -2 \end{vmatrix} = -84$$

$$|A_2| = \begin{vmatrix} 5 & 30 \\ 6 & 8 \end{vmatrix} = -140$$

Por lo tanto, en virtud de (5.18), escribimos de inmediato

$$x_1^* = \frac{|A_1|}{|A|} = \frac{-84}{-28} = 3 \qquad \text{y} \qquad x_2^* = \frac{|A_2|}{|A|} = \frac{-140}{-28} = 5$$

Ejemplo 2 Obtenga la solución del sistema de ecuaciones

$$\begin{aligned} 7x_1 - x_2 - x_3 &= 0 \\ 10x_1 - 2x_2 + x_3 &= 8 \\ 6x_1 + 3x_2 - 2x_3 &= 7 \end{aligned}$$

Los determinantes importantes $|A|$ y $|A_j|$ son

$$|A| = \begin{vmatrix} 7 & -1 & -1 \\ 10 & -2 & 1 \\ 6 & 3 & -2 \end{vmatrix} = -61 \qquad |A_1| = \begin{vmatrix} 0 & -1 & -1 \\ 8 & -2 & 1 \\ 7 & 3 & -2 \end{vmatrix} = -61$$

$$|A_2| = \begin{vmatrix} 7 & 0 & -1 \\ 10 & 8 & 1 \\ 6 & 7 & -2 \end{vmatrix} = -183 \qquad |A_3| = \begin{vmatrix} 7 & -1 & 0 \\ 10 & -2 & 8 \\ 6 & 3 & 7 \end{vmatrix} = -244$$

de manera que los valores solución de las variables son

$$x_1^* = \frac{|A_1|}{|A|} = \frac{-61}{-61} = 1 \qquad x_2^* = \frac{|A_2|}{|A|} = \frac{-183}{-61} = 3 \qquad x_3^* = \frac{|A_3|}{|A|} = \frac{-244}{-61} = 4$$

Observe que en cada uno de estos ejemplos se encuentra que $|A| \neq 0$. Ésta es una condición necesaria para la aplicación de la regla de Cramer, como lo es para la existencia de la inversa A^{-1}. La regla de Cramer, después de todo, se basa en el concepto de la matriz inversa, aunque en la práctica evita el proceso de inversión de matriz.

Nota acerca de los sistemas de ecuaciones homogéneos

Los sistemas de ecuaciones $Ax = d$ considerados antes pueden tener constantes cualesquiera en el vector d. Sin embargo, si $d = 0$, es decir, si $d_1 = d_2 = \cdots = d_n = 0$, el sistema de ecuaciones se convierte en

$$Ax = 0$$

donde 0 es un vector nulo. Este caso especial se denomina *sistema de ecuaciones homogéneo*. La palabra *homogéneo* se relaciona con la propiedad de que cuando las variables $x_1, \ldots, x_n$ se multiplican por el mismo número, el sistema de ecuaciones aún es válido. Esto es posible sólo si los términos constantes del sistema, los que no están unidos a ninguna x_i, son cero.

Si la matriz A es no singular, un sistema de ecuaciones homogéneo puede producir sólo una "solución trivial", a saber, $x_1^* = x_2^* = \cdots = x_n^* = 0$. Esto se deduce del hecho de que la solución $x^* = A^{-1}d$ en este caso se convierte en

$$\underset{(n\times 1)}{x^*} = \underset{(n\times n)}{A^{-1}} \underset{(n\times 1)}{0} = \underset{(n\times 1)}{0}$$

Otra posibilidad es que este resultado se obtenga de la regla de Cramer. El hecho de que $d = 0$ significa que $|A_j|$, para toda j, debe contener una columna completa de ceros y, por lo tanto, la solución resulta ser

$$x_j^* = \frac{|A_j|}{|A|} = \frac{0}{|A|} = 0 \qquad (j = 1, 2, \ldots, n)$$

¡Qué curioso, la *única* forma de obtener una solución *no* trivial de un sistema de ecuaciones homogéneo es tener $|A| = 0$, es decir, tener una matriz de coeficiente *singular A*! En ese caso, se tiene

$$x_j^* = \frac{|A_j|}{|A|} = \frac{0}{0}$$

donde la expresión $0/0$ no es igual a cero, sino más bien algo indefinido. En consecuencia, la regla de Cramer no es aplicable. Esto no significa que no podamos obtener soluciones; significa sólo que no es posible obtener una solución única.

Considere el sistema de ecuaciones homogéneo

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 &= 0 \\ a_{21}x_1 + a_{22}x_2 &= 0 \end{aligned} \tag{5.19}$$

Resulta evidente que $x_1^* = x_2^* = 0$ es una solución, pero esa solución es trivial. Ahora, suponga que la matriz de coeficientes A es singular, así que $|A| = 0$. Esto significa que el vector renglón $[a_{11} \quad a_{12}]$ es un múltiplo del vector renglón $[a_{21} \quad a_{22}]$; en consecuencia, una de las dos ecuaciones es redundante. Si se elimina, por ejemplo, la segunda ecuación de (5.19), tenemos una ecuación (la primera) con dos variables, cuya solución es $x_1^* = (-a_{12}/a_{11})x_2^*$. Esta solución es no trivial y está bien definida si $a_{11} \neq 0$, pero en realidad representa un número infinito de soluciones, porque para todo valor posible de x_2^*, hay un valor correspondiente x_1^* tal que el par constituye una solución. Así que no existe solución no trivial única para este sistema de ecuaciones homogéneo. Esta última afirmación es generalmente válida también para el caso de n variables.

Tipos de solución para un sistema de ecuaciones lineales

Nuestro análisis sobre las distintas variantes del sistema de ecuaciones lineales $Ax = d$ revela que son posibles cuatro tipos de soluciones. Para una mejor visualización de estas variantes, listamos todas las variantes en la tabla 5.1.

Como una primera posibilidad, el sistema tiene una solución no trivial única. Este tipo de resultado surge sólo cuando se tiene un sistema no homogéneo con una matriz de coeficientes

TABLA 5.1
Tipos de solución de un sistema de ecuaciones lineales $Ax = d$

	Vector d	
Determinante $\lvert A\rvert$	**$d \neq 0$ (sistema no homogéneo)**	**$d = 0$ (sistema homogéneo)**
$\lvert A\rvert \neq 0$ (matriz A no singular)	Existe una solución no trivial única $x^* \neq 0$.	Existe una solución trivial única $x^* = 0$.
$\lvert A\rvert = 0$ (matriz A singular)		
Ecuaciones dependientes	Existe un número infinito de soluciones (sin incluir la solución trivial).	Hay un número infinito de soluciones (incluida la solución trivial).
Ecuaciones inconsistentes	No existe solución.	[No es posible.]

no singular A. El segundo resultado posible es una solución trivial única, y esto se relaciona con un sistema homogéneo con una matriz no singular A. Como una tercera posibilidad, el sistema tiene un número infinito de soluciones. Esta posibilidad se relaciona exclusivamente con un sistema en el que las ecuaciones son dependientes (es decir, en el que hay ecuaciones redundantes). Dependiendo de si el sistema es homogéneo, la solución trivial puede ser incluida o no en el conjunto de número infinito de soluciones. Por último, en el caso de un sistema de ecuaciones inconsistente, no existe solución en absoluto. Desde el punto de vista de un constructor de modelos, el resultado más útil y deseable es, por supuesto, el de una solución no trivial única $x^* \neq 0$.

EJERCICIO 5.5

1. Use la regla de Cramer para resolver los siguientes sistemas de ecuaciones:

(*a*) $\begin{aligned} 3x_1 - 2x_2 &= 6 \\ 2x_1 + x_2 &= 11 \end{aligned}$ (*c*) $\begin{aligned} 8x_1 - 7x_2 &= 9 \\ x_1 + x_2 &= 3 \end{aligned}$

(*b*) $\begin{aligned} -x_1 + 3x_2 &= -3 \\ 4x_1 - x_2 &= 12 \end{aligned}$ (*d*) $\begin{aligned} 5x_1 + 9x_2 &= 14 \\ 7x_1 - 3x_2 &= 4 \end{aligned}$

2. Para cada uno de los sistemas de ecuaciones del problema 1, encuentre la inversa de la matriz de coeficientes y obtenga la solución por la fórmula $x^* = A^{-1}d$.

3. Utilice la regla de Cramer para resolver los siguientes sistemas de ecuaciones:

(*a*) $\begin{aligned} 8x_1 - x_2 \qquad\quad &= 16 \\ 2x_2 + 5x_3 &= 5 \\ 2x_1 \qquad + 3x_3 &= 7 \end{aligned}$ (*c*) $\begin{aligned} 4x + 3y - 2z &= 1 \\ x + 2y \qquad &= 6 \\ 3x \qquad + z &= 4 \end{aligned}$

(*b*) $\begin{aligned} -x_1 + 3x_2 + 2x_3 &= 24 \\ x_1 \qquad + x_3 &= 6 \\ 5x_2 - x_3 &= 8 \end{aligned}$ (*d*) $\begin{aligned} -x + y + z &= a \\ x - y + z &= b \\ x + y - z &= c \end{aligned}$

4. Muestre que la regla de Cramer se puede obtener de otra manera mediante el siguiente procedimiento. Multiplique ambos lados de la primera ecuación del sistema $Ax = d$ por el cofactor $|C_{1j}|$, y después multiplique ambos lados de la segunda ecuación por el cofactor $|C_{2j}|$, etc. Sume las ecuaciones recién obtenidas. Luego, asigne los valores $1, 2, \ldots, n$ al índice j, de forma sucesiva, para obtener los valores solución $x_1^*, x_2^*, \ldots, x_n^*$ como se muestra en (5.17).

5.6 Aplicación a modelos de mercado y de ingreso nacional

Los modelos de equilibrio simple como los descritos en el capítulo 3 se resuelven con facilidad mediante la regla de Cramer o por inversión de matriz.

Modelo de mercado

El modelo de dos artículos descrito en (3.12) se puede escribir (después de eliminar las variables de cantidad) como un sistema de dos ecuaciones lineales, como en (3.13′):

$$\begin{aligned} c_1 P_1 + c_2 P_2 &= -c_0 \\ \gamma_1 P_1 + \gamma_2 P_2 &= -\gamma_0 \end{aligned}$$

Los tres determinantes necesarios: $|A|$, $|A_1|$ y $|A_2|$, tienen los siguientes valores:

$$|A| = \begin{vmatrix} c_1 & c_2 \\ \gamma_1 & \gamma_2 \end{vmatrix} = c_1\gamma_2 - c_2\gamma_1$$

$$|A_1| = \begin{vmatrix} -c_0 & c_2 \\ -\gamma_0 & \gamma_2 \end{vmatrix} = -c_0\gamma_2 + c_2\gamma_0$$

$$|A_2| = \begin{vmatrix} c_1 & -c_0 \\ \gamma_1 & -\gamma_0 \end{vmatrix} = -c_1\gamma_0 + c_0\gamma_1$$

Por lo tanto, los precios de equilibrio deben ser

$$P_1^* = \frac{|A_1|}{|A|} = \frac{c_2\gamma_0 - c_0\gamma_2}{c_1\gamma_2 - c_2\gamma_1} \qquad P_2^* = \frac{|A_2|}{|A|} = \frac{c_0\gamma_1 - c_1\gamma_0}{c_1\gamma_2 - c_2\gamma_1}$$

estos precios son precisamente los obtenidos en (3.14) y (3.15). Las cantidades de equilibrio se obtienen, como antes, al fijar $P_1 = P_1^*$ y $P_2 = P_2^*$ en las funciones de oferta y demanda.

Modelo de ingreso nacional

El modelo simple de ingreso nacional citado en (3.23) se puede resolver también mediante el uso de la regla de Cramer. De la manera como se escribe en (3.23), el modelo consiste en las dos ecuaciones simultáneas siguientes:

$$\begin{aligned} Y &= C + I_0 + G_0 \\ C &= a + bY \qquad (a > 0, \quad 0 < b < 1) \end{aligned}$$

Éstas se pueden disponer en la forma

$$\begin{aligned} Y - C &= I_0 + G_0 \\ -bY + C &= a \end{aligned}$$

así que las variables endógenas Y y C aparecen sólo a la izquierda de los signos de igualdad, mientras que las variables exógenas y el parámetro aislado aparecen sólo a la derecha. La matriz de coeficientes ahora toma la forma $\begin{bmatrix} 1 & -1 \\ -b & 1 \end{bmatrix}$, y el vector columna de constantes (datos), $\begin{bmatrix} I_0 + G_0 \\ a \end{bmatrix}$. Note que la suma $I_0 + G_0$ se considera como una sola entidad, es decir, un solo elemento en el vector constante.

La regla de Cramer ahora conduce de inmediato a la solución siguiente:

$$Y^* = \frac{\begin{vmatrix} (I_0 + G_0) & -1 \\ a & 1 \end{vmatrix}}{\begin{vmatrix} 1 & -1 \\ -b & 1 \end{vmatrix}} = \frac{I_0 + G_0 + a}{1 - b}$$

$$C^* = \frac{\begin{vmatrix} 1 & (I_0 + G_0) \\ -b & a \end{vmatrix}}{\begin{vmatrix} 1 & -1 \\ -b & 1 \end{vmatrix}} = \frac{a + b(I_0 + G_0)}{1 - b}$$

El lector debe comprobar que los valores solución recién obtenidos son idénticos a los mostrados en (3.24) y (3.25).

A continuación se intenta resolver este modelo invirtiendo la matriz de coeficientes. Puesto que la matriz de coeficientes es $A = \begin{bmatrix} 1 & -1 \\ -b & 1 \end{bmatrix}$, su matriz de cofactores es $\begin{bmatrix} 1 & b \\ 1 & 1 \end{bmatrix}$ y, por lo tanto, se tiene adj $A = \begin{bmatrix} 1 & 1 \\ b & 1 \end{bmatrix}$. Se deduce que la matriz inversa es

$$A^{-1} = \frac{1}{|A|}\text{adj}A = \frac{1}{1-b}\begin{bmatrix} 1 & 1 \\ b & 1 \end{bmatrix}$$

Como sabemos que, para el sistema de ecuaciones $Ax = d$, la solución se puede expresar como $x^* = A^{-1}d$, este modelo tiene la solución

$$\begin{bmatrix} Y^* \\ C^* \end{bmatrix} = \frac{1}{1-b}\begin{bmatrix} 1 & 1 \\ b & 1 \end{bmatrix}\begin{bmatrix} I_0 + G_0 \\ a \end{bmatrix} = \frac{1}{1-b}\begin{bmatrix} I_0 + G_0 + a \\ b(I_0 + G_0) + a \end{bmatrix}$$

Es fácil ver de nuevo que es la misma solución obtenida antes.

Modelo IS-LM: economía cerrada

Como otro modelo lineal de economía, se puede considerar que la economía está constituida por dos sectores: el sector de los bienes reales y el sector monetario.

El mercado de bienes tiene que ver con las siguientes ecuaciones:

$$\begin{aligned} Y &= C + I + G \\ C &= a + b(1-t)Y \\ I &= d - ei \\ G &= G_0 \end{aligned}$$

Las variables endógenas son Y, C, I e i (donde i es la tasa de interés). La variable exógena es G_0, mientras que a, d, e, b y t son parámetros estructurales.

Ahora se introduce el mercado de dinero mediante las ecuaciones siguientes:

$$\text{Condición de equilibrio: } M_d = M_s$$

$$\text{Demanda de dinero: } M_d = kY - li$$

$$\text{Oferta de dinero: } M_s = M_0$$

donde M_0 es el dinero exógeno existente y k y l son parámetros. Estas tres ecuaciones se pueden condensar en:

$$M_0 = kY - li$$

Juntos, los dos sectores proporcionan el siguiente sistema de ecuaciones:

$$\begin{aligned} Y - C - I &= G_0 \\ b(1-t)Y - C &= -a \\ I + ei &= d \\ kY - li &= M_0 \end{aligned}$$

Note que al sustituir más el sistema se podría reducir a un sistema de ecuaciones de 2×2. Por ahora, se deja como un sistema de 4×4. En forma matricial, se tiene

$$\begin{bmatrix} 1 & -1 & -1 & 0 \\ b(1-t) & -1 & 0 & 0 \\ 0 & 0 & 1 & e \\ k & 0 & 0 & -l \end{bmatrix}\begin{bmatrix} Y \\ C \\ I \\ i \end{bmatrix} = \begin{bmatrix} G_0 \\ -a \\ d \\ M_0 \end{bmatrix}$$

Para hallar el determinante de la matriz de coeficientes, se usa la expansión de Laplace en una de las columnas (de preferencia con la mayor cantidad de ceros). Al expandir la cuarta columna, se encuentra

$$
\begin{aligned}
|A| &= (-e)\begin{vmatrix} 1 & -1 & -1 \\ b(1-t) & -1 & 0 \\ k & 0 & 0 \end{vmatrix} - l\begin{vmatrix} 1 & -1 & -1 \\ b(1-t) & -1 & 0 \\ 0 & 0 & 1 \end{vmatrix} \\
&= (-e)(k)\begin{vmatrix} -1 & -1 \\ -1 & 0 \end{vmatrix} - l\begin{vmatrix} 1 & -1 \\ b(1-t) & -1 \end{vmatrix} \\
&= ek - l[(-1) - (-1)b(1-t)] \\
&= ek + l[1 - b(1-t)]
\end{aligned}
$$

Se puede usar la regla de Cramer para hallar el ingreso de equilibrio Y^*. Esto se logra sustituyendo la primera columna de la matriz de coeficientes A por el vector de variables exógenas y tomando el cociente del determinante de la nueva matriz entre el determinante original, es decir

$$
Y^* = \frac{|A_1|}{|A|} = \frac{\begin{vmatrix} G_0 & -1 & -1 & 0 \\ -a & -1 & 0 & 0 \\ d & 0 & 1 & e \\ M_0 & 0 & 0 & -l \end{vmatrix}}{ek + l[1 - b(1-t)]}
$$

Al usar la expansión de Laplace en la segunda columna del numerador se obtiene

$$
\begin{aligned}
Y^* &= \frac{(-1)(-1)^3\begin{vmatrix} -a & 0 & 0 \\ d & 1 & e \\ M_0 & 0 & -l \end{vmatrix}}{ek + l[1 - b(1-t)]} + \frac{(-1)(-1)^4\begin{vmatrix} G_0 & -1 & 0 \\ d & 1 & e \\ M_0 & 0 & -l \end{vmatrix}}{ek + l[1 - b(1-t)]} \\
&= \frac{\begin{vmatrix} -a & 0 & 0 \\ d & 1 & e \\ M_0 & 0 & -l \end{vmatrix} - \begin{vmatrix} G_0 & -1 & 0 \\ d & 1 & e \\ M_0 & 0 & -l \end{vmatrix}}{ek + l[1 - b(1-t)]}
\end{aligned}
$$

Mediante una expansión ulterior, tenemos

$$
\begin{aligned}
Y^* &= \frac{(1)\begin{vmatrix} -a & 0 \\ M_0 & -l \end{vmatrix} - \left\{(-1)(-1)^3\begin{vmatrix} d & e \\ M_0 & -l \end{vmatrix} + (-1)^4\begin{vmatrix} G_0 & 0 \\ M_0 & -l \end{vmatrix}\right\}}{ek + l[1 - b(1-t)]} \\
&= \frac{al - [d(-l) - eM_0] - G_0(-l)}{ek + l[1 - b(1-t)]} \\
&= \frac{l(a + d + G_0) + eM_0}{ek + l[1 - b(1-t)]}
\end{aligned}
$$

Puesto que la solución para Y^* es lineal con respecto a las variables exógenas, Y^* se puede reescribir como

$$
Y^* = \left(\frac{e}{ek + l[1 - b(1-t)]}\right) M_0 + \left(\frac{l}{ek + l[1 - b(1-t)]}\right)(a + d + G_0)
$$

En esta forma, se puede ver que los multiplicadores de política keynesianos con respecto a la oferta de dinero y al gasto público son los coeficientes de M_0 y G_0, es decir,

Multiplicador de oferta de dinero: $$\frac{e}{ek + l[1 - b(1 - t)]}$$

y

Multiplicador de gastos públicos: $$\frac{l}{ek + l[l - b(1 - t)]}$$

Álgebra de matrices *versus* eliminación de variables

Los modelos económicos recién utilizados para ejemplificar tienen que ver sólo con dos o cuatro ecuaciones y, por consiguiente, sólo es necesario evaluar determinantes de cuarto orden o menor. Para sistemas de ecuaciones grandes, aparecerán determinantes de orden superior, y su evaluación será más complicada; lo mismo que la inversión de matrices grandes. Desde el punto de vista del cálculo, la inversión de matriz y la regla de Cramer no son necesariamente más eficaces que el método de eliminaciones sucesivas de variables.

Sin embargo, los métodos matriciales tienen otros méritos. Como vimos en páginas anteriores, el álgebra de matrices proporciona una notación compacta para cualquier sistema de ecuaciones lineales y también provee un criterio de determinantes para probar la existencia de una solución única. Éstas son ventajas inalcanzables por otra vía. Además, se debe notar que, a diferencia del método de eliminación de variable, que no ofrece medios para expresar la solución en forma analítica, el método de inversión de matriz y la regla de Cramer proporcionan las expresiones solución convenientes $x^* = A^{-1}d$ y $x_j^* = |A_j|/|A|$. Esta clase de expresiones analíticas de la solución son útiles no sólo porque son en sí mismas una expresión concisa del procedimiento de solución real, sino también porque hacen posible la realización de más operaciones matemáticas en la solución, si se requiere.

En ciertas circunstancias, los métodos matriciales pueden presentar incluso una ventaja computacional, como cuando la tarea es resolver al mismo tiempo varios sistemas de ecuaciones que tienen una matriz de coeficientes idéntica A, pero diferentes vectores de término constante. En esos casos, el método de eliminación de variable requeriría que el procedimiento de cálculo se repitiera cada vez que se considera un nuevo sistema de ecuaciones. Sin embargo, con el método de inversión de matriz, se requiere hallar *sólo una vez* la matriz inversa común A^{-1}; entonces la misma inversa se puede usar para premultiplicar los vectores de término constante que pertenecen a los distintos sistemas de ecuaciones en cuestión, a fin de obtener sus soluciones respectivas. Esta ventaja de cálculo particular adquirirá gran significado práctico al considerar la solución de los modelos de Leontief de insumo-producto en la sección 5.7.

EJERCICIO 5.6

1. Resuelva el modelo de ingreso nacional del ejercicio 3.5-1:
   (*a*) Por inversión de matriz (*b*) Por la regla de Cramer
   (Enumere las variables en el orden *Y*, *C*, *T*.)
2. Resuelva el modelo de ingreso nacional del ejercicio 3.5-2:
   (*a*) Por inversión de matriz (*b*) Por la regla de Cramer
   (Enumere las variables en el orden *Y*, *C*, *G*.)

3. Sea la ecuación para IS (*investment saving*)

$$Y = \frac{A}{1-b} - \frac{g}{1-b}i$$

donde $1 - b$ es la propensión marginal a ahorrar, g es la sensibilidad de inversión en relación con las tasas de interés y A es un agregado de variables exógenas. Sea la ecuación para LM (*liquity market*)

$$Y = \frac{M_0}{k} + \frac{l}{k}i$$

donde k y l son la sensibilidad de demanda de dinero respecto al ingreso y a la tasa de interés, respectivamente, y M_0 son los saldos reales en efectivo.

Si $b = 0.7$, $g = 100$, $A = 252$, $k = 0.25$, $l = 200$ y $M_0 = 176$, entonces

(*a*) Escriba el sistema IS-LM en forma matricial.

(*b*) Determine Y e i mediante la inversión de matriz.

5.7 Modelos de Leontief de insumo-producto

En su versión "estática", el análisis de insumo-producto del profesor Wassily Leontief, ganador de un premio Nobel,[3] trata con esta pregunta particular: "¿qué nivel de producción debe tener cada una de las n industrias en una economía a fin de satisfacer la demanda total de un determinado producto?".

Es bastante simple ver el fundamento para el término *análisis de insumo-producto*. La producción de cualquier industria (por ejemplo, la del acero) se requiere como insumos en los productos de muchas otras industrias, e incluso para la producción de esa misma industria; por lo tanto, el nivel "correcto" (es decir, sin déficit ni excedente) de producción de acero dependerá de los requerimientos de acero como insumo de las n industrias. A su vez, la producción de muchas otras industrias entrará a la industria del acero como insumo y, en consecuencia, los niveles "correctos" de los otros productos dependerán en parte de los requerimientos de insumos de la industria del acero. En vista de esta dependencia entre industrias, cualquier conjunto de niveles de producción "correctos" para las n industrias debe ser consistente con los requerimientos de insumos en la economía, para que en ninguna parte surjan cuellos de botella. Desde esta perspectiva, resulta claro que el análisis de insumo-producto debe ser muy útil en la planificación de la producción, así como en la planificación del desarrollo económico de un país o en un programa de defensa nacional.

En términos estrictos, el análisis de insumo-producto no es una forma del análisis de equilibrio general analizado en el capítulo 3. Aunque se remarca la interdependencia de las distintas industrias, los niveles de producción "correctos" previstos son los que satisfacen las relaciones técnicas de insumo-producto y no las condiciones de equilibrio del mercado. No obstante, el problema planteado en el análisis de insumo-producto también se reduce a uno de resolver un sistema de ecuaciones simultáneas, y de nuevo puede servir el álgebra de matrices.

Estructura de un modelo de insumo-producto

Puesto que un modelo de insumo-producto comprende normalmente un gran número de industrias, su marco de trabajo es por necesidad bastante complejo. Para ejemplificar el problema, se adoptan como regla las siguientes suposiciones: (1) cada industria produce sólo un

[3] Wassily W. Leontief, *The Structure of American Economy 1919-1939*, 2a. ed., Oxford University Press, Fair Lawn, N.J., 1951.

TABLA 5.2
Matriz de coeficientes de insumo

$$\begin{array}{c|c}
 & \text{Producto} \\
\text{Insumo} & \begin{array}{ccccc} \text{I} & \text{II} & \text{III} & \cdots & N \end{array} \\
\begin{array}{c} \text{I} \\ \text{II} \\ \text{III} \\ \vdots \\ N \end{array} & \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix}
\end{array}$$

artículo homogéneo (si se interpreta esto en términos generales, el modelo permite considerar el caso de dos o más artículos producidos de manera conjunta, siempre que se produzcan en una proporción fija entre sí); (2) cada industria utiliza una relación de insumos fija (o combinación de factores) para la obtención de su producto, y (3) la producción en cada industria está sujeta a rendimientos constantes a escala, de modo que un cambio de k veces en todo insumo producirá un cambio de exactamente k veces en el producto. Por supuesto, estas suposiciones son irreales. Una característica que salva al modelo es que, si una industria produce dos artículos distintos o emplea dos combinaciones distintas de factores posibles, entonces esa industria se puede descomponer, al menos de manera conceptual, en dos industrias separadas.

De estas suposiciones, vemos que, a fin de producir cada unidad del j-ésimo artículo, el insumo necesario para el i-ésimo artículo debe ser una cantidad fija, que se denotará por a_{ij}. En especial, la producción de cada unidad del j-ésimo artículo requerirá a_{1j} (cantidad) del primer artículo, a_{2j} del segundo artículo,..., y a_{nj} del n-ésimo artículo. (El orden de los subíndices en a_{ij} es fácil de recordar: el primer subíndice se refiere al insumo y el segundo al producto, así que a_{ij} indica cuánto del i-ésimo artículo se usa para la producción de cada unidad del j-ésimo artículo.) Para los fines que aquí perseguimos, podemos suponer que los precios están dados y, por lo tanto, adoptamos "la cantidad con valor de un dólar" de cada artículo como su unidad. Entonces la expresión $a_{32} = 0.35$ significa que se requieren 35 centavos del tercer artículo como insumo para producir el valor de un dólar del segundo artículo. El símbolo a_{ij} se denomina *coeficiente de insumo*.

Para una economía de n industrias, los coeficientes de insumo se pueden ordenar en una matriz $A = [a_{ij}]$, como en la tabla 5.2, en la cual cada *columna* especifica los requerimientos de insumos para la producción de una unidad del producto de una determinada industria. La segunda columna, por ejemplo, establece que para producir una unidad (el valor de un dólar) del artículo II, los insumos necesarios son: a_{12} unidades del artículo I, a_{22} unidades del artículo II, etc. Si ninguna industria emplea su propio producto como insumo, entonces los elementos de la diagonal principal de la matriz A serán cero.

Modelo abierto

Si las n industrias de la tabla 5.2 constituyen la totalidad de la economía, entonces sus productos serían para el único propósito de satisfacer la *demanda de insumos* de las mismas n industrias (que se emplearán para producir más) en oposición a la *demanda final* (por ejemplo, la demanda del consumidor, no para producción). Al mismo tiempo, todos los insumos utilizados en la economía estarían clasificados como *insumos intermedios* (los suministrados por las n industrias) en oposición a los *insumos primarios* (por ejemplo, mano de obra, que no es un producto industrial). Para permitir la presencia de demanda final e insumos primarios, se debe incluir en el modelo un *sector abierto* fuera de la red de n industrias. Este sector abierto puede acomodar las actividades de los consumidores domésticos, el sector gubernamental e incluso a países extranjeros.

En vista de la presencia del sector abierto, la suma de elementos en cada columna de la matriz de coeficientes de insumo A (o *matriz de insumos* A, para abreviar) debe ser menor que 1. Cada suma de columna representa el costo de insumos *parcial* (sin incluir el costo de los insumos primarios) en que se incurre para producir una cantidad con valor de un dólar de algún artículo; si esta suma es mayor que o igual a \$1, entonces la producción no será justificable desde el punto de vista económico. En símbolos este hecho se puede expresar así:

$$\sum_{i=1}^{n} a_{ij} < 1 \qquad (j = 1, 2, \dots, n)$$

donde la suma es sobre i, es decir, sobre los elementos que aparecen en los distintos *renglones* de una columna específica j. Si se continúa este razonamiento, se puede establecer también que, como el valor del producto (\$1) debe ser absorbido totalmente por los pagos a todos los factores de producción, la cantidad por la que la suma de columna no llega a \$1 debe representar el pago para los insumos primarios del sector abierto. Así, el valor de los insumos primarios necesarios para producir una unidad del j-ésimo artículo debe ser $1 - \sum_{i=1}^{n} a_{ij}$.

Si la industria I va a producir apenas suficiente para satisfacer los requerimientos de insumos de las n industrias, así como la demanda final del sector abierto, su nivel de producción x_1 debe satisfacer las siguiente ecuación:

$$x_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n + d_1$$

donde d_1 denota la demanda final de su producto y $a_{1j}x_j$ representa la demanda de insumo de la j-ésima industria.[4] De la misma manera, los niveles de producción de las otras industrias deben satisfacer las ecuaciones

$$\begin{aligned} x_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n + d_2 \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x_n &= a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n + d_n \end{aligned}$$

Después de mover a la izquierda del signo igual los términos en los que intervienen las variables x_j y dejar a la derecha sólo las demandas finales d_j determinadas exógenamente, se pueden expresar los niveles de producción "correctos" de las n industrias mediante el siguiente sistema de n ecuaciones lineales:

$$\begin{aligned} (1 - a_{11})x_1 - \quad a_{12}x_2 - \cdots - \quad a_{1n}x_n &= d_1 \\ -a_{21}x_1 + (1 - a_{22})x_2 - \cdots - \quad a_{2n}x_n &= d_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ -a_{n1}x_1 - \quad a_{n2}x_2 - \cdots + (1 - a_{nn})x_n &= d_n \end{aligned} \tag{5.20}$$

En notación de matrices, esto se puede escribir como

$$\begin{bmatrix} (1 - a_{11}) & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & (1 - a_{22}) & \cdots & -a_{2n} \\ \vdots & \vdots & & \vdots \\ -a_{n1} & -a_{n2} & \cdots & (1 - a_{nn}) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix} \tag{5.20'}$$

Si se ignoran los unos de la diagonal principal de la matriz de la izquierda, la matriz es simplemente $-A = [-a_{ij}]$. Tal como está, por otro lado, la matriz es la *suma* de la matriz identi-

[4] Nunca sume los coeficientes de insumo en un renglón; tal suma, por ejemplo, $a_{11} + a_{12} + \cdots + a_{1n}$, carece de significado económico útil. Por otro lado, la suma de los productos $a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n$ tiene significado económico; éste representa la cantidad total de x_1 necesaria como insumo para las n industrias.

dad I_n (con unos en su diagonal principal y con ceros en cualquier otra parte) y la matriz $-A$. Así, (5.20′) se puede escribir también como

$$(I - A)x = d \tag{5.20″}$$

donde x y d son, respectivamente, el vector variable y el vector de demanda final (término constante). La matriz $I - A$ se llama *matriz de Leontief.* Siempre que $I - A$ sea no singular, se podrá hallar su inversa $(I - A)^{-1}$, y obtener la solución única del sistema a partir de la ecuación

$$x^* = (I - A)^{-1}d \tag{5.21}$$

Un ejemplo numérico

Para fines de ilustración, suponga que sólo hay tres industrias en la economía y un insumo primario, y que la matriz de coeficientes de insumo es como sigue (esta vez se usarán valores decimales):

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} 0.2 & 0.3 & 0.2 \\ 0.4 & 0.1 & 0.2 \\ 0.1 & 0.3 & 0.2 \end{bmatrix} \tag{5.22}$$

Note que cada suma de columna en A es menor que 1, como debe ser. Además, si se denota mediante a_{0j} la cantidad de dólares del insumo primario usado para producir el valor de un dólar del j-ésimo artículo, se puede escribir [al restar cada suma de columna en (5.22) de 1]:

$$a_{01} = 0.3 \qquad a_{02} = 0.3 \qquad \text{y} \qquad a_{03} = 0.4 \tag{5.23}$$

Con la matriz A de (5.22), el sistema abierto de insumo-producto se puede expresar en la forma $(I - A)x = d$ como sigue:

$$\begin{bmatrix} 0.8 & -0.3 & -0.2 \\ -0.4 & 0.9 & -0.2 \\ -0.1 & -0.3 & 0.8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} \tag{5.24}$$

De manera deliberada no se han especificado valores para las demandas finales d_1, d_2 y d_3. De esta manera, al mantener el vector d en forma paramétrica, la solución aparecerá como una "fórmula" en la que es posible alimentar varios vectores específicos d para obtener varias soluciones específicas correspondientes.

Al invertir una matriz de Leontief de 3×3, se encuentra que la solución aproximada de (5.24) (debido al redondeo de cifras decimales) es

$$\begin{bmatrix} x_1^* \\ x_2^* \\ x_3^* \end{bmatrix} = (I - A)^{-1}d = \frac{1}{0.384} \begin{bmatrix} 0.66 & 0.30 & 0.24 \\ 0.34 & 0.62 & 0.24 \\ 0.21 & 0.27 & 0.60 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix}$$

Si sucede que el vector específico de demanda final (por ejemplo, el objetivo de producto final de un programa de desarrollo) es $d = \begin{bmatrix} 10 \\ 5 \\ 6 \end{bmatrix}$, en miles de millones de dólares, entonces surgirán los siguientes valores solución específicos (de nuevo en miles de millones de dólares):

$$x_1^* = \frac{1}{0.384}[0.66(10) + 0.30(5) + 0.24(6)] = \frac{9.54}{0.384} = 24.84$$

y, de manera similar,

$$x_2^* = \frac{7.94}{0.384} = 20.68 \qquad \text{y} \qquad x_3^* = \frac{7.05}{0.384} = 18.36$$

Ahora surge una pregunta importante. La producción de la mezcla de producto x_1^*, x_2^* y x_3^* debe implicar una cantidad requerida definida del insumo primario. ¿La cantidad *requerida podría* ser consistente con lo que está *disponible* en la economía? Con base en (5.23), el insumo primario requerido se puede calcular como sigue:

$$\sum_{j=1}^{3} a_{0j}x_j^* = 0.3(24.84) + 0.3(20.68) + 0.4(18.36) = 21 \times 10^9 \text{ dólares.}$$

Por lo tanto, la demanda final específica $d = \begin{bmatrix} 10 \\ 5 \\ 6 \end{bmatrix}$ será factible si y sólo si la cantidad disponible del insumo primario es por lo menos 21×10^9 dólares. Si no se alcanza la cantidad disponible, entonces se tendrá que revisar en consecuencia ese objetivo de producción particular.

Una característica notable del análisis previo es que, mientras no cambien los coeficientes de insumo, no cambiará la inversa $(I - A)^{-1}$; por lo tanto, sólo se necesita llevar a cabo *una* inversión de matriz, incluso si se van a considerar cientos o miles de vectores distintos de demanda final, por ejemplo un espectro de objetivos de desarrollo alternativos. Esto economiza el esfuerzo de cálculo comparado con el método de eliminación de variable. Sin embargo, la regla de Cramer descrita en (5.18) no comparte esta ventaja, porque cada vez que se usa un vector distinto de demanda final d, se debe calcular un nuevo determinante como el numerador en (5.18), que no es tan simple como multiplicar una matriz inversa conocida $(I - A)^{-1}$ por un nuevo vector d.

Existencia de soluciones no negativas

En el ejemplo numérico anterior, sucede que la matriz de Leontief $I - A$ es no singular; por lo tanto, existen valores solución de las variables de producto x_j. Además, los valores solución x_j^* resultan ser no negativos, como dictaría el sentido económico. Sin embargo, no se puede esperar que tales resultados deseados surjan de manera automática; ocurren sólo cuando la matriz de Leontief posee ciertas propiedades. Estas propiedades se describen en la denominada *condición de Hawkins-Simon*.[5]

Para explicar esta condición, es necesario introducir el concepto matemático de *menores principales* de una matriz, porque los signos algebraicos de los menores principales proporcionan pistas importantes que sirven como guías para llegar a conclusiones analíticas. Ya se sabe que, dada una matriz cuadrada, por ejemplo, B, con determinante $|B|$, un menor es un subdeterminante obtenido al eliminar el i-ésimo renglón y la j-ésima columna de $|B|$, donde i y j no son necesariamente iguales. Si ahora se impone la restricción de que $i = j$, entonces el menor resultante se conoce como menor *principal*. Por ejemplo, dada una matriz B de 3×3, su determinante se puede escribir como

$$|B| = \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix} \tag{5.25}$$

[5] David Hawkins y Herbert A. Simon, "Note: Some Conditions of Macroeconomic Stability", *Econometrica*, julio-octubre, 1949, pp. 245-248.

La eliminación simultánea del i-ésimo renglón y la i-ésima columna ($i = 3, 2, 1$, sucesivamente) da como resultado los tres menores principales de 2×2 siguientes:

$$\begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix} \qquad \begin{vmatrix} b_{11} & b_{13} \\ b_{31} & b_{33} \end{vmatrix} \qquad \begin{vmatrix} b_{22} & b_{23} \\ b_{32} & b_{33} \end{vmatrix} \tag{5.26}$$

En vista de sus dimensiones 2×2, éstos se conocen como *menores principales de segundo orden*. También se pueden generar los *menores principales de primer orden* (1×1) si se eliminan de $|B|$ dos renglones cualesquiera y las columnas con el mismo número que los renglones. Éstos son

$$|b_{11}| = b_{11} \qquad |b_{22}| = b_{22} \qquad |b_{33}| = b_{33} \tag{5.27}$$

Por último, para completar el cuadro, se puede considerar a $|B|$ como el *menor principal de tercer orden* de $|B|$. Note que en los menores listados en (5.25) a (5.27), sus elementos de la diagonal principal consisten exclusivamente de los elementos de la diagonal principal de B. En esto radica el fundamento del nombre "menores principales".[6]

Si bien ciertas aplicaciones económicas requieren la comprobación de los signos algebraicos de *todos* los menores principales de una matriz B, con bastante frecuencia la conclusión depende sólo del patrón de signos de un subconjunto particular de los menores principales a los que se denomina *menores principales directores*, *menores principales ordenados de manera natural* o *menores principales sucesivos*. En el caso 3×3, este subconjunto consiste sólo en los primeros elementos de (5.25) a (5.27):

$$|B_1| \equiv |b_{11}| \qquad |B_2| \equiv \begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix} \qquad |B_3| \equiv \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix} \tag{5.28}$$

Aquí, el único subíndice m en el símbolo $|B_m|$, a diferencia del uso de subíndice en el contexto de la regla de Cramer, se emplea para indicar que el menor principal director es de dimensión $m \times m$. Una forma fácil de obtener los menores principales directores es seccionar el determinante $|B|$ con líneas discontinuas sucesivas como se ilustra:

$$\begin{vmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix} \tag{5.29}$$

Si se toma el elemento superior de la diagonal principal de $|B|$ se obtiene $|B_1|$; si se toman los dos primeros elementos de la diagonal principal, b_{11} y b_{22}, junto con sus elementos acompañantes fuera de la diagonal, se obtiene $|B_2|$, y así sucesivamente.

[6] Una definición alternativa de menores principales tomaría en cuenta las distintas permutaciones de los subíndices i, j y k. Esto significaría, en el contexto de insumo-producto, renumerar las industrias (por ejemplo, la primera industria se convierte en la segunda, y viceversa, de modo que el subíndice 11 se convierte en 22, y el subíndice 22 se transforma en 11, etc.). Como resultado, además de los menores principales de 2×2 en (5.26), se tendría también

$$\begin{vmatrix} b_{22} & b_{21} \\ b_{12} & b_{11} \end{vmatrix} \qquad \begin{vmatrix} b_{33} & b_{31} \\ b_{13} & b_{11} \end{vmatrix} \qquad \text{y} \qquad \begin{vmatrix} b_{33} & b_{32} \\ b_{23} & b_{22} \end{vmatrix}$$

Pero estos tres últimos, en el orden mostrado, corresponden con los tres listados en (5.26) en valor y signo algebraico; así que es posible no considerarlos más para los fines que buscamos. De manera similar, aunque la permutación de los subíndices puede generar más menores principales de 3×3, sólo duplican el de (5.25) en valor y signo y, por lo tanto, se pueden omitir.

Dado un determinante de dimensión superior, por ejemplo, $n \times n$, habrá por supuesto un número más grande de menores principales, pero su patrón de construcción es el mismo. Un menor principal de *k*-ésimo orden se obtiene siempre eliminando de $|B|$, $n - k$ renglones cualesquiera y las columnas con el mismo número. Y sus menores principales directores $|B_m|$ (con $m = 1, 2, \ldots, n$) se forman siempre tomando los m primeros elementos de la diagonal principal en $|B|$ junto con sus elementos acompañantes fuera de la diagonal.

Con esta introducción, estamos listos para expresar el siguiente teorema importante, debido a Hawkins y Simon:

> Dada (a) una matriz B de $n \times n$, con $b_{ij} \leq 0$ $(i \neq j)$ (es decir, con los elementos no positivos fuera de la diagonal) y (b) un vector $d \geq 0$ de dimensión $n \times 1$ (todos los elementos no negativos), existe un vector $x^* \geq 0$ de dimensión $n \times 1$ tal que $Bx^* = d$, si y sólo si
>
> $$|B_m| > 0 \qquad (m = 1, 2, \ldots, n)$$
>
> es decir, si y sólo si los menores principales directores de B son todos positivos.

La importancia de este teorema para el análisis de insumo-producto se aclara cuando se establece que B represente la matriz de Leontief $I - A$ (donde $b_{ij} = -a_{ij}$ para $i \neq j$ son de hecho no positivos), y d, el vector de demanda final (donde todos los elementos son, de hecho, no negativos). Entonces $Bx^* = d$ es equivalente a $(I - A)x^* = d$, y la existencia de la solución $x^* \geq 0$ garantiza niveles de producción no negativos. La condición necesaria y suficiente para esto, conocida como *condición de Hawkins-Simon*, es que todos los menores principales de la matriz de Leontief $I - A$ sean positivos.

La demostración de este teorema es demasiado extensa para presentarla aquí,[7] pero debe ser importante explorar sus significados económicos, que es relativamente fácil de ver en el caso simple de dos industrias $(n = 2)$.

Significado económico de la condición de Hawkins-Simon

Para el caso de dos industrias, la matriz de Leontief es

$$I - A = \begin{bmatrix} 1 - a_{11} & -a_{12} \\ -a_{21} & 1 - a_{22} \end{bmatrix}$$

La primera parte de la condición de Hawkins-Simon, $|B_1| > 0$, requiere que

$$1 - a_{11} > 0 \qquad \text{o} \qquad a_{11} < 1$$

Desde el punto de vista económico, esto requiere que la cantidad del primer artículo usado en la producción del valor de un dólar del primer artículo sea menor que un dólar. La otra parte de la condición, $|B_2| > 0$, requiere que

$$(1 - a_{11})(1 - a_{22}) - a_{12}a_{21} > 0$$

[7] Una explicación completa se encuentra en Akira Takayama, *Mathematical Economics*, 2a. ed., Cambridge University Press, 1985, pp. 380 a 385.

Algunos autores usan una versión alternativa de la condición de Hawkins-Simon, la cual requiere que *todos* los menores principales de *B* (no sólo los directores) sean positivos. Sin embargo, como muestra Takayama, en el presente caso, con la restricción especial en |*B*|, la positividad de los menores principales directores (una condición menos rigurosa) es suficiente para lograr el mismo resultado. No obstante, se debe remarcar que, como regla general, el hecho de que los menores principales directores satisfagan un requerimiento de signo particular no garantiza que todos los menores principales satisfagan también de forma automática ese requerimiento. Por consiguiente, una condición expresada en términos de *todos* los menores principales se debe comprobar tomando en cuenta *todos* los menores principales, y no sólo los directores.

o, equivalentemente,

$$a_{11} + a_{12}a_{21} + (1 - a_{11})\, a_{22} < 1$$

Además, puesto que $(1 - a_{11})\, a_{22}$ es positivo, la desigualdad previa indica que

$$a_{11} + a_{12}a_{21} < 1$$

En un sentido económico, a_{11} mide el uso *directo* del primer artículo como insumo en la producción del primer artículo, y $a_{12}a_{21}$ mide el uso *indirecto*, éste proporciona la cantidad del primer artículo necesaria para producir la cantidad específica del segundo artículo que se destina a la producción de una cantidad con valor de un dólar del primer artículo. Así, la última desigualdad exige que la cantidad del primer artículo usado como insumo directo e indirecto para producir una cantidad con valor de un dólar del artículo, debe ser menor que un dólar. Por lo tanto, lo que hace la condición de Hawkins-Simon es especificar ciertas restricciones de factibilidad y viabilidad para el proceso de producción. El proceso de producción es factible y viable desde el punto de vista económico si y sólo si éste puede dar soluciones con niveles de producción no negativos.

Modelo cerrado

Si el sector exógeno del modelo abierto de insumo-producto es absorbido en el sistema simplemente como otra *industria*, el modelo se convierte en un *modelo cerrado*. En tal modelo no aparecen la demanda final y el insumo primario; en su lugar estarán los requerimientos de insumo y el producto de la industria recién concebida. Ahora todos los bienes serán de naturaleza *intermedia*, porque todo lo que se produce es sólo para satisfacer los requerimientos de insumo de las $(n + 1)$ industrias en el modelo.

A primera vista, la conversión del sector abierto en una industria adicional no parecería crear ningún cambio importante en el análisis. En realidad, puesto que se supone que la nueva industria tiene una relación de insumos fija como cualquier otra industria, el suministro de lo que solía ser el insumo primario ahora debe mantener una proporción fija en relación con lo que solía llamarse *demanda final*. De modo más concreto, esto podría significar, por ejemplo, que en los hogares se consumirá cada artículo en una proporción fija respecto al servicio de mano de obra que suministran. Esto de hecho constituye un cambio importante en el marco analítico en cuestión.

Desde el punto de vista matemático, la desaparición de las demandas finales significa que ahora se tendrá un sistema de ecuaciones homogéneo. Suponiendo sólo cuatro industrias (incluyendo la nueva, designada mediante el subíndice 0), los niveles de producción "correctos", por analogía con (5.20′), serán los que satisfacen el sistema de ecuaciones:

$$\begin{bmatrix} (1-a_{00}) & -a_{01} & -a_{02} & -a_{03} \\ -a_{10} & (1-a_{11}) & -a_{12} & -a_{13} \\ -a_{20} & -a_{21} & (1-a_{22}) & -a_{23} \\ -a_{30} & -a_{31} & -a_{32} & (1-a_{33}) \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Debido a que este sistema de ecuaciones es homogéneo, puede tener una solución no trivial si y sólo si la matriz de Leontief $I - A$ de 4×4 tiene un determinante nulo. La última condición de hecho siempre se satisface: en un modelo cerrado, no existe insumo primario; por consiguiente, cada suma de columna en la matriz A de coeficientes de insumo debe ser ahora exactamente igual a (en vez de menor que) 1; es decir, $a_{0j} + a_{1j} + a_{2j} + a_{3j} = 1$, o bien

$$a_{0j} = 1 - a_{1j} - a_{2j} - a_{3j}$$

Pero esto significa que, en cada columna de la matriz $I - A$, proporcionada antes, el elemento superior siempre es igual al negativo de la suma de los otros tres elementos. En consecuencia, los cuatro renglones son linealmente dependientes, y se debe determinar $|I - A| = 0$. Esto garantiza que el sistema posee soluciones no triviales; de hecho, como se indica en la tabla 5.1, tiene un número infinito de soluciones. Esto significa que un modelo cerrado, con un sistema de ecuaciones homogéneo, no existe combinación única de producción "correcta". Se puede determinar los niveles de producción $x_1^*, \ldots, x_4^*$ en proporción entre sí, pero no se puede fijar sus niveles absolutos a menos que se impongan otras restricciones al modelo.

EJERCICIO 5.7

1. Con base en el modelo en (5.24), si las demandas finales son $d_1 = 30$, $d_2 = 15$ y $d_3 = 10$ (todas en miles de millones de dólares), ¿cuáles son los niveles de producción correctos para las tres industrias? (Redondee las respuestas a dos decimales.)
2. Con la información en (5.23), calcule la cantidad total de insumo primario requerido para producir los niveles de producción correctos del problema 1.
3. En una economía de dos industrias, se sabe que la industria I utiliza 10 centavos de su propio producto y 60 centavos del artículo II para producir una cantidad con valor de un dólar del artículo I; la industria II no utiliza su propio producto pero emplea 50 centavos del artículo I para producir una cantidad con valor de un dólar del artículo II, y el sector abierto demanda 1 000 miles de millones de dólares del artículo I y 2 000 miles de millones del artículo II.
   (*a*) Escriba la matriz de insumos, la matriz de Leontief y la ecuación matricial específica de insumo-producto para esta economía.
   (*b*) Compruebe si los datos de este problema satisfacen la condición de Hawkins-Simon.
   (*c*) Determine los niveles de producción correctos mediante la regla de Cramer.
4. Dados la matriz de insumos y el vector de demanda final
$$A = \begin{bmatrix} 0.05 & 0.25 & 0.34 \\ 0.33 & 0.10 & 0.12 \\ 0.19 & 0.38 & 0 \end{bmatrix} \qquad d = \begin{bmatrix} 1\,800 \\ 200 \\ 900 \end{bmatrix}$$
   (*a*) Explique el significado económico de los elementos 0.33, 0 y 200.
   (*b*) Explique el significado económico (si existe) de la suma de la tercera columna.
   (*c*) Explique el significado económico (si existe) de la suma de la tercera columna.
   (*d*) Escriba la ecuación matricial específica de insumo-producto para este modelo.
   (*e*) Compruebe si los datos de este problema satisfacen la condición de Hawkins-Simon.
5. (*a*) Dada una matriz $B = [b_{ij}]$ de 4×4, escriba los menores principales.
   (*b*) Escriba los menores principales directores.
6. Muestre que, por sí misma (sin otras restricciones en la matriz B), la condición de Hawkins-Simon garantiza la existencia de un vector solución único x^*, aunque no necesariamente no negativa.

5.8 Limitaciones del análisis estático

En la explicación de equilibrio estático en el mercado o en el ingreso nacional, el interés principal ha sido hallar valores de equilibrio de las variables endógenas del modelo. Un punto fundamental que se ignoró en tal análisis es el proceso real de ajustes y reajustes de las variables que en última instancia conducen al estado de equilibrio (si en algún modo es alcanzable). Sólo se preguntó acerca de dónde se llegará pero no cuándo o qué podría suceder mientras tanto.

Por lo tanto, el tipo de análisis estático no toma en cuenta dos problemas importantes. Uno es que, como el proceso de ajuste puede tomar un tiempo largo para completarse, un estado de equilibrio determinado dentro de un marco particular de análisis estático podría haber perdido su importancia incluso antes de lograrse, si mientras tanto las fuerzas exógenas del modelo han experimentado algunos cambios. Éste es el problema de cambios del estado de equilibrio. El segundo es que, aun cuando se permite que el proceso de ajuste transcurra sin perturbaciones, el estado de equilibrio previsto en un análisis estático pueda ser completamente inalcanzable. Éste sería el caso de un equilibrio conocido como inestable, que se caracteriza por el hecho de que el proceso de ajuste alejará a las variables de ese estado de equilibrio, en vez de acercarlas cada vez más. Por lo tanto, ignorar el proceso de ajuste es dejar de lado el problema de accesibilidad del equilibrio.

Los cambios del estado de equilibrio (en respuesta a cambios exógenos) pertenecen a un tipo de análisis llamado *estática comparativa*, y la pregunta de accesibilidad y estabilidad del equilibrio cae dentro del dominio del *análisis dinámico*. Cada una de estas áreas llena un espacio importante en el análisis económico y, por lo tanto, es imperativo examinar también esas áreas de análisis. El estudio del análisis dinámico se deja para la parte 5 del libro y a continuación se centra la atención en el problema de la estática comparativa.

Parte 3

Análisis estático comparativo

Capítulo 6

Estática comparativa y el concepto de derivada

Este capítulo, así como el 7 y el 8, se dedica a los métodos de análisis estático comparativo.

6.1 Naturaleza de la estática comparativa

La estática comparativa, como sugiere el nombre, tiene que ver con la comparación de distintos estados de equilibrio relacionados con diferentes conjuntos de valores de parámetros y variables exógenas. Para los fines de tal comparación, empezamos siempre por suponer un determinado estado de equilibrio inicial. En el modelo de mercado aislado, por ejemplo, tal equilibrio inicial estará representado por un precio establecido P^* y una cantidad correspondiente Q^*. De manera similar, en el modelo simple de ingreso nacional de (3.23), el equilibrio inicial se especificará mediante una Y^* fija y una C^* correspondiente. Entonces, si dejamos que ocurra un cambio desequilibrante en el modelo —en la forma de un cambio en el valor de algún parámetro o variable exógena—, por supuesto que se perturbará el equilibrio inicial. Como resultado, las distintas variables endógenas deben experimentar ciertos ajustes. Si se supone que se puede definir y lograr un nuevo estado de equilibrio pertinente a los nuevos valores de los datos, la pregunta planteada en el análisis estático comparativo es: ¿cómo se compararía el nuevo equilibrio con el anterior?

Se debe notar que en la estática comparativa aún no se toma en cuenta el proceso de ajuste de las variables; solamente se compara el equilibrio inicial (*pre*cambio) con el estado de equilibrio final (*pos*cambio). Asimismo, se excluye aún la posibilidad de que el nuevo equilibrio sea inestable, porque suponemos que se puede alcanzar, tal como lo hicimos en el anterior.

Un análisis estático comparativo puede ser de naturaleza cualitativa o cuantitativa. Si sólo nos interesa la pregunta de, por ejemplo, si un incremento en la inversión I_0 aumentará o disminuirá el ingreso de equilibrio Y^*, el análisis será cualitativo porque la *dirección* de cambio es el único elemento considerado. Pero si nuestro interés se centra en la *magnitud* del cambio en Y^* que resulta de un determinado cambio en I_0 (es decir, el tamaño del multiplicador de inversión), es evidente que el análisis será cuantitativo. Sin embargo, al obtener una respuesta cuantitativa, podemos indicar de forma automática la dirección de cambio a partir de su signo algebraico. Por consiguiente, el análisis cuantitativo abarca siempre al cualitativo.

Debe quedar claro que el problema en consideración es básicamente el de hallar una *tasa de cambio*: la tasa de cambio del valor de equilibrio de una variable endógena respecto al cambio en un parámetro particular o variable exógena. Por esta razón, el concepto matemático de *derivada* adquiere importancia preponderante en la estática comparativa, porque ese concepto —el más fundamental en la rama de la matemática conocida como *cálculo diferencial*— ¡se relaciona directamente con el concepto de tasa de cambio! Además, después veremos que el concepto de derivada es de importancia extrema en problemas de optimización.

6.2 La tasa de cambio y la derivada

Aun cuando el presente contexto se relaciona sólo con las tasas de cambio de los valores de equilibrio de las variables en un modelo, podemos continuar el análisis de una manera más general considerando la tasa de cambio de cualquier variable y como respuesta a un cambio en otra variable x, donde las dos variables se relacionan entre sí mediante la función

$$y = f(x)$$

Aplicada al contexto estático comparativo, la variable y representará el valor de equilibrio de una variable endógena, y x será algún parámetro. Note que, en primer lugar, restringimos el análisis al caso simple donde hay un solo parámetro o variable exógena en el modelo. No obstante, una vez que hayamos comprendido este caso simplificado, la extensión al caso de más parámetros resultará relativamente fácil.

Cociente de diferencias

Puesto que el concepto de "cambio" figura de forma prominente en el presente contexto, se necesita un símbolo especial para representarlo. Cuando la variable x cambia del valor x_0 a un nuevo valor x_1, el cambio se mide por la diferencia $x_1 - x_0$. Por consiguiente, si se usa el símbolo Δ (la letra griega mayúscula delta, para "diferencia") para denotar el cambio, escribimos $\Delta x = x_1 - x_0$. También se requiere una forma para denotar el valor de la función $f(x)$ en varios valores de x. Lo normal es usar la notación $f(x_i)$ para representar el valor de $f(x)$ cuando $x = x_i$. Así, para la función $f(x) = 5 + x^2$, tenemos $f(0) = 5 + 0^2 = 5$, y de manera similar, $f(2) = 5 + 2^2 = 9$, etcétera.

Cuando x cambia de un valor inicial x_0 a un nuevo valor $(x_0 + \Delta x)$, el valor de la función $y = f(x)$ cambia de $f(x_0)$ a $f(x_0 + \Delta x)$. El cambio en y por unidad de cambio en x se representa mediante el *cociente de diferencias*.

$$\frac{\Delta y}{\Delta x} = \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} \tag{6.1}$$

Este cociente, que mide la tasa promedio de cambio de y, se puede calcular si conocemos el valor inicial de x, o x_0, y la magnitud de cambio en x, o Δx. Es decir, $\Delta y/\Delta x$ es una función de x_0 y Δx.

Ejemplo 1

Dada $y = f(x) = 3x^2 - 4$, podemos escribir

$$f(x_0) = 3(x_0)^2 - 4 \qquad f(x_0 + \Delta x) = 3(x_0 + \Delta x)^2 - 4$$

Por lo tanto, el cociente de diferencias es

$$\frac{\Delta y}{\Delta x} = \frac{3(x_0 + \Delta x)^2 - 4 - \left(3x_0^2 - 4\right)}{\Delta x} = \frac{6x_0 \, \Delta x + 3(\Delta x)^2}{\Delta x}$$
$$= 6x_0 + 3 \, \Delta x \qquad \textbf{(6.2)}$$

que se puede evaluar si tenemos x_0 y Δx. Sea $x_0 = 3$ y $\Delta x = 4$; entonces, la tasa promedio de cambio de y es $6(3) + 3(4) = 30$. Esto significa que, en promedio, cuando x cambia de 3 a 7, el cambio en y es 30 unidades por cambio unitario en x.

Derivada

Con frecuencia, nos interesa la tasa de cambio de y cuando Δx es muy pequeña. En tal caso, es posible obtener una aproximación de $\Delta y/\Delta x$ eliminando los términos del cociente de diferencias en el que interviene la expresión Δx. En (6.2), por ejemplo, si Δx es muy pequeña, podemos tomar simplemente el término $6x_0$ de la derecha como una aproximación de $\Delta y/\Delta x$. Mientras más pequeño sea el valor de Δx, más cercana, por supuesto, es la aproximación al valor verdadero de $\Delta y/\Delta x$.

Cuando Δx tiende a cero (lo cual significa que se aproxima cada vez más, pero en realidad nunca llega, a cero), $(6x_0 + 3 \, \Delta x)$ se aproxima al valor $6x_0$, y de la misma manera, $\Delta y/\Delta x$ tenderá a $6x_0$ también. En símbolos, este hecho se expresa ya sea mediante la expresión $\Delta y/\Delta x \to 6x_0$ cuando $\Delta x \to 0$, o por medio de la ecuación

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} (6x_0 + 3\Delta x) = 6x_0 \qquad \textbf{(6.3)}$$

donde el símbolo $\lim\limits_{\Delta x \to 0}$ se lee como "el límite de . . . cuando Δx tiende a cero". Cuando $\Delta x \to 0$, si existe en realidad el límite del cociente de diferencias $\Delta y/\Delta x$, ese límite se llama derivada de la función $y = f(x)$.

En caso de que haya la derivada, conviene resaltar varios puntos acerca de ella. Primero, una derivada es una *función*; de hecho, en esta acepción la palabra *derivada* en realidad significa una función derivada. La función original $y = f(x)$ es una *función primitiva*, y la derivada es otra función derivada de ésta. Mientras el cociente de diferencia es una función de x_0 y Δx, debe observar —de (6.3), por ejemplo— que la derivada es solamente una función de x_0. Esto es porque Δx ya está forzada a tender a cero y, por lo tanto, no se debe considerar como otra variable en la función. Asimismo, debemos agregar que hasta el momento hemos usado el símbolo con subíndice x_0 sólo para remarcar el hecho de que un cambio en x debe empezar desde algún valor específico de x. Una vez comprendido lo anterior, podemos eliminar el subíndice y expresar simplemente que la derivada, como la función primitiva, es por sí misma una función de la variable independiente x; es decir, para cada valor de x, hay un valor único correspondiente para la función derivada.

Segundo, puesto que la derivada es solamente un límite del cociente de diferencias, que mide una tasa de cambio de y, la derivada debe ser también por necesidad una medida de alguna tasa de cambio. En vista de que el cambio en x contemplado en el concepto de derivada es infinitesimal (es decir, $\Delta x \to 0$), la tasa medida por la derivada tiene la naturaleza de una tasa de cambio *instantánea*.

Tercero, hay un asunto de notación. Las funciones derivadas, por lo común, se denotan de dos maneras. Dada una función primitiva $y = f(x)$, una forma de denotar su derivada (si existe) es usar el símbolo $f'(x)$, o simplemente f'; esta notación se atribuye al matemático Lagrange. La otra notación común es dy/dx, diseñada por el matemático Leibniz. [En realidad, hay una tercera notación, Dy, o $Df(x)$, pero no la usaremos en la siguiente explicación.] La notación $f'(x)$, que se asemeja a la notación para la función primitiva $f(x)$, tiene la ventaja de transmitir la idea de que la derivada es por sí misma una función de x. La razón para expresarla como $f'(x)$, y no como, por ejemplo, $\phi(x)$, es destacar que la función f' se deriva de la función primitiva f. La notación alternativa dy/dx sirve en cambio para remarcar que el valor de una derivada mide una tasa de cambio. La letra d es la contraparte de la letra griega mayúscula Δ, y dy/dx difiere de $\Delta y/\Delta x$, sobre todo, en que la primera es el límite de la última cuando Δx tiende a cero. En el análisis posterior se usarán ambas notaciones, dependiendo de cuál parezca más conveniente en un determinado contexto.

Con estas dos notaciones, se puede definir la derivada de una función $y = f(x)$ como sigue:

$$\frac{dy}{dx} \equiv f'(x) \equiv \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$$

Ejemplo 2

En relación con la función $y = 3x^2 - 4$, se ha mostrado que su cociente de diferencias es (6.2), y el límite de ese cociente es (6.3). Con base en esto último, ahora podemos escribir (sustituyendo x_0 con x):

$$\frac{dy}{dx} = 6x, \quad \text{o bien, } f'(x) = 6x$$

Note que distintos valores de x darán a la derivada distintos valores correspondientes. Por ejemplo, cuando $x = 3$, encontramos, sustituyendo $x = 3$ en la expresión $f'(x)$, que $f'(3) = 6(3) = 18$; de manera similar, cuando $x = 4$, tenemos $f'(4) = 6(4) = 24$. Así, mientras $f'(x)$ denota una *función derivada*, las expresiones $f'(3)$ y $f'(4)$ representan cada una un *valor de derivada* específico.

EJERCICIO 6.2

1. Dada la función $y = 4x^2 + 9$:
   (*a*) Encuentre el cociente de diferencias como una función de x y Δx. (Use x en lugar de x_0.)
   (*b*) Obtenga la derivada dy/dx.
   (*c*) Determine $f'(3)$ y $f'(4)$.
2. Dada la función $y = 5x^2 - 4x$:
   (*a*) Encuentre el cociente de diferencias como una función de x y Δx.
   (*b*) Obtenga la derivada dy/dx.
   (*c*) Determine $f'(2)$ y $f'(3)$.
3. Dada la función $y = 5x - 2$:
   (*a*) Encuentre el cociente de diferencias $\Delta y/\Delta x$. ¿De qué tipo de función se trata?
   (*b*) Puesto que la expresión Δx no aparece en la función $\Delta y/\Delta x$ en el inciso (*a*), ¿sería importante para el valor de $\Delta y/\Delta x$ si Δx fuera grande o pequeña? En consecuencia, ¿cuál es el límite del cociente de diferencias cuando Δx tiende a cero?

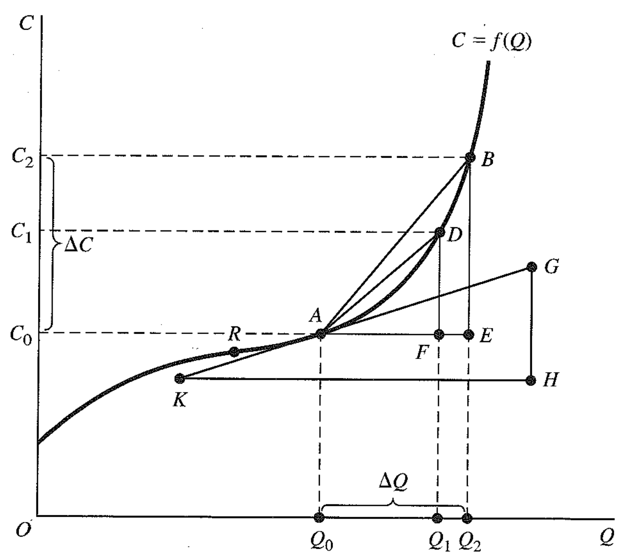
6.3 Derivada y pendiente de una curva

La economía elemental sostiene que, dada una función de costo total $C = f(Q)$, donde C denota el costo total y Q el producto, el costo marginal (CM) se define como el cambio del costo total que resulta de un incremento unitario en el producto; es decir, $\text{CM} = \Delta C/\Delta Q$. Se entiende que ΔQ es un cambio extremadamente pequeño. Para el caso de un producto que tiene unidades discretas (enteros únicamente), un cambio de una unidad es el cambio más pequeño posible; pero para el caso de un producto cuya cantidad es una variable continua, ΔQ se puede referir a un cambio infinitesimal. En este último caso, es bien sabido que mediante la pendiente de la curva de costo total se puede medir el costo marginal. Pero la pendiente de la curva de costo total no es sino el límite de la relación $\Delta C/\Delta Q$, cuando ΔQ tiende a cero. Así, el concepto de la pendiente de una curva es solamente la contraparte geométrica del concepto de la derivada. Ambos tienen que ver con la noción "marginal", de uso tan extendido en economía.

En la figura 6.1 se ha dibujado una curva de costo total C, que es la gráfica de la función (primitiva) $C = f(Q)$. Suponga que se considera a Q_0 como el nivel de producción inicial a partir del cual se mide un incremento de producción; entonces el punto pertinente en la curva de costo es el punto A. Si la producción se elevará a $Q_0 + \Delta Q = Q_2$, el costo total se incrementará de C_0 a $C_0 + \Delta C = C_2$; así, $\Delta C/\Delta Q = (C_2 - C_0)/(Q_2 - Q_0)$. Desde el punto de vista geométrico, ésta es la relación de dos segmentos de recta, EB/AE, o la *pendiente* de la recta AB. Esta relación particular mide una tasa de cambio promedio, el costo marginal *promedio* para la ΔQ particular ilustrada, y representa un cociente de diferencias. Como tal, es una función del valor inicial Q_0 y la cantidad de cambio ΔQ.

¿Qué sucede cuando modificamos la magnitud de ΔQ? Si se contempla un incremento de producción más pequeño (por ejemplo, de Q_0 a Q_1 solamente), entonces la pendiente de la recta AD medirá el costo marginal promedio. Además, a medida que se reduce más y más el incremento de producción, resultarán rectas cada vez más planas hasta que, en el límite (cuando

FIGURA 6.1



$\Delta Q \to 0$), se obtiene la recta KG (que es la *recta tangente* a la curva de costo en el punto A) como la recta pertinente. La pendiente de KG $(= HG/KH)$ mide la pendiente de la curva de costo total en el punto A y representa el límite de $\Delta C/\Delta Q$, cuando $\Delta Q \to 0$, cuando la producción inicial está en $Q = Q_0$. Por lo tanto, en términos de la derivada, la pendiente de la curva $C = f(Q)$ en el punto A corresponde al valor de la derivada particular $f'(Q_0)$.

¿Qué sucede si el nivel de producción inicial se cambia de Q_0 a, por ejemplo, Q_2? En ese caso, el punto B sobre la curva reemplaza al punto A como el punto pertinente, y la pendiente de la curva en el nuevo punto B proporciona el valor de derivada $f'(Q_2)$. Para otros niveles de producción iniciales se obtienen resultados análogos. En general, la derivada $f'(Q)$, como función de Q, varía cuando Q cambia.

6.4 Concepto de límite

La derivada dy/dx se ha definido como el límite del cociente de diferencias $\Delta y/\Delta x$ cuando $\Delta x \to 0$. Si se adoptan los símbolos $q \equiv \Delta y/\Delta x$ (q es el cociente) y $v \equiv \Delta x$ (v por variación en el valor de x), se tiene

$$\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{v \to 0} q$$

En vista del hecho de que el concepto de derivada depende en gran medida del concepto de límite, es imperativo que se tenga una idea clara acerca de este concepto.

Límite izquierdo y límite derecho

El concepto de límite tiene que ver con la pregunta: "¿A qué valor se aproxima una variable (por ejemplo, q) cuando otra variable (por ejemplo, v) tiende a un valor específico (por ejemplo, cero)?" Para que esta pregunta tenga sentido, q debe ser, por supuesto, una función de v; por ejemplo, $q = g(v)$. El interés inmediato es hallar el límite de q cuando $v \to 0$, pero de igual modo se puede explorar fácilmente el caso más general de $v \to N$, donde N es cualquier número real finito. Entonces, $\lim\limits_{v \to 0} q$ será solamente un caso especial de $\lim\limits_{v \to N} q$, donde $N = 0$. En el transcurso de la explicación, en realidad se considerará también el límite de q cuando $v \to +\infty$ (más infinito) o cuando $v \to -\infty$ (menos infinito).

Cuando se dice que $v \to N$, la variable v se aproxima al número N, ya sea desde valores mayores que N, o desde valores menores que N. Si, cuando $v \to N$ por la izquierda (de valores menores que N), q se aproxima a un número finito L, se dice que L es el *límite izquierdo* de q. Por otro lado, si L es el número al que tiende q cuando $v \to N$ por la derecha (desde valores mayores que N), llamamos a L el *límite derecho* de q. Los límites izquierdo y derecho pueden ser iguales o no.

El límite izquierdo de q se simboliza por $\lim\limits_{v \to N^-} q$ (el signo menos se refiere a valores menores que N) y el límite derecho se escribe como $\lim\limits_{v \to N^+} q$. Cuando, y sólo cuando, los dos límites tienen el mismo valor finito (por ejemplo, L), consideramos que existe el límite de q y lo escribimos como $\lim\limits_{v \to N} q = L$. Note que L debe ser un número *finito*. Si se tiene la situación de $\lim\limits_{v \to N} q = \infty$ (o $-\infty$), podemos considerar que q *no* tiene límite, porque $\lim\limits_{v \to N} q = \infty$ significa que $q \to \infty$ cuando $v \to N$, y si q asumiera valores *cada vez mayores* cuando v tiende a N, sería contradictorio decir que q tiene un límite. Sin embargo, como una forma conveniente de expresar el hecho de que $q \to \infty$ cuando $v \to N$, algunas personas de hecho escriben $\lim\limits_{v \to N} q = \infty$ y hablan de que q tiene "límite infinito".

En ciertos casos, sólo se requiere considerar el límite de un lado. Al tomar el límite de q cuando $v \to +\infty$, por ejemplo, sólo es importante el límite izquierdo de q, porque v se puede aproximar a $+\infty$ sólo desde la izquierda. De manera similar, para el caso de $v \to -\infty$, sólo es relevante el límite derecho. En estos casos, el límite existe sólo si q tiende a un valor finito cuando $v \to +\infty$, o cuando $v \to -\infty$.

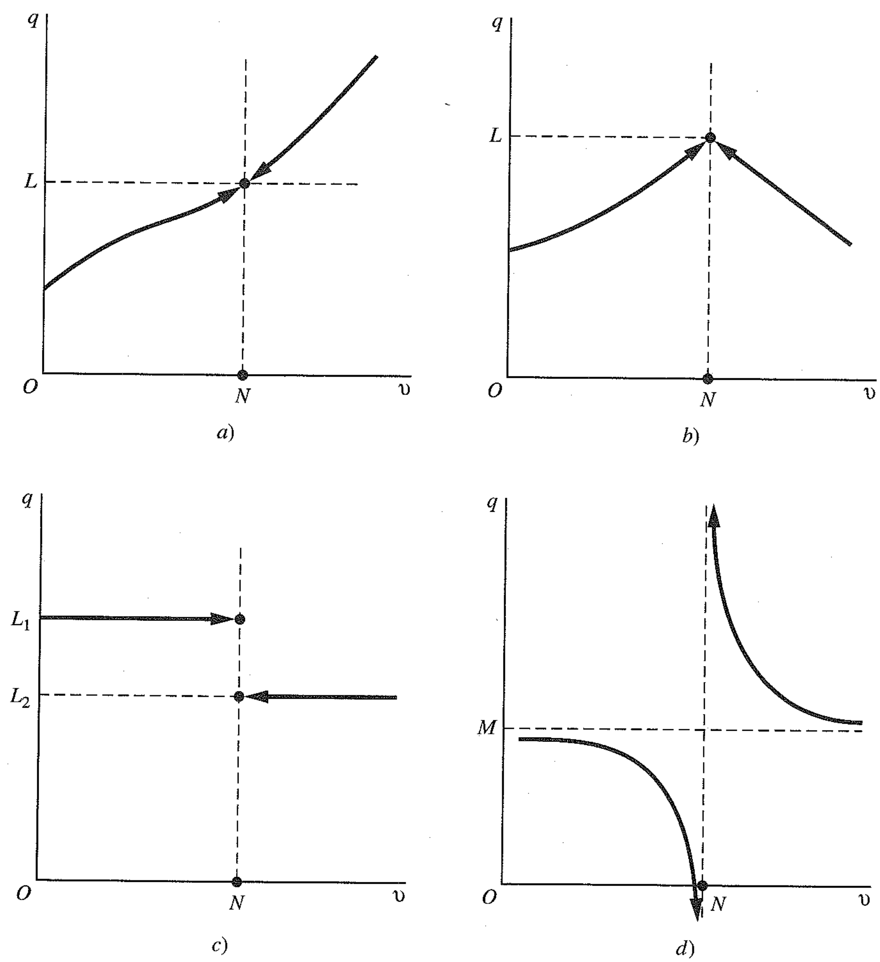
Es importante entender que el símbolo ∞ (infinito) no es un número y, por lo tanto, no puede estar sujeto a las operaciones algebraicas usuales. No es posible tener $3 + \infty$ o $1/\infty$; tampoco se puede escribir $q = \infty$, que es diferente de $q \to \infty$. Sin embargo, es aceptable expresar el *límite* de q como "=" (en oposición a $\to$) ∞, porque esto sólo indica que $q \to \infty$.

Ilustraciones gráficas

A continuación se ilustran, en la figura 6.2, varias situaciones posibles en relación con el límite de una función $q = g(v)$.

En la figura 6.2*a* se muestra una curva lisa. Cuando la variable v tiende al valor N desde *cualquier* lado en el eje horizontal, la variable q tiende al valor L. En este caso, el límite del lado izquierdo es idéntico al del lado derecho; por lo tanto, se puede escribir $\lim\limits_{v \to N} q = L$.

FIGURA 6.2



La curva dibujada en la figura 6.2*b* no es uniforme; tiene un punto de cambio bien definido directamente arriba del punto N. Sin embargo, cuando v tiende a N desde cualquier lado, q de nuevo tiende a un valor idéntico L. De nuevo, el límite de q existe y es igual a L.

La figura 6.2*c* muestra lo que se conoce como una *función escalón*.[1] En este caso, cuando v tiende a N, el límite izquierdo de q es L_1, pero el límite del lado derecho es L_2, un número diferente. Por consiguiente, q no tiene un límite cuando $v \to N$.

Por último, en la figura 6.2*d*, cuando v tiende a N, el límite izquierdo de q es $-\infty$, en tanto que el límite derecho es $+\infty$, porque las dos partes de la curva (hiperbólica) disminuirán y aumentarán de modo indefinido mientras se aproximan a la recta vertical discontinua como una asíntota. De nuevo, $\lim_{v \to N} q$ no existe. Por otro lado, si consideramos un tipo de límite diferente en el diagrama d, a saber, $\lim_{v \to +\infty} q$, entonces sólo el límite izquierdo tiene importancia, y se encuentra que el límite existe: $\lim_{v \to +\infty} q = M$. De modo análogo, se puede comprobar que $\lim_{v \to -\infty} q = M$ existe.

También es posible aplicar los conceptos de límite izquierdo y derecho al análisis del costo marginal en la figura 6.1. En ese contexto, las variables q y v se referirán, respectivamente, al cociente $\Delta C/\Delta Q$ y a la magnitud de ΔQ, con los cambios medidos desde el punto A sobre la curva. En otras palabras, q se referirá a la pendiente de rectas como AB, AD y KG, mientras que v se referirá a la longitud de rectas como Q_0Q_2 (= recta AE) y Q_0Q_1 (= recta AF). Ya hemos visto que, cuando v tiende a cero desde un valor positivo, q tenderá a un valor igual a la pendiente de la recta KG. De manera similar, se puede establecer que, si ΔQ tiende a cero desde un valor negativo (es decir, cuando la *disminución* de producción se vuelve cada vez menor), el cociente $\Delta C/\Delta Q$, medido por la pendiente de rectas como RA (no dibujada), tenderá también a un valor igual a la pendiente de la recta KG. De hecho, la situación aquí es muy semejante a la ilustrada en la figura 6.2*a*. Así, la pendiente de KG en la figura 6.1 (la contraparte de L en la figura 6.2) es el límite del cociente q cuando v tiende a cero, y como tal da el costo marginal al nivel de producción $Q = Q_0$.

Evaluación de un límite

Ahora procederemos a ilustrar la evaluación algebraica de un límite de una función dada $q = g(v)$.

Ejemplo 1

Dada $q = 2 + v^2$, determine el $\lim_{v \to 0} q$. Para tomar el límite izquierdo, se sustituye la serie de valores negativos $-1, -\frac{1}{10}, -\frac{1}{100}, \ldots$ (en ese orden) para v y se encuentra que $(2 + v^2)$ disminuye en forma constante y tiende a 2 (porque v^2 se aproxima poco a poco a 0). A continuación, para el límite del lado derecho, sustituimos la serie de valores positivos $1, \frac{1}{10}, \frac{1}{100}, \ldots$ (en ese orden) para v y se encuentra el mismo límite que antes. En vista de que los dos límites son idénticos, se considera que existe el límite de q y se escribe $\lim_{v \to 0} q = 2$.

[1] Este nombre se explica fácilmente por la forma de la curva. Pero las funciones escalón también se pueden expresar de modo algebraico. La que se ilustra en la figura 6.2*c* se expresa mediante la ecuación

$$q = \begin{cases} L_1 \text{ (para } 0 \le v < N) \\ L_2 \text{ (para } N \le v) \end{cases}$$

Note que, en cada subconjunto de su dominio como está descrito, la función aparece como una función constante distinta, lo que constituye un "escalón" en la gráfica.

En economía, las funciones escalón se pueden usar, por ejemplo, para mostrar los distintos precios facturados por diferentes cantidades compradas (la curva mostrada en la figura 6.2*c* ilustra el *descuento de cantidad*) o las distintas tasas de impuestos aplicables a diferentes escalas de ingresos.

Es tentador considerar la respuesta obtenida en el ejemplo 1 como el resultado de fijar $v = 0$ en la ecuación $q = 2 + v^2$, pero en general se debe resistir esta tentación. Al evaluar $\lim_{v \to N} q$, sólo se permite que v *tienda a N*, pero, como regla, no se permite que $v = N$. De hecho, se puede hablar con bastante legitimidad del límite de q cuando $v \to N$, incluso si N *no* está en el dominio de la función $q = g(v)$. En este último caso, si se intenta establecer $v = N$, es obvio que q estará indefinida.

Ejemplo 2

Dada $q = (1 - v^2)/(1 - v)$ encuentre $\lim_{v \to 1} q$. Aquí, $N = 1$ no está en el dominio de la función, y no se puede establecer $v = 1$ porque eso tendría que ver con la división entre cero. Además, incluso el procedimiento de evaluación de límite de permitir que $v \to 1$, como se usó en el ejemplo 1, causará dificultad, porque el denominador $(1 - v)$ tenderá a cero cuando $v \to 1$, y todavía no se tiene manera de llevar a cabo la división en el límite.

Una manera de evitar esta dificultad es tratar de transformar la relación dada a una forma en la cual v no aparezca en el denominador. Puesto que $v \to 1$ significa que $v \neq 1$, de modo que $(1 - v)$ es distinta de cero, es legítimo dividir la expresión $(1 - v^2)$ entre $(1 - v)$ y escribir[2]

$$q = \frac{1 - v^2}{1 - v} = 1 + v \qquad (v \neq 1)$$

En esta nueva expresión para q, ya no hay un denominador con v en él. Puesto que $(1 + v) \to 2$ cuando $v \to 1$ desde *cualquier* lado, se puede concluir entonces que $\lim_{v \to 1} q = 2$.

Ejemplo 3

Dada $q = (2v + 5)/(v + 1)$, encuentre $\lim_{v \to +\infty} q$. De nuevo aparece la variable v *tanto* en el numerador *como* en el denominador. Si establecemos que $v \to +\infty$ en ambos, el resultado será una relación entre dos números infinitamente grandes, lo cual no tiene un significado claro. Para salir de la dificultad, esta vez se intenta transformar la relación dada a una forma en la cual la variable v no aparecerá en el numerador.[3] Esto, de nuevo, se puede llevar a cabo dividiendo la relación dada. Sin embargo, puesto que $(2v + 5)$ no es exactamente divisible entre $(v + 1)$, el resultado contendrá un residuo como sigue:

$$q = \frac{2v + 5}{v + 1} = 2 + \frac{3}{v + 1}$$

Pero, de todos modos, esta nueva expresión para q ya no tiene un numerador con v en él. Al observar que el residuo $3/(v + 1) \to 0$ cuando $v \to +\infty$, se puede concluir entonces que $\lim_{v \to +\infty} q = 2$.

También existen varios teoremas útiles en la evaluación de límites. Éstos se estudiarán en la sección 6.6.

[2] La división se puede efectuar, como en el caso de números, de la siguiente manera:

$$\begin{array}{r@{\,}l} & 1 + v \\ 1 - v \,\big) & 1 \quad - v^2 \\ & \underline{1 - v} \\ & \quad v - v^2 \\ & \quad \underline{v - v^2} \end{array}$$

Por otro lado, podemos recurrir a la factorización como sigue:

$$\frac{1 - v^2}{1 - v} = \frac{(1 + v)(1 - v)}{1 - v} = 1 + v \qquad (v \neq 1)$$

[3] Note que, a diferencia del caso $v \to 0$, donde se desea sacar a v del *denominador* a fin de evitar la división entre cero, el caso $v \to \infty$ se presenta mejor al sacar a v del *numerador*. Cuando $v \to \infty$, una expresión que contiene a v en el numerador se volverá infinita, pero una expresión con v en el denominador tenderá a cero y desaparecerá suavemente de la escena, lo cual es más conveniente.

Punto de vista formal del concepto de límite

El análisis previo debe haber transmitido algunas ideas generales acerca del concepto de límite. Ahora daremos una definición más precisa. Como en esta definición se hace uso del concepto de *vecindad* en un punto sobre una recta (en particular, un número específico como un punto en la recta de números reales), primero explicaremos el último término.

Para un determinado número L, siempre se puede hallar un número $(L - a_1) < L$ y otro número $(L + a_2) > L$, donde a_1 y a_2 son ciertos números positivos arbitrarios. El conjunto de los números que caen entre $(L - a_1)$ y $(L + a_2)$ se llama el *intervalo* entre esos dos números. Si los números $(L - a_1)$ y $(L + a_2)$ se incluyen en el conjunto, éste es un *intervalo cerrado*; si se excluyen, el conjunto es un *intervalo abierto*. Un intervalo cerrado entre $(L - a_1)$ y $(L + a_2)$ se denota mediante la expresión entre corchetes

$$[L - a_1, L + a_2] \equiv \{q \mid L - a_1 \leq q \leq L + a_2\}$$

y el intervalo *abierto* correspondiente se denota con paréntesis:

$$(L - a_1, L + a_2) \equiv \{q \mid L - a_1 < q < L + a_2\} \tag{6.4}$$

Así, [ ] se relaciona con el signo de desigualdad débil $\leq$, mientras que ( ) se relaciona con el signo de desigualdad estricto $<$. En ambos tipos de intervalos, el número más pequeño $(L - a_1)$ siempre se lista primero. Después, tendremos ocasión de referirnos a intervalos *semiabiertos* y *semicerrados* como $(3, 5]$ y $[6, \infty)$, que tienen los siguientes significados:

$$(3, 5] \equiv \{x \mid 3 < x \leq 5\} \qquad [6, \infty) \equiv \{x \mid 6 \leq x < \infty\}$$

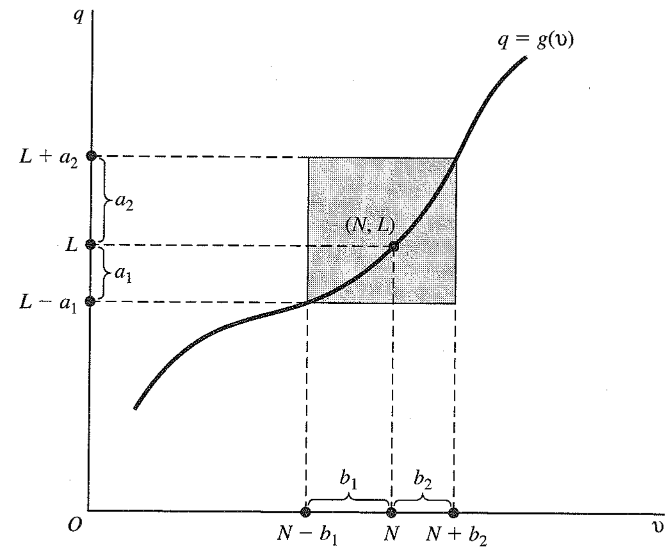
Ahora podemos definir una *vecindad* de L como un intervalo abierto según se define en (6.4), que es un intervalo que "abarca" al número L.[4] Dependiendo de las magnitudes de los números arbitrarios a_1 y a_2, se pueden construir varias vecindades para un determinado número L. Si se utiliza el concepto de vecindad, el límite de una función se puede definir entonces como sigue:

> Cuando v se aproxima a un número N, el límite de $q = g(v)$ es el número L, si, para toda vecindad de L que puede ser elegida, *sin importar cuán pequeña sea*, es posible hallar una vecindad correspondiente de N (excluyendo el punto $v = N$) en el dominio de la función tal que, para todo valor de v en esa vecindad de N su imagen se encuentra en la vecindad de L que fue elegida.

Esta afirmación se aclara con la ayuda de la figura 6.3, que se asemeja a la figura 6.2*a*. De lo que se aprendió acerca de la figura 6.2*a*, sabemos que $\lim_{v \to N} q = L$ en la figura 6.3. Ahora demostraremos que L de hecho satisface la nueva definición de límite. En primer lugar, seleccione una vecindad pequeña arbitraria de L, por ejemplo, $(L - a_1, L + a_2)$. (Esta elección pudo haber sido incluso más pequeña, pero la tomamos relativamente grande para facilitar la exposición.) Ahora construya una vecindad de N, por ejemplo $(N - b_1, N + b_2)$, de modo que las dos vecindades (cuando se amplían al cuadrante I) definan un rectángulo (sombreado en el diagrama) con dos de sus esquinas sobre la curva. Se puede comprobar entonces que, para todo valor de v en esta vecindad de N (sin contar $v = N$), el valor correspondiente de $q = g(v)$ se encuentra en la vecindad elegida de L. De hecho, no importa cuán *pequeña* sea la

[4] La identificación de un intervalo abierto como la vecindad de un punto es válida sólo cuando se considera un punto sobre una recta (un espacio unidimensional). En el caso de un punto en un plano (espacio bidimensional), su vecindad se debe considerar como un área, por ejemplo, un área circular que incluye al punto.

FIGURA 6.3



vecindad de L que se elija, se puede hallar una vecindad de N (correspondientemente pequeña) con la propiedad antes citada. Así, L satisface la definición de límite, como se tenía que demostrar.

Ahora podemos aplicar también la definición dada a la función escalón de la figura 6.2*c* a fin de mostrar que ni L_1 ni L_2 califican como $\lim\limits_{v \to N} q$. Si elegimos una vecindad muy pequeña de L_1 —por ejemplo, apenas el ancho de un cabello en cada lado de L_1—, entonces, sin importar cuál vecindad se elija para N, es posible que el rectángulo relacionado con las dos vecindades no abarque al escalón inferior de la función. En consecuencia, para cualquier valor de $v > N$, el valor correspondiente de q (localizado en el escalón inferior) no estará en la vecindad de L_1 y, por lo tanto, L_1 no pasa la prueba para un límite. Mediante un razonamiento similar, L_2 debe ser descartado como candidato para $\lim\limits_{v \to N} q$. De hecho, en este caso no existe límite para q cuando $v \to N$.

El cumplimiento de la definición se comprueba también de modo algebraico en vez de gráfico. Por ejemplo, considere de nuevo la función

$$q = \frac{1 - v^2}{1 - v} = 1 + v \qquad (v \neq 1) \tag{6.5}$$

Vimos en el ejemplo 2 que $\lim\limits_{v \to 1} q = 2$; por lo tanto, aquí se tiene $N = 1$ y $L = 2$. Para comprobar que $L = 2$ es de hecho el límite de q, debemos demostrar que, para cada vecindad elegida de L, $(2 - a_1, 2 + a_2)$, existe una vecindad de N, $(1 - b_1, 1 + b_2)$, tal que, siempre que v esté en esta vecindad de N, q debe estar en la vecindad elegida de L. En esencia, esto significa que, para valores dados de a_1 y a_2, no importa cuán pequeños sean, se debe hallar dos números b_1 y b_2 tales que, siempre que la desigualdad

$$1 - b_1 < v < 1 + b_2 \qquad (v \neq 1) \tag{6.6}$$

se satisfaga, otra desigualdad de la forma

$$2 - a_1 < q < 2 + a_2 \tag{6.7}$$

también se debe satisfacer. Para hallar este par de números b_1 y b_2, se reescribe primero (6.7) sustituyendo (6.5):

$$2 - a_1 < 1 + v < 2 + a_2 \tag{6.7'}$$

Esto, a su vez, se puede transformar (al restar 1 de cada lado) en la desigualdad

$$1 - a_1 < v < 1 + a_2 \tag{6.7''}$$

Una comparación de (6.7″), una variante de (6.7), con (6.6) indica que si se eligen los dos números b_1 y b_2 como $b_1 = a_1$ y $b_2 = a_2$, las dos desigualdades (6.6) y (6.7) siempre se satisfarán de forma simultánea. Así, la vecindad de N, $(1 - b_1, 1 + b_2)$, como se requiere en definición de límite, se puede encontrar de hecho para el caso $L = 2$, y esto establece a $L = 2$ como el límite.

Ahora utilizaremos la definición de un límite en el sentido opuesto, para mostrar que otro valor (por ejemplo, 3) no puede calificar como $\lim\limits_{v \to 1} q$ para la función en (6.5). Si 3 fuera ese límite, tendría que ser cierto que, para cada vecindad elegida de 3, $(3 - a_1, 3 + a_2)$, existe una vecindad de 1, $(1 - b_1, 1 + b_2)$, tal que, siempre que v esté en la última vecindad, q debe estar en la primera vecindad. Es decir, siempre que se satisfaga la desigualdad

$$1 - b_1 < v < 1 + b_2$$

se debe satisfacer otra desigualdad de la forma

$$3 - a_1 < 1 + v < 3 + a_2$$

o bien,
$$2 - a_1 < v < 2 + a_2$$

La *única* forma de lograr este resultado es elegir $b_1 = a_1 - 1$ y $b_2 = a_2 + 1$. Esto implicaría que la vecindad de 1 será el intervalo abierto $(2 - a_1, 2 + a_2)$. De acuerdo con la definición de límite, a_1 y a_2 pueden hacerse arbitrariamente pequeños, por ejemplo, $a_1 = a_2 = 0.1$. En ese caso, el último intervalo mencionado resultará ser (1.9, 2.1) que se encuentra por completo a la derecha del punto $v = 1$ en el eje horizontal y, por consiguiente, no es una vecindad de 1. Así que el número 3 no satisface la definición de límite. Se puede emplear un procedimiento similar para mostrar que *cualquier* número distinto de 2 contradice la definición de límite en el caso presente.

En general, si un número satisface la definición de límite de q cuando $v \to N$, entonces ningún otro lo puede hacer. Si existe un límite, es único.

EJERCICIO 6.4

1. Dada la función $q = (v^2 + v - 56)/(v - 7)$, $(v \neq 7)$, halle el límite izquierdo y el límite derecho de q cuando v tiende a 7. ¿Se puede concluir de estas respuestas que q tiene un límite cuando v se aproxima a 7?
2. Dada $q = [(v + 2)^3 - 8]/v$, $(v \neq 0)$, encuentre:
   (*a*) $\lim\limits_{v \to 0} q$ (*b*) $\lim\limits_{v \to 2} q$ (*c*) $\lim\limits_{v \to a} q$
3. Dada $q = 5 - 1/v$, $(v \neq 0)$, encuentre:
   (*a*) $\lim\limits_{v \to +\infty} q$ (*b*) $\lim\limits_{v \to -\infty} q$
4. Use la figura 6.3 para mostrar que *no se puede* considerar al número $(L + a_2)$ como el límite de q cuando v tiende a N.

6.5 Digresión acerca de desigualdades y valores absolutos

Ya antes hemos encontrado muchos signos de desigualdad. En la explicación de la sección 6.4, también se aplicaron operaciones matemáticas a desigualdades. Al transformar (6.7′) en (6.7″), por ejemplo, se resta 1 de cada lado de la desigualdad. ¿Qué reglas de operaciones se aplican por lo general a las desigualdades (a diferencia de las ecuaciones)?

Reglas de desigualdades

Para comenzar, expresamos una propiedad importante de las desigualdades: las desigualdades son *transitivas*. Esto significa que, si $a > b$ y si $b > c$, entonces $a > c$. Puesto que las igualdades (ecuaciones) son también transitivas, la propiedad transitiva se debe aplicar a desigualdades "débiles" ($\geq$ o $\leq$) así como a las "estrictas" ($>$ o $<$). Así, se tiene

$$a > b, b > c \;\Rightarrow\; a > c$$
$$a \geq b, b \geq c \;\Rightarrow\; a \geq c$$

Esta propiedad es la que hace posible la escritura de una *desigualdad continuada*, como $3 < a < b < 8$ o $7 \leq x \leq 24$. (Al escribir una desigualdad continuada, los signos de desigualdad se disponen, en general, en la misma dirección, normalmente con el número más pequeño a la izquierda.)

Las reglas más importantes de las desigualdades son las que gobiernan la suma (resta) de un número a (de) una desigualdad, la multiplicación o división por un número, y elevar al cuadrado una desigualdad. En particular, estas reglas son como sigue.

Regla I (suma y resta) $a > b \Rightarrow a \pm k > b \pm k$

Una desigualdad aún es válida si se suma o resta una cantidad igual de cada lado. Esta regla se puede generalizar así: si $a > b > c$, entonces $a \pm k > b \pm k > c \pm k$.

Regla II (multiplicación y división)

$$a > b \Rightarrow \begin{cases} ka > kb & (k > 0) \\ ka < kb & (k < 0) \end{cases}$$

La multiplicación de ambos lados por un número *positivo* conserva la desigualdad, pero un número *negativo* causa que se invierta el *sentido* (o *dirección*) de la desigualdad.

Ejemplo 1 Puesto que $6 > 5$, la multiplicación por 3 produce $3(6) > 3(5)$, o $18 > 15$; pero la multiplicación por -3 da como resultado $(-3)6 < (-3)5$, o $-18 < -15$.

La división de una desigualdad entre un número n es equivalente a la multiplicación por el número $1/n$; por lo tanto, la regla sobre la división se incluye en la regla sobre la multiplicación.

Regla III (elevar al cuadrado) $a > b, (b \geq 0) \Rightarrow a^2 > b^2$

Si sus dos lados no son negativos, la desigualdad se mantendrá cuando ambos lados se eleven al cuadrado.

Ejemplo 2 Puesto que $4 > 3$ y como ambos lados son positivos, se tiene $4^2 > 3^2$, o bien $16 > 9$. De manera similar, puesto que $2 > 0$, se deduce que $2^2 > 0^2$, o bien $4 > 0$.

Las reglas I a III se expresaron en términos de desigualdades estrictas, pero su validez no se ve afectada si los signos $>$ se reemplazan con signos $\geq$.

Valores absolutos y desigualdades

Cuando el dominio de una variable x es un intervalo abierto (a, b), el dominio se puede denotar mediante el conjunto $\{x \mid a < x < b\}$ o, en forma más simple, mediante la desigualdad $a < x < b$. De manera similar, si es un intervalo cerrado $[a, b]$, se puede expresar mediante la desigualdad $a \leq x \leq b$. En el caso especial de un intervalo de la forma $(-a, a)$ —por ejemplo, $(-10, 10)$—, se puede representar ya sea mediante la desigualdad $-10 < x < 10$ o, como alternativa, mediante la desigualdad

$$|x| < 10$$

donde el símbolo $|x|$ denota el *valor absoluto* (o *valor numérico*) de x.

Para cualquier número real n, el valor absoluto de n se define como sigue:[5]

$$|n| \equiv \begin{cases} n & (\text{si } n > 0) \\ -n & (\text{si } n < 0) \\ 0 & (\text{si } n = 0) \end{cases} \qquad \textbf{(6.8)}$$

Note que, si $n = 15$, entonces $|15| = 15$; pero si $n = -15$, se encuentra también que

$$|-15| = -(-15) = 15$$

Por lo tanto, en efecto el valor absoluto de cualquier número real es simplemente su valor numérico después que se elimina el signo. Por esta razón, se tiene siempre $|n| = |-n|$. El valor absoluto de n se llama también *módulo* de n.

Dada la expresión $|x| = 10$, podemos concluir de (6.8) que x debe ser 10 o -10. De la misma manera, la expresión $|x| < 10$ significa que (1) si $x > 0$, entonces $x \equiv |x| < 10$, de modo que x debe ser menor que 10; pero también (2) si $x < 0$, entonces de acuerdo con (6.8) se tiene $-x \equiv |x| < 10$, o bien $x > -10$, así que x debe ser mayor que -10. Por consiguiente, al combinar las dos partes de este resultado, se ve que x debe estar dentro del intervalo abierto $(-10, 10)$. En general, se puede escribir

$$|x| < n \Leftrightarrow -n < x < n \qquad (n > 0) \qquad \textbf{(6.9)}$$

que también se puede extender a desigualdades débiles como sigue:

$$|x| \leq n \Leftrightarrow -n \leq x \leq n \qquad (n \geq 0) \qquad \textbf{(6.10)}$$

Debido a que por sí mismos son números, los valores absolutos de dos números m y n se pueden sumar, restar, multiplicar y dividir. Las siguientes propiedades caracterizan a los valores absolutos:

$$|m| + |n| \geq |m + n|$$
$$|m| \cdot |n| = |m \cdot n|$$
$$\frac{|m|}{|n|} = \left|\frac{m}{n}\right|$$

Es interesante ver que la primera de estas propiedades tiene que ver con una desigualdad y no con una ecuación. La razón de esto se ve con facilidad: si bien la expresión del lado izquierdo

[5] De nuevo se advierte que, aunque la notación de valor absoluto es similar a la de un determinante de primer orden, estos dos conceptos son por completo diferentes. La definición de un determinante de primer orden es $|a_{ij}| \equiv a_{ij}$, sin importar el signo de a_{ij}. Por otro lado, en la definición de valor absoluto $|n|$, el signo de n marca la diferencia. El contexto de la explicación debe aclarar por lo común si un valor absoluto o si un determinante de primer orden está bajo consideración.

$|m| + |n|$ es en definitiva una *suma* de dos valores numéricos (ambos tomados como positivos), la expresión $|m + n|$ es el valor numérico de una suma (si m y n son, por ejemplo, positivos) o una diferencia (si m y n tienen signos opuestos). Así, el lado izquierdo puede exceder al derecho.

Ejemplo 3 Si $m = 5$ y $n = 3$, entonces $|m| + |n| = |m + n| = 8$. Pero si $m = 5$ y $n = -3$, entonces $|m| + |n| = 5 + 3 = 8$, mientras que

$$|m + n| = |5 - 3| = 2$$

es un número más pequeño.

Por otro lado, en las otras dos propiedades no importa si m y n tienen signos idénticos u opuestos, ya que, al tomar el valor absoluto del producto o el cociente en el lado derecho, el signo del último término se elimina en cualquier caso.

Ejemplo 4 Si $m = 7$ y $n = 8$, entonces $|m| \cdot |n| = |m \cdot n| = 7(8) = 56$. Pero incluso si $m = -7$ y $n = 8$ (signos opuestos), aún se obtiene el mismo resultado de

$$|m| \cdot |n| = |-7| \cdot |8| = 7(8) = 56$$

y

$$|m \cdot n| = |-7(8)| = 7(8) = 56$$

Solución de una desigualdad

Al igual que una ecuación, una desigualdad que contiene una variable (por ejemplo, x) puede tener una solución; la solución, si existe, es un conjunto de valores de x que hacen que la desigualdad sea un enunciado verdadero. Por lo común, la solución por sí misma estará en la forma de una desigualdad.

Ejemplo 5 Obtenga la solución de la desigualdad

$$3x - 3 > x + 1$$

Como cuando se resuelve una ecuación, los términos variables se deben reunir primero en un lado de la desigualdad. Al sumar $(3 - x)$ a ambos lados, se obtiene

$$3x - 3 + 3 - x > x + 1 + 3 - x$$

o bien,

$$2x > 4$$

Multiplicar ambos lados por $\frac{1}{2}$ (lo cual no cambia el sentido de la desigualdad, porque $\frac{1}{2} > 0$) producirá entonces la solución

$$x > 2$$

que por sí misma es una desigualdad. Esta solución no es un solo número, sino un conjunto de números. Por lo tanto, la solución se puede expresar también como el conjunto $\{x \mid x > 2\}$ o como el intervalo abierto $(2, \infty)$.

Ejemplo 6 Resuelva la desigualdad $|1 - x| \leq 3$. Primero, se elimina la notación de valor absoluto utilizando (6.10). La desigualdad dada es equivalente a la afirmación de que

$$-3 \leq 1 - x \leq 3$$

o bien, después de restar 1 de cada lado,

$$-4 \leq -x \leq 2$$

Al multiplicar cada lado por (-1), se obtiene

$$4 \geq x \geq -2$$

donde el sentido de la desigualdad se ha invertido apropiadamente. Escribiendo primero el número más pequeño, podemos expresar la solución en la forma de la desigualdad

$$-2 \leq x \leq 4$$

o en la forma del conjunto $\{x \mid -2 \leq x \leq 4\}$ o el intervalo cerrado $[-2, 4]$.

En ocasiones, un problema podría requerir el cumplimiento de varias desigualdades en varias variables al mismo tiempo; entonces, se debe resolver un sistema de desigualdades simultáneas. Este problema surge, por ejemplo, en programación no lineal, que analizaremos en el capítulo 13.

EJERCICIO 6.5

1. Resuelva las siguientes desigualdades:
   (*a*) $3x - 1 < 7x + 2$ (*c*) $5x + 1 < x + 3$
   (*b*) $2x + 5 < x - 4$ (*d*) $2x - 1 < 6x + 5$
2. Si $8x - 3 < 0$ y $8x > 0$, exprese estas desigualdades en una desigualdad continua y encuentre su solución.
3. Resuelva lo siguiente:
   (*a*) $|x + 1| < 6$ (*b*) $|4 - 3x| < 2$ (*c*) $|2x + 3| \leq 5$

6.6 Teoremas de límites

El interés en relación con las tasas de cambio nos llevó a considerar el concepto de derivada, que, por su naturaleza de límite de un cociente de diferencias, nos motivó a su vez a estudiar preguntas acerca de la existencia y evaluación de un límite. El proceso básico de evaluación de límite como se ilustró en la sección 6.4, tiene que ver con hacer que la variable v tienda a un número particular (por ejemplo, N) y observar el valor al que se aproxima q. Sin embargo, cuando en realidad se evalúa el límite de una función, se puede hacer uso de ciertos teoremas de límites establecidos, que desde el punto de vista matemático simplifican la tarea, en particular para funciones complicadas.

Teoremas en los que interviene una sola función

Cuando se tiene una sola función $q = g(v)$, son pertinentes los siguientes teoremas.

Teorema I Si $q = av + b$, entonces $\lim\limits_{v \to N} q = aN + b$ (a y b son constantes).

Ejemplo 1 Dada $q = 5v + 7$, se tiene $\lim\limits_{v \to 2} q = 5(2) + 7 = 17$. De manera similar, $\lim\limits_{v \to 0} q = 5(0) + 7 = 7$.

Teorema II Si $q = g(v) = b$, entonces $\lim\limits_{v \to N} q = b$.

Este teorema, que afirma que el límite de una función constante es la constante en esa función, es solamente un caso especial del teorema I, con $a = 0$. (En el ejercicio 6.2-3 se encontró un ejemplo de este caso.)

Teorema III Si $q = v$, entonces $\lim\limits_{v \to N} q = N$.

Si $q = v^k$, entonces $\lim\limits_{v \to N} q = N^k$.

Ejemplo 2 Dada $q = v^3$, se tiene $\lim_{v\to 2} q = (2)^3 = 8$.

Tal vez notó que, en los teoremas I al III, lo que se hizo para hallar el límite de q cuando $v \to N$ es de hecho hacer $v = N$. Pero éstos son casos especiales y no alteran la regla general de que "$v \to N$" no significa "$v = N$".

Teoremas en los que intervienen dos funciones

Si tenemos dos funciones de la misma variable independiente v, $q_1 = g(v)$ y $q_2 = h(v)$, y si *ambas* funciones poseen límites como sigue:

$$\lim_{v\to N} q_1 = L_1 \qquad \lim_{v\to N} q_2 = L_2$$

donde L_1 y L_2 son dos números *finitos*, los siguientes teoremas son apropiados.

Teorema IV (teorema del límite de suma-diferencia)

$$\lim_{v\to N}(q_1 \pm q_2) = L_1 \pm L_2$$

El límite de una suma (diferencia) de dos funciones es la suma (diferencia) se sus respectivos límites.

En particular, se observa que

$$\lim_{v\to N} 2q_1 = \lim_{v\to N}(q_1 + q_1) = L_1 + L_1 = 2L_1$$

que concuerda con el teorema I.

Teorema V (teorema del límite de producto)

$$\lim_{v\to N}(q_1 q_2) = L_1 L_2$$

El límite de un producto de dos funciones es el producto de sus límites.

Aplicado al cuadrado de una función, esto da

$$\lim_{v\to N}(q_1 q_1) = L_1 L_1 = L_1^2$$

lo cual coincide con el teorema III.

Teorema VI (teorema del límite del cociente)

$$\lim_{v\to N} \frac{q_1}{q_2} = \frac{L_1}{L_2} \qquad (L_2 \neq 0)$$

El límite de un cociente de dos funciones es el cociente de sus límites. Por supuesto, el límite L_2 está restringido a ser distinto de cero, de lo contrario el cociente está indefinido.

Ejemplo 3 Determine $\lim_{v\to 0}(1 + v)/(2 + v)$. Puesto que aquí se tiene $\lim_{v\to 0}(1 + v) = 1$ y $\lim_{v\to 0}(2 + v) = 2$, el límite deseado es $\frac{1}{2}$.

Recuerde que L_1 y L_2 representan números finitos; de lo contrario, estos teoremas no se aplican. Además, en el caso del teorema VI, L_2 no debe ser cero. Si no se cumplen estas restricciones, se debe recurrir al método de evaluación de límite ilustrado en los ejemplos 2 y 3 de la sección 6.4, que se relacionan con los casos, respectivamente, de L_2 con valor cero y de L_2 que es infinito.

Límite de una función polinomial

Ahora que ya tenemos los teoremas de límites, podemos evaluar fácilmente el límite de cualquier función polinomial

$$q = g(v) = a_0 + a_1 v + a_2 v^2 + \cdots + a_n v^n \tag{6.11}$$

cuando v tiende al número N. Puesto que los límites en los términos separados son, respectivamente,

$$\lim_{v\to N} a_0 = a_0 \qquad \lim_{v\to N} a_1 v = a_1 N \qquad \lim_{v\to N} a_2 v^2 = a_2 N^2 \qquad \text{(etc.)}$$

el límite de la función polinomial es (por el teorema del límite de suma)

$$\lim_{v\to N} q = a_0 + a_1 N + a_2 N^2 + \cdots + a_n N^n \tag{6.12}$$

Este límite es también, según observamos, igual a $g(N)$, es decir, igual al valor de la función (6.11) cuando $v = N$. La importancia de este resultado particular se observa al analizar el concepto de *continuidad* de la función polinomial.

EJERCICIO 6.6

1. Encuentre los límites de la función $q = 7 - 9v + v^2$:
   (*a*) Cuando $v \to 0$ (*b*) Cuando $v \to 3$ (*c*) Cuando $v \to -1$
2. Halle los límites de $q = (v + 2)(v - 3)$:
   (*a*) Cuando $v \to -1$ (*b*) Cuando $v \to 0$ (*c*) Cuando $v \to 5$
3. Obtenga los límites de $q = (3v + 5)/(v + 2)$:
   (*a*) Cuando $v \to 0$ (*b*) Cuando $v \to 5$ (*c*) Cuando $v \to -1$

6.7 Continuidad y diferenciabilidad de una función

El análisis anterior del concepto de límite y su evaluación se puede usar ahora para definir la continuidad y diferenciabilidad de una función. Estos conceptos tratan directamente de la derivada de la función, que es lo que nos interesa aquí.

Continuidad de una función

Cuando una función $q = g(v)$ posee un límite mientras v tiende al punto N en el dominio, y cuando el límite es también igual a $g(N)$, es decir, igual al valor de la función en $v = N$, se dice que la función es *continua* en N. Como se define aquí, el término *continuidad* tiene que ver con no menos de tres requerimientos: (1) el punto N debe estar en el dominio de la función, es decir, $g(N)$ está definida; (2) la función debe tener un límite cuando $v \to N$, esto es, $\lim_{v\to N} g(v)$ existe, y (3) que el límite debe tener el mismo valor que $g(N)$; es decir, $\lim_{v\to N} g(v) = g(N)$.

Es importante notar que si bien el punto (N, L) no se consideró al analizar el límite de la curva en la figura 6.3, ya no lo excluimos en el presente contexto. Además, como se afirma en especial en el tercer requerimiento, el punto (N, L) debe estar sobre la gráfica de la función antes de que ésta se pueda considerar como continua en un punto N.

A continuación, comprobamos si las funciones mostradas en la figura 6.2 son continuas. En el diagrama a, los tres requerimientos se satisfacen en el punto N. El punto N está en el dominio; q tiene el límite L cuando $v \to N$; y el límite L resulta ser también el valor de la función en N. Así, la función representada por esa curva es continua en N. Lo mismo se cumple para la función ilustrada en la figura 6.2b, puesto que L es el límite de la función cuando v se aproxima al valor N en el dominio, y puesto que L es también el valor de la función en N. Este último ejemplo gráfico debe ser suficiente para establecer que la continuidad de una función en el punto N *no* significa necesariamente que la gráfica de la función sea suave en $v = N$, porque el punto (N, L) en la figura 6.2b es en realidad un punto "picudo" y, sin embargo, la función es continua en ese valor de v.

Cuando una función $q = g(v)$ es continua en todos los valores de v en el intervalo (a, b), se dice que es continua en ese intervalo. Si la función es continua en todos los puntos de un subconjunto S del dominio (donde el subconjunto S puede ser la unión de varios intervalos disjuntos), se dice que es continuo en S. Y, por último, si la función es continua en todos los puntos de su dominio, se dice que es continua en su dominio. Sin embargo, incluso en este último caso, la gráfica de la función podría mostrar una discontinuidad (una abertura) en algún valor de v, por ejemplo, $v = 5$, si el valor de v *no* está en su dominio.

De nuevo, en relación con la figura 6.2, vemos que en el diagrama c la función es *discontinua* en N porque el límite no existe en ese punto, lo cual viola el segundo requerimiento de continuidad. No obstante, la función satisface los requerimientos de continuidad en el intervalo $(0, N)$ del dominio, así como en el intervalo $[N, \infty)$. Resulta claro que el diagrama d también es discontinuo en $v = N$. Esta vez la discontinuidad surge del hecho de que N se excluye del dominio, lo cual viola el primer requerimiento de continuidad.

Con base en las gráficas de la figura 6.2, se ve que los puntos en donde la gráfica presente un "pico" son consistentes con la continuidad, como en el diagrama b, pero que los huecos o saltos en la gráfica son discontinuidades, como en los diagramas c y d. Éste es en realidad el caso. En términos generales, una función que es continua en un intervalo particular es aquella cuya gráfica se puede trazar para dicho intervalo sin levantar el lápiz del papel, una proeza que es posible incluso si hay picos en la gráfica, pero imposible cuando hay huecos o saltos.

Funciones polinomiales y racionales

Consideremos ahora la continuidad de ciertas funciones halladas con frecuencia. Para cualquier función polinomial, como $q = g(v)$ en (6.11), se encontró en (6.12) que $\lim_{v \to N} q$ existe y es igual al valor de la función en N. Como N es un punto (cualquier punto) en el dominio de la función, se concluye que cualquier función polinomial es continua en su dominio. Ésta es una información muy útil, porque es común hallar funciones polinomiales.

¿Qué se puede decir acerca de las funciones racionales? Respecto a la continuidad, existe un teorema interesante (el teorema de continuidad) que afirma que la suma, diferencia, producto y cociente de algún número finito de funciones que son continuas en el dominio también son continuas, respectivamente, en el dominio. Como resultado, cualquier función racional (un cociente de dos funciones polinomiales) también debe ser continua en su dominio.

Ejemplo 1

La función racional

$$q = g(v) = \frac{4v^2}{v^2 + 1}$$

se define para todos los números reales finitos; así que su dominio consiste en el intervalo $(-\infty, \infty)$. Para cualquier número N en el dominio, el límite de q es (por el teorema de límite de cociente)

$$\lim_{v \to N} q = \frac{\lim_{v \to N}(4v^2)}{\lim_{v \to N}(v^2 + 1)} = \frac{4N^2}{N^2 + 1}$$

que es igual a $g(N)$. Por lo tanto, los tres requerimientos de continuidad se satisfacen en N. Además, se observa que N puede representar cualquier punto en el dominio de esta función; en consecuencia, esta función es continua en su dominio.

Ejemplo 2 La función racional

$$q = \frac{v^3 + v^2 - 4v - 4}{v^2 - 4}$$

no está definida en $v = 2$ y en $v = -2$. Puesto que esos dos valores de v no están en el dominio, la función es discontinua en $v = -2$ y $v = 2$, a pesar del hecho de que existe un límite de q cuando $v \to -2$ o 2. En una gráfica, esta función muestra un hueco en cada uno de estos dos valores de v. Pero para otros valores de v (los que *están* en el dominio), esta función es continua.

Diferenciabilidad de una función

Con el análisis previo obtuvimos las herramientas para determinar si alguna función tiene un límite cuando su variable independiente se aproxima a un valor específico. De esta manera, podemos intentar tomar el límite de alguna función $y = f(x)$ cuando x tiende a algún valor elegido, por ejemplo x_0. Sin embargo, también se puede aplicar el concepto de "límite" a un nivel diferente y tomar el límite del cociente de diferencias de esa función, $\Delta y/\Delta x$, cuando Δx tiende a cero. Los resultados de tomar el límite en estos dos niveles distintos se relacionan con dos propiedades distintas, aunque vinculadas, de la función f.

Tomando el límite de la función $y = f(x)$, podemos, de acuerdo con la explicación de la subsección anterior, se puede examinar si la función f es *continua* en $x = x_0$. Las condiciones para continuidad son (1) $x = x_0$ debe estar en el dominio de la función f, (2) y debe tener un límite cuando $x \to x_0$ y (3) dicho límite debe ser igual a $f(x_0)$. Cuando éstas se satisfacen, se puede escribir

$$\lim_{x \to x_0} f(x) = f(x_0) \qquad \text{[condición de continuidad]} \qquad \textbf{(6.13)}$$

En contraste, cuando se aplica el concepto de "límite" al cociente de diferencias $\Delta y/\Delta x$ cuando $\Delta x \to 0$, tratamos en cambio con la pregunta de si la función f es *diferenciable* en $x = x_0$, es decir, si la derivada dy/dx existe en $x = x_0$, o si $f'(x_0)$ existe. El término *diferenciable* se usa aquí porque el proceso de obtener la derivada dy/dx se conoce como *diferenciación* (también llamada *derivación*). Puesto que $f'(x_0)$ existe si y sólo si el límite de $\Delta y/\Delta x$ existe en $x = x_0$ cuando $\Delta x \to 0$, la expresión simbólica de la diferenciabilidad de f es

$$\begin{aligned} f'(x_0) &= \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} \\ &\equiv \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} \qquad \text{[condición de diferenciabilidad]} \qquad \textbf{(6.14)} \end{aligned}$$

Estas dos propiedades, continuidad y diferenciabilidad, están muy relacionadas entre sí —la continuidad de f es una condición *necesaria* para su diferenciabilidad (aunque, como se verá después, esta condición *no es suficiente*)—. Lo que esto significa es que, para ser diferenciable en $x = x_0$, la función debe pasar primero la prueba de ser continua en $x = x_0$. Para probar esto, debemos demostrar que, dada una función $y = f(x)$, su continuidad en $x = x_0$ se deduce de su diferenciabilidad en $x = x_0$; es decir, la condición (6.13) se deduce de la condición (6.14). Sin embargo, antes de hacer esto, simplificaremos un poco la notación (1) reemplazando x_0 por el símbolo N y (2) sustituyendo $(x_0 + \Delta x)$ con el símbolo x. Esto último es justificable porque el nuevo valor de x puede ser cualquier número (dependiendo de la magnitud del cambio) y, por lo tanto, es una variable que se puede denotar por x. La equivalencia de los dos sistemas de notación se muestra en la figura 6.4, donde las notaciones anteriores aparecen (entre corchetes) junto a la nueva. Observe que, con el cambio de notación, Δx ahora se convierte en $(x - N)$, de modo que la expresión "$\Delta x \to 0$" se transforma en "$x \to N$", que es análoga a la expresión $v \to N$ usada antes en relación con la función $q = g(v)$. En consecuencia, (6.13) y (6.14) ahora se pueden reescribir, respectivamente, como

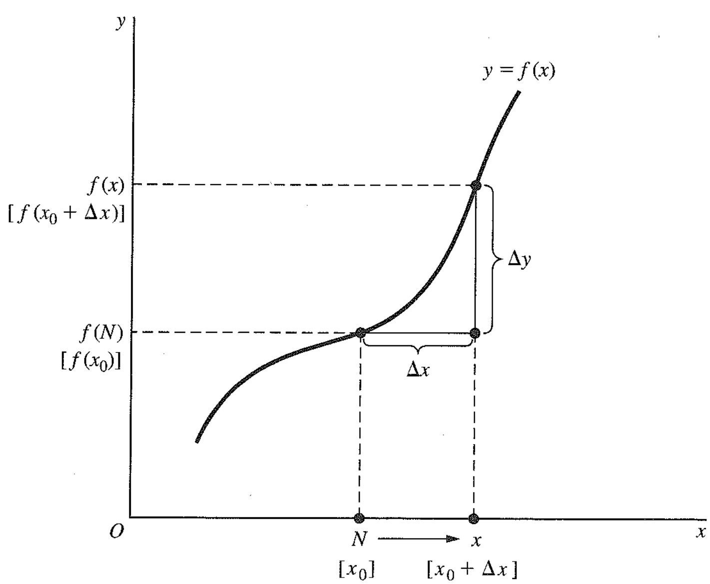
$$\lim_{x \to N} f(x) = f(N) \tag{6.13'}$$

$$f'(N) = \lim_{x \to N} \frac{f(x) - f(N)}{x - N} \tag{6.14'}$$

Lo que queremos demostrar es, por lo tanto, que la condición de continuidad (6.13′) se deduce de la condición de diferenciabilidad (6.14′). Primero, puesto que la notación $x \to N$ implica que $x \neq N$, de modo que $x - N$ es un número no cero, es aceptable escribir la siguiente identidad:

$$f(x) - f(N) \equiv \frac{f(x) - f(N)}{x - N}(x - N) \tag{6.15}$$

FIGURA 6.4



Tomando el límite en cada lado de (6.15) cuando $x \to N$ se obtienen los resultados siguientes:

$$\text{Lado izquierdo} = \lim_{x \to N} f(x) - \lim_{x \to N} f(N) \qquad \text{[teorema del límite de diferencias]}$$
$$= \lim_{x \to N} f(x) - f(N) \qquad [f(N) \text{ es una constante}]$$
$$\text{Lado derecho} = \lim_{x \to N} \frac{f(x) - f(N)}{x - N} \lim_{x \to N} (x - N) \qquad \text{[teorema del límite de producto]}$$
$$= f'(N)(\lim_{x \to N} x - \lim_{x \to N} N) \qquad \text{[por (6.14') y el teorema del límite de diferencias]}$$
$$= f'(N)(N - N) = 0$$

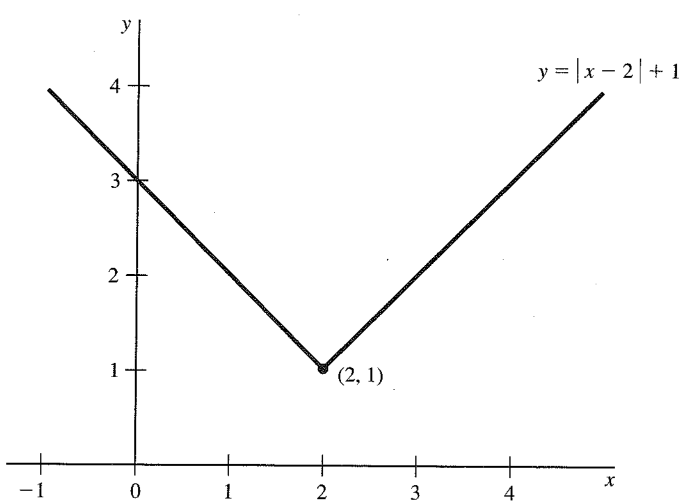
Note que no habría sido posible escribir estos resultados si no se hubiera concedido la condición (6.14′), porque si $f'(N)$ no existiera, entonces la expresión del lado derecho (y, por lo tanto, también la expresión del lado izquierdo) en (6.15) no poseería un límite. Sin embargo, si $f'(N)$ existe, los dos lados tendrán límites como se muestra en las ecuaciones anteriores. Además, cuando se igualan el resultado del lado izquierdo y el del lado derecho, obtenemos $\lim_{x \to N} f(x) - f(N) = 0$, que es idéntica a (6.13′). Así, hemos probado que la continuidad, como se ilustra en (6.13′), se deduce de la diferenciabilidad, según se muestra en (6.14′). En general, si una función es diferenciable en todo punto de su dominio, concluimos que debe ser continua en su dominio.

Aunque la diferenciabilidad implica continuidad, lo contrario no es cierto. Es decir, la continuidad es una condición *necesaria*, pero *no suficiente* para la diferenciabilidad. Para demostrar esto, solamente tenemos que hacer un contraejemplo. Considere la función

$$y = f(x) = |x - 2| + 1 \qquad \textbf{(6.16)}$$

que se grafica en la figura 6.5. Como se puede demostrar fácilmente, esta función no es diferenciable, aunque continua, cuando $x = 2$. Es fácil establecer que la función sea continua en $x = 2$. Primero, $x = 2$ está en el dominio de la función. Segundo, el límite de y existe cuando x tiende a 2; para ser específicos, $\lim_{x \to 2^+} y = \lim_{x \to 2^-} y = 1$. Tercero, se encuentra también que

FIGURA 6.5



$f(2)$ es 1. Así, se satisfacen los tres requerimientos de continuidad. Para mostrar que la función f *no* es diferenciable en $x = 2$, se debe mostrar que el límite del cociente de diferencias

$$\lim_{x\to 2} \frac{f(x) - f(2)}{x - 2} = \lim_{x\to 2} \frac{|x-2| + 1 - 1}{x - 2} = \lim_{x\to 2} \frac{|x-2|}{x-2}$$

no existe. Esto implica la demostración de una disparidad entre los límites izquierdo y derecho. Puesto que, al considerar el límite derecho, x debe ser mayor que 2, según la definición de valor absoluto en (6.8) tenemos $|x - 2| = x - 2$. Así, el límite derecho es

$$\lim_{x\to 2^+} \frac{|x-2|}{x-2} = \lim_{x\to 2^+} \frac{x-2}{x-2} = \lim_{x\to 2^+} 1 = 1$$

Por otro lado, al considerar el límite izquierdo, x debe ser menor que 2; por lo tanto, de acuerdo con (6.8), $|x - 2| = -(x - 2)$. En consecuencia, el límite izquierdo es

$$\lim_{x\to 2^-} = \frac{|x-2|}{x-2} = \lim_{x\to 2^-} \frac{-(x-2)}{x-2} = \lim_{x\to 2^-} (-1) = -1$$

que es diferente del límite derecho. Esto demuestra que la continuidad no garantiza diferenciabilidad. En suma, todas las funciones diferenciables son continuas, pero no todas las funciones continuas son diferenciables.

En la figura 6.5, la no diferenciabilidad de la función en $x = 2$ se manifiesta en el hecho de que el punto (2, 1) no tiene recta tangente definida y, por lo tanto, no se puede asignar ninguna pendiente definida a ese punto. En particular, a la izquierda de ese punto, la curva tiene una pendiente de -1, pero a la derecha tiene una pendiente de $+1$, y las pendientes en los dos lados no muestran tendencia a aproximarse a una magnitud común en $x = 2$. El punto (2, 1) es, por supuesto, un punto especial; es el único punto definido en la curva. En otros puntos sobre la curva, la derivada está definida y la función es diferenciable. De modo más específico, la función en (6.16) se puede dividir en dos funciones lineales como sigue:

Parte izquierda: $\quad y = -(x-2) + 1 = 3 - x \quad (x \leq 2)$

Parte derecha: $\quad y = (x-2) + 1 = x - 1 \quad (x > 2)$

La parte izquierda es diferenciable en el intervalo $(-\infty, 2)$, y la parte derecha es diferenciable en el intervalo $(2, \infty)$ en el dominio.

En general, la diferenciabilidad es una condición más restrictiva que la continuidad, porque requiere algo además de continuidad. La continuidad en un punto sólo descarta la presencia de un hueco o salto, mientras que la diferenciabilidad también elimina los "picos". Por lo tanto, la diferenciabilidad requiere "suavidad " de la gráfica de la función (curva), además de la continuidad. La mayor parte de las funciones *específicas* empleadas en economía tienen la propiedad de que son diferenciables en todas partes. Cuando se usan funciones *generales*, se supone que son diferenciables en cualquier punto, como se verá en la explicación posterior.

EJERCICIO 6.7

1. Una función $y = f(x)$ es discontinua en $x = x_0$ cuando se viola *cualquiera* de los tres requerimientos para continuidad en $x = x_0$. Construya tres gráficas para ilustrar la violación de cada uno de esos requerimientos.

2. Tomando el conjunto de los números reales finitos como el dominio de la función $q = g(v) = v^2 - 5v - 2$:
   (*a*) Encuentre el límite de q cuando v tiende a N (un número real finito).
   (*b*) Compruebe si este límite es igual a $g(N)$.
   (*c*) Compruebe si la función es continua en N y continua en su dominio.
3. Dada la función $q = g(v) = \dfrac{v+2}{v^2+2}$:
   (*a*) Use los teoremas de límite para hallar $\lim\limits_{v \to N} q$, siendo N un número real finito.
   (*b*) Compruebe si este límite es igual a $g(N)$.
   (*c*) Verifique la continuidad de la función $g(v)$ en N y en su dominio $(-\infty, \infty)$.
4. Dada $y = f(x) = \dfrac{x^2 - 9x + 20}{x - 4}$:
   (*a*) ¿Es posible aplicar el teorema del límite del cociente para hallar el límite de esta función cuando $x \to 4$?
   (*b*) ¿Esta función es continua en $x = 4$? ¿Por qué?
   (*c*) Obtenga una función que, para $x \neq 4$, es equivalente a la función dada, y obtenga de la función equivalente el límite de y cuando $x \to 4$.
5. En la función racional del ejemplo 2, el numerador es divisible por el denominador de modo uniforme, y el cociente es $v + 1$. Con base en lo anterior, ¿se puede reemplazar en el acto la función por $q = v + 1$? ¿Por qué sí o por qué no?
6. Con base en las gráficas de las seis funciones de la figura 2.8, ¿concluiría que cada una de esas funciones es diferenciable en todo punto de su dominio? Explique.

Capítulo 7

Reglas de diferenciación y su uso en estática comparativa

El problema central del análisis estático comparativo: encontrar una tasa de cambio, se puede identificar con el problema de hallar la derivada de alguna función $y = f(x)$, siempre que sólo se considere un cambio infinitesimal en x. Aun cuando la derivada dy/dx se define como el límite del cociente de diferencias $q = g(v)$ cuando $v \to 0$, no es necesario emprender el proceso de tomar el límite cada vez que se busque la derivada de una función, porque existen varias reglas de diferenciación (derivación) que nos permitirán obtener de modo directo las derivadas deseadas. En lugar de ir de inmediato a los modelos de estática comparativa, comenzaremos por aprender algunas reglas de diferenciación.

7.1 Reglas de diferenciación para una función de una variable

Primero, analizaremos tres reglas que se aplican, respectivamente, a los siguientes tipos de función de una sola variable independiente: $y = k$ (función constante) y $y = x^n$ y $y = cx^n$ (funciones de potencia); todas tienen gráficas suaves y continuas y son, por lo tanto, diferenciables en todas partes.

Regla de función constante

La derivada de una función constante $y = k$, o $f(x) = k$ es idénticamente cero, por ejemplo, es cero para todos lo valores de x. En símbolos, esta regla se puede expresar como: dada $y = f(x) = k$, la derivada es

$$\frac{dy}{dx} = \frac{dk}{dx} = 0 \quad \text{o bien,} \quad f'(x) = 0$$

Por otro lado, se puede expresar la regla como: dada $y = f(x) = k$, la derivada es

$$\frac{d}{dx}y = \frac{d}{dx}f(x) = \frac{d}{dx}k = 0$$

donde el símbolo de derivada se ha separado en dos partes, d/dx por un lado, y y [o $f(x)$ o k] por el otro. La primera parte, d/dx, es un *símbolo operador*, y señala que se debe efectuar una operación matemática particular. Al igual que el símbolo operador $\sqrt{}$ indica que se debe sacar la raíz cuadrada, el símbolo d/dx representa una instrucción para tomar la derivada de, o diferenciar (alguna función) respecto a la variable x. La función por operar (o diferenciar) se indica en la segunda parte; en este caso es $y = f(x) = k$.

La prueba de la regla es la siguiente: dada $f(x) = k$, se tiene $f(N) = k$ para algún valor de N. Así, el valor de $f'(N)$, el valor de la derivada en $x = N$, como se define en (6.13) es

$$f'(N) = \lim_{x \to N} \frac{f(x) - f(N)}{x - N} = \lim_{x \to N} \frac{k - k}{x - N} = \lim_{x \to N} 0 = 0$$

Además, puesto que N representa cualquier valor de x, el resultado $f'(N) = 0$ se puede generalizar de inmediato a $f'(x) = 0$. Esto prueba la regla.

Es importante distinguir con claridad entre la afirmación $f'(x) = 0$ y la proposición que parece similar, pero que es diferente, $f'(x_0) = 0$. Por $f'(x) = 0$, entendemos que la función derivada f' tiene un valor cero para *todos* los valores de x; por otro lado, al escribir $f'(x_0) = 0$ solamente asociamos el valor cero a la derivada con un valor particular de x, a saber, $x = x_0$.

Como se explicó antes, la derivada de una función tiene su contraparte geométrica en la pendiente de la curva. La gráfica de una función constante, por ejemplo, una función de costo fijo $C_F = f(Q) = \$1\,200$, es una recta horizontal con una pendiente cero. Correspondientemente, la derivada también debe ser cero para todos los valores de Q:

$$\frac{d}{dQ}C_F = \frac{d}{dQ}1\,200 = 0$$

Regla de la función potencia

La derivada de una función potencia $y = f(x) = x^n$ es nx^{n-1}. En símbolos, esto se expresa como

$$\frac{d}{dx}x^n = nx^{n-1} \quad \text{o bien,} \quad f'(x) = nx^{n-1} \tag{7.1}$$

Ejemplo 1 La derivada de $y = x^3$ es $\frac{dy}{dx} = \frac{d}{dx}x^3 = 3x^2$.

Ejemplo 2 La derivada de $y = x^9$ es $\frac{d}{dx}x^9 = 9x^8$.

Esta regla es válida para cualquier potencia real de x, es decir, el exponente puede ser cualquier número real; sin embargo, se probará sólo para el caso donde n es algún entero positivo. En el caso más simple, el de $n = 1$, la función es $f(x) = x$ y, de acuerdo con la regla, la derivada es

$$f'(x) = \frac{d}{dx}x = 1(x^0) = 1$$

La prueba de este resultado se deduce fácilmente a partir de la definición de $f'(N)$ especificada en (6.14′). Dada $f(x) = x$, el valor de la derivada para todo valor de x, por ejemplo, $x = N$, es

$$f'(N) = \lim_{x \to N} \frac{f(x) - f(N)}{x - N} = \lim_{x \to N} \frac{x - N}{x - N} = \lim_{x \to N} 1 = 1$$

Puesto que N representa cualquier valor de x, se puede escribir $f'(x) = 1$. Esto prueba la regla para el caso de $n = 1$. Como la contraparte gráfica de este resultado, se ve que la función $y = f(x) = x$ se grafica como una recta de 45°, y tiene una pendiente de $+1$.

Para los casos de enteros más grandes, $n = 2, 3, \ldots,$ se deben notar primero las siguientes identidades:

$$\frac{x^2 - N^2}{x - N} = x + N \qquad \text{[en el miembro derecho hay 2 términos]}$$

$$\frac{x^3 - N^3}{x - N} = x^2 + Nx + N^2 \qquad \text{[en el miembro derecho hay 3 términos]}$$

$$\vdots$$

$$\frac{x^n - N^n}{x - N} = x^{n-1} + Nx^{n-2} + N^2x^{n-3} + \cdots + N^{n-1}$$

$$\text{[en el miembro derecho hay } n \text{ términos]} \qquad \textbf{(7.2)}$$

Con base en (7.2), se puede expresar la derivada de una función de potencia $f(x) = x^n$ en $x = N$ como sigue:

$$\begin{aligned} f'(N) &= \lim_{x \to N} \frac{f(x) - f(N)}{x - N} = \lim_{x \to N} \frac{x^n - N^n}{x - N} \\ &= \lim_{x \to N} (x^{n-1} + Nx^{n-2} + \cdots + N^{n-1}) && \text{[por (7.2)]} \\ &= \lim_{x \to N} x^{n-1} + \lim_{x \to N} Nx^{n-2} + \cdots + \lim_{x \to N} N^{n-1} && \text{[teorema del límite de suma]} \\ &= N^{n-1} + N^{n-1} + \cdots + N^{n-1} && \text{[un total de } n \text{ términos]} \\ &= nN^{n-1} && \textbf{(7.3)} \end{aligned}$$

De nuevo, N es cualquier valor de x; así que este último resultado se puede generalizar a

$$f'(x) = nx^{n-1}$$

lo cual prueba la regla para cualquier entero positivo n.

Como se mencionó, esta regla se aplica aun cuando el exponente n en la expresión x^n no sea un entero positivo. Los siguientes ejemplos sirven para ilustrar su aplicación a los últimos casos.

Ejemplo 3 Encuentre la derivada de $y = x^0$. Al aplicar (7.1) se encuentra

$$\frac{d}{dx} x^0 = 0(x^{-1}) = 0$$

Ejemplo 4 Halle la derivada de $y = 1/x^3$. Esto tiene que ver con el recíproco de una potencia, pero al reescribir la función como $y = x^{-3}$, podemos utilizar de nuevo (7.1) para obtener la derivada:

$$\frac{d}{dx} x^{-3} = -3x^{-4} \qquad \left[= \frac{-3}{x^4}\right]$$

Ejemplo 5 Encuentre la derivada de $y = \sqrt{x}$. En este caso se tiene una raíz cuadrada, pero como $\sqrt{x} = x^{1/2}$, la derivada se puede hallar como sigue:

$$\frac{d}{dx}x^{1/2} = \frac{1}{2}x^{-1/2} \qquad \left[= \frac{1}{2\sqrt{x}} = \frac{\sqrt{x}}{2x}\right]$$

Las derivadas son por sí mismas funciones de la variable independiente x. En el ejemplo 1, la derivada es $dy/dx = 3x^2$, o bien, $f'(x) = 3x^2$, de modo que un valor diferente de x dará como resultado un valor distinto de la derivada, como

$$f'(1) = 3(1)^2 = 3 \qquad f'(2) = 3(2)^2 = 12$$

Estos valores específicos de la derivada se pueden expresar de otra forma, como

$$\left.\frac{dy}{dx}\right|_{x=1} = 3 \qquad \left.\frac{dy}{dx}\right|_{x=2} = 12$$

pero son preferibles las notaciones $f'(1)$ y $f'(2)$ debido a su simplicidad.

Es muy importante entender que, para hallar los valores de la derivada $f'(1)$, $f'(2)$, etcétera, se debe diferenciar *primero* la función $f(x)$, obtener la función derivada $f'(x)$ y *luego* dejar que x asuma valores específicos en $f'(x)$. En definitiva, no se permiten sustituir valores específicos de x en la función primitiva $f(x)$ antes de la diferenciación. Como ilustración, si se sustituye $x = 1$ en la función del ejemplo 1 antes de la diferenciación, la función degenerará en $y = x^3 = (1)^3 = 1$, una función constante, que producirá una derivada cero en vez de la respuesta correcta de $f'(x) = 3x^2$.

Regla generalizada de la función potencia

Cuando aparece una constante multiplicativa c en la función potencia, de manera que $f(x) = cx^n$, su derivada es

$$\frac{d}{dx}cx^n = cnx^{n-1} \qquad \text{o} \qquad f'(x) = cnx^{n-1}$$

Este resultado muestra que, al diferenciar cx^n, podemos mantener intacta la constante multiplicativa c y luego diferenciar el término x^n de acuerdo con (7.1).

Ejemplo 6 Dada $y = 2x$, tenemos $dy/dx = 2x^0 = 2$.

Ejemplo 7 Dada $f(x) = 4x^3$, la derivada es $f'(x) = 12x^2$.

Ejemplo 8 La derivada de $f(x) = 3x^{-2}$ es $f'(x) = -6x^{-3}$.

Para probar esta nueva regla, considere el hecho de que para cualquier valor de x, por ejemplo $x = N$, el valor de la derivada de $f(x) = cx^n$ es

$$\begin{aligned} f'(N) &= \lim_{x\to N}\frac{f(x)-f(N)}{x-N} = \lim_{x\to N}\frac{cx^n - cN^n}{x-N} = \lim_{x\to N} c\left(\frac{x^n - N^n}{x-N}\right) \\ &= \lim_{x\to N} c \lim_{x\to N}\frac{x^n - N^n}{x-N} && \text{[teorema del límite del producto]} \\ &= c\lim_{x\to N}\frac{x^n - N^n}{x-N} && \text{[teorema del límite de una constante]} \\ &= cnN^{n-1} && \text{[de (7.3)]} \end{aligned}$$

En vista de que N es cualquier valor de x, este último resultado se puede generalizar de inmediato a $f'(x) = cnx^{n-1}$, lo cual prueba la regla.

EJERCICIO 7.1

1. Obtenga la derivada de cada una de las siguientes funciones:
   (*a*) $y = x^{12}$ (*c*) $y = 7x^5$ (*e*) $w = -4u^{1/2}$
   (*b*) $y = 63$ (*d*) $w = 3u^{-1}$ (*f*) $w = 4u^{1/4}$
2. Encuentre lo siguiente:
   (*a*) $\frac{d}{dx}(-x^{-4})$ (*c*) $\frac{d}{dw}5w^4$ (*e*) $\frac{d}{du}au^b$
   (*b*) $\frac{d}{dx}9x^{1/3}$ (*d*) $\frac{d}{dx}cx^2$ (*f*) $\frac{d}{du} - au^{-b}$
3. Halle $f'(1)$ y $f'(2)$ de las siguientes funciones:
   (*a*) $y = f(x) = 18x$ (*c*) $f(x) = -5x^{-2}$ (*e*) $f(w) = 6w^{1/3}$
   (*b*) $y = f(x) = cx^3$ (*d*) $f(x) = \frac{3}{4}x^{4/3}$ (*f*) $f(w) = -3w^{-1/6}$
4. Grafique una función $f(x)$ que dé lugar a la función derivada $f'(x) = 0$. Luego, grafique una función $g(x)$ caracterizada por $g'(x_0) = 0$.

7.2 Reglas de diferenciación con dos o más funciones de la misma variable

Las tres reglas presentadas en la sección 7.1 tienen que ver con una sola función dada $f(x)$. Ahora suponga que se tienen dos funciones *diferenciables* de la misma variable x, por ejemplo, $f(x)$ y $g(x)$, y que se quiere diferenciar la suma, diferencia, producto o cociente formados con estas dos funciones. En esas circunstancias, ¿hay reglas apropiadas? Concretamente, dadas dos funciones, por ejemplo, $f(x) = 3x^2$ y $g(x) = 9x^{12}$, ¿cómo obtenemos la derivada de, por ejemplo, $3x^2 + 9x^{12}$, o la derivada de $(3x^2)(9x^{12})$?

Regla de la suma o de la diferencia

La derivada de una suma (diferencia) de dos funciones es la suma (diferencia) de las derivadas de las dos funciones:

$$\frac{d}{dx}[f(x) \pm g(x)] = \frac{d}{dx}f(x) \pm \frac{d}{dx}g(x) = f'(x) \pm g'(x)$$

La prueba de esto de nuevo tiene que ver con la aplicación de la definición de una derivada y de los distintos teoremas de límites. Omitiremos la prueba y, en cambio, solamente comprobaremos su validez e ilustraremos su aplicación.

Ejemplo 1

De la función $y = 14x^3$ se puede obtener la derivada $dy/dx = 42x^2$. Pero $14x^3 = 5x^3 + 9x^3$, de modo que y se puede considerar como la suma de dos funciones $f(x) = 5x^3$ y $g(x) = 9x^3$. De acuerdo con la regla de la suma, se tiene entonces

$$\frac{dy}{dx} = \frac{d}{dx}(5x^3 + 9x^3) = \frac{d}{dx}5x^3 + \frac{d}{dx}9x^3 = 15x^2 + 27x^2 = 42x^2$$

que es idéntica a nuestro resultado anterior.

Esta regla, que se expresa en términos de dos funciones, se puede ampliar fácilmente a más funciones. Así, también es válido escribir

$$\frac{d}{dx}[f(x) \pm g(x) \pm h(x)] = f'(x) \pm g'(x) \pm h'(x)$$

Ejemplo 2

La función citada en el ejemplo 1, $y = 14x^3$, se puede escribir como $y = 2x^3 + 13x^3 - x^3$. La derivada de esta última, de acuerdo con la regla de la suma o de la diferencia, es

$$\frac{dy}{dx} = \frac{d}{dx}(2x^3 + 13x^3 - x^3) = 6x^2 + 39x^2 - 3x^2 = 42x^2$$

que de nuevo se comprueba con la respuesta previa.

Esta regla es de gran importancia práctica. Ahora que se cuenta con esto, se puede hallar la derivada de cualquier función polinomial, puesto que la última no es sino una suma de funciones potencia.

Ejemplo 3

$$\frac{d}{dx}(ax^2 + bx + c) = 2ax + b$$

Ejemplo 4

$$\frac{d}{dx}(7x^4 + 2x^3 - 3x + 37) = 28x^3 + 6x^2 - 3 + 0 = 28x^3 + 6x^2 - 3$$

Note que en los ejemplos 3 y 4, las constantes c y 37 en realidad no producen ningún efecto en la derivada, porque la derivada de un término constante es cero. En contraste con la constante *multiplicativa*, que se retiene durante la diferenciación, la constante *aditiva* se elimina. Esto proporciona la explicación matemática del principio económico bien conocido de que el costo fijo de una empresa no afecta su costo marginal. Dada una función de costo total de corto plazo

$$C = Q^3 - 4Q^2 + 10Q + 75$$

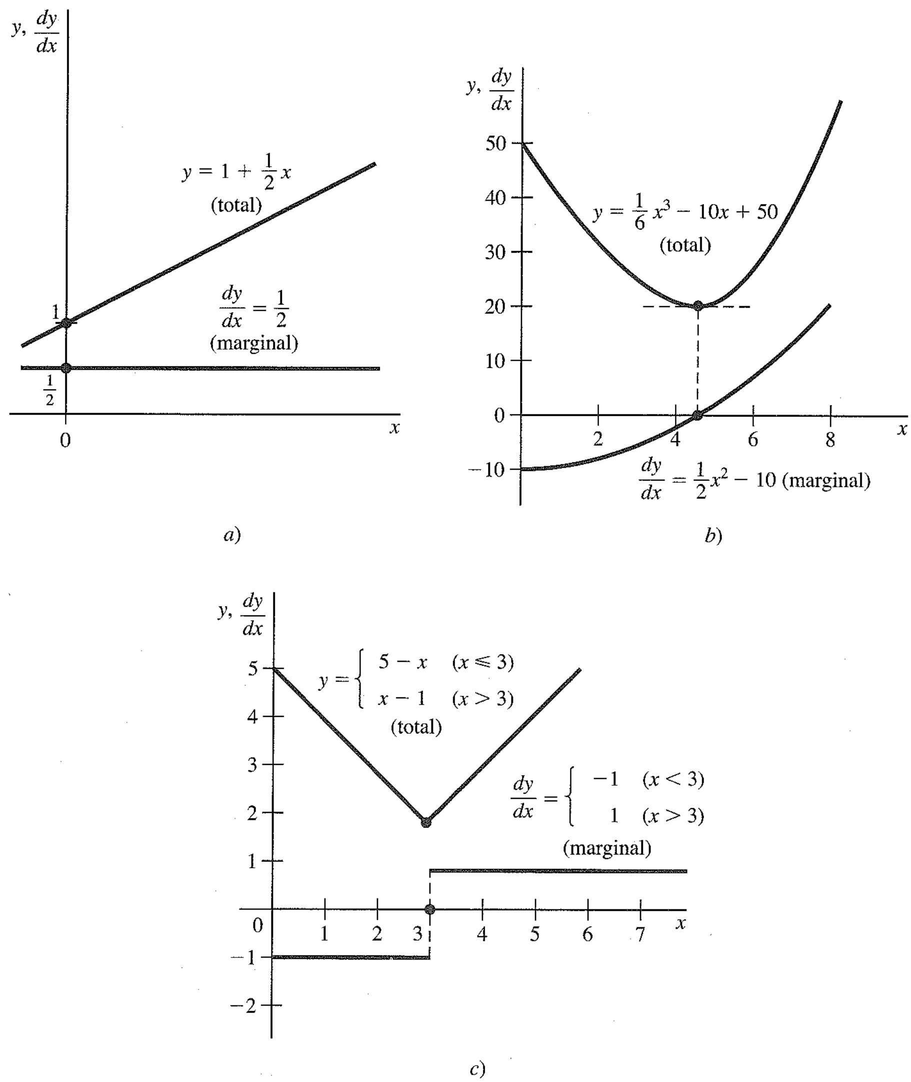
la función de costo marginal (para el cambio de producción infinitesimal) es el límite del cociente $\Delta C/\Delta Q$, o la derivada de la función C:

$$\frac{dC}{dQ} = 3Q^2 - 8Q + 10$$

mientras que el costo fijo se representa por la constante aditiva 75. Puesto que esta última se elimina durante el proceso de obtener dC/dQ, resulta obvio que la magnitud del costo fijo no puede afectar el costo marginal.

En general, si una función primitiva $y = f(x)$ representa un *total*, entonces la función derivada dy/dx es su función *marginal*. Por supuesto, ambas funciones se pueden representar en forma gráfica contra la variable x; y como resultado de la correspondencia entre la derivada de una función y la pendiente de su curva, para cada valor de x la función marginal debe mostrar la pendiente de la función total en ese valor de x. En la figura 7.1a se ve que una función total lineal (de pendiente constante) tiene una función marginal constante. Por otro lado, la función total no lineal (de pendiente variable) de la figura 7.1b da lugar a una función marginal curva, la cual yace abajo (arriba) del eje horizontal cuando la función total adquiere una pendiente negativa (positiva). Por último, el lector puede ver en la figura 7.1c (véase figura 6.5) que la

FIGURA 7.1



"no suavidad" de una función total dará como resultado un hueco o salto (discontinuidad) en la función marginal o derivada. Esto contrasta con la función total suave en todas partes de la figura 7.1*b* que da lugar a una función marginal continua. Por esta razón, la *suavidad* de una función *primitiva* se puede relacionar con la *continuidad* de su función *derivada*. En particular, en lugar de decir que cierta función es suave (y diferenciable) en todas partes, la podemos caracterizar como una función con una función derivada continua, y nos referimos a ella como una función *continuamente diferenciable.*

Las notaciones siguientes se usan con frecuencia para expresar la continuidad y lo continuamente diferenciable de una función f:

$f \in C^{(0)}$ o $f \in C$: f es continua

$f \in C^{(1)}$ o $f \in C'$: f es diferenciable y su derivada es continua donde $C^{(0)}$, o simplemente C, es el símbolo para el conjunto de todas las funciones continuas, y $C^{(1)}$ o C', es el símbolo para el conjunto de todas las funciones diferenciables con derivada continua.

Regla del producto

La derivada del producto de dos funciones (diferenciables) es igual a la primera función por la derivada de la segunda función más la segunda función por la derivada de la primera función:

$$\begin{aligned}\frac{d}{dx}[f(x)g(x)] &= f(x)\frac{d}{dx}g(x) + g(x)\frac{d}{dx}f(x) \\ &= f(x)g'(x) + g(x)f'(x) \qquad \textbf{(7.4)}\end{aligned}$$

Por supuesto, también se pueden reacomodar los términos y expresar la regla como

$$\frac{d}{dx}[f(x)g(x)] = f'(x)g(x) + f(x)g'(x) \qquad \textbf{(7.4')}$$

Ejemplo 5

Encuentre la derivada de $y = (2x+3)(3x^2)$. Sean $f(x) = 2x+3$ y $g(x) = 3x^2$. Entonces, se deduce que $f'(x) = 2$ y $g'(x) = 6x$ y, de acuerdo con (7.4), la derivada deseada es

$$\frac{d}{dx}[(2x+3)(3x^2)] = (2x+3)(6x) + (3x^2)(2) = 18x^2 + 18x$$

Este resultado se comprueba al multiplicar $f(x)g(x)$ y después tomar la derivada del producto de los polinomios. El producto en este caso es $f(x)g(x) = (2x+3)(3x^2) = 6x^3 + 9x^2$, y la diferenciación directa produce la misma derivada, $18x^2 + 18x$.

El punto importante que se debe recordar es que la derivada de un producto de dos funciones *no* es el simple producto de las dos derivadas separadas. En cambio, es una suma ponderada de $f'(x)$ y $g'(x)$, donde las ponderaciones son $g(x)$ y $f(x)$, respectivamente. Puesto que esto difiere de lo que se esperaría de la generalización intuitiva, se procede a producir una prueba para (7.4). De acuerdo con (6.13), el valor de la derivada de $f(x)g(x)$ cuando $x = N$ debe ser

$$\left.\frac{d}{dx}[f(x)g(x)]\right|_{x=N} = \lim_{x\to N} \frac{f(x)g(x) - f(N)g(N)}{x - N} \qquad \textbf{(7.5)}$$

Pero al sumar *y* restar $f(x)g(N)$ en el numerador (y, por lo tanto, dejar sin cambio la magnitud original), se puede transformar el cociente de la derecha de (7.5) como sigue:

$$\begin{aligned}&\frac{f(x)g(x) - f(x)g(N) + f(x)g(N) - f(N)g(N)}{x - N} \\ &= f(x)\frac{g(x) - g(N)}{x - N} + g(N)\frac{f(x) - f(N)}{x - N}\end{aligned}$$

Al sustituir esto por el cociente de la derecha de (7.5) y tomar su límite, se obtiene

$$\begin{aligned}\left.\frac{d}{dx}[f(x)g(x)]\right|_{x=N} &= \lim_{x\to N} f(x) \lim_{x\to N} \frac{g(x) - g(N)}{x - N} \\ &\quad + \lim_{x\to N} g(N) \lim_{x\to N} \frac{f(x) - f(N)}{x - N} \qquad \textbf{(7.5')}\end{aligned}$$

Las cuatro expresiones de límite de (7.5′) se evalúan fácilmente. La primera es $f(N)$ y la tercera es $g(N)$ (límite de una constante). Las dos restantes son, de acuerdo con (6.13), respectivamente, $g'(N)$ y $f'(N)$. Así, (7.5′) se reduce a

$$\frac{d}{dx}[f(x)g(x)]\bigg|_{x=N} = f(N)g'(N) + g(N)f'(N) \qquad \textbf{(7.5″)}$$

Y, puesto que N representa cualquier valor de x, (7.5″) sigue siendo válida si el símbolo N se reemplaza por x. Esto prueba la regla.

Como una extensión de la regla al caso de *tres* funciones, se tiene

$$\frac{d}{dx}[f(x)g(x)h(x)] = f'(x)g(x)h(x) + f(x)g'(x)h(x) + f(x)g(x)h'(x) \qquad [\text{cf. } (7.4')] \qquad \textbf{(7.6)}$$

Expresado en palabras se diría: la derivada del producto de tres funciones es igual al producto de las funciones segunda y tercera por la derivada de la primera, más el producto de las funciones primera y tercera por la derivada de la segunda, más el producto de las funciones primera y segunda por la derivada de la tercera. Este resultado se puede obtener mediante la aplicación repetida de (7.4). Primero trate el producto $g(x)h(x)$ como una sola función, por ejemplo, $\phi(x)$, de modo que el producto original de las tres funciones se convierta en un producto de *dos* funciones, $f(x)\phi(x)$. Para esto, (7.4) es aplicable. Después que se obtiene la derivada de $f(x)\phi(x)$, se puede aplicar de nuevo (7.4) al producto $g(x)h(x) \equiv \phi(x)$ para obtener $\phi'(x)$. Entonces, se deduce (7.6). Le dejamos que realice los detalles como ejercicio.

La validez de una regla es una cosa; su utilidad es algo más. ¿Por qué necesitamos la regla del producto cuando podemos recurrir al procedimiento alternativo de multiplicar las dos funciones $f(x)$ y $g(x)$ y después tomar directamente la derivada del producto resultante? Una respuesta es que el procedimiento alternativo es aplicable sólo a funciones definidas mediante fórmulas *específicas* (numéricas o paramétricas), mientras que la regla del producto es aplicable incluso cuando las funciones se dan en forma *genérica*. A continuación se ilustra con un ejemplo económico.

Determinación de la función de ingreso marginal a partir de la función de ingreso promedio

Si se tiene una función de ingreso promedio (AR) en forma específica,

$$\text{AR} = 15 - Q$$

para obtener la función de ingreso marginal (MR), primero se determina la función de ingreso total (R) como el producto del ingreso promedio por Q:

$$R \equiv \text{AR} \cdot Q = (15 - Q)Q = 15Q - Q^2$$

y después se deriva R:

$$\text{MR} \equiv \frac{dR}{dQ} = 15 - 2Q$$

Pero si la función AR se da en forma genérica AR $= f(Q)$, entonces la función de ingreso total también estará en forma genérica:

$$R \equiv \text{AR} \cdot Q = f(Q) \cdot Q$$

por lo tanto, el método de "multiplicar antes de derivar" no será útil. Sin embargo, debido a que R es un producto de dos funciones de Q, a saber, $f(Q)$ y Q, se puede poner en práctica la regla del producto. Así, podemos diferenciar R para obtener la función MR como sigue:

$$\text{MR} \equiv \frac{dR}{dQ} = f(Q) \cdot 1 + Q \cdot f'(Q) = f(Q) + Qf'(Q) \qquad \textbf{(7.7)}$$

Sin embargo, ¿puede un resultado tan general decir algo importante acerca del MR? En realidad, sí. Recordando que $f(Q)$ denota al ingreso promedio AR, se reordena (7.7) y se escribe

$$\text{MR} - \text{AR} = \text{MR} - f(Q) = Qf'(Q) \qquad \textbf{(7.7')}$$

Esto nos da una relación importante entre el ingreso marginal (MR) y el ingreso promedio (AR): siempre diferirán por la cantidad $Qf'(Q)$.

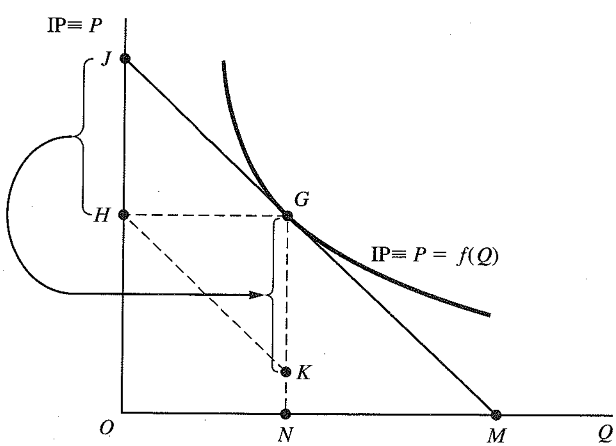
Aún falta examinar la expresión $Qf'(Q)$. Su primer componente Q denota producto y nunca es negativo. El otro componente, $f'(Q)$, representa la pendiente de la curva *AR* graficada contra Q. Puesto que "ingreso promedio" y "precio" no son sino nombres diferentes para la misma cosa:

$$\text{IP} \equiv \frac{\text{I}}{Q} \equiv \frac{PQ}{Q} \equiv P$$

la curva AR se puede considerar también como una curva que relaciona el precio P con la producción Q: $P = f(Q)$. Desde este punto de vista, la curva AR es simplemente el *inverso* de la curva de demanda para el producto de la empresa, es decir, la curva de demanda graficada después de invertir los ejes P y Q. En la competencia pura, la curva AR es una recta horizontal, de modo que $f'(Q) = 0$ y, de (7.7'), MR − AR = 0 para todos los valores posibles de Q. Así, la curva MR y la curva AR deben coincidir. Por otro lado, en la competencia imperfecta la curva AR, por lo común, tiene pendiente descendente, como en la figura 7.2, de tal manera que $f'(Q) < 0$ y, de (7.7'), MR − AR < 0 para todos los niveles positivos de producción. En este caso, la curva MR debe quedar debajo de la curva AR.

La conclusión antes mencionada es de naturaleza *cualitativa*; sólo tiene que ver con posiciones relativas de las dos curvas. Pero (7.7') también ofrece la información *cuantitativa* de que la curva MR queda por debajo de la curva AR por precisamente la cantidad $Qf'(Q)$ en cualquier nivel de producción Q. Examine de nuevo la figura 7.2 y considere el nivel de pro-

FIGURA 7.2



ducción particular N. Para esa producción, la expresión $Qf'(Q)$ se convierte en $Nf'(N)$; si se puede determinar la magnitud de $Nf'(N)$ en el diagrama, se sabrá cuán abajo del punto G de ingreso promedio queda el punto correspondiente al ingreso marginal.

La magnitud de N ya está especificada. Y $f'(N)$ es simplemente la pendiente de la curva AR en el punto G (donde $Q = N$), es decir, la pendiente de la recta tangente JM medida por la relación de dos distancias OJ/OM. Sin embargo, se ve que $OJ/OM = HJ/HG$; además, la distancia HG es la cantidad de producto en consideración, N. Así, la distancia $Nf'(N)$, por la cual la curva MR debe quedar debajo de la curva AR en la producción N, es

$$Nf'(N) = HG\frac{HJ}{HG} = HJ$$

En consecuencia, si se marca una distancia vertical $KG = HJ$ directamente debajo del punto G, entonces K debe ser un punto sobre la curva MR. (Una forma sencilla de graficar con precisión KG es dibujar una línea recta que pasa por el punto H y es paralela a JG; el punto K es donde esa recta interseca la recta vertical NG.)

El mismo procedimiento se puede usar para localizar otros puntos sobre la curva MR. Todo lo que se debe hacer, para algún punto G' elegido sobre la curva, es dibujar primero una tangente a la curva AR en G' que se encontrará con el eje vertical en algún punto J'. Luego, dibuje una recta horizontal de G' al eje vertical, e identifique la intersección con el eje como H'. Si se marca una distancia vertical $K'G' = H'J'$ directamente debajo del punto G', entonces el punto K' será un punto sobre la curva MR. Ésta es la forma gráfica de obtener una curva MR a partir de una curva dada AR. En sentido estricto, el trazo preciso de una recta tangente requiere conocer el valor de la derivada en la producción pertinente, es decir, $f'(N)$; por consiguiente, el método gráfico recién descrito no puede por sí mismo tener aplicación práctica. Una excepción importante es el caso de una curva AR lineal, donde la tangente a algún punto sobre la curva es simplemente la misma recta, de tal manera que no hay necesidad de trazar ninguna tangente. Entonces, el método gráfico se aplicará de modo directo.

Regla del cociente

La derivada del cociente de dos funciones, $f(x)/g(x)$, es

$$\frac{d}{dx}\frac{f(x)}{g(x)} = \frac{f'(x)g(x) - f(x)g'(x)}{g^2(x)}$$

En el numerador de la expresión del lado derecho se encuentran dos términos de producto, cada uno con la derivada de sólo una de las dos funciones originales. Note que $f'(x)$ aparece en el término positivo, y $g'(x)$ en el término negativo. El denominador está formado por el cuadrado de la función $g(x)$; es decir, $g^2(x) \equiv [g(x)]^2$.

Ejemplo 6

$$\frac{d}{dx}\left(\frac{2x-3}{x+1}\right) = \frac{2(x+1) - (2x-3)(1)}{(x+1)^2} = \frac{5}{(x+1)^2}$$

Ejemplo 7

$$\frac{d}{dx}\left(\frac{5x}{x^2+1}\right) = \frac{5(x^2+1) - 5x(2x)}{(x^2+1)^2} = \frac{5(1-x^2)}{(x^2+1)^2}$$

Ejemplo 8

$$\frac{d}{dx}\left(\frac{ax^2+b}{cx}\right) = \frac{2ax(cx) - (ax^2+b)(c)}{(cx)^2}$$

$$= \frac{c(ax^2-b)}{(cx)^2} = \frac{ax^2-b}{cx^2}$$

Esta regla se puede probar como sigue. Para cualquier valor de $x = N$, tenemos

$$\left.\frac{d}{dx}\frac{f(x)}{g(x)}\right|_{x=N} = \lim_{x\to N}\frac{f(x)/g(x) - f(N)/g(N)}{x - N} \tag{7.8}$$

La expresión de cociente después del signo de límite se puede reescribir en la forma

$$\frac{f(x)g(N) - f(N)g(x)}{g(x)g(N)}\,\frac{1}{x - N}$$

Al sumar y restar $f(N)g(N)$ en el numerador y reordenar, podemos además transformar la expresión en

$$\begin{aligned}&\frac{1}{g(x)g(N)}\left[\frac{f(x)g(N) - f(N)g(N) + f(N)g(N) - f(N)g(x)}{x - N}\right]\\ &= \frac{1}{g(x)g(N)}\left[g(N)\,\frac{f(x) - f(N)}{x - N} - f(N)\frac{g(x) - g(N)}{x - N}\right]\end{aligned}$$

Al sustituir este resultado en (7.8) y tomar el límite, tenemos

$$\begin{aligned}\left.\frac{d}{dx}\frac{f(x)}{g(x)}\right|_{x=N} &= \lim_{x\to N}\frac{1}{g(x)g(N)}\left[\lim_{x\to N} g(N)\lim_{x\to N}\frac{f(x) - f(N)}{x - N}\right.\\ &\qquad \left. - \lim_{x\to N} f(N)\lim_{x\to N}\frac{g(x) - g(N)}{x - N}\right]\\ &= \frac{1}{g^2(N)}[g(N)f'(N) - f(N)g'(N)] \qquad \text{[por (6.13)]}\end{aligned}$$

que se puede generalizar reemplazando el símbolo N con x, porque N representa cualquier valor de x. Esto prueba la regla del cociente.

Relación entre las funciones de costo marginal y costo promedio

Como una aplicación económica de la regla del cociente, debemos considerar la tasa de cambio de costo promedio cuando varía la producción.

Dada una función de costo total $C = C(Q)$, la función de costo promedio (AC) es un cociente de dos funciones de Q, puesto que $\text{AC} \equiv C(Q)/Q$, definida siempre y cuando $Q > 0$. Por lo tanto, la tasa de cambio de AC respecto a Q se determina derivando AC:

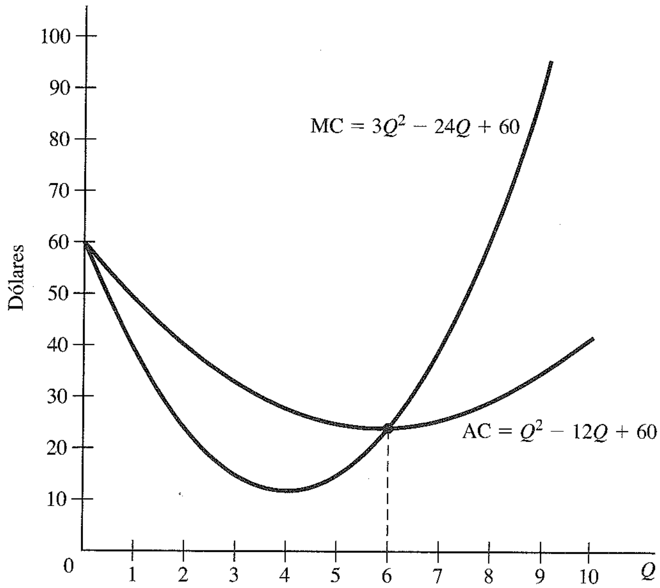
$$\frac{d}{dQ}\frac{C(Q)}{Q} = \frac{[C'(Q)\cdot Q - C(Q)\cdot 1]}{Q^2} = \frac{1}{Q}\left[C'(Q) - \frac{C(Q)}{Q}\right] \tag{7.9}$$

De esto se deduce que, para $Q > 0$,

$$\frac{d}{dQ}\frac{C(Q)}{Q} \gtreqqless 0 \qquad \text{si} \qquad C'(Q) \gtreqqless \frac{C(Q)}{Q} \tag{7.10}$$

Puesto que la derivada $C'(Q)$ representa la función de costo marginal (MC), y $C(Q)/Q$ representa la función AC, el significado económico de (7.10) es: la pendiente de la curva AC se-

FIGURA 7.3



rá positiva, cero o negativa si y sólo si la curva de costo marginal se encuentra arriba en el cruce o está debajo de la curva AC. Esto se ilustra en la figura 7.3, donde las funciones MC y AC graficadas se basan en la función específica de costo total

$$C = Q^3 - 12Q^2 + 60Q$$

A la izquierda de $Q = 6$, AC tiene pendiente descendente y, por lo tanto, MC se encuentra abajo; a la derecha, resulta cierto lo contrario. En $Q = 6$, AC tiene una pendiente de cero, y MC y AC tienen el mismo valor.[1]

La conclusión cualitativa de (7.10) se expresa de forma explícita en términos de funciones de costo. Sin embargo, su validez no se afecta si se interpreta a $C(Q)$ como *cualquier otra* función total diferenciable, con $C(Q)/Q$ y $C'(Q)$ como sus funciones promedio y marginal correspondientes. De esta forma, el resultado proporciona una relación *general* entre marginal y promedio. En particular, podemos señalar que, el hecho de que el ingreso marginal (MR) esté debajo del AR cuando éste tiene pendiente descendente, como se explicó en relación con la figura 7.2, no es sino un caso especial del resultado general expresado en (7.10).

[1] Observe que (7.10) *no* establece que, cuando AC tiene pendiente negativa, el MC también debe tener pendiente negativa; esto sólo indica que el AC debe exceder al MC en esa circunstancia. Por ejemplo, en Q = 5 de la figura 7.3, el AC está en declive pero el MC tiende a elevarse, de modo que sus pendientes tendrán signos opuestos.

EJERCICIO 7.2

1. Dada la función de costo total $C = Q^3 - 5Q^2 + 12Q + 75$, escriba la función de costo variable (VC). Encuentre la derivada de la función VC e interprete el significado económico de esa derivada.
2. Dada la función de costo promedio $AC = Q^2 - 4Q + 174$, encuentre la función MC. ¿La función dada es más apropiada como una función de largo o corto plazo? ¿Por qué?

3. Diferencie las siguientes funciones por medio de la regla del producto:
   (*a*) $(9x^2-2)(3x+1)$ (*c*) $x^2(4x+6)$ (*e*) $(2-3x)(1+x)(x+2)$
   (*b*) $(3x+10)(6x^2-7x)$ (*d*) $(ax-b)(cx^2)$ (*f*) $(x^2+3)x^{-1}$
4. (*a*) Dado un $\text{AR} = 60 - 3Q$, grafique la curva de ingreso promedio; después, determine la curva de MR por el método usado en la figura 7.2.
   (*b*) Encuentre matemáticamente la función de ingreso total y la función de ingreso marginal correspondientes a la función dada AR.
   (*c*) ¿La curva MR obtenida de forma gráfica en (*a*) coincide con la función MR obtenida matemáticamente en (*b*)?
   (*d*) Comparando las funciones AR y MR, ¿qué se puede concluir acerca de sus pendientes relativas?
5. Proporcione una prueba matemática para el resultado general de que, dada una curva *lineal* que representa un promedio, la curva marginal correspondiente debe tener la misma ordenada al origen, pero su pendiente será el doble de la pendiente de la curva promedio.
6. Pruebe el resultado de (7.6) tratando primero a $g(x)h(x)$ como una sola función, $g(x)h(x) \equiv \phi(x)$, y después aplicando la regla del producto (7.4).
7. Encuentre las derivadas de:
   (*a*) $(x^2+3)/x$ (*c*) $6x/(x+5)$
   (*b*) $(x+9)/x$ (*d*) $(ax^2+b)/(cx+d)$
8. Dada la función $f(x) = ax + b$, encuentre las derivadas de:
   (*a*) $f(x)$ (*b*) $xf(x)$ (*c*) $1/f(x)$ (*d*) $f(x)/x$
9. (*a*) ¿Es cierto que $f \in C' \Rightarrow f \in C$?
   (*b*) ¿Es verdad que $f \in C \Rightarrow f \in C'$?
10. Encuentre las funciones marginal y promedio de las siguientes funciones totales y grafique los resultados.
    Función de costo total:
    (*a*) $C = 3Q^2 + 7Q + 12$
    Función de ingreso total:
    (*b*) $R = 10Q - Q^2$
    Función de producto total:
    (*c*) $Q = aL + bL^2 - cL^3 \quad (a, b, c > 0)$

7.3 Reglas de diferenciación para funciones de variables diferentes

En la sección 7.2 analizamos las reglas de diferenciación de una suma, diferencia, producto o cociente de dos (o más) funciones diferenciables de la misma variable. Ahora consideraremos casos donde hay dos o más funciones diferenciables, cada una de las cuales tiene una variable independiente *distinta*.

Regla de la cadena

Si tenemos una función diferenciable $z = f(y)$, donde y es a su vez una función diferenciable de otra variable x, por ejemplo $y = g(x)$, entonces la derivada de z respecto a x es igual a la derivada de z respecto a y, por la derivada de y respecto a x. Expresada en símbolos,

$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx} = f'(y)g'(x) \tag{7.11}$$

Esta regla, conocida como *regla de la cadena*, apela fácilmente a la intuición. Dada una Δx, debe resultar una Δy correspondiente vía la función $y = g(x)$, pero ésta Δy da lugar a su vez a una Δz vía la función $z = f(y)$. Por lo tanto, hay una "reacción en cadena", como sigue:

$$\Delta x \xrightarrow{\text{vía } g} \Delta y \xrightarrow{\text{vía } f} \Delta z$$

Los dos enlaces en esta cadena conllevan dos cocientes de diferencias, $\Delta y/\Delta x$ y $\Delta z/\Delta y$, pero cuando se multiplican, la Δy se cancela y se termina con

$$\frac{\Delta z}{\Delta y}\frac{\Delta y}{\Delta x} = \frac{\Delta z}{\Delta x}$$

un cociente de diferencias que relaciona a Δz con Δx. Si se toma el límite de estos cocientes de diferencias cuando $\Delta x \to 0$ (lo cual significa que $\Delta y \to 0$), cada cociente de diferencias se transforma en una derivada; es decir, se tendrá $(dz/dy)(dy/dx) = dz/dx$. Éste es precisamente el resultado de (7.11).

Tomando en cuenta la función $y = g(x)$, podemos expresar la función $z = f(y)$ como $z = f[g(x)]$, donde la presencia contigua de los dos símbolos de función f y g indican que ésta es una *función compuesta* (función de funciones). Por esta razón, la regla de la cadena se conoce también como *regla de función compuesta* o *regla de una función de funciones.*

La extensión de la regla de la cadena a tres o más funciones es directa. Si tenemos $z = f(y)$, $y = g(x)$ y $x = h(w)$, entonces

$$\frac{dz}{dw} = \frac{dz}{dy}\frac{dy}{dx}\frac{dx}{dw} = f'(y)g'(x)h'(w)$$

y de manera similar para casos en los que intervienen más de dos funciones.

Ejemplo 1

Si $z = 3y^2$, donde $y = 2x + 5$, entonces

$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx} = 6y(2) = 12y = 12(2x+5)$$

Ejemplo 2

Si $z = y - 3$, donde $y = x^3$, entonces

$$\frac{dz}{dx} = 1(3x^2) = 3x^2$$

Ejemplo 3

La utilidad de esta regla se aprecia mejor cuando debemos diferenciar una función como $z = (x^2 + 3x - 2)^{17}$. Sin la regla de la cadena a la mano, dz/dx se obtiene sólo a través de la laboriosa ruta de desarrollar primero la expresión a la decimoséptima potencia. Sin embargo, con la regla de la cadena, podemos tomar un atajo definiendo una nueva variable *intermedia* $y = x^2 + 3x - 2$, de tal manera que se obtienen dos funciones enlazadas en una cadena:

$$z = y^{17} \qquad \text{y} \qquad y = x^2 + 3x - 2$$

Entonces, la derivada dz/dx se puede hallar como sigue:

$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx} = 17y^{16}(2x+3) = 17(x^2+3x-2)^{16}(2x+3)$$

Ejemplo 4 Dada una función de ingreso total de una empresa $R = f(Q)$, donde la producción Q es una función de insumo de mano de obra L, o $Q = g(L)$, determine dR/dL. Por la regla de la cadena, se tiene

$$\frac{dR}{dL} = \frac{dR}{dQ}\frac{dQ}{dL} = f'(Q)g'(L)$$

Traducida en términos económicos, dR/dQ es la función MR y dQ/dL es la función de producto físico marginal de mano de obra (MPP_L). De manera similar, dR/dL tiene la connotación de la función de producto de ingreso marginal de mano de obra (MRP_L). Así, el resultado mostrado constituye la expresión matemática del resultado bien conocido en economía de que $MRP_L = MR \cdot MPP_L$.

Regla de la función inversa

Si la función $y = f(x)$ representa un mapeo uno a uno, es decir, si es tal que cada valor de y se relaciona con un único valor de x, la función f tendrá una *función inversa* $x = f^{-1}(y)$ (léase: "x es una función inversa de y"). Aquí, el símbolo f^{-1} es un símbolo de función que, como el símbolo de derivada de una función f', significa una función relacionada con la función f; esto es, *no* significa el recíproco de la función $f(x)$.

Lo que en esencia significa la existencia de una función inversa es, en este caso, que no sólo cada valor de x produce un único valor de y [es decir, $y = f(x)$], sino también cada valor de y corresponde a un valor único de x. Tomando un caso no numérico, se puede ejemplificar la aplicación uno a uno mediante el envío del conjunto de todos los esposos al conjunto de todas las esposas en una sociedad monógama. Cada esposo tiene una esposa única, y cada esposa tiene un esposo único. En contraste, la aplicación del conjunto de todos los padres al conjunto de todos los hijos no es uno a uno, porque un padre puede tener más de un hijo, aunque cada hijo tiene un único padre.

Cuando x y y se refieren particularmente a números, se ve que la propiedad de la aplicación uno a uno es única para la clase de funciones conocidas como *funciones estrictamente monotónicas* (o *monótonas*). Dada una función $f(x)$, si valores sucesivamente más grandes de la variable independiente x conducen *siempre* a valores sucesivamente más grandes de $f(x)$, es decir, si

$$x_1 > x_2 \Rightarrow f(x_1) > f(x_2)$$

entonces, se dice que la función f es una *estrictamente creciente*. Por otro lado, cuando incrementos sucesivos de x conducen siempre a disminuciones sucesivas en $f(x)$, es decir, si

$$x_1 > x_2 \Rightarrow f(x_1) < f(x_2)$$

se dice que la función es *estrictamente decreciente*. En cualquiera de estos casos, existe una función inversa f^{-1}.[2]

Una forma práctica de determinar la monotonía estricta de una determinada función $y = f(x)$ es comprobar si la derivada $f'(x)$ conserva siempre el mismo signo algebraico (no cero) para todos los valores de x. Desde el punto de vista geométrico, esto significa que su pen-

[2] Si se omite el adverbio *estrictamente*, las funciones *monotónicas* (o *monótonas*) se pueden definir como: una *función creciente* es una función con la propiedad que

$$x_1 > x_2 \Rightarrow f(x_1) \geq f(x_2) \qquad \text{[con la desigualdad débil } \geq]$$

y una *función decreciente* es una con la propiedad que

$$x_1 > x_2 \Rightarrow f(x_1) \leq f(x_2) \qquad \text{[con la desigualdad débil } \leq]$$

Note que, en esta definición, una función escalón ascendente (descendente) califica como una función creciente (decreciente), a pesar del hecho de que su gráfica contiene segmentos horizontales. Puesto que tales función no tienen una aplicación uno a uno, no tienen funciones inversas.

diente siempre es ascendente o descendente. Así, la curva de demanda de una empresa $Q = f(P)$ que tiene una pendiente negativa en cada punto es estrictamente decreciente. Como tal, tiene una función inversa $P = f^{-1}(Q)$, la cual, como ya mencionamos previamente, da la curva de ingreso promedio de la empresa, puesto que $P \equiv \text{AR}$.

Ejemplo 5

La función

$$y = 5x + 25$$

tiene la derivada $dy/dx = 5$, que es positiva sin importar el valor de x; por lo tanto, la función es estrictamente creciente. Se deduce que su función inversa existe. En el caso presente, la función inversa se determina con facilidad resolviendo la ecuación $y = 5x + 25$ para x. El resultado es la función

$$x = \tfrac{1}{5}y - 5$$

Es interesante notar que esta función inversa es también estrictamente creciente, porque $dx/dy = \frac{1}{5} > 0$ para todos los valores de y.

En términos generales, si existe una función inversa, las funciones original e inversa deben ser estrictamente monotónicas. Además, si f^{-1} es la función inversa de f, entonces f debe ser la función inversa de f^{-1}; es decir, f y f^{-1} deben ser funciones inversas una respecto a la otra.

Es fácil comprobar que la gráfica de $y = f(x)$ y que $x = f^{-1}(y)$ son lo mismo, sólo con los ejes invertidos. Si se coloca el eje x de la gráfica de f^{-1} sobre el eje x de la gráfica de f (y de manera similar para el eje y), coincidirán las dos curvas. Por otro lado, si el eje x de la gráfica de f^{-1} se coloca sobre el eje y de la gráfica de f (y viceversa), las dos curvas serán *imágenes especulares* entre sí en relación con la recta de 45° que pasa por el origen. Esta relación de imagen especular proporciona una forma fácil de graficar la función inversa f^{-1}, una vez que se da la gráfica de la función original f. (Usted puede probar esto con las dos funciones del ejemplo 5.)

Para funciones inversas, la regla de diferenciación es

$$\frac{dx}{dy} = \frac{1}{dy/dx}$$

Esto significa que la derivada de la función inversa es la recíproca de la derivada de la función original; como tal, dx/dy debe llevar el mismo signo que dy/dx, de manera que si f es estrictamente creciente (decreciente), entonces también lo debe ser f^{-1}.

Como comprobación de esta regla, se puede retomar el ejemplo 5, donde se encontró que dy/dx es 5, y dx/dy es igual a $\frac{1}{5}$. Estas dos derivadas son, de hecho, recíprocas entre sí y tienen el mismo signo.

En este ejemplo sencillo, es relativamente fácil obtener la función inversa, de modo que su derivada dx/dy se encuentra de modo directo a partir de la función inversa. Sin embargo, como se muestra en el ejemplo 6, a veces es difícil expresar de forma explícita la función inversa y, por lo tanto, no es posible llevar a cabo la diferenciación directa. La utilidad de la regla de la función inversa se vuelve por completo evidente.

Ejemplo 6

Dada $y = x^5 + x$, determine dx/dy. En primer lugar, puesto que

$$\frac{dy}{dx} = 5x^4 + 1 > 0$$

para cualquier valor de x, la función dada es estrictamente creciente, y existe una función inversa. Resolver la ecuación para x podría ser difícil, pero por medio de la regla de la función inversa se puede hallar con rapidez la derivada de la función inversa:

$$\frac{dx}{dy} = \frac{1}{dy/dx} = \frac{1}{5x^4 + 1}$$

La regla de la función inversa, en términos estrictos, es aplicable sólo cuando la función en cuestión es una aplicación uno a uno; sin embargo, en realidad tenemos cierta libertad. Por ejemplo, al tratar con una curva en forma de U (no estrictamente monotónica), podemos considerar que los segmentos de la curva con pendiente descendente y ascendente representan dos funciones *separadas*, cada una con dominio restringido, y estrictamente monótonas en el dominio restringido. Para cada una de éstas, se puede aplicar de nuevo la regla de la función inversa.

EJERCICIO 7.3

1. Dada $y = u^3 + 2u$, donde $u = 5 - x^2$, encuentre dy/dx por la regla de la cadena.
2. Dada $w = ay^2$ y $y = bx^2 + cx$, obtenga dw/dx por la regla de la cadena.
3. Use la regla de la cadena para hallar dy/dx para las siguientes funciones:
   (*a*) $y = (3x^2 - 13)^3$ (*b*) $y = (7x^3 - 5)^9$ (*c*) $y = (ax + b)^5$
4. Dada $y = (16x + 3)^{-2}$, use la regla de la cadena para hallar dy/dx. Después, exprese la función como $y = 1/(16x + 3)^2$ y encuentre dy/dx por la regla del cociente. ¿Son idénticas las respuestas?
5. Dada $y = 7x + 21$, determine su función inversa. Luego, halle dy/dx y dx/dy, y compruebe la regla de la derivada de la función inversa. Asimismo, verifique que las gráficas de las dos funciones guardan una relación de imagen especular entre sí.
6. ¿Las siguientes funciones son estrictamente monótonas?
   (*a*) $y = -x^6 + 5$ $(x > 0)$
   (*b*) $y = 4x^5 + x^3 + 3x$
   Para cada función estrictamente monótona, determine dx/dy mediante la regla de la función inversa.

7.4 Diferenciación parcial

Hasta ahora, sólo hemos considerado las derivadas de funciones de una sola variable independiente; sin embargo, en el análisis estático comparativo es probable que hallemos la situación en la cual aparecen varios parámetros en el modelo, de manera que el valor de equilibrio de cada variable endógena puede ser una función de más de un parámetro. Por lo tanto, como preparación final a la aplicación del concepto de derivada a la estática comparativa, debemos aprender cómo hallar la derivada de una función de más de una variable.

Derivadas parciales

Consideremos una función

$$y = f(x_1, x_2, \ldots, x_n) \tag{7.12}$$

donde las variables x_i $(i = 1, 2, \ldots, n)$ son todas *independientes* entre sí, de modo que cada una puede variar por sí misma sin afectar a las otras. Si la variable x_1 experimenta un cambio

Δx_1 mientras $x_2, \ldots, x_n$ permanecen fijas, habrá un cambio correspondiente en y, a saber, Δy. El cociente de diferencias se puede expresar como

$$\frac{\Delta y}{\Delta x_1} = \frac{f(x_1 + \Delta x_1, x_2, \ldots, x_n) - f(x_1, x_2, \ldots, x_n)}{\Delta x_1} \tag{7.13}$$

Si se toma el límite de $\Delta y/\Delta x_1$ cuando $\Delta x_1 \to 0$, ese límite constituirá una derivada, a la que llamaremos *derivada parcial* de y con respecto a x_1, para indicar que las otras variables independientes en la función se mantienen constantes al tomar esta derivada particular. Se pueden definir derivadas parciales similares para cambios infinitesimales en variables independientes. El proceso de tomar derivadas parciales se llama *diferenciación parcial*.

A las derivadas parciales se les asignan símbolos distintos. En lugar de la letra d (como en dy/dx), se emplea el símbolo ∂, que es una variante de la letra griega δ (delta minúscula). Por lo tanto, ahora debemos escribir $\partial y/\partial x_i$, que se lee: "la derivada parcial de y con respecto a x_i". A veces el símbolo de derivada parcial se escribe también como < $\frac{\partial}{\partial x_i} y$; en ese caso, la parte $\partial/\partial x_i$ se considera como un símbolo de operador que instruye a tomar la derivada parcial de (alguna función) con respecto a la variable x_i. Puesto que la función en cuestión se denota por f en (7.12), también es permisible escribir $\partial f/\partial x_i$.

¿Existe también una contraparte de derivada parcial para el símbolo $f'(x)$ que hayamos usado antes? La respuesta es afirmativa. Sin embargo, en lugar de f' ahora se usa f_1, f_2, etc., donde el subíndice indica la variable independiente (única) a la que se le permite variar. Si la función (7.12) se escribe en términos de variables sin subíndice, como $y = f(u, v, w)$, entonces las derivadas parciales se pueden denotar mediante f_u, f_v, y f_w en vez de f_1, f_2 y f_3.

De acuerdo con estas notaciones, y con base en (7.12) y (7.13), ahora se puede definir

$$f_1 \equiv \frac{\partial y}{\partial x_1} \equiv \lim_{\Delta x_1 \to 0} \frac{\Delta y}{\Delta x_1}$$

como la primera en el conjunto de n derivadas parciales de la función f.

Técnicas de diferenciación parcial

La diferenciación parcial difiere de la diferenciación descrita antes, sobre todo, en que se deben mantener *constantes* $(n - 1)$ variables independientes mientras se permite que cambie una variable. En vista de que ya aprendimos a manejar *constantes* en la diferenciación, en la práctica, la diferenciación parcial debe presentar poco problema.

Ejemplo 1

Dada $y = f(x_1, x_2) = 3x_1^2 + x_1x_2 + 4x_2^2$, encuentre las derivadas parciales. Al determinar $\partial y/\partial x_1$ (o f_1), debemos recordar que x_2 se tratará como constante durante la diferenciación. Como tal, x_2 se elimina del proceso si es una constante *aditiva* (por ejemplo, el término $4x_2^2$), pero se retiene si es una constante *multiplicativa* (por ejemplo, el término x_1x_2). Así, se tiene

$$\frac{\partial y}{\partial x_1} \equiv f_1 = 6x_1 + x_2$$

De manera similar, al tratar x_1 como una constante, se encuentra que

$$\frac{\partial y}{\partial x_2} \equiv f_2 = x_1 + 8x_2$$

Note que, como la función primitiva f, ambas derivadas parciales son por sí mismas funciones de las variables x_1 y x_2; es decir, es posible escribirlas como dos funciones derivadas

$$f_1 = f_1(x_1, x_2) \qquad \text{y} \qquad f_2 = f_2(x_1, x_2)$$

Para el punto $(x_1, x_2) = (1, 3)$ en el dominio de la función f, por ejemplo, las derivadas parciales tomarán los siguientes valores específicos:

$$f_1(1, 3) = 6(1) + 3 = 9 \qquad \text{y} \qquad f_2(1, 3) = 1 + 8(3) = 25$$

Ejemplo 2 Dada $y = f(u, v) = (u + 4)(3u + 2v)$, las derivadas parciales se determinan mediante la regla del producto. Si se mantiene v constante, se tiene

$$f_u = (u + 4)(3) + 1(3u + 2v) = 2(3u + v + 6)$$

De manera similar, si se mantiene u constante, se encuentra que

$$f_v = (u + 4)(2) + 0(3u + 2v) = 2(u + 4)$$

Cuando $u = 2$ y $v = 1$, estas derivadas toman los siguientes valores:

$$f_u(2, 1) = 2(13) = 26 \qquad \text{y} \qquad f_v(2, 1) = 2(6) = 12$$

Ejemplo 3 Dada $y = (3u - 2v)/(u^2 + 3v)$, las derivadas parciales se determinan mediante el uso de la regla del cociente:

$$\frac{\partial y}{\partial u} = \frac{3(u^2 + 3v) - 2u(3u - 2v)}{(u^2 + 3v)^2} = \frac{-3u^2 + 4uv + 9v}{(u^2 + 3v)^2}$$

$$\frac{\partial y}{\partial v} = \frac{-2(u^2 + 3v) - 3(3u - 2v)}{(u^2 + 3v)^2} = \frac{-u(2u + 9)}{(u^2 + 3v)^2}$$

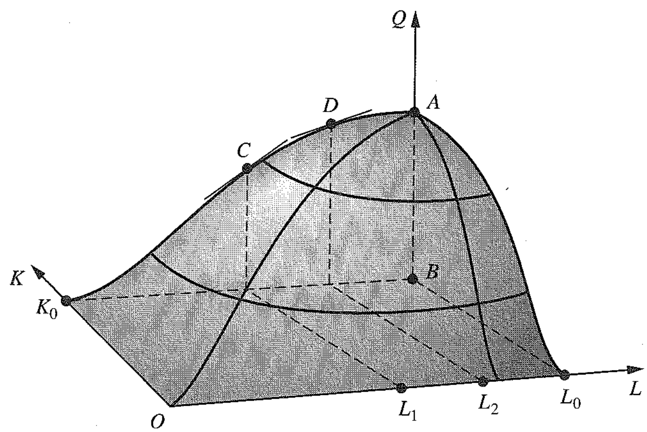
Interpretación geométrica de las derivadas parciales

Como un tipo especial de derivada, una derivada parcial es una medida de las razones instantáneas de cambio de alguna variable y, como tal, de nuevo tiene una contraparte geométrica en la pendiente de una curva particular.

Consideremos una función de producción $Q = Q(K, L)$, donde Q, K y L denotan producción, aportación de capital y de mano de obra, respectivamente. Esta función es una versión particular de dos variables de (7.12), con $n = 2$. Por lo tanto, podemos definir dos derivadas parciales $\partial Q/\partial K$ (o bien, Q_K) y $\partial Q/\partial L$ (o Q_L). La derivada parcial Q_K se relaciona con la razón de cambio de la producción respecto a cambios infinitesimales de capital, mientras se mantiene constante la mano de obra. Por consiguiente, Q_K simboliza la función de producto físico marginal de capital (MPP_K). De manera similar, la derivada parcial Q_L es la representación matemática de la función MPP_L.

En términos geométricos, la función de producción $Q = Q(K, L)$ se ilustra mediante una *superficie de producción* en un espacio tridimensional como se muestra en la figura 7.4. La variable Q se grafica verticalmente, de manera que para cualquier punto (K, L) en el plano base (plano KL), la altura de la superficie indicará la producción Q. El dominio de la función debe abarcar todo el cuadrante no negativo del plano base, pero para los fines que aquí se persiguen

FIGURA 7.4



es suficiente considerar un subconjunto de él, el rectángulo OK_0BL_0. Como una consecuencia, sólo se muestra una pequeña porción de la superficie de producción en la figura.

Ahora mantendremos fijo el capital en el nivel K_0 y consideraremos sólo variaciones en el insumo L. Al establecer $K = K_0$, todos los puntos en el dominio (restringido) se vuelven irrelevantes, excepto los del segmento de recta K_0B. De la misma manera, sólo la curva K_0CDA (una sección transversal de la superficie de producción) se relaciona con la presente explicación. Esta curva representa una curva de producto físico total de mano de obra (TPP_L) para una cantidad fija de capital $K = K_0$; por lo tanto, podemos leer en su pendiente la tasa de cambio de Q respecto a cambios en L mientras K se mantiene constante. En consecuencia, resulta claro que la pendiente de una curva como K_0CDA representa la expresión geométrica de la derivada parcial Q_L. De nuevo, se advierte que la pendiente de una curva que representa un total (TPP_L) es su curva marginal correspondiente ($\text{MPP}_L \equiv Q_L$).

Como se mencionó, una derivada parcial es una función de todas las variables independientes de la función primitiva. Que Q_L sea una función de L es inmediatamente obvio en la curva K_0CDA. Cuando $L = L_1$, el valor de Q_L es igual a la pendiente de la curva en el punto C; pero cuando $L = L_2$, la pendiente que importa es la del punto D. ¿Por qué Q_L es también una función de K? La respuesta es que K se puede fijar en varios niveles, y para cada nivel fijo de K aparece una curva TPP_L distinta (una sección transversal diferente de la superficie de producción), con repercusiones inevitables en la derivada Q_L. Por consiguiente, Q_L es también una función de K.

La derivada parcial Q_k se puede interpretar análogamente. Si se mantiene constante la aportación de mano de obra en lugar de K (por ejemplo, en el nivel de L_0), el segmento de recta L_0B será el subconjunto adecuado del dominio, y la curva L_0A indicará el subconjunto pertinente de la superficie de producción. La derivada parcial Q_K se puede interpretar, entonces, como la pendiente de la curva L_0A, sin olvidar que el eje K va de sudeste a noreste en la figura 7.4. Se debe observar que Q_K es de nuevo una función de las variables L y K.

Vector gradiente

Las derivadas parciales de una función $y = f(x_1, x_2, \ldots, x_n)$ se pueden reunir bajo una sola entidad matemática llamada *vector gradiente*, o simplemente *gradiente*, de la función f:

$$\text{grad } f(x_1, x_2, \ldots, x_n) = (f_1, f_2, \ldots, f_n)$$

donde $f_i \equiv \partial y/\partial x_i$. Note que aquí se usan paréntesis y no corchetes para escribir el vector. Por otro lado, el gradiente se puede denotar por $\nabla f(x_1, x_2, \ldots, x_n)$, donde ∇ (léase "nabla") es la versión invertida de la letra griega Δ.

Puesto que la función f tiene n argumentos, hay en total n derivadas parciales; por lo tanto, grad f es un vector n. Cuando estas derivadas se evalúan en un punto específico $(x_{10}, x_{20}, \ldots, x_{n0})$ en el dominio, se obtiene grad $f(x_{10}, x_{20}, \ldots, x_{n0})$, un vector de *valores* de derivadas específicas.

Ejemplo 4 El vector gradiente de la función de producción $Q = Q(K, L)$ es

$$\nabla Q = \nabla Q(K, L) = (Q_K, Q_L)$$

EJERCICIO 7.4

1. Encuentre $\partial y/\partial x_1$ y $\partial y/\partial x_2$ para cada una de las siguientes funciones:
   (*a*) $y = 2x_1^3 - 11x_1^2x_2 + 3x_2^2$
   (*b*) $y = 7x_1 + 6x_1x_2^2 - 9x_2^3$
   (*c*) $y = (2x_1 + 3)(x_2 - 2)$
   (*d*) $y = (5x_1 + 3)/(x_2 - 2)$
2. Determine f_x y f_y a partir de las siguientes funciones:
   (*a*) $f(x, y) = x^2 + 5xy - y^3$
   (*b*) $f(x, y) = (x^2 - 3y)(x - 2)$
   (*c*) $f(x, y) = \dfrac{2x - 3y}{x + y}$
   (*d*) $f(x, y) = \dfrac{x^2 - 1}{xy}$
3. De las respuestas al problema 2, determine $f_x(1, 2)$, el valor de la derivada parcial f_x cuando $x = 1$ y $y = 2$, para cada función.
4. Dada la función de producción $Q = 96K^{0.3}L^{0.7}$, encuentre las funciones MPP_K y MPP_L. ¿MPP_K es una función de K solamente, o de K y L? ¿Qué se puede decir acerca de MPP_L?
5. Si la función de utilidad de un individuo toma la forma
   $$U = U(x_1, x_2) = (x_1 + 2)^2(x_2 + 3)^3$$
   donde U es la utilidad total, y x_1 y x_2 son las cantidades de dos artículos consumidos.
   (*a*) Halle la función de utilidad marginal de cada uno de los dos artículos.
   (*b*) Encuentre el valor de la utilidad marginal del primer artículo cuando se han consumido tres unidades de cada artículo.
6. La oferta de dinero total M tiene dos componentes: depósitos bancarios D y tenencias de efectivo C, que se supone que exhiben una relación constante $C/D = c$, $0 < c < 1$. El dinero de alto poder expansivo o base monetaria H se define como la suma de tenencias de efectivo que mantiene el público y las reservas que tienen los bancos. Las reservas de los bancos son una fracción de depósitos bancarios, determinados por el coeficiente de reservas r, $0 < r < 1$.
   (*a*) Exprese la oferta de dinero M como una función de la base monetaria H.
   (*b*) ¿Un incremento en la relación de reservas r aumenta o disminuye la oferta de dinero?
   (*c*) ¿Cómo afectaría un incremento en el cociente de efectivo sobre depósito c a la oferta monetaria?
7. Escriba los gradientes de las funciones siguientes:
   (*a*) $f(x, y, z) = x^2 + y^3 + z^4$
   (*b*) $f(x, y, z) = xyz$

7.5 Aplicaciones al análisis estático comparativo

Ahora que ya contamos con conocimientos acerca de las distintas reglas de diferenciación, podemos, al menos, enfrentar el problema que se presenta en el análisis estático comparativo: cómo cambiará el valor de equilibrio de una variable endógena cuando hay un cambio en alguna de las variables exógenas o parámetros.

Modelo de mercado

Primero, consideremos de nuevo el modelo de mercado simple de un único artículo de (3.1). Ese modelo se puede escribir en la forma de dos ecuaciones:

$$Q = a - bP \qquad (a, b > 0) \qquad \text{[demanda]}$$
$$Q = -c + dP \qquad (c, d > 0) \qquad \text{[oferta]}$$

con soluciones

$$P^* = \frac{a + c}{b + d} \tag{7.14}$$

$$Q^* = \frac{ad - bc}{b + d} \tag{7.15}$$

Se hará referencia a estas soluciones como la *forma reducida*: las dos variables endógenas han sido reducidas a expresiones explícitas de los cuatro parámetros mutuamente independientes a, b, c y d.

Para determinar cómo afectará al valor de P^* a un cambio infinitesimal en uno de los parámetros, sólo se tiene que diferenciar parcialmente la ecuación (7.14) respecto a cada uno de los parámetros. Si se puede determinar el *signo* de una derivada parcial, por ejemplo, $\partial P^*/\partial a$, a partir de la información dada acerca de los parámetros, se sabrá la dirección en la cual se moverá P^* cuando cambia el parámetro a; esto constituye una conclusión cualitativa. Si se puede determinar la magnitud de $\partial P^*/\partial a$, esto constituirá una conclusión cuantitativa.

De manera similar, se sacan conclusiones cualitativas o cuantitativas de las derivadas parciales de Q^* con respecto a cada parámetro, como $\partial Q^*/\partial a$. Sin embargo, para evitar una interpretación incorrecta, se debe distinguir claramente entre las dos derivadas $\partial Q^*/\partial a$ y $\partial Q/\partial a$. La última derivada es un concepto apropiado para la función de demanda sola, y sin considerar la función de oferta. Por otro lado, la derivada $\partial Q^*/\partial a$ toma en cuenta la interacción de oferta y demanda, puesto que (7.15) es una solución del modelo y pertenece a la cantidad de equilibrio. Para remarcar esta distinción, haremos referencia a las derivadas parciales de P^* y Q^*, respecto a los parámetros como *derivadas estáticas comparativas*. La posibilidad de confusión entre $\partial Q^*/\partial a$ y $\partial Q/\partial a$ es precisamente la razón por la que eligimos usar la notación de asterisco, como en Q^* para denotar el valor de equilibrio.

Si centramos nuestra atención en P^*, obtenemos las siguientes cuatro derivadas parciales de (7.14):

$$\frac{\partial P^*}{\partial a} = \frac{1}{b+d} \qquad \left[\text{el parámetro } a \text{ tiene el coeficiente } \frac{1}{b+d}\right]$$

$$\frac{\partial P^*}{\partial b} = \frac{0(b+d) - 1(a+c)}{(b+d)^2} = \frac{-(a+c)}{(b+d)^2} \qquad \text{[regla del cociente]}$$

$$\frac{\partial P^*}{\partial c} = \frac{1}{b+d} \left(= \frac{\partial P^*}{\partial a} \right)$$

$$\frac{\partial P^*}{\partial d} = \frac{0(b+d) - 1(a+c)}{(b+d)^2} = \frac{-(a+c)}{(b+d)^2} \left(= \frac{\partial P^*}{\partial b} \right)$$

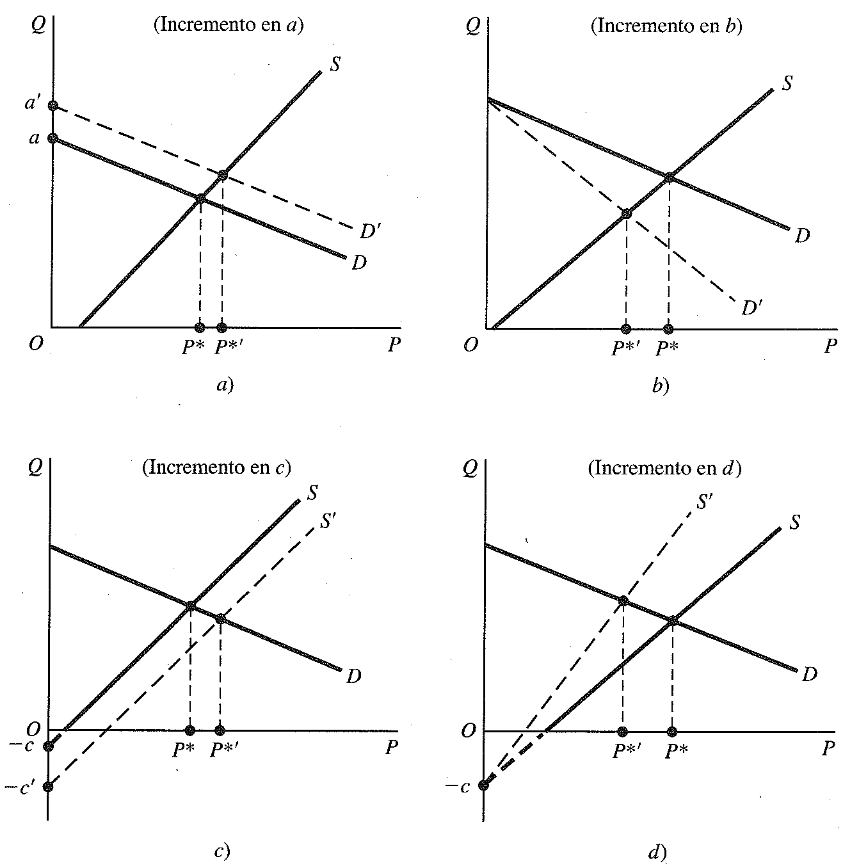
Puesto que todos los parámetros están restringidos a ser positivos en el modelo actual, se concluye que

$$\frac{\partial P^*}{\partial a} = \frac{\partial P^*}{\partial c} > 0 \qquad \text{y} \qquad \frac{\partial P^*}{\partial b} = \frac{\partial P^*}{\partial d} < 0 \tag{7.16}$$

Para una apreciación completa de los resultados de (7.16), examine la figura 7.5, donde cada diagrama muestra un cambio en *uno* de los parámetros. Como antes, se grafica Q (y no P) en el eje vertical.

La figura 7.5a ilustra un incremento en el parámetro a (desde a hasta a'). Esto significa una ordenada al origen mayor para la curva de demanda y, en vista de que el parámetro b (el parámetro de la pendiente) no cambia, el incremento en a da como resultado un desplazamiento paralelo hacia arriba de la curva de demanda de D a D'. La intersección de D' y la curva de oferta S determina un precio de equilibrio $P^{*\prime}$, el cual es mayor que el precio de equilibrio an-

FIGURA 7.5



terior P^*. Esto corrobora el resultado de que $\partial P^*/\partial a > 0$, aunque por razones de exposición se ha mostrado en la figura 7.5*a* un cambio mucho más grande en el parámetro a de lo que implica el concepto de derivada.

La situación que muestra la figura 7.5*c* tiene una interpretación similar; sin embargo, como el incremento toma lugar en el parámetro c, el resultado, en este caso, es un desplazamiento paralelo de la curva de oferta. Note que este desplazamiento es hacia abajo porque la curva de oferta tiene un intercepto vertical de $-c$; por lo tanto, un incremento en c significaría un cambio en el intercepto, por ejemplo, de -2 a -4. El resultado estático comparativo gráfico de que $P^{*\prime}$ excede a P^* de nuevo cumple con lo que el signo positivo de la derivada $\partial P^*/\partial c$ llevaría a esperar.

En la figura 7.5*b* y *d* se ilustra los efectos de cambios en los parámetros de las pendientes b y d de las dos funciones del modelo. Un incremento en b significa que la pendiente de la curva de demanda asumirá un valor numérico más grande (absoluto); es decir, se volverá más inclinada. De acuerdo con el resultado $\partial P^*/\partial b < 0$, encontramos una disminución de P^* en este diagrama. El incremento en d que hace que la curva de la oferta se vuelva más inclinada también da como resultado una disminución en el precio de equilibrio. Esto, por supuesto, concuerda de nuevo con el signo negativo de la derivada estática comparativa $\partial P^*/\partial d$.

Hasta aquí, todos los resultados de (7.16) parece que se pueden obtener de forma gráfica. Si es así, ¿por qué preocuparse por usar la diferenciación? La respuesta es que el método de diferenciación tiene al menos dos grandes ventajas: en primer lugar, la técnica gráfica está sujeta a una restricción dimensional, mientras que con la diferenciación no se tiene tal restricción. Incluso cuando el número de variables endógenas y parámetros es tal que el estado de equilibrio no se puede mostrar gráficamente, se pueden aplicar las técnicas de diferenciación al problema. En segundo lugar, el método de diferenciación produce resultados que están en un nivel de generalidad superior. Los resultados de (7.16) aún son válidos, sin importar los valores específicos que tomen los parámetros a, b, c y d, siempre y cuando satisfagan las restricciones de signo. De esta manera, las conclusiones estáticas comparativas de este modelo son aplicables a un número infinito de combinaciones de funciones de oferta y demanda (lineales). Por el contrario, el método gráfico trata sólo con algunos miembros específicos de la familia de curvas de oferta y demanda, y el resultado analítico obtenido es aplicable, en términos estrictos, sólo a las funciones específicas ilustradas.

Esta explicación sirve para ilustrar la aplicación de la diferenciación parcial al análisis estático comparativo del modelo de mercado simple, pero en realidad sólo se ha completado la mitad de la tarea, porque también es posible hallar las derivadas estáticas comparativas que pertenecen a Q^*; le dejamos esta tarea como ejercicio.

Modelo de ingreso nacional

En lugar del modelo simple de ingreso nacional estudiado en el capítulo 3, ahora se trabajará con un modelo un poco más grande con tres variables endógenas, Y (ingreso nacional), C (consumo) y T (impuestos):

$$\begin{aligned} Y &= C + I_0 + G_0 \\ C &= \alpha + \beta(Y - T) \qquad (\alpha > 0; \quad 0 < \beta < 1) \\ T &= \gamma + \delta Y \qquad (\gamma > 0; \quad 0 < \delta < 1) \end{aligned} \tag{7.17}$$

La primera ecuación de este sistema proporciona la condición de equilibrio para el ingreso nacional, mientras que las ecuaciones segunda y tercera muestran, respectivamente, cómo se determinan C y T en el modelo.

Las restricciones en los valores de los parámetros α, β, γ y δ se pueden explicar así: α es positiva debido a que el consumo es positivo, incluso si el ingreso disponible $(Y - T)$ es cero; β es una fracción positiva porque representa la propensión marginal al consumo; γ es positiva porque, incluso si Y es cero, el gobierno aún tiene un ingreso de impuestos positivo (de bases impositivas aparte del ingreso), y por último, δ es una fracción positiva porque representa una tasa de impuestos sobre la renta, y como tal no puede ser mayor de 100 por ciento. Las variables exógenas I_0 (inversión) y G_0 (gasto público) son no negativas. Se supone que todos los parámetros y variables exógenas son independientes entre sí, de modo que a cualquiera se le puede asignar un nuevo valor sin afectar a las demás.

De este modelo se puede despejar Y^* al sustituir la tercera ecuación de (7.17) en la segunda y sustituir después la ecuación resultante en la primera. El ingreso de equilibrio (en forma reducida) es

$$Y^* = \frac{\alpha - \beta\gamma + I_0 + G_0}{1 - \beta + \beta\delta} \tag{7.18}$$

También es posible determinar valores de equilibrio similares para las variables endógenas C y T, pero se centrará la atención en el ingreso de equilibrio.

De (7.18) se pueden obtener seis derivadas estáticas comparativas. Entre éstas, las tres siguientes tienen importancia especial de plan de acción:

$$\frac{\partial Y^*}{\partial G_0} = \frac{1}{1 - \beta + \beta\delta} > 0 \tag{7.19}$$

$$\frac{\partial Y^*}{\partial \gamma} = \frac{-\beta}{1 - \beta + \beta\delta} < 0 \tag{7.20}$$

$$\frac{\partial Y^*}{\partial \delta} = \frac{-\beta(\alpha - \beta\gamma + I_0 + G_0)}{(1 - \beta + \beta\delta)^2} = \frac{-\beta Y^*}{1 - \beta + \beta\delta} < 0 \qquad [\text{por (7.18)}] \tag{7.21}$$

La derivada parcial de (7.19) proporciona el *multiplicador de gasto público*. Aquí tiene un signo positivo porque β es menor que 1, y $\beta\delta$ es mayor que cero. Si se dan valores numéricos a los parámetros β y δ, se puede determinar también el valor numérico de este multiplicador a partir de (7.19). La derivada de (7.20) se podría llamar *multiplicador de ingreso no gravable*, porque muestra cómo un cambio en γ, parámetro que representa el ingreso público de fuentes no gravables, afectará el ingreso de equilibrio. Este multiplicador es negativo en el modelo presente porque el denominador de (7.20) es positivo y el numerador es negativo. Por último, la derivada parcial de (7.21), que no es de la naturaleza de un multiplicador, puesto que no relaciona un cambio de dinero con otro cambio de dinero como lo hacen las derivadas de (7.19) y (7.20), nos indica el grado al que un incremento en la tasa de impuesto sobre la renta δ disminuirá el ingreso de equilibrio.

De nuevo, advierta la diferencia entre las dos derivadas $\partial Y^*/\partial G_0$ y $\partial Y/\partial G_0$. La primera se deduce de (7.18), la expresión para el ingreso de equilibrio. La última, obtenible de la primera ecuación (7.17), es $\partial Y/\partial G_0 = 1$, que es diferente por completo en magnitud y en concepto.

Modelo de insumo producto

La solución de un modelo abierto de insumo-producto aparece como una ecuación matricial $x^* = (I - A)^{-1}d$. Si la matriz inversa $(I - A)^{-1}$ se denota por $V = [v_{ij}]$, entonces, la solución para una economía de tres industrias se puede escribir como $x^* = Vd$, o bien

$$\begin{bmatrix} x_1^* \\ x_2^* \\ x_3^* \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} \tag{7.22}$$

¿Cuáles son las tasas de cambio de las soluciones x_j^* con respecto a las demandas finales exógenas d_1, d_2 y d_3? La respuesta general es que

$$\frac{\partial x_j^*}{\partial d_k} = v_{jk} \qquad (j, k = 1, 2, 3) \tag{7.23}$$

Para ver esto, multiplcamos Vd de (7.22) y expresamos la solución como

$$\begin{bmatrix} x_1^* \\ x_2^* \\ x_3^* \end{bmatrix} = \begin{bmatrix} v_{11}d_1 + v_{12}d_2 + v_{13}d_3 \\ v_{21}d_1 + v_{22}d_2 + v_{23}d_3 \\ v_{31}d_1 + v_{32}d_2 + v_{33}d_3 \end{bmatrix}$$

En este sistema de tres ecuaciones, cada una produce una solución particular como función de las demandas finales exógenas. La diferenciación parcial de éstas produce un total de nueve derivadas estáticas comparativas:

$$\begin{array}{lll} \dfrac{\partial x_1^*}{\partial d_1} = v_{11} & \dfrac{\partial x_1^*}{\partial d_2} = v_{12} & \dfrac{\partial x_1^*}{\partial d_3} = v_{13} \\[2ex] \dfrac{\partial x_2^*}{\partial d_1} = v_{21} & \dfrac{\partial x_2^*}{\partial d_2} = v_{22} & \dfrac{\partial x_2^*}{\partial d_3} = v_{23} \\[2ex] \dfrac{\partial x_3^*}{\partial d_1} = v_{31} & \dfrac{\partial x_3^*}{\partial d_2} = v_{32} & \dfrac{\partial x_3^*}{\partial d_3} = v_{33} \end{array} \tag{7.23'}$$

Ésta es simplemente la versión desarrollada de (7.23).

Al leer (7.23′) como tres columnas distintas, se pueden combinar las tres derivadas de cada columna en una derivada matricial (vectorial).

$$\frac{\partial x^*}{\partial d_1} \equiv \frac{\partial}{\partial d_1}\begin{bmatrix} x_1^* \\ x_2^* \\ x_3^* \end{bmatrix} = \begin{bmatrix} v_{11} \\ v_{21} \\ v_{31} \end{bmatrix} \qquad \frac{\partial x^*}{\partial d_2} = \begin{bmatrix} v_{12} \\ v_{22} \\ v_{32} \end{bmatrix} \qquad \frac{\partial x^*}{\partial d_3} = \begin{bmatrix} v_{13} \\ v_{23} \\ v_{33} \end{bmatrix} \tag{7.23''}$$

Puesto que los tres vectores columna de (7.23″) son solamente las columnas de la matriz V, mediante una mayor consolidación podemos resumir las nueve derivadas en una sola derivada matricial $\partial x^*/\partial d$. Dada $x^* = Vd$, se puede escribir simplemente

$$\frac{\partial x^*}{\partial d} = \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix} = V \equiv (I - A)^{-1}$$

Así, $(I - A)^{-1}$, la inversa de la matriz de Leontief nos proporciona una presentación ordenada de las derivadas estáticas comparativas del modelo abierto de insumo-producto. Es evidente que esta derivada matricial se puede extender con facilidad del presente modelo de tres industrias al caso general de n industrias.

Las derivadas estáticas comparativas del modelo de insumo-producto son herramientas de la planificación económica, porque dan la respuesta a esta pregunta: si se revisan los objetivos de planificación, según se reflejan en $(d_1, d_2, \ldots, d_n)$, y si se desea atender todos los requerimientos directos e indirectos en la economía para estar libre por completo de cuellos de botella, ¿cómo se deben cambiar los objetivos de producción de las n industrias?

EJERCICIO 7.5

1. Examine las propiedades estáticas comparativas de la cantidad de equilibrio de (7.15) y compruebe los resultados mediante análisis gráfico.
2. Con base en (7.18), halle las derivadas parciales $\partial Y^*/\partial I_0$, $\partial Y^*/\partial \alpha$ y $\partial Y^*/\partial \beta$. Interprete el significado de cada una y determine sus signos.
3. El modelo numérico de insumo-producto (5.21) se resolvió en la sección 5.7.
   (*a*) ¿Cuántas derivadas estáticas comparativas se pueden deducir?
   (*b*) Escriba estas derivadas en la forma de (7.23′) y (7.23″).

7.6 Nota acerca de los determinantes jacobianos

El interés por las derivadas parciales se originó sólo por consideraciones estáticas comparativas. Pero las derivadas parciales también proporcionan un medio para probar si existe dependencia funcional (lineal *o* no lineal) entre un conjunto de n funciones en n variables. Esto se relaciona con la noción de determinantes jacobianos (en honor a Jacobi).

Considere las dos funciones

$$\begin{aligned} y_1 &= 2x_1 + 3x_2 \\ y_2 &= 4x_1^2 + 12x_1x_2 + 9x_2^2 \end{aligned} \tag{7.24}$$

Si se obtienen las cuatro derivadas parciales

$$\frac{\partial y_1}{\partial x_1} = 2 \qquad \frac{\partial y_1}{\partial x_2} = 3 \qquad \frac{\partial y_2}{\partial x_1} = 8x_1 + 12x_2 \qquad \frac{\partial y_2}{\partial x_2} = 12x_1 + 18x_2$$

y se disponen en una matriz cuadrada en un orden prescrito, conocida como matriz jacobiana y denotada por J, y luego se toma su determinante, el resultado será lo que se conoce como *determinante jacobiano* (o sólo *jacobiano*, para abreviar), denotado por $|J|$:

$$|J| \equiv \begin{vmatrix} \dfrac{\partial y_1}{\partial x_1} & \dfrac{\partial y_1}{\partial x_2} \\ \dfrac{\partial y_2}{\partial x_1} & \dfrac{\partial y_2}{\partial x_2} \end{vmatrix} = \begin{vmatrix} 2 & 3 \\ (8x_1 + 12x_2) & (12x_1 + 18x_2) \end{vmatrix} \tag{7.25}$$

Por economía de espacio, este jacobiano a veces se expresa también como

$$|J| \equiv \left| \frac{\partial(y_1, y_2)}{\partial(x_1, x_2)} \right|$$

En términos más generales, si se tienen n funciones diferenciables en n variables, no necesariamente lineales,

$$\begin{aligned} y_1 &= f^1(x_1, x_2, \ldots, x_n) \\ y_2 &= f^2(x_1, x_2, \ldots, x_n) \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_n &= f^n(x_1, x_2, \ldots, x_n) \end{aligned} \tag{7.26}$$

donde el símbolo f^n denota la n-ésima función (y *no* la función elevada a la n-ésima potencia), podemos obtener un total de n^2 derivadas parciales. Si se adopta la notación $f_j^i \equiv \partial y^i/\partial x_j$, se puede escribir el jacobiano

$$|J| \equiv \left|\frac{\partial(y_1, y_2, \ldots, y_n)}{\partial(x_1, x_2, \ldots, x_n)}\right| \equiv \begin{vmatrix} \partial y_1/\partial x_1 & \cdots & \partial y_1/\partial x_n \\ \vdots & & \vdots \\ \partial y_n/\partial x_1 & \cdots & \partial y_n/\partial x_n \end{vmatrix} \equiv \begin{vmatrix} f_1^1 & \cdots & f_n^1 \\ \vdots & & \vdots \\ f_1^n & \cdots & f_n^n \end{vmatrix} \tag{7.27}$$

El siguiente teorema proporciona un criterio jacobiano para la existencia de dependencia funcional entre un conjunto de n funciones: el jacobiano $|J|$ definido en (7.27) será igual a cero para todos los valores de $x_1, \ldots, x_n$ si y sólo si las n funciones $f^1, \ldots, f^n$ en (7.26) son dependientes desde el punto de vista funcional (lineal o no linealmente).

Como ejemplo, para las dos funciones de (7.24) el jacobiano de (7.25) tiene el valor

$$|J| = (24x_1 + 36x_2) - (24x_1 + 36x_2) = 0$$

Es decir, el jacobiano se anula para todos los valores de x_1 y x_2. Por lo tanto, de acuerdo con el teorema, las dos funciones de (7.24) deben ser dependientes. Se puede comprobar que y_2 es simplemente y_1 al cuadrado; por lo tanto, son de hecho funcionalmente dependientes, aquí la dependencia es *no* lineal.

Consideremos el caso especial de funciones *lineales*. Ya mostramos que los renglones de la matriz de coeficientes A de un sistema de ecuaciones lineales

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= d_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= d_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= d_n \end{aligned} \tag{7.28}$$

son linealmente dependientes si y sólo si el determinante $|A| = 0$. Este resultado se puede interpretar ahora como una aplicación especial del criterio jacobiano de dependencia funcional.

Tome la parte izquierda de cada ecuación de (7.28) como una función separada de las n variables $x_1, \ldots, x_n$, y denote estas funciones por $y_1, \ldots, y_n$. Las derivadas parciales de estas funciones resultan ser $\partial y_1/\partial x_1 = a_{11}$, $\partial y_1/\partial x_2 = a_{12}$, etcétera, de tal manera que se podría escribir, en general, $\partial y_i/\partial x_j = a_{ij}$. En vista de esto, los elementos del jacobiano de estas n funciones serán precisamente los elementos de la matriz de coeficientes A, ya dispuestos en el orden correcto. Es decir, se tiene $|J| = |A|$ y, por lo tanto, el criterio jacobiano de dependencia funcional entre $y_1, \ldots, y_n$, —o bien, lo que es lo mismo, dependencia lineal entre los renglones de la matriz de coeficientes A—, equivalente al criterio $|A| = 0$ en el presente caso lineal.

Ya analizamos el jacobiano en el contexto de un sistema de n funciones de n variables. Sin embargo, se debe señalar que el jacobiano de (7.27) se define incluso si cada función de (7.26) contiene más que n variables, por ejemplo, $n + 2$ variables:

$$y_i = f^i(x_1, \ldots, x_n, x_{n+1}, x_{n+2}) \qquad (i = 1, 2, \ldots, n)$$

En tal caso, si se mantiene constante cualquiera de las dos variables (p. ej., x_{n+1} y x_{n+2}), o se tratan como parámetros, de nuevo se tienen n funciones con n variables y se puede formar un jacobiano. Además, si se mantiene constante un par diferente de las variables x, se puede formar un jacobiano diferente. De hecho, esta situación se encuentra en el capítulo 8, junto con la explicación del teorema de función implícita.

EJERCICIO 7.6

1. Use determinantes jacobianos para probar la existencia de dependencia funcional entre los pares de funciones.
   (*a*) $y_1 = 3x_1^2 + x_2$
   $y_2 = 9x_1^4 + 6x_1^2(x_2 + 4) + x_2(x_2 + 8) + 12$
   (*b*) $y_1 = 3x_1^2 + 2x_2^2$
   $y_2 = 5x_1 + 1$
2. Considere a (7.22) como un conjunto de tres funciones $x_i^* = f^i(d_1, d_2, d_3)$ (con $i = 1, 2, 3$).
   (*a*) Escriba el jacobiano de 3×3. ¿Tiene éste alguna relación con (7.23′)? ¿Se puede escribir $|J| = |V|$?
   (*b*) Puesto que $V \equiv (I - A)^{-1}$, se puede concluir que $| \, |V| \neq 0$? ¿Qué se puede inferir de esto acerca de las tres ecuaciones de (7.22)?

Capítulo 8

Análisis estático comparativo de modelos con funciones generales

El estudio de derivadas parciales nos permitió en el capítulo 7 manejar el tipo más sencillo de problemas de estática comparativa, en los cuales la solución de equilibrio del modelo podemos expresarla en forma reducida. En ese caso, la diferenciación parcial de la solución producirá de modo directo la información estática comparativa deseada. Recuerde que la definición de derivada parcial requiere la ausencia de cualquier relación funcional entre variables independientes (por ejemplo, x_i), de tal manera que x_1 pueda variar sin afectar los valores de $x_2, x_3, \ldots, x_n$. Aplicado al análisis estático comparativo, esto significa que los parámetros o variables exógenas, o ambos, que aparecen en la solución de forma reducida deben ser mutuamente independientes. Puesto que éstos se definen de hecho como datos predeterminados para propósitos del modelo, la posibilidad de su mutua afectación está inherentemente descartada. Por lo tanto, el procedimiento de diferenciación parcial que adoptamos en el capítulo 7 es completamente justificable.

Sin embargo, no debemos esperar esta simplicidad cuando, debido a la inclusión de funciones generales en un modelo, no podemos obtener solución explícita en forma reducida. En tales casos, tendremos que hallar las derivadas estáticas comparativas directamente de las ecuaciones originales del modelo. Toma, por ejemplo, un modelo de ingreso nacional simple con dos variables endógenas Y y C.

$$\begin{aligned} Y &= C + I_0 + G_0 \\ C &= C(Y, T_0) \qquad\qquad [T_0\text{: impuestos exógenos}] \end{aligned}$$

la cual podemos reducirla a una sola ecuación (una condición de equilibrio)

$$Y = C(Y, T_0) + I_0 + G_0$$

que resolveremos para Y^*. Sin embargo, como resultado de la forma general de la función C, no está disponible ninguna solución explícita. Por lo tanto, debemos hallar las derivadas estáticas comparativas directamente de esta ecuación. ¿Cómo podríamos enfocar el problema? ¿Qué dificultad especial podríamos encontrar?

Supongamos que existe una solución de equilibrio Y^*. Entonces, bajo ciertas condiciones bastante generales (que analizaremos en la sección 8.5), podríamos tomar Y^* como una función diferenciable de las variables exógenas I_0, G_0 y T_0. Por lo tanto, la ecuación podemos escribirla como

$$Y^* = Y^*(I_0, G_0, T_0)$$

aun cuando no es posible determinar de manera explícita la forma que toma esta función. Además, en alguna vecindad del valor de equilibrio Y^*, se cumplirá la siguiente identidad:

$$Y^* \equiv C(Y^*, T_0) + I_0 + G_0$$

A este tipo de identidad le llamaremos *identidad de equilibrio,* porque es la condición de equilibrio con la variable Y remplazada por su valor de equilibrio Y^*. Ahora que está presente Y^*, a primera vista nos podría parecer que la diferenciación parcial de esta identidad producirá alguna derivada estática comparativa deseada, por ejemplo, $\partial Y^*/\partial T_0$. Éste, desafortunadamente, no es el caso. Puesto que Y^* es una función de T_0, los dos argumentos de la función C *no* son independientes. En particular, en este caso T_0 puede afectar a C no sólo de forma *directa*, sino también *indirectamente* vía Y^*. En consecuencia, la diferenciación parcial ya no es apropiada para los fines que aquí perseguimos. ¿Cómo enfrentamos esta situación?

La respuesta es que debemos recurrir a la *diferenciación total* (en oposición a la diferenciación parcial). Con base en el concepto de *diferenciales totales*, el proceso de diferenciación total puede conducir al concepto relacionado de *derivada total*, lo cual significa la tasa de cambio de una función tal como $C(Y^*, T_0)$ respecto al argumento T_0, cuando T_0 afecta también al otro argumento, Y^*. De esta manera, una vez que estos conceptos son familiares, podremos tratar con funciones cuyos argumentos no son todos independientes, y eso eliminaría el obstáculo que hemos encontrado hasta aquí en el estudio de la estática comparativa de un modelo de función general. Como introducción al estudio de estos conceptos, se debe especificar primero el concepto de *diferenciales*.

8.1 Diferenciales

El símbolo dy/dx, para la derivada de la función $y = f(x)$, lo hemos considerado hasta el momento como una sola identidad. Ahora lo reinterpretaremos como una relación de dos cantidades, dy y dx.

Diferenciales y derivadas

Por definición, la derivada $dy/dx = f'(x)$ es el límite de un cociente de diferencias:

$$\frac{dy}{dx} = f'(x) = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} \tag{8.1}$$

Por sí mismo, $\Delta y/\Delta x$ (sin requerir que $\Delta x \to 0$) no es igual a dy/dx. Si señalamos con δ la discrepancia entre los dos cocientes, podemos escribir

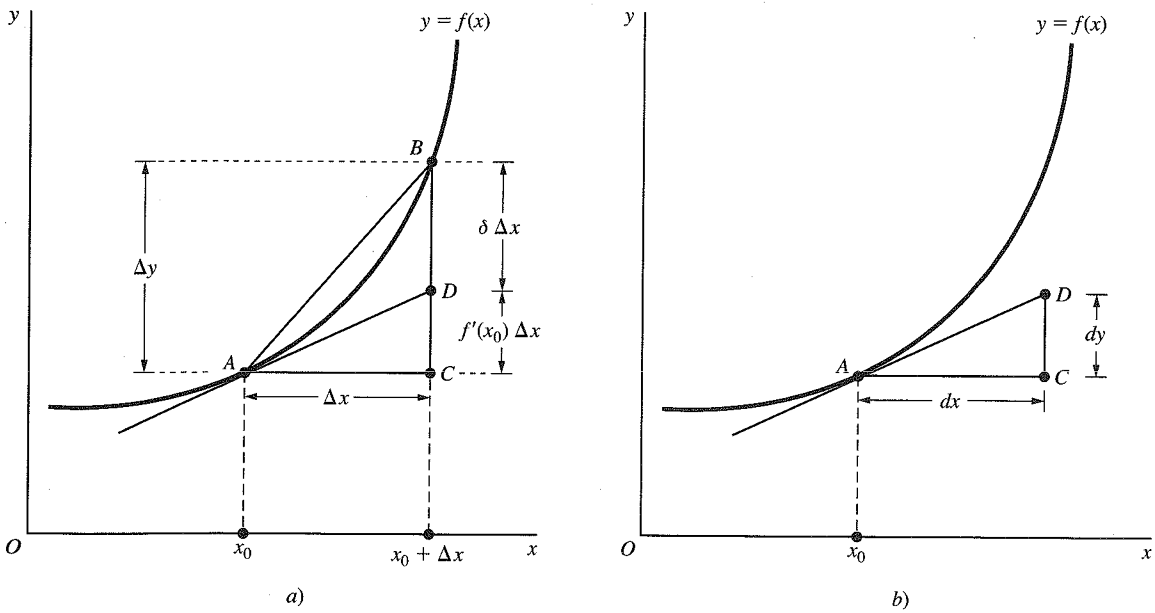
$$\frac{\Delta y}{\Delta x} - \frac{dy}{dx} = \delta \qquad \text{donde} \qquad \delta \to 0 \text{ cuando } \Delta x \to 0 \qquad [\text{por } (8.1)] \tag{8.2}$$

Al multiplicar (8.2) por Δx, y reordenar, obtenemos

$$\Delta y = \frac{dy}{dx}\Delta x + \delta\,\Delta x \quad \text{o bien} \qquad \Delta y = f'(x)\Delta x + \delta\,\Delta x \tag{8.3}$$

Esta ecuación describe el cambio en y (Δy) que resulta de un cambio específico, no necesariamente pequeño, en x (Δx) a partir de algún valor inicial de x en el dominio de la función

FIGURA 8.1



$y = f(x)$. Pero esto también indica que podemos usar, ignorando el término de discrepancia $\delta\, \Delta x$, el término $f'(x)\, \Delta x$ como una aproximación al valor verdadero Δy, donde la aproximación es cada vez mejor a medida que Δx se hace cada vez más pequeña.

En la figura 8.1*a*, cuando x cambia de x_0 a $x_0 + \Delta x$, ocurre un movimiento del punto A al punto B en la gráfica de $y = f(x)$. La Δy verdadera se mide por la distancia *CB*, y la relación de las dos distancias $CB/AC = \Delta y/\Delta x$ se lee de la pendiente del segmento de recta AB. Pero si trazamos una recta tangente AD por el punto A, y usamos AD en lugar de AB para aproximar el valor de Δy, obtenemos la distancia *CD*, que deja la distancia *DB* como la discrepancia o error de aproximación. Puesto que la pendiente de *AD* es $f'(x_0)$, la distancia *CD* es igual a $f'(x_0)\, \Delta x$ y, por (8.3), la distancia *DB* es igual a $\delta\, \Delta x$. Es evidente que, cuando Δx disminuye, el punto B se deslizaría a lo largo de la curva hacia el punto A, reduciendo de este modo la discrepancia y haciendo que $f'(x)$ o dy/dx sea una mejor aproximación para $\Delta y/\Delta x$.

A fin de etiquetar de nuevo las distancias *AC* y *CD* mediante *dx* y *dy*, respectivamente, como en la figura 8.1*b*, centramos ahora la atención en la recta tangente *AD* y tomamos la distancia *CD* como una aproximación a *CB*. Entonces,

$$\frac{dy}{dx} = \text{pendiente de la tangente } AD = f'(x)$$

y, después de multiplicar por *dx*, obtenemos

$$dy = f'(x)\, dx \qquad \textbf{(8.4)}$$

La derivada $f'(x)$ podemos interpretarla entonces como el factor de proporcionalidad entre los dos cambios finitos *dy* y *dx*. En consecuencia, dado un valor específico de *dx*, podemos multiplicarlo por $f'(x)$ para obtener dy como una aproximación a Δy, con el entendido de que mientras más pequeña sea Δx, mejor es la aproximación. Las cantidades dx y dy se llaman *diferenciales* de x y y, respectivamente.

En relación con las diferenciales como entidades matemáticas son pertinentes algunas observaciones. Primero, si bien dx es una variable independiente, dy es una variable dependiente. En particular, dy es una función de x así como de dx: depende de x porque una posición diferente para x_0 en la figura 8.1 significaría una ubicación diferente para el punto A y para su recta tangente; depende de dx porque una magnitud diferente de dx significaría una posición distinta para el punto C como también una distancia diferente CD. Segundo, si $dx = 0$, entonces $dy = 0$, porque el punto B en ese caso coincidiría con el punto A. Pero si $dx \neq 0$, entonces se puede dividir dy entre dx para obtener $f'(x)$, del mismo modo que se puede multiplicar dx por $f'(x)$ para obtener dy. Tercero, la diferencial dy se puede expresar sólo en términos de alguna otra diferencial, en este caso, dx. Esto se debe a que el presente contexto requiere el acoplamiento de un cambio dependiente dy con un cambio independiente dx. Si bien es lógico escribir $dy = f'(x)\,dx$, no tiene sentido quitar el término dx de la derecha y escribir $dy = f'(x)$. El acoplamiento de los dos cambios se lleva a cabo por medio de la derivada $f'(x)$, que se puede considerar como un "convertidor" que sirve para trasladar un determinado cambio dx a un cambio equivalente dy.

El proceso de hallar la diferencial dy de una determinada función $y = f(x)$ se llama *diferenciación*. Recuerde que hemos estado usando este término como sinónimo de derivación, sin haber dado una explicación adecuada. En vista de la interpretación de una derivada dada aquí como un cociente de dos diferenciales, la razón fundamental del término se vuelve evidente por sí misma. Sin embargo, aún es un poco ambiguo usar el término simple "diferenciación" para referirse al proceso de hallar la diferencial dy, así como al de encontrar la derivada dy/dx. Para evitar esta confusión, la práctica común es calificar la palabra *diferenciación* con la frase "respecto a x" cuando se toma la derivada dy/dx.

Diferenciales y elasticidad puntual

Para ilustrar la aplicación económica de las diferenciales, consideremos la noción de elasticidad de una función. Dada una función de demanda $Q = f(P)$, por ejemplo, su elasticidad se define como $(\Delta Q/Q)/(\Delta P/P)$. Si usamos la idea de aproximación explicada en la figura 8.1, podemos reemplazar el cambio independiente ΔP y el cambio dependiente ΔQ con las diferenciales dP y dQ, respectivamente, para obtener la medida de elasticidad de aproximación, conocida como la *elasticidad puntual*, o punto, de la demanda y denotada por ε_d (la letra minúscula griega épsilon, para "elasticidad"):[1]

$$\varepsilon_d \equiv \frac{dQ/Q}{dP/P} = \frac{dQ/dP}{Q/P} \tag{8.5}$$

Observe que en el extremo derecho de la expresión se han dispuesto las diferenciales dQ y dP en una relación dQ/dP, que se puede interpretar como la derivada, o la función *marginal*, de la función de demanda $Q = f(P)$. Puesto que podemos interpretar de manera similar la relación Q/P del denominador como la función *promedio* de la función de demanda, la elasticidad puntual de demanda ε_d en (8.5) se considera como la relación entre la función marginal y la función promedio de la función de demanda.

[1] La medida de elasticidad puntual se puede interpretar alternativamente como el límite de $\dfrac{\Delta Q/Q}{\Delta P/P} = \dfrac{\Delta Q/\Delta P}{Q/P}$ cuando $\Delta P \to 0$, que da el mismo resultado que (8.5).

De hecho, esta relación recién descrita es válida no sólo para la función de demanda, sino también para cualquier otra función, porque para cualquier función *total* $y = f(x)$ podemos escribir la fórmula para la elasticidad puntual de y respecto a x como

$$\varepsilon_{yx} = \frac{dy/dx}{y/x} = \frac{\text{función marginal}}{\text{función promedio}} \tag{8.6}$$

Por convención, el valor *absoluto* de la medida de elasticidad se usa para decidir si la función es elástica en un punto particular. En el caso de una función de demanda, por ejemplo, se estipula:

$$\text{La demanda es} \left\{ \begin{array}{l} \text{elástica} \\ \text{de elasticidad unitaria} \\ \text{inelástica} \end{array} \right\} \text{en un punto cuando } |\varepsilon_d| \gtreqless 1.$$

Ejemplo 1

Encuentre ε_d si la función de demanda es $Q = 100 - 2P$. La función marginal y la función promedio de la demanda dada son

$$\frac{dQ}{dP} = -2 \qquad \text{y} \qquad \frac{Q}{P} = \frac{100 - 2P}{P}$$

de modo que su relación produce

$$\varepsilon_d = \frac{-P}{50 - P}$$

Como está escrita, la elasticidad se muestra como una función de P. Sin embargo, tan pronto como se elige un precio específico, se determina la magnitud de la elasticidad puntual. Cuando $P = 25$, por ejemplo, tenemos $e_d = -1$, o bien, $|e_d| = 1$, de modo que la elasticidad de demanda es unitaria en ese punto. Cuando $P = 30$, en contraste, tenemos $|e_d| = 1.5$; por consiguiente, la demanda es elástica en ese precio. En términos generales, podemos comprobar que tenemos $|e_d| > 1$ para $25 < P < 50$ y $|e_d| < 1$ para $0 < P < 25$ en el presente ejemplo. (¿Podemos considerar significativo un precio $P > 50$ en este caso?)

Ejemplo 2

Encuentre la elasticidad puntual de la oferta ε_s a partir de la función $Q = P^2 + 7P$, y determine si la oferta es elástica en $P = 2$. Puesto que las funciones marginal y promedio son, respectivamente,

$$\frac{dQ}{dP} = 2P + 7 \qquad \text{y} \qquad \frac{Q}{P} = P + 7$$

su relación proporciona la elasticidad de la oferta

$$\varepsilon_s = \frac{2P + 7}{P + 7}$$

Cuando $P = 2$, esta elasticidad tiene el valor $11/9 > 1$; por lo tanto, la oferta es elástica en $P = 2$.

A riesgo de apartarse un poco del tema, podemos agregar también que la interpretación de la relación de dos diferenciales como una derivada, y la transformación resultante de la fórmula de elasticidad de una función en una relación de su función marginal a su función promedio, posibilita una forma rápida de determinar la elasticidad puntual de forma gráfica. Los dos diagramas de la figura 8.2 ilustran los casos de una curva con pendiente negativa y una con pendiente positiva, respectivamente. En cada caso, el valor de la función marginal en el punto A sobre la curva, o en $x = x_0$ en el dominio, se mide por la pendiente de la recta tangente AB. Por

FIGURA 8.2

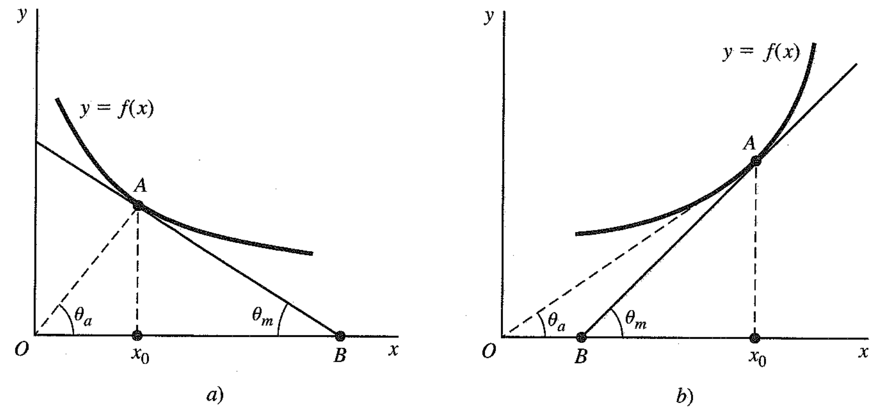
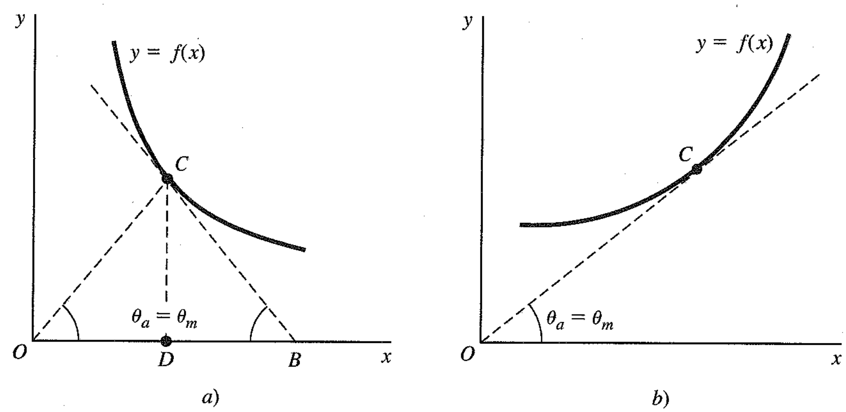


FIGURA 8.3



otro lado, el valor de la función promedio se mide en cada caso por la pendiente de la recta OA (la recta que une al origen con el punto A sobre la curva, como un radio vector), porque en el punto A tenemos $y = x_0A$ y $x = Ox_0$, de modo que el promedio es $y/x = x_0A/Ox_0$ = pendiente de OA. La elasticidad en el punto A se determina fácilmente al comparar los valores *numéricos* de las dos pendientes en cuestión: si AB es más inclinada que OA, la función es elástica en el punto A; en el caso opuesto, es inelástica en A. En consecuencia, la función ilustrada en la figura 8.2*a* es inelástica en A (o en $x = x_0$), mientras que la de la figura 8.2*b* es elástica en A.

Además, las dos pendientes en comparación dependen de forma directa de los tamaños respectivos de los dos ángulos θ_m y θ_a (los subíndices m y a de la letra griega theta indican marginal y promedio, respectivamente). Así, opcionalmente, podemos comparar estos dos ángulos en lugar de las dos pendientes correspondientes. En relación con la figura 8.2, vemos que $\theta_m < \theta_a$ en el punto A del diagrama a, lo cual indica que la función marginal no alcanza el promedio en valor numérico; por lo tanto, la función es inelástica en el punto A. En la figura 8.2*b* se cumple lo contrario.

A veces se tiene interés en localizar un punto de elasticidad unitaria en una curva determinada. Esto podemos realizarlo ahora sin dificultad. Si la curva es de pendiente negativa, como en la figura 8.3*a*, debemos hallar un punto C tal que la recta OC y la tangente BC formen un ángulo del mismo tamaño con el eje x, aunque en dirección opuesta. En el caso de una curva con pendiente positiva, como en la figura 8.3*b*, sólo tenemos que hallar un punto C tal que la recta tangente en C, cuando se extiende de manera adecuada, pase por el origen.

Se advierte que el método gráfico recién descrito se basa en la suposición de que la función $y = f(x)$ se grafica con la variable dependiente y en el eje vertical. En particular, al aplicar el método a una curva de demanda, debemos estar seguros de que Q está sobre el eje vertical. (Ahora, suponga que Q se grafica en realidad sobre el eje horizontal. ¿Cómo debemos modificar el método de leer la elasticidad puntual?)

EJERCICIO 8.1

1. Determine la diferencial dy, dada:
   (*a*) $y = -x(x^2 + 3)$ (*b*) $y = (x - 8)(7x + 5)$ (*c*) $y = \dfrac{x}{x^2 + 1}$
2. Dada la función de importación $M = f(Y)$, donde M es importaciones y Y es ingreso nacional, exprese la elasticidad del ingreso de importaciones ε_{MY} en términos de las propensiones a importar.
3. Dada la función de consumo $C = a + bY$ (con $a > 0$; $0 < b < 1$):
   (*a*) Encuentre su función marginal y su función promedio.
   (*b*) Halle la elasticidad del ingreso de consumo ε_{CY}, y determine su signo, suponiendo que $Y > 0$.
   (*c*) Demuestre que esta función de consumo es inelástica en todos los niveles de ingreso positivo.
4. Encuentre la elasticidad puntual de demanda, dada $Q = k/P^n$, donde k y n son constantes positivas.
   (*a*) En este caso, ¿la elasticidad depende del precio?
   (*b*) En el caso especial donde $n = 1$, ¿cuál es la forma de la curva de demanda? ¿Cuál es la elasticidad puntual de demanda?
5. (*a*) Halle una curva de pendiente positiva con una elasticidad puntual constante en cualquier parte sobre la curva.
   (*b*) Escriba la ecuación de la curva y compruebe mediante (8.6) que la elasticidad es de hecho una constante.
6. Dada $Q = 100 - 2P + 0.02Y$, donde Q es la cantidad demandada, P es el precio y Y el ingreso, y dada $P = 20$ y $Y = 5\,000$, determine:
   (*a*) la elasticidad de demanda en los precios.
   (*b*) la elasticidad de demanda en el ingreso.

8.2 Diferenciales totales

El concepto de diferenciales podemos ampliarlo fácilmente a una función de dos o más variables independientes. Considere una función de ahorro

$$S = S(Y, i) \tag{8.7}$$

donde S son los ahorros, Y es el ingreso nacional e i es la tasa de interés. Se supone que esta función, como se supondrá para todas las funciones que usaremos aquí, es continua y posee derivadas (parciales) continuas, o en símbolos, $f \in C'$. La derivada parcial $\partial S/\partial Y$ mide la propensión marginal a ahorrar. Así, para cualquier cambio en Y, dY, el cambio resultante en S se puede aproximar por medio de la cantidad $(\partial S/\partial Y)\, dY$, que es comparable a la expresión del lado derecho de (8.4). De manera similar, dado un cambio en i, di, se puede tomar $(\partial S/\partial i)\, di$

como la aproximación al cambio resultante en S. El cambio total en S se aproxima, entonces, mediante la diferencial

$$dS = \frac{\partial S}{\partial Y}\, dY + \frac{\partial S}{\partial i}\, di \tag{8.8}$$

o bien, en otra notación,

$$dS = S_Y\, dY + S_i\, di$$

Note que las dos derivadas parciales S_Y y S_i desempeñan de nuevo el papel de "convertidores", que sirven para convertir los cambios dY y di, respectivamente, en un cambio correspondiente dS. La expresión dS, que es la *suma* de los cambios aproximados de ambas fuentes, se llama la *diferencial total* de la función de ahorro. Y el proceso de hallar tal diferencial total se denomina *diferenciación total*. En contraste, los dos términos de la derecha del signo igual en (8.8) se llaman *diferenciales parciales* de la función de ahorro.

Es posible, por supuesto, que Y cambie mientras i permanece constante. En ese caso, $di = 0$, y la diferencial total se reducirá a $dS = (\partial S/\partial Y)\, dY$. Dividiendo ambos lados entre dY, obtenemos

$$\frac{\partial S}{\partial Y} = \left(\frac{dS}{dY}\right)_{i \text{ constante}}$$

Es claro que la derivada parcial $\partial S/\partial Y$ podemos interpretarla también, de la misma manera que la figura 8.1b, como la razón entre dos diferenciales dS y dY, con la condición de que i, la otra variable independiente de la función, se mantenga constante. De manera análoga, podemos interpretar la derivada parcial $\partial S/\partial i$ como la relación de la diferencial dS (con Y constante) a la diferencial di. Note que aunque ahora dS y di pueden aparecer cada una como una diferencial, la expresión $\partial S/\partial i$ permanece como una sola entidad.

El caso más general de una función de n variables independientes podemos ejemplificarla, por ejemplo, mediante una función de utilidad en la forma general

$$U = U(x_1, x_2, \ldots, x_n) \tag{8.9}$$

La diferencial total de esta función podemos escribirla como

$$dU = \frac{\partial U}{\partial x_1}\, dx_1 + \frac{\partial U}{\partial x_2}\, dx_2 + \cdots + \frac{\partial U}{\partial x_n}\, dx_n \tag{8.10}$$

o bien, $$dU = U_1\, dx_1 + U_2\, dx_2 + \cdots + U_n\, dx_n = \sum_{i=1}^{n} U_i\, dx_i$$

en la cual cada término de la derecha indica el cambio aproximado en U, que resulta de un cambio en una de las variables independientes. Desde el punto de vista económico, el primer término, $U_1\, dx_1$, significa la utilidad marginal del primer artículo multiplicada por el incremento en el consumo de ese artículo, y de manera similar para los otros términos. La suma de todos los términos es dU y representa el cambio aproximado total de utilidad proveniente de todas las fuentes posibles de cambio. Como muestra el razonamiento de (8.3), dU es una aproximación y tiende hacia el cambio verdadero ΔU cuando todos los términos dx_i tienden a cero.

Igual que con cualquier otra función, podemos esperar que tanto la función de ahorro (8.7) como la función de utilidad (8.9) den lugar a medidas de elasticidad puntual similares a la definida en (8.6). En estos casos, sin embargo, cada medida de elasticidad debemos definirla sólo en términos del cambio en *una* de las variables independientes; por lo tanto, habrá *dos* medidas de elasticidad para la función de ahorro, y n de ellas para la función de utilidad. En consecuencia, éstas se denominan *elasticidades parciales*. Para la función de ahorro, las elasticidades parciales podemos escribirlas como

$$\varepsilon_{SY} = \frac{\partial S/\partial Y}{S/Y} = \frac{\partial S}{\partial Y}\frac{Y}{S} \qquad \text{y} \qquad \varepsilon_{Si} = \frac{\partial S/\partial i}{S/i} = \frac{\partial S}{\partial i}\frac{i}{S}$$

Para la función de utilidad, las n elasticidades parciales podemos expresarlas de forma concisa como sigue:

$$\varepsilon_{Ux_i} = \frac{\partial U}{\partial x_i}\frac{x_i}{U} \qquad (i = 1, 2, \ldots, n)$$

Ejemplo 1 Encuentre la diferencial total para las siguientes funciones de utilidad, donde $a, b > 0$:

(*a*) $U(x_1, x_2) = ax_1 + bx_2$
(*b*) $U(x_1, x_2) = x_1^2 + x_2^3 + x_1x_2$
(*c*) $U(x_1, x_2) = x_1^a x_2^b$

Las diferenciales totales son las siguientes:

(*a*)
$$\frac{\partial U}{\partial x_1} = U_1 = a \qquad \frac{\partial U}{\partial x_2} = U_2 = b$$
y
$$dU = U_1\,dx_1 + U_2\,dx_2 = a\,dx_1 + b\,dx_2$$

(*b*)
$$\frac{\partial U}{\partial x_1} = U_1 = 2x_1 + x_2 \qquad \frac{\partial U}{\partial x_2} = U_2 = 3x_2^2 + x_1$$
y
$$dU = U_1\,dx_1 + U_2\,dx_2 = (2x_1 + x_2)\,dx_1 + \left(3x_2^2 + x_1\right)dx_2$$

(*c*)
$$\frac{\partial U}{\partial x_1} = U_1 = ax_1^{a-1}x_2^b = \frac{ax_1^a x_2^b}{x_1} \qquad \frac{\partial U}{\partial x_2} = U_2 = bx_1^a x_2^{b-1} = \frac{bx_1^a x_2^b}{x_2}$$
y
$$dU = \left(\frac{ax_1^a x_2^b}{x_1}\right)dx_1 + \left(\frac{bx_1^a x_2^b}{x_2}\right)dx_2$$

EJERCICIO 8.2

1. Exprese la diferencial total dU por medio del vector gradiente ∇U.
2. Encuentre la diferencial total, dada
   (*a*) $z = 3x^2 + xy - 2y^3$
   (*b*) $U = 2x_1 + 9x_1x_2 + x_2^2$
3. Determine la diferencial total, dada
   (*a*) $y = \dfrac{x_1}{x_1 + x_2}$ (*b*) $y = \dfrac{2x_1x_2}{x_1 + x_2}$
4. La función de oferta de cierto artículo es
   $Q = a + bP^2 + R^{1/2} \qquad (a < 0,\ b > 0) \qquad [R\text{: lluvia}]$

Determine la elasticidad de la oferta con relación a los precios ε_{QP}, y la elasticidad de la oferta con relación a la lluvia ε_{QR}.

5. ¿Cómo varían las dos elasticidades parciales del problema 4 respecto a P y R? En un modo estrictamente monótono (suponiendo que P y R son positivas)?
6. La demanda exterior para nuestras exportaciones X depende del ingreso exterior Y_f y de nuestro nivel de precios P: $X = Y_f^{1/2} + P^{-2}$. Encuentre la elasticidad parcial de la demanda exterior para nuestras exportaciones respecto a nuestro nivel de precios.
7. Determine la diferencial total para cada una de las siguientes funciones:
   (*a*) $U = -5x^3 - 12xy - 6y^5$
   (*b*) $U = 7x^2y^3$
   (*c*) $U = 3x^3(8x - 7y)$
   (*d*) $U = (5x^2 + 7y)(2x - 4y^3)$
   (*e*) $U = \dfrac{9y^3}{x - y}$
   (*f*) $U = (x - 3y)^3$

8.3 Reglas de diferenciales

Una forma directa de hallar la diferencial total dy, dada una función

$$y = f(x_1, x_2)$$

es determinar las derivadas parciales f_1 y f_2 y sustituirlas en la ecuación

$$dy = f_1\,dx_1 + f_2\,dx_2$$

A veces, sin embargo, podría ser más conveniente aplicar ciertas reglas de diferenciales que, en vista de su sorprendente parecido con las fórmulas de derivadas estudiadas antes, son muy fáciles de recordar.

Sea k una constante y u y v dos funciones de las variables x_1 y x_2. Entonces, las siguientes reglas son válidas: [2]

Regla I	$dk = 0$	(véase regla de la función constante)
Regla II	$d(cu^n) = cnu^{n-1}\,du$	(véase regla de la función de potencia)
Regla III	$d(u \pm v) = du \pm dv$	(véase regla de la suma-diferencia)
Regla IV	$d(uv) = v\,du + u\,dv$	(véase regla del producto)
Regla V	$d\left(\dfrac{u}{v}\right) = \dfrac{1}{v^2}(v\,du - u\,dv)$	(véase regla del cociente)

En lugar de demostrar aquí estas reglas, solamente se ilustra su aplicación práctica.

[2] Todas las reglas de diferenciales descritas en esta sección son aplicables también cuando u y v son por sí mismas variables independientes (en vez de funciones de algunas otras variables x_1 y x_2).

Ejemplo 1 Encuentre la diferencial total dy de la función

$$y = 5x_1^2 + 3x_2$$

El método directo requiere la evaluación de las derivadas parciales $f_1 = 10x_1$ y $f_2 = 3$, las cuales permitirán escribir

$$dy = f_1\, dx_1 + f_2\, dx_2 = 10x_1\, dx_1 + 3\, dx_2$$

Sin embargo, se podría permitir que $u = 5x_1^2$ y $v = 3x_2$ y aplicar las reglas anteriores para obtener la respuesta idéntica como sigue:

$$\begin{aligned} dy &= d\left(5x_1^2\right) + d(3x_2) && \text{[por la regla III]} \\ &= 10x_1\, dx_1 + 3\, dx_2 && \text{[por la regla II]} \end{aligned}$$

Ejemplo 2 Encuentre la diferencial total de la función

$$y = 3x_1^2 + x_1 x_2^2$$

Puesto que $f_1 = 6x_1 + x_2^2$ y $f_2 = 2x_1 x_2$, la diferencial deseada es

$$dy = \left(6x_1 + x_2^2\right) dx_1 + 2x_1 x_2\, dx_2$$

Al aplicar las reglas dadas, podemos llegar al mismo resultado:

$$\begin{aligned} dy &= d\left(3x_1^2\right) + d\left(x_1 x_2^2\right) && \text{[por la regla III]} \\ &= 6x_1\, dx_1 + x_2^2\, dx_1 + x_1\, d\left(x_2^2\right) && \text{[por las reglas II y IV]} \\ &= \left(6x_1 + x_2^2\right) dx_1 + 2x_1 x_2\, dx_2 && \text{[por la regla II]} \end{aligned}$$

Ejemplo 3 Encuentre la diferencial total de la función

$$y = \frac{x_1 + x_2}{2x_1^2}$$

En vista del hecho de que las derivadas parciales en este caso son

$$f_1 = \frac{-(x_1 + 2x_2)}{2x_1^3} \qquad \text{y} \qquad f_2 = \frac{1}{2x_1^2}$$

(compruébelas a modo de un ejercicio), la diferencial deseada es

$$dy = \frac{-(x_1 + 2x_2)}{2x_1^3} dx_1 + \frac{1}{2x_1^2} dx_2$$

Sin embargo, podríamos obtener el mismo resultado al aplicar las reglas como sigue:

$$\begin{aligned} dy &= \frac{1}{4x_1^4}\left[2x_1^2 d(x_1 + x_2) - (x_1 + x_2)\, d\left(2x_1^2\right)\right] && \text{[por la regla V]} \\ &= \frac{1}{4x_1^4}\left[2x_1^2(dx_1 + dx_2) - (x_1 + x_2)4x_1\, dx_1\right] && \text{[por las reglas III y II]} \\ &= \frac{1}{4x_1^4}\left[-2x_1(x_1 + 2x_2)\, dx_1 + 2x_1^2\, dx_2\right] \\ &= \frac{-(x_1 + 2x_2)}{2x_1^3} dx_1 + \frac{1}{2x_1^2} dx_2 \end{aligned}$$

Estas reglas podemos ampliarlas a casos donde intervienen más de dos funciones de x_1 y x_2. En particular, a la colección previa podemos añadir las dos reglas siguientes:

Regla VI $\quad d(u \pm v \pm w) = du \pm dv \pm dw$

Regla VII $\quad d(uvw) = vwdu + uwdv + uv\,dw$

Para derivar la regla VII, podemos emplear el truco conocido de permitir que $z = vw$, de modo que

$$d(uvw) = d(uz) = z\,du + u\,dz \qquad \text{[por la regla IV]}$$

Luego, aplicando la regla IV de nuevo a dz, obtenemos el resultado intermedio

$$dz = d(vw) = w\,dv + v\,dw$$

que, cuando lo sustituimos en la ecuación anterior, produce

$$d(uvw) = vw\,du + u(w\,dv + v\,dw) = vw\,du + uw\,dv + uv\,dw$$

como el resultado final deseado. Para deducir la regla VI podemos emplear un procedimiento similar.

EJERCICIO 8.3

1. Use las reglas de diferenciales para hallar (*a*) dz a partir de $z = 3x^2 + xy - 2y^3$ y (*b*) dU a partir de $U = 2x_1 + 9x_1x_2 + x_2^2$. Compruebe las respuestas contra las obtenidas en el ejercicio 8.2-2.
2. Use las reglas de diferenciales para hallar dy de las siguientes funciones:

   (*a*) $y = \dfrac{x_1}{x_1 + x_2}$ $\qquad$ (*b*) $y = \dfrac{2x_1x_2}{x_1 + x_2}$

   Compruebe las respuestas contra las obtenidas en el ejercicio 8.2-3.
3. Dada $y = 3x_1(2x_2 - 1)(x_3 + 5)$
   (*a*) Determine dy por la regla VII.
   (*b*) Encuentre la diferencial de y, si $dx_2 = dx_3 = 0$.
4. Pruebe las reglas II, III, IV y V, suponiendo que u y v son las variables independientes (en vez de funciones de algunas otras variables).

8.4 Derivadas totales

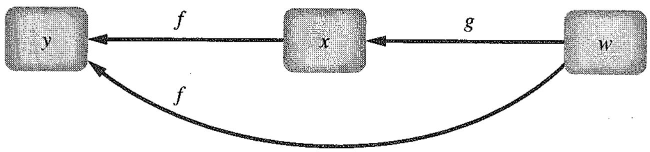
Ahora abordamos la pregunta planteada al comienzo del capítulo: ¿cómo podemos hallar la tasa de cambio de la función $C(Y^*, T_0)$ respecto a T_0, cuando Y^* y T_0 están relacionadas? Como se mencionó, la respuesta radica en el concepto de derivada total. A diferencia de una derivada *parcial*, una derivada *total* no requiere que el argumento Y^* permanezca constante cuando T_0 varía y, por lo tanto, puede tomar en consideración la relación postulada entre los dos argumentos.

Determinación de la derivada total

Para continuar la explicación en el marco general, consideremos cualquier función

$$y = f(x, w) \qquad \text{donde} \qquad x = g(w) \tag{8.11}$$

FIGURA 8.4



Las dos funciones f y g podemos combinarlas en una función compuesta

$$y = f[g(w), w] \tag{8.11'}$$

Las tres variables y, x y w se relacionan entre sí como se ilustra en la figura 8.4. En esta figura, que se denominará *mapa de canal*, vemos con claridad que w, la última fuente de cambio, puede afectar a y a través de dos canales separados: (1) *indirectamente*, vía la función g y f (las flechas rectas) y (2) *directamente*, vía la función f (la flecha curva). El efecto directo se puede representar simplemente por la derivada parcial f_w. Pero el efecto indirecto sólo se puede expresar mediante un producto de dos derivadas, $f_x \dfrac{dx}{dw}$, o bien $\dfrac{\partial y}{\partial x}\dfrac{dx}{dw}$, por la regla de la cadena para una función compuesta. Al sumar los dos efectos obtenemos la derivada total deseada de y respecto a w:

$$\begin{aligned}\frac{dy}{dw} &= f_x \frac{dx}{dw} + f_w \\ &= \frac{\partial y}{\partial x}\frac{dx}{dw} + \frac{\partial y}{\partial w}\end{aligned} \tag{8.12}$$

Esta derivada total podemos obtenerla también mediante otro método: podemos diferenciar primero por completo la función $y = f(x, w)$ para obtener la diferencial total

$$dy = f_x\, dx + f_w\, dw$$

y luego dividir toda la ecuación entre dw. El resultado es idéntico a (8.12). De cualquier modo, el proceso de hallar la derivada total dy/dw se conoce como *diferenciación total de* y *respecto a* w.

Es muy importante distinguir entre los dos símbolos parecidos dy/dw y $\partial y/\partial w$ de (8.12). El primero es la derivada *total*, y el último, una derivada *parcial*. De hecho, el último es solamente un componente del primero.

Ejemplo 1

Encuentre la derivada total dy/dw, dada la función

$$y = f(x, w) = 3x - w^2 \qquad \text{donde} \qquad x = g(w) = 2w^2 + w + 4$$

En virtud de (8.12), la derivada total debe ser

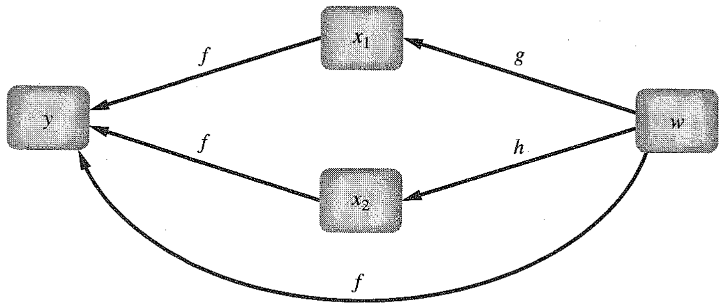
$$\frac{dy}{dw} = 3(4w + 1) + (-2w) = 10w + 3$$

Como comprobación, podemos sustituir la función g en la función f, para obtener

$$y = 3(2w^2 + w + 4) - w^2 = 5w^2 + 3w + 12$$

que ahora es sólo una función de w. La derivada de dy/dw se determina sin dificultad como $10w + 3$, la respuesta idéntica.

FIGURA 8.5



Ejemplo 2

Si tenemos una función de utilidad $U = U(c, s)$, donde c es la cantidad de café consumido y s es la cantidad de azúcar consumida, y otra función $s = g(c)$ que indica la complementariedad entre estos dos artículos, entonces podemos escribir la función compuesta

$$U = U[c, g(c)]$$

de la cual se deduce que

$$\frac{dU}{dc} = \frac{\partial U}{\partial c} + \frac{\partial U}{\partial g(c)} g'(c)$$

Una variación sobre el mismo tema

La situación es sólo ligeramente más complicada cuando tenemos

$$y = f(x_1, x_2, w) \qquad \text{donde} \begin{cases} x_1 = g(w) \\ x_2 = h(w) \end{cases} \tag{8.13}$$

El mapa de canales aparecerá ahora como en la figura 8.5. Esta vez, la variable w puede afectar a y por tres canales: (1) de modo indirecto, vía la función g y luego f; (2) otra vez de manera indirecta, vía la función h y luego f, y (3) de forma directa vía f. De la experiencia previa, se espera que estos tres efectos sean expresables, respectivamente, como $\frac{\partial y}{\partial x_1}\frac{dx_1}{dw}$, $\frac{\partial y}{\partial x_2}\frac{dx_2}{dw}$, y $\frac{\partial y}{\partial w}$. Al agruparlos, obtenemos la derivada total

$$\begin{aligned} \frac{dy}{dw} &= \frac{\partial y}{\partial x_1}\frac{dx_1}{dw} + \frac{\partial y}{\partial x_2}\frac{dx_2}{dw} + \frac{\partial y}{\partial w} \\ &= f_1 \frac{dx_1}{dw} + f_2 \frac{dx_2}{dw} + f_w \end{aligned} \tag{8.14}$$

que es comparable con (8.12). Si tomamos la diferencial total dy, y luego la dividimos entre dw, llegamos al mismo resultado.

Ejemplo 3

Sea la función de producción

$$Q = Q(K, L, t)$$

donde, aparte de los dos insumos K y L, hay un tercer argumento, que denota tiempo. La presencia del argumento t indica que la función de producción puede cambiar con el tiempo en respuesta a cambios tecnológicos. Así, ésta es una función de producción dinámica en vez de estática. Puesto que el capital y la mano de obra pueden cambiar también con el tiempo, podemos escribir

$$K = K(t) \qquad \text{y} \qquad L = L(t)$$

Entonces, la tasa de cambio de la producción respecto al tiempo podemos expresarla, de acuerdo con la fórmula de derivada total (8.14), como

$$\frac{dQ}{dt} = \frac{\partial Q}{\partial K}\frac{dK}{dt} + \frac{\partial Q}{\partial L}\frac{dL}{dt} + \frac{\partial Q}{\partial t}$$

o bien, con otra notación,

$$\frac{dQ}{dt} = Q_K K'(t) + Q_L L'(t) + Q_t$$

Otra variación sobre el mismo tema

Cuando reemplazamos la última fuente de cambio, w en (8.13), por dos fuentes coexistentes, u y v, la situación es la siguiente:

$$y = f(x_1, x_2, u, v) \qquad \text{donde} \begin{cases} x_1 = g(u, v) \\ x_2 = h(u, v) \end{cases} \tag{8.15}$$

Si bien el mapa de canal contiene ahora más flechas, el principio de su construcción es el mismo; por lo tanto, el trazo se deja al lector. Para hallar la derivada total de y respecto a u (mientras se mantiene constante v), tomamos la diferencial total de y, y luego la dividimos entre la diferencial du, con el resultado:

$$\begin{aligned}\frac{dy}{du} &= \frac{\partial y}{\partial x_1}\frac{dx_1}{du} + \frac{\partial y}{\partial x_2}\frac{dx_2}{du} + \frac{\partial y}{\partial u}\frac{du}{du} + \frac{\partial y}{\partial v}\frac{dv}{du} \\ &= \frac{\partial y}{\partial x_1}\frac{dx_1}{du} + \frac{\partial y}{\partial x_2}\frac{dx_2}{du} + \frac{\partial y}{\partial u} \qquad \left[\frac{dv}{du} = 0 \quad \text{puesto que } v \text{ se mantiene constante}\right]\end{aligned}$$

En vista del hecho de que u varía mientras v se mantiene constante (ya que una sola derivada no puede manejar cambios tanto en u como en v), no obstante, el resultado debe ser modificado en dos formas: (1) las derivadas dx_1/du y dx_2/du de la derecha se deben reescribir con el signo de derivada parcial como $\partial x_1/\partial u$ y $\partial x_2/\partial u$, que concuerda con las funciones g y h en (8.15); y (2) la relación dy/du de la izquierda se debe interpretar también como una derivada *parcial*, aunque, se obtiene por el proceso de diferenciación total de y, en realidad es de la naturaleza de una derivada *total*. Por esta razón, nos referiremos a ella con el nombre explícito de *derivada parcial total*, y la denotaremos mediante $\S y/\S u$ (con § en lugar de ∂) a fin de distinguirla de la derivada parcial simple $\partial y/\partial u$ que, como observamos en el resultado, no es sino uno de los tres términos componentes que dan como resultado la derivada parcial total.[3]

Con estas modificaciones, el resultado es

$$\frac{\S y}{\S u} = \frac{\partial y}{\partial x_1}\frac{\partial x_1}{\partial u} + \frac{\partial y}{\partial x_2}\frac{\partial x_2}{\partial u} + \frac{\partial y}{\partial u} \tag{8.16}$$

que es comparable con (8.14). Note la presencia del símbolo $\partial y/\partial u$ a la derecha, que necesita la adopción del nuevo símbolo $\S y/\S u$ a la izquierda para indicar el concepto más amplio de

[3] Otra forma de denotar esta derivada parcial total es

$$\left.\frac{dy}{du}\right|_{v \text{ (v constante}} \quad \text{o bien} \quad \left.\frac{dy}{du}\right|_{dv=0}$$

derivada parcial total. De una manera completamente análoga, podemos deducir la otra derivada parcial total, $\S y/\S v$. Sin embargo, en vista de que los papeles de u y v son simétricos en (8.15), tenemos una alternativa más simple. Todo lo que debemos hacer para obtener $\S y/\S v$ es reemplazar el símbolo u en (8.16) por el símbolo v.

El uso de los nuevos símbolos $\S y/\S u$ y $\S y/\S v$ para las derivadas parciales totales, si bien poco convencional, contribuye al buen propósito de evitar confusión con las derivadas parciales simples $\partial y/\partial u$ y $\partial y/\partial v$ que pueden surgir de la función *f* sola en (8.15). No obstante, en el caso especial donde la función *f* toma la forma de $y = f(x_1, x_2)$, sin los argumentos u y v, las derivadas parciales simples $\partial y/\partial u$ y $\partial y/\partial v$ no están definidas. Por consiguiente, en tal caso podría ser apropiado usar los últimos símbolos para las derivadas parciales totales de y respecto a u y v, puesto que es probable que no surja ninguna confusión; aunque, incluso en ese caso, se recomienda el uso de un símbolo especial para mayor claridad.

Algunas observaciones generales

Para concluir esta sección, ofrecemos tres observaciones generales en relación con la derivada total y la diferenciación total:

1. En los casos analizados, la situación tiene que ver, sin excepción, con una variable que es funcionalmente dependiente de una segunda variable, que a su vez tiene dependencia funcional en una tercera variable. Como consecuencia, de forma inevitable entra en acción el concepto de *cadena*, como lo muestra la presencia de un producto (o productos), de dos expresiones de derivadas como el(los) término(s) de una derivada total. Por esta razón, las fórmulas de derivada total de (8.12), (8.14) y (8.16) podemos considerarlas también como expresiones de la regla de la cadena, o la regla de la función compuesta, una versión más compleja de la regla de la cadena presentada en la sección 7.3.
2. La cadena de derivadas no tiene que limitarse a sólo dos "eslabones" (dos derivadas que se multiplican); el concepto de derivada total se debe poder ampliar a casos donde hay tres o más eslabones en la función compuesta.
3. En todos los casos explicados, las derivadas totales, incluso las llamadas *derivadas parciales totales*, miden las tasas de cambio respecto a algunas variables *últimas* en la cadena o, en otras palabras, respecto a ciertas variables que son en un sentido *exógenas* y que *no* se expresan como funciones algunas otras variables. La esencia de la derivada total y el proceso de diferenciación total es tomar en cuenta *todos* los canales, tanto los indirectos como los directos, mediante los cuales sea posible tomar en cuenta todos los efectos del cambio en una variable independiente *última* a la variable dependiente particular bajo estudio.

EJERCICIO 8.4

1. Encuentre la derivada total dz/dy, a partir de
   (*a*) $z = f(x, y) = 5x + xy - y^2$, donde $x = g(y) = 3y^2$
   (*b*) $z = 4x^2 - 3xy + 2y^2$, donde $x = 1/y$
   (*c*) $z = (x + y)(x - 2y)$, donde $x = 2 - 7y$
2. Determine la derivada total dz/dt, a partir de
   (*a*) $z = x^2 - 8xy - y^3$, donde $x = 3t$ y $y = 1 - t$
   (*b*) $z = 7u + vt$, donde $u = 2t^2$ y $v = t + 1$

(*c*) $z = f(x, y, t)$, donde $x = a + bt$ y $y = c + kt$

3. Halle la tasa de cambio de producción respecto al tiempo, si la función de producción es $Q = A(t)K^{\alpha}L^{\beta}$, donde $A(t)$ es una función creciente de t, y $K = K_0 + at$, y $L = L_0 + bt$.
4. Obtenga las derivadas parciales totales $\S W/\S u$ y $\S W/\S v$ si
   (*a*) $W = ax^2 + bxy + cu$, donde $x = \alpha u + \beta v$ y $y = \gamma u$
   (*b*) $W = f(x_1, x_2)$, donde $x_1 = 5u^2 + 3v$ y $x_2 = u - 4v^3$
5. Dibuje el mapa de canal apropiado para el caso de (8.15).
6. Deduzca formalmente la expresión para $\S y/\S v$ a partir de (8.15) tomando la diferencial total de y y dividiendo después toda la expresión entre dv.

8.5 Derivadas de funciones implícitas

El concepto de diferenciales totales puede permitir también hallar las derivadas de las denominadas funciones implícitas.

Funciones implícitas

Una función dada en la forma $y = f(x)$, por ejemplo,

$$y = f(x) = 3x^4 \tag{8.17}$$

se llama *función explícita*, porque la variable y se expresa explícitamente como una función de x. Sin embargo, si escribimos esta función en la forma equivalente

$$y - 3x^4 = 0 \tag{8.17$'$}$$

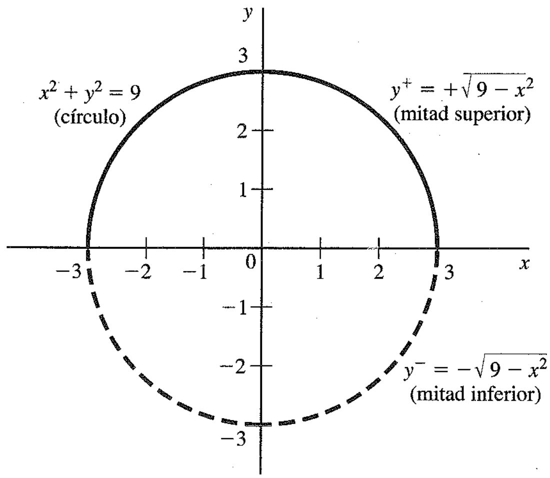
ya no tenemos una función explícita. Entonces, la función (8.17) solamente está definida en forma *implícita* por la ecuación (8.17′). Por lo tanto, cuando sólo tenemos una ecuación en la forma (8.17′), la función $y = f(x)$ implicada , y cuya forma específica no siempre es posible conocer, se denomina *función implícita*.

En general, una ecuación en la forma de (8.17′) se denota mediante $F(y, x) = 0$, porque su lado izquierdo es una función de las dos variables y y x. Note que estamos usando la letra mayúscula F para distinguirla de la función f; la función F, que representa la expresión del lado izquierdo en (8.17′), tiene dos argumentos, y y x, mientras que la función f, que representa la función implícita, tiene sólo un argumento, x. Por supuesto que puede haber más de dos argumentos en la función F. Por ejemplo, podríamos encontrar una ecuación $F(y, x_1, \ldots, x_m) = 0$. Tal ecuación *podría* definir también una función implícita $y = f(x_1, \ldots, x_m)$.

En el último enunciado, la palabra ambigua "*podría*" la usamos de manera intencional. Porque, mientras que una función explícita, por ejemplo, $y = f(x)$, se puede transformar siempre en una ecuación $F(y, x) = 0$ al pasar la expresión $f(x)$ del lado izquierdo del signo igual, la transformación inversa no siempre es posible. De hecho, en ciertos casos, es posible que una ecuación dada en la forma de $F(y, x) = 0$ no defina de forma explícita una función $y = f(x)$. Por ejemplo, la ecuación $x^2 + y^2 = 0$ se satisface sólo en el punto de origen (0, 0) y, por lo tanto, no produce una función representativa . Como otro ejemplo, la ecuación

$$F(y, x) = x^2 + y^2 - 9 = 0 \tag{8.18}$$

FIGURA 8.6



no implica una función, sino una relación, porque (8.18) representa un círculo, como se ilustra en la figura 8.6, así que ningún valor de y corresponde a cada valor de x. Sin embargo, advierte que si se restringe y a valores no negativos, entonces tendremos solamente la mitad superior del círculo, y eso constituye una función, a saber, $y = +\sqrt{9 - x^2}$. De manera similar, la mitad inferior del círculo, con y valores no positivos, constituye otra función, $y = -\sqrt{9 - x^2}$. Por el contrario, ni la mitad izquierda ni la derecha del círculo califican como una función.

Ante esta incertidumbre, se vuelve interesante preguntar si hay condiciones generales conocidas en las que podamos estar seguros de que una ecuación dada en la forma de

$$F(y, x_1, \ldots, x_m) = 0 \tag{8.19}$$

define una función implícita

$$y = f(x_1, \ldots, x_m) \tag{8.20}$$

localmente, es decir, alrededor de un punto específico en el dominio. La respuesta la encontramos en el teorema de la función implícita, el cual establece que:

> Dada (8.19), si (*a*) la función F tiene derivadas parciales continuas $F_y, F_1, \ldots, F_m$ y si (*b*) en un punto $(y_0, x_{10}, \ldots, x_{m0})$ que satisface la ecuación (8.19), F_y no es cero, entonces existe una vecindad N que resulta m-dimensional alrededor de $(x_{10}, \ldots, x_{m0})$, N, en la cual y es una función definida implícitamente de las variables $x_1, \ldots, x_m$, en la forma de (8.20). Esta función implícita satisface $y_0 = f(x_{10}, \ldots, x_{m0})$. Asimismo, satisface la ecuación (8.19) para *cada m*-tupla $(x_1, \ldots, x_m)$ en la vecindad N; por consiguiente, da a (8.19) el estatus de una *identidad* en esa vecindad. Además, la función implícita f es continua y tiene derivadas parciales continuas $f_1, \ldots, f_m$.

La aplicación de este teorema a la ecuación del círculo (8.18), que contiene sólo una variable x, es de la siguiente forma. Primero, puede comprobar formalmente que $F_y = 2y$ y $F_x = 2x$ son continuas, tal como se requiere. Luego, se observa que F_y no es cero excepto cuando $y = 0$, es decir, excepto en el punto del extremo izquierdo $(-3, 0)$ y el punto del extremo derecho $(3, 0)$ del círculo. Así, alrededor de cualquier punto del círculo excepto $(-3, 0)$ y $(3, 0)$, podemos construir una vecindad en la que la ecuación (8.18) define una función implícita $y = f(x)$. Esto lo comprobamos fácilmente en la figura 8.6, donde de hecho podemos trazar una vecindad, por ejemplo, un rectángulo alrededor de cualquier punto del círculo, excepto $(-3, 0)$ y $(3, 0)$, tal que la porción del círculo encerrada constituirá la gráfica de una función, con un valor único y para cada valor de x en ese rectángulo.

Acerca del teorema de la función implícita se deben observar varios puntos. Primero, las condiciones citadas en el teorema son de la naturaleza de las condiciones suficientes (pero no necesarias). Esto significa que si encontramos $F_y = 0$ en un punto que satisface (8.19), no podemos usar el teorema para negar la existencia de una función implícita alrededor de ese punto. Porque, de hecho, podría existir tal función (véase el ejercicio 8.5-7).[4] Segundo, incluso si aseguramos que existe una función implícita f, el teorema no indica la forma específica que toma la función f. Tampoco indica el tamaño exacto de la vecindad en la cual está definida la función implícita. Sin embargo, a pesar de tales limitaciones, este teorema es de gran importancia. Siempre que se satisfagan las condiciones del teorema, tiene sentido hablar y hacer uso de una función como (8.20), incluso si el modelo contiene una ecuación (8.19) difícil o imposible de resolver de forma explícita para y en términos de las variables x. Además, puesto que el teorema garantiza también la existencia de las derivadas parciales $f_1, \ldots, f_m$, es correcto hablar de estas derivadas de la función implícita.

Derivadas de funciones implícitas

Si de la ecuación $F(y, x_1, \ldots, x_m) = 0$ es posible despejar y, podemos escribir la función $y = f(x_1, \ldots, x_m)$ en forma explícita y hallar sus derivadas por medio de los métodos aprendidos antes. Por ejemplo, (8.18) se puede resolver para producir dos funciones separadas

$$\begin{aligned} y^+ &= +\sqrt{9-x^2} \qquad \text{[mitad superior del círculo]} \\ y^- &= -\sqrt{9-x^2} \qquad \text{[mitad inferior del círculo]} \end{aligned} \tag{8.18'}$$

y sus derivadas se encuentran como sigue:

$$\begin{aligned} \frac{dy^+}{dx} &= \frac{d}{dx}(9-x^2)^{1/2} = \tfrac{1}{2}(9-x^2)^{-1/2}(-2x) \\ &= \frac{-x}{\sqrt{9-x^2}} = \frac{-x}{y^+} \qquad (y^+ \neq 0) \\ \frac{dy^-}{dx} &= \frac{d}{dx}[-(9-x^2)^{1/2}] = -\tfrac{1}{2}(9-x^2)^{-1/2}(-2x) \\ &= \frac{x}{\sqrt{9-x^2}} = \frac{-x}{y^-} \qquad (y^- \neq 0) \end{aligned} \tag{8.21}$$

Pero, ¿qué pasa si la ecuación dada $F(y, x_1, \ldots, x_m) = 0$ no se puede resolver para y en forma explícita? En este caso, si sabemos que existe una función implícita en los términos del teorema de la función implícita, aún es posible obtener las derivadas deseadas sin tener que resolver primero para y. Para lograr esto, se utiliza la regla de la función implícita, que permite obtener las derivadas de *toda* función implícita definida por la ecuación dada. El desarrollo de esta regla depende de los siguientes hechos básicos: (1) si dos expresiones son *idénticamente*

[4] Por otro lado, si $F_y = 0$ en una vecindad completa, entonces podemos concluir que ninguna función implícita está definida en esa vecindad. De la misma manera, si F_y se anula idénticamente, $F_y \equiv 0$, entonces en ninguna parte existe función implícita.

iguales, sus respectivas diferenciales totales deben ser iguales;[5] (2) la diferenciación de una expresión que tiene que ver con $y, x_1, \ldots, x_m$ producirá una expresión en la que intervienen las diferenciales $dy, dx_1, \ldots, dx_m$, y (3) la diferencial de y, dy, se puede sustituir, así que no importa el hecho de que no se pueda expresar y de forma explícita.

Al aplicar estos hechos a la ecuación $F(y, x_1, \ldots, x_m) = 0$, que tiene el estatus de una *identidad* en la vecindad de N en la cual está definida la función implícita, podemos escribir $dF = d0$, o bien

$$F_y\, dy + F_1\, dx_1 + F_2\, dx_2 + \cdots + F_m\, dx_m = 0 \qquad \textbf{(8.22)}$$

Puesto que la función implícita $y = f(x_1, x_2, \ldots, x_m)$ tiene la diferencial total

$$dy = f_1\, dx_1 + f_2\, dx_2 + \cdots + f_n\, dx_n$$

podemos sustituir esta expresión dy en (8.22) para obtener (después de reunir términos)

$$(F_y\, f_1 + F_1)\, dx_1 + (F_y\, f_2 + F_2)\, dx_2 + \cdots + (F_y\, f_m + F_m)\, dx_m = 0 \qquad \textbf{(8.22')}$$

El hecho de que todas las dx_i puedan variar de modo independiente de una a otra significa que, para que se cumpla la ecuación (8.22′), cada expresión entre paréntesis se debe anular; es decir, se debe tener

$$F_y\, f_i + F_i = 0 \qquad \text{(para toda } i\text{)}$$

Al dividir entre F_y y despejar f_i, obtenemos la denominada regla de la función implícita para hallar la derivada parcial f_i de la función implícita $y = f(x_1, x_2, \ldots, x_m)$:

$$f_i \equiv \frac{\partial y}{\partial x_i} = -\frac{F_i}{F_y} \qquad (i = 1, 2, \ldots, m) \qquad \textbf{(8.23)}$$

En el caso sencillo donde la ecuación dada es $F(y, x) = 0$, la regla produce

$$\frac{dy}{dx} = -\frac{F_x}{F_y} \qquad \textbf{(8.23')}$$

[5] Tome, por ejemplo, la identidad

$$x^2 - y^2 \equiv (x + y)(x - y)$$

Ésta es una identidad porque los dos lados son iguales para *cualquier* valor de x y y que uno pudiera asignar. Tomando la diferencial total de cada lado, tenemos

$$\begin{aligned} d(\text{lado izquierdo})&: 2x\, dx - 2y\, dy \\ d(\text{lado derecho})&: (x - y)\, d(x + y) + (x + y)\, d(x - y) \\ &= (x - y)(dx + dy) + (x + y)(dx - dy) \\ &= 2x\, dx - 2y\, dy \end{aligned}$$

Los dos resultados son de hecho iguales. Si dos expresiones *no* son idénticamente iguales, pero son iguales sólo para ciertos valores específicos de las variables, sus diferenciales totales *no* serán iguales. La ecuación

$$x^2 - y^2 = x^2 + y^2 - 2$$

por ejemplo, es válida sólo para $y = \pm 1$. Las diferenciales totales de los dos lados son

$$\begin{aligned} d(\text{lado izquierdo})\quad & 2x\, dx - 2y\, dy \\ d(\text{lado derecho})\quad & 2x\, dx + 2y\, dy \end{aligned}$$

que no son iguales. Advierta, en particular, que no son iguales incluso en $y = \pm 1$.

Lo que expresa esta regla es que, aun cuando no se conoce la forma específica de la función implícita, se puede hallar su derivada o derivadas al tomar el *negativo* de la razón de un par de derivadas parciales de la función F que aparece en la ecuación dada que define la función implícita. Observe que F_y aparece siempre en el denominador de la expresión. Siendo este el caso, no es admisible tener $F_y = 0$. Puesto que el teorema de la función implícita especifica que $F_y \neq 0$ en el punto alrededor del cual se define la función implícita, el problema de un denominador cero queda resuelto de forma automática en la vecindad pertinente de ese punto.

Ejemplo 1

Encuentre dy/dx para la función implícita definida por (8.17′). Puesto que $F(y, x)$ toma la forma de $y - 3x^4$, tenemos, por (8.23′),

$$\frac{dy}{dx} = -\frac{F_x}{F_y} = -\frac{-12x^3}{1} = 12x^3$$

En este caso particular, podemos despejar y de la ecuación para obtener $y = 3x^4$. Por lo tanto, se comprueba sin dificultad que la derivada es la correcta.

Ejemplo 2

Halle dy/dx para las funciones implícitas definidas por la ecuación del círculo (8.18). Esta vez tenemos $F(y, x) = x^2 + y^2 - 9$; por lo tanto, $F_y = 2y$ y $F_x = 2x$. Por (8.23′), la derivada deseada es

$$\frac{dy}{dx} = -\frac{2x}{2y} = -\frac{x}{y} \qquad (y \neq 0)$$[6]

Antes, afirmamos que la regla de la función implícita da la derivada de *toda* función explícita definida por una determinada ecuación. Esto lo comprobaremos con las dos funciones de (8.18′) y sus derivadas de (8.21). Si y se sustituye por y^+ en el resultado de la regla de la función implícita $dy/dx = -x/y$, obtenemos la derivada dy^+/dx como se muestra en (8.21); de manera similar, la sustitución de y^- produce la otra derivada en (8.21). Así, comprobamos la afirmación anterior.

Ejemplo 3

Determine $\partial y/\partial x$ para cualquier función implícita que se pueda definir mediante la ecuación $F(y, x, w) = y^3x^2 + w^3 + yxw - 3 = 0$. Despejar y de esta ecuación resulta difícil. Pero puesto que F_y, F_x y F_w son continuas, y como $F_y = 3y^2x^2 + xw$ es distinta de cero en un punto como (1, 1, 1) que satisface la ecuación dada, con seguridad por lo menos existe una función implícita $y = f(x, w)$ alrededor de ese punto. Así, tiene sentido hablar de la derivada $\partial y/\partial x$. Además, por (8.23) podemos escribir de inmediato

$$\frac{\partial y}{\partial x} = -\frac{F_x}{F_y} = -\frac{2y^3x + yw}{3y^2x^2 + xw}$$

En el punto (1, 1, 1) esta derivada tiene el valor $-\frac{3}{4}$.

Ejemplo 4

Suponga que la ecuación $F(Q, K, L) = 0$ define implícitamente una función de producción $Q = f(K, L)$. Encuentre una forma de expresar los productos físicos marginales MPP_K y MPP_L en relación con la función F. Puesto que los productos marginales son las derivadas parciales $\partial Q/\partial K$ y $\partial Q/\partial L$, podemos aplicar la regla de la función implícita y escribir

$$\text{MPP}_K \equiv \frac{\partial Q}{\partial K} = -\frac{F_K}{F_Q} \qquad \text{y} \qquad \text{MPP}_L \equiv \frac{\partial Q}{\partial L} = -\frac{F_L}{F_Q}$$

[6] La restricción $y \neq 0$ es, por supuesto, perfectamente consistente con la explicación anterior de la ecuación (8.18) de la que se deduce el enunciado del teorema de la función implícita.

Aparte de éstas, podemos obtener todavía otra derivada parcial,

$$\frac{\partial K}{\partial L} = -\frac{F_L}{F_K}$$

de la ecuación $F(Q, K, L) = 0$. ¿Cuál es el significado económico de $\partial K/\partial L$? El signo parcial indica que se mantiene constante la otra variable Q; se deduce que los cambios en K y L descritos por esta derivada son de la naturaleza de cambios "compensatorios" diseñados para mantener la producción Q constante en un nivel especificado. Por lo tanto, éstos son el tipo de cambios que pertenecen a movimientos *a lo largo* de una *isocuanta* de producción trazada con la variable K en el eje vertical y la variable L en el eje horizontal. De hecho, la derivada $\partial K/\partial L$ es la medida de la pendiente de tal isocuanta, la cual es negativa en el caso normal. El valor absoluto de $\partial K/\partial L$, por otro lado, es la medida de la *tasa marginal de sustitución técnica* entre los dos insumos.

Extensión al caso de ecuaciones simultáneas

El teorema de la función implícita viene también en una versión más general y eficaz que trata con las condiciones en las que un conjunto de ecuaciones simultáneas

$$\begin{aligned} F^1(y_1, \ldots, y_n; x_1, \ldots, x_m) &= 0 \\ F^2(y_1, \ldots, y_n; x_1, \ldots, x_m) &= 0 \\ \cdots\cdots\cdots&\cdots\cdots \\ F^n(y_1, \ldots, y_n; x_1, \ldots, x_m) &= 0 \end{aligned} \tag{8.24}$$

define con toda seguridad un conjunto de funciones implícitas[7]

$$\begin{aligned} y_1 &= f^1(x_1, \ldots, x_m) \\ y_2 &= f^2(x_1, \ldots, x_m) \\ \cdots&\cdots\cdots \\ y_n &= f^n(x_1, \ldots, x_m) \end{aligned} \tag{8.25}$$

La versión generalizada del teorema establece que:

> Dado el sistema de ecuaciones (8.24), si (a) todas las funciones $F^1, \ldots, F^n$ tienen derivadas parciales continuas respecto a todas las variables y y x, y si (b) en un punto $(y_{10}, \ldots, y_{n0}; x_{10}, \ldots, x_{m0})$ que satisface a (8.24), el siguiente determinante jacobiano es no cero:

$$|J| \equiv \left|\frac{\partial(F^1, \ldots, F^n)}{\partial(y_1, \ldots, y_n)}\right| \equiv \begin{vmatrix} \frac{\partial F^1}{\partial y_1} & \frac{\partial F^1}{\partial y_2} & \cdots & \frac{\partial F^1}{\partial y_n} \\ \frac{\partial F^2}{\partial y_1} & \frac{\partial F^2}{\partial y_2} & \cdots & \frac{\partial F^2}{\partial y_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial F^n}{\partial y_1} & \frac{\partial F^n}{\partial y_2} & \cdots & \frac{\partial F^n}{\partial y_n} \end{vmatrix} \neq 0$$

[7] Visto de otra manera, la finalidad de estas condiciones es asegurar que las n ecuaciones en (8.24) pueden *en principio* resolverse para las n variables, $y_1, \ldots, y_n$, incluso si no pudiéramos obtener la solución (8.25) de forma explícita.

entonces, existe una vecindad m-dimensional de $(x_{10}, \ldots, x_{m0})$, N, en la cual las variables $y_1, \ldots, y_n$ son funciones de las variables $x_1, \ldots, x_m$ en la forma de (8.25). Estas funciones implícitas satisfacen

$$\begin{aligned} y_{10} &= f^1(x_{10}, \ldots, x_{m0}) \\ &\cdots\cdots\cdots\cdots\cdots \\ y_{n0} &= f^n(x_{10}, \ldots, x_{m0}) \end{aligned}$$

También satisfacen (8.24) para *toda* m-tupla $(x_1, \ldots, x_m)$ en la vecindad N; por lo tanto, dan a (8.24) el estatus de un conjunto de *identidades* en lo que respecta a esta vecindad. Además, las funciones implícitas $f^1, \ldots, f^n$ son continuas y tienen derivadas parciales continuas respecto a las variables x.

Como en el caso de una sola ecuación, se pueden hallar las derivadas parciales de las funciones implícitas directamente de las n ecuaciones en (8.24), sin tener que despejar las variables y. Si aprovechamos el hecho de que, en la vecindad N, las ecuaciones de (8.24) tienen el estatus de identidades, podemos tomar la diferencial total de cada una de éstas y escribir $dF^j = 0$ $(j = 1, 2, \ldots, n)$. El resultado es un conjunto de ecuaciones con las diferenciales $dy_1, \ldots, dy_n$ y $dx_1, \ldots, dx_m$. En particular, después de pasar los términos dx_i a la derecha de los signos de igualdad, tenemos

$$\begin{aligned} \frac{\partial F^1}{\partial y_1}dy_1 + \frac{\partial F^1}{\partial y_2}dy_2 + \cdots + \frac{\partial F^1}{\partial y_n}dy_n &= -\left(\frac{\partial F^1}{\partial x_1}dx_1 + \cdots + \frac{\partial F^1}{\partial x_m}dx_m\right) \\ \frac{\partial F^2}{\partial y_1}dy_1 + \frac{\partial F^2}{\partial y_2}dy_2 + \cdots + \frac{\partial F^2}{\partial y_n}dy_n &= -\left(\frac{\partial F^2}{\partial x_1}dx_1 + \cdots + \frac{\partial F^2}{\partial x_m}dx_m\right) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \frac{\partial F^n}{\partial y_1}dy_1 + \frac{\partial F^n}{\partial y_2}dy_2 + \cdots + \frac{\partial F^n}{\partial y_n}dy_n &= -\left(\frac{\partial F^n}{\partial x_1}dx_1 + \cdots + \frac{\partial F^n}{\partial x_m}dx_m\right) \end{aligned} \tag{8.26}$$

Además, partiendo de (8.25) podemos escribir las diferenciales de las variables y_j como

$$\begin{aligned} dy_1 &= \frac{\partial y_1}{\partial x_1}dx_1 + \frac{\partial y_1}{\partial x_2}dx_2 + \cdots + \frac{\partial y_1}{\partial x_m}dx_m \\ dy_2 &= \frac{\partial y_2}{\partial x_1}dx_1 + \frac{\partial y_2}{\partial x_2}dx_2 + \cdots + \frac{\partial y_2}{\partial x_m}dx_m \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ dy_n &= \frac{\partial y_n}{\partial x_1}dx_1 + \frac{\partial y_n}{\partial x_2}dx_2 + \cdots + \frac{\partial y_n}{\partial x_m}dx_m \end{aligned} \tag{8.27}$$

y podemos usarlas para eliminar las expresiones dy_j de (8.26). Pero como el resultado de la sustitución sería difícil de manejar, simplificamos considerando sólo lo que sucedería cuando x_1 cambia mientras se mantienen constantes las otras variables $x_2, \ldots, x_m$. Si $dx_1 \neq 0$, pero $dx_2 = \cdots = dx_m = 0$ en (8.26) y (8.27), entonces al sustituir (8.27) en (8.26) y dividir todo entre $dx_1 \neq 0$, obtenemos el sistema de ecuaciones

$$\begin{aligned}
\frac{\partial F^1}{\partial y_1}\left(\frac{\partial y_1}{\partial x_1}\right)+\frac{\partial F^1}{\partial y_2}\left(\frac{\partial y_2}{\partial x_1}\right)+\cdots+\frac{\partial F^1}{\partial y_n}\left(\frac{\partial y_n}{\partial x_1}\right)&=-\frac{\partial F^1}{\partial x_1}\\
\frac{\partial F^2}{\partial y_1}\left(\frac{\partial y_1}{\partial x_1}\right)+\frac{\partial F^2}{\partial y_2}\left(\frac{\partial y_2}{\partial x_1}\right)+\cdots+\frac{\partial F^2}{\partial y_n}\left(\frac{\partial y_n}{\partial x_1}\right)&=-\frac{\partial F^2}{\partial x_1}\\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots&\cdots\cdots\\
\frac{\partial F^n}{\partial y_1}\left(\frac{\partial y_1}{\partial x_1}\right)+\frac{\partial F^n}{\partial y_2}\left(\frac{\partial y_2}{\partial x_1}\right)+\cdots+\frac{\partial F^n}{\partial y_n}\left(\frac{\partial y_n}{\partial x_1}\right)&=-\frac{\partial F^n}{\partial x_1}
\end{aligned}\qquad \textbf{(8.28)}$$

Incluso este resultado, para el caso donde sólo cambia x_1, parece demasiado complejo, porque está lleno de derivadas. Sin embargo, en realidad su estructura es bastante fácil de comprender, una vez que hayamos aprendido a distinguir entre dos tipos de derivadas que aparecen en (8.28). Un tipo que hemos escrito entre paréntesis para distinguirlo visualmente consiste en las derivadas parciales de las funciones implícitas respecto a x_1 que se están buscando. A éstas, por lo tanto, debemos considerarlas como las "variables" por determinar en (8.28). Por otro lado, el otro tipo consiste en las derivadas parciales de las funciones F^j dadas en (8.24). Puesto que tomarían valores específicos cuando se evalúan en el punto $(y_{10}, \ldots, y_{n0}; x_{10}, \ldots, x_{m0})$, el punto alrededor del cual están definidas las funciones implícitas, aparecen aquí no como funciones de derivadas sino como valores de derivadas. Como tales, podemos tratarlas como constantes. Este hecho convierte a (8.28) en un *sistema lineal*, con una estructura similar a (4.1). Resulta interesante notar que tal sistema lineal ha surgido durante el proceso de análisis de un problema que no es necesariamente lineal por sí mismo, puesto que no se han colocado restricciones de linealidad en el sistema de ecuaciones (8.24). Por lo tanto, tenemos una ilustración de cómo puede entrar en acción el álgebra lineal incluso en problemas no lineales.

En vista de que (8.28) es un sistema de ecuaciones lineales, podemos escribir en notación matricial como

$$\begin{bmatrix}
\frac{\partial F^1}{\partial y_1} & \frac{\partial F^1}{\partial y_2} & \cdots & \frac{\partial F^1}{\partial y_n}\\
\frac{\partial F^2}{\partial y_1} & \frac{\partial F^2}{\partial y_2} & \cdots & \frac{\partial F^2}{\partial y_n}\\
\cdots & \cdots & \cdots & \cdots\\
\frac{\partial F^n}{\partial y_1} & \frac{\partial F^n}{\partial y_2} & \cdots & \frac{\partial F^n}{\partial y_n}
\end{bmatrix}
\begin{bmatrix}
\left(\frac{\partial y_1}{\partial x_1}\right)\\
\left(\frac{\partial y_2}{\partial x_1}\right)\\
\vdots\\
\left(\frac{\partial y_n}{\partial x_1}\right)
\end{bmatrix}
=
\begin{bmatrix}
-\frac{\partial F^1}{\partial x_1}\\
-\frac{\partial F^2}{\partial x_1}\\
\vdots\\
-\frac{\partial F^n}{\partial x_1}
\end{bmatrix}\qquad \textbf{(8.28')}$$

Puesto que el determinante de la matriz de coeficientes en (8.28′) no es sino el determinante jacobiano particular $|J|$ que sabemos que es no cero bajo las condiciones del teorema de la función implícita, y puesto que el sistema no debe ser homogéneo (¿por qué?), debe haber una solución no trivial única para (8.28′). Por la regla de Cramer, esta solución podemos expresarla en forma analítica como:

$$\left(\frac{\partial y_j}{\partial x_1}\right)=\frac{|J_j|}{|J|}\qquad (j=1,2,\ldots,n)\qquad [\text{véase } (5.18)]\qquad \textbf{(8.29)}$$

Mediante una adaptación adecuada de este procedimiento, también podemos obtener las derivadas parciales de las funciones implícitas respecto a las otras variables, $x_2, \ldots, x_m$. Una característica agradable de este procedimiento es que, cada vez que permitimos que cambie una variable x_i particular, podemos obtener en un instante las derivadas parciales de las funciones implícitas $f^1, \ldots, f^n$ respecto a esa variable x_i particular.

De manera similar, en relación con la regla de la función implícita (8.23) para el caso de una sola ecuación, el procedimiento recién descrito requiere el uso de las derivadas parciales de las funciones F, evaluadas en el punto $(y_{10}, \ldots, y_{n0}; x_{10}, \ldots, x_{m0})$, en el cálculo de las derivadas parciales de las funciones implícitas $f^1, \ldots, f^n$. Así, la ecuación matricial (8.28′) y su solución analítica (8.29) son de hecho un enunciado de la versión de ecuaciones simultáneas de la regla de la función implícita.

Note que el requerimiento $|J| \neq 0$ descarta un denominador cero en (8.29), al igual que con el requerimiento $F_y \neq 0$ en la regla de la función implícita (8.23) y (8.23′). Asimismo, el papel que desempeña la condición $|J| \neq 0$ en garantizar una solución única (aunque implícita) (8.25) para el sistema general (posiblemente *no lineal*) (8.24) es muy similar al papel de la condición de no singularidad $|A| \neq 0$ en un sistema *lineal* $Ax = d$.

Ejemplo 5

Las siguientes tres ecuaciones

$$\begin{aligned} xy - w &= 0 & \qquad F^1 &= (x, y, w; z) = 0 \\ y - w^3 - 3z &= 0 & F^2 &= (x, y, w; z) = 0 \\ w^3 + z^3 - 2zw &= 0 & F^3 &= (x, y, w; z) = 0 \end{aligned}$$

se cumplen en el punto P: $(x, y, w; z) = (\frac{1}{4}, 4, 1, 1)$. Las funciones F^i poseen, obviamente, derivadas continuas. Así, si el jacobiano $|J|$ es no cero en el punto P, podemos usar el teorema de la función implícita para hallar la derivada estática comparativa $(\partial x/\partial z)$.

Para hacer esto, tomamos primero la diferencial total del sistema:

$$\begin{aligned} y\,dx + x\,dy - dw &= 0 \\ dy - 3w^2\,dw - 3\,dz &= 0 \\ (3w^2 - 2z)\,dw + (3z^2 - 2w)\,dz &= 0 \end{aligned}$$

Al pasar la diferencial exógena (y sus coeficientes) al lado derecho y escribir en forma matricial, obtenemos

$$\begin{bmatrix} y & x & -1 \\ 0 & 1 & -3w^2 \\ 0 & 0 & (3w^2 - 2z) \end{bmatrix} \begin{bmatrix} dx \\ dy \\ dw \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 2w - 3z^2 \end{bmatrix} dz$$

donde la matriz de coeficientes del lado izquierdo es el jacobiano

$$|J| = \begin{vmatrix} F_x^1 & F_y^1 & F_w^1 \\ F_x^2 & F_y^2 & F_w^2 \\ F_x^3 & F_y^3 & F_w^3 \end{vmatrix} = \begin{vmatrix} y & x & -1 \\ 0 & 1 & -3w^2 \\ 0 & 0 & (3w^2 - 2z) \end{vmatrix} = y(3w^2 - 2z)$$

En el punto P, el determinante jacobiano $|J| = 4$ $(\neq 0)$. Por lo tanto, aplicamos la regla de la función implícita y

$$\begin{bmatrix} y & x & -1 \\ 0 & 1 & -3w^2 \\ 0 & 0 & (3w^2 - 2z) \end{bmatrix} \begin{bmatrix} \left(\dfrac{\partial x}{\partial z}\right) \\ \left(\dfrac{\partial y}{\partial z}\right) \\ \left(\dfrac{\partial w}{\partial z}\right) \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 2w - 3z^2 \end{bmatrix}$$

Usando la regla de Cramer para hallar la expresión para $(\partial x/\partial z)$, obtenemos

$$\left(\frac{\partial x}{\partial z}\right) = \frac{\begin{vmatrix} 0 & x & -1 \\ 3 & 1 & -3w^2 \\ 2w-3z^2 & 0 & (3w^2-2z) \end{vmatrix}}{|J|} = \frac{\begin{vmatrix} 0 & \frac{1}{4} & -1 \\ 3 & 1 & -3 \\ -1 & 0 & 1 \end{vmatrix}}{4}$$

$$= 0 + (-3)\frac{\begin{vmatrix} \frac{1}{4} & -1 \\ 0 & 1 \end{vmatrix}}{4} + (-1)\frac{\begin{vmatrix} \frac{1}{4} & -1 \\ 1 & -3 \end{vmatrix}}{4}$$

$$= \frac{-3}{16} + \frac{-1}{16}$$

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

Ejemplo 6 Sea el modelo de ingreso nacional (7.17) expresado en la forma

$$\begin{aligned} Y - C - I_0 - G_0 &= 0 \\ C - \alpha - \beta(Y - T) &= 0 \\ T - \gamma - \delta Y &= 0 \end{aligned} \tag{8.30}$$

Si tomamos las variables endógenas (Y, C, T) como (y_1, y_2, y_3), y las variables exógenas y parámetros $(I_0, G_0, \alpha, \beta, \gamma, \delta)$ como $(x_1, x_2, \ldots, x_6)$, entonces la expresión del lado izquierdo de cada ecuación podemos considerarla como una función específica F, en la forma de $F^j(Y, C, T; I_0, G_0, \alpha, \beta, \gamma, \delta)$. Por lo tanto, (8.30) es un caso específico de (8.24), con $n = 3$ y $m = 6$. Puesto que las funciones F^1, F^2 y F^3 tienen derivadas parciales continuas, y como el determinante jacobiano pertinente (el que sólo tiene variables endógenas),

$$|J| = \begin{vmatrix} \frac{\partial F^1}{\partial Y} & \frac{\partial F^1}{\partial C} & \frac{\partial F^1}{\partial T} \\ \frac{\partial F^2}{\partial Y} & \frac{\partial F^2}{\partial C} & \frac{\partial F^2}{\partial T} \\ \frac{\partial F^3}{\partial Y} & \frac{\partial F^3}{\partial C} & \frac{\partial F^3}{\partial T} \end{vmatrix} = \begin{vmatrix} 1 & -1 & 0 \\ -\beta & 1 & \beta \\ -\delta & 0 & 1 \end{vmatrix} = 1 - \beta + \beta\delta \tag{8.31}$$

nunca es cero (tanto β como δ están restringidas a ser fracciones positivas), se puede tomar Y, C y T como funciones implícitas de $(I_0, G_0, \alpha, \beta, \gamma, \delta)$ *alrededor* de cualquier punto que cumple (8.30) y en el punto mismo. Pero un punto que cumple (8.30) es una solución de equilibrio, en relación con Y^*, C^* y T^*. Por consiguiente, el teorema de la función implícita dice que se justifica escribir

$$\begin{aligned} Y^* &= f^1(I_0, G_0, \alpha, \beta, \gamma, \delta) \\ C^* &= f^2(I_0, G_0, \alpha, \beta, \gamma, \delta) \\ T^* &= f^3(I_0, G_0, \alpha, \beta, \gamma, \delta) \end{aligned}$$

lo cual indica que los valores de equilibrio de las variables endógenas son funciones implícitas de las variables exógenas y los parámetros.

Las derivadas parciales de las funciones implícitas, como $\partial Y^*/\partial I_0$ y $\partial Y^*/\partial G_0$, son de la naturaleza de las derivadas estáticas comparativas. Para hallar éstas, se requieren sólo derivadas parciales de las funciones F, evaluadas en el estado de equilibrio del modelo. Además, puesto que $n = 3$, en una operación se pueden hallar tres de estas derivadas. Suponga que ahora mantenemos fijas todas las variables exógenas y también los parámetros excepto G_0. Entonces, adaptando el resultado de (8.28′), podemos escribir la ecuación

$$\begin{bmatrix} 1 & -1 & 0 \\ -\beta & 1 & \beta \\ -\delta & 0 & 1 \end{bmatrix} \begin{bmatrix} \partial Y^*/\partial G_0 \\ \partial C^*/\partial G_0 \\ \partial T^*/\partial G_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

a partir de la cual podemos calcular tres derivadas estáticas comparativas (todas respecto a G_0). La primera, que representa al multiplicador de gasto público, será

$$\frac{\partial Y^*}{\partial G_0} = \frac{\begin{vmatrix} 1 & -1 & 0 \\ 0 & 1 & \beta \\ 0 & 0 & 1 \end{vmatrix}}{|J|} = \frac{1}{1-\beta+\beta\delta} \qquad \text{[por (8.31)]}$$

Éste es nada menos que el resultado obtenido antes en (7.19). Sin embargo, nota que en el presente método sólo hemos trabajado con funciones implícitas, y omitimos por completo el paso de resolver en forma explícita el sistema (8.30) para Y^*, C^* y T^*. Esta característica particular del método es la que nos permitirá afrontar la estática comparativa de los modelos de función general, que por naturaleza propia no pueden producir una solución explícita.

EJERCICIO 8.5

1. Para cada $F(x, y) = 0$, encuentre dy/dx para cada una de las siguientes ecuaciones:
   (*a*) $y - 6x + 7 = 0$
   (*b*) $3y + 12x + 17 = 0$
   (*c*) $x^2 + 6x - 13 - y = 0$
2. Para cada $F(x, y) = 0$ use la regla de la función implícita para hallar dy/dx:
   (*a*) $F(x, y) = 3x^2 + 2xy + 4y^3 = 0$
   (*b*) $F(x, y) = 12x^5 - 2y = 0$
   (*c*) $F(x, y) = 7x^2 + 2xy^2 + 9y^4 = 0$
   (*d*) $F(x, y) = 6x^3 - 3y = 0$
3. Para cada $F(x, y, z) = 0$ use la regla de la función implícita para hallar $\partial y/\partial x$ y $\partial y/\partial z$:
   (*a*) $F(x, y, z) = x^2y^3 + z^2 + xyz = 0$
   (*b*) $F(x, y, z) = x^3z^2 + y^3 + 4xyz = 0$
   (*c*) $F(x, y, z) = 3x^2y^3 + xz^2y^2 + y^3zx^4 + y^2z = 0$
4. Suponiendo que la ecuación $F(U, x_1, x_2, \ldots, x_n) = 0$ define de forma implícita una función de utilidad $U = f(x_1, x_2, \ldots, x_n)$:
   (*a*) Encuentre las expresiones para $\partial U/\partial x_2$, $\partial U/\partial x_n$, $\partial x_3/\partial x_2$ y $\partial x_4/\partial x_n$.
   (*b*) Interprete sus respectivos significados económicos.
5. ¿Para cada una de las ecuaciones dadas $F(y, x) = 0$, es una función implícita $y = f(x)$ definida alrededor del punto $(y = 3, x = 1)$?
   (*a*) $x^3 - 2x^2y + 3xy^2 - 22 = 0$
   (*b*) $2x^2 + 4xy - y^4 + 67 = 0$

Si su respuesta es afirmativa, determine dy/dx mediante la regla de la función implícita, y evalúela en el punto mencionado.

6. Dada $x^2 + 3xy + 2yz + y^2 + z^2 - 11 = 0$, ¿es una función implícita $z = f(x, y)$ definida alrededor del punto $(x = 1, y = 2, z = 0)$? En caso afirmativo, determine $\partial z/\partial x$ y $\partial z/\partial y$ mediante la regla de la función implícita, y evalúelas en ese punto.
7. Al considerar la ecuación $F(y, x) = (x - y)^3 = 0$ en una vecindad alrededor del punto de origen, compruebe que las condiciones citadas en el teorema de la función implícita *no* son de la naturaleza de las condiciones *necesarias*.
8. Si la ecuación $F(x, y, z) = 0$ define implícitamente cada una de las tres variables como una función de las otras dos, y si las derivadas en cuestión existen, determine el valor de $\dfrac{\partial z}{\partial x}\dfrac{\partial x}{\partial y}\dfrac{\partial y}{\partial z}$.
9. Justifique la afirmación de que el sistema de ecuaciones (8.28′) debe ser no homogéneo.
10. Del modelo de ingreso nacional (8.30), encuentre el multiplicador de ingreso no gravable mediante la regla de la función implícita. Compruebe sus resultados contra (7.20).

8.6 Estática comparativa de modelos de funciones generales

Cuando consideramos por primera vez el problema de análisis estático comparativo en el capítulo 7, tratamos con el caso donde los valores de equilibrio de las variables endógenas del modelo son expresables explícitamente en términos de las variables exógenas y parámetros. Todo lo que necesitamos allí fue la técnica de diferenciación parcial simple. Sin embargo, cuando un modelo contiene funciones expresadas en la forma general, esa técnica se vuelve inaplicable como resultado de la no disponibilidad de soluciones explícitas. En cambio, debemos emplear una nueva técnica que hace uso de conceptos como diferenciales totales, derivadas totales, así como del teorema de la función implícita y la regla de la función implícita. Esto lo ilustraremos primero con un modelo de mercado, y luego lo continuaremos con los modelos de ingreso nacional.

Modelo de mercado

Tomemos en cuenta un mercado de un solo artículo, donde la cantidad demandada Q_d es una función no sólo del precio P sino también de un ingreso determinado de forma exógena Y_0. Por otro lado, la cantidad suministrada Q_s es solamente una función del precio. Si estas funciones no se dan en formas específicas, el modelo podemos escribirlo como sigue:

$$\begin{aligned} Q_d &= Q_s \\ Q_d &= D(P, Y_0) \qquad (\partial D/\partial P < 0;\ \partial D/\partial Y_0 > 0) \\ Q_s &= S(P) \qquad (dS/dP > 0) \end{aligned} \tag{8.32}$$

Se supone que las dos funciones D y S poseen derivadas continuas o, en otras palabras, tienen gráficas suaves. Además, a fin de asegurar relevancia económica, hemos impuesto restricciones definidas a los signos de estas derivadas. Por la restricción $dS/dP > 0$, se estipula que la función de oferta es creciente en sentido estricto, aunque se permite que sea lineal o no lineal. De manera similar, por las restricciones sobre las dos derivadas parciales de la función de demanda, se indica que es una función de precio estrictamente decreciente, pero una función de ingreso estrictamente creciente. Para simplificar la notación, las restricciones de signo en las derivadas de una función las indicamos a veces con signos + o – colocados debajo de las variables independientes. Así, las funciones D y S en (8.32) podemos representarlas de otro modo como

$$Q_d = D(\underset{-}{P}, \underset{+}{Y_0}) \qquad Q_s = S(\underset{+}{P})$$

Estas restricciones sirven para confinar el análisis al caso "normal" que esperamos encontrar.

Al dibujar el tipo usual de curva de demanda bidimensional se supone que el nivel de ingreso se mantiene fijo. Cuando cambia el ingreso, desestabilizará un determinado equilibrio causando un cambio en la curva de demanda. De manera similar, en (8.32), Y_0 puede provocar un desequilibrio en la función de demanda. Aquí, Y_0 es la única variable exógena o parámetro; así, el análisis estático comparativo de este modelo tendrá que ver exclusivamente con la forma en que un cambio en Y_0 afectará la posición de equilibrio del modelo.

La posición de equilibrio del mercado se define por la condición de equilibrio $Q_d = Q_s$, la cual, después de sustituir y reordenar, podemos expresarla mediante

$$D(P, Y_0) - S(P) = 0 \tag{8.33}$$

Aun cuando de esta ecuación no se puede expresar de forma explícita para el precio de equilibrio P^*, suponemos que existe un equilibrio estático, porque de lo contrario no tendría sentido plantear la pregunta de la estática comparativa. De la experiencia con modelos de función expresados a través de una fórmula específica se espera que P^* sea una función de la variable exógena Y_0:

$$P^* = P^*(Y_0) \tag{8.34}$$

Ahora podemos proporcionar un fundamento riguroso para esta conjetura al recurrir al teorema de la función implícita. En vista de que (8.33) está en la forma de $F(P, Y_0) = 0$, el cumplimiento de las condiciones del teorema de la función implícita garantizará que todo valor de Y_0 produzca un valor único de P^* en la vecindad de un punto que satisface a (8.33), es decir, en la vecindad de una solución de equilibrio (inicial o "anterior"). En ese caso, podemos escribir de hecho la función implícita $P^* = P^*(Y_0)$ y analizar su derivada, dP^*/dY_0, la derivada estática comparativa específica que deseamos, la cual sabemos que existe. Por lo tanto, procedemos a comprobar esas condiciones. Primero, la función $F(P, Y_0)$ posee de hecho derivadas continuas; esto es porque, por suposición, sus dos términos $D(P, Y_0)$ y $S(P)$ tienen derivadas continuas. Segundo, la derivada parcial de F con respecto a P, a saber, $F_P = \partial D/\partial P - dS/dP$, es negativa, y por lo tanto diferente de cero, sin importar dónde se evalúe. Así, aplicamos el teorema de la función implícita, y (8.34) es de hecho legítima.

De acuerdo con el mismo teorema, la condición de equilibrio (8.33) podemos tomarla ahora como una identidad en alguna vecindad de la solución de equilibrio. En consecuencia, la identidad de equilibrio podemos escribirla como

$$\underbrace{D(P^*, Y_0) - S(P^*)}_{F(P^*, Y_0)} \equiv 0 \quad \text{[Demanda excedente} \equiv 0 \text{ en el equilibrio]} \tag{8.35}$$

Requerimos entonces sólo una aplicación directa de la regla de la función implícita para producir la derivada estática comparativa, dP^*/dY_0. A fin de tener claridad visual, en adelante pondremos entre paréntesis las derivadas estáticas comparativas para distinguirlas de las expresiones de derivadas normales que solamente forman parte de la especificación del modelo. El resultado de la regla de función implícita es

$$\left(\frac{dP^*}{dY_0}\right) = -\frac{\partial F/\partial Y_0}{\partial F/\partial P^*} = -\frac{\partial D/\partial Y_0}{\partial D/\partial P^* - dS/dP^*} > 0 \qquad \textbf{(8.36)}$$

En este resultado, la expresión $\partial D/\partial P^*$ se refiere a la derivada $\partial D/\partial P$ evaluada en el equilibrio inicial, es decir, en $P = P^*$; a dS/dP^* se le da una interpretación similar. De hecho, $\partial D/\partial Y_0$ se debe evaluar también en el punto de equilibrio. En virtud de las especificaciones de signo de (8.32), (dP^*/dY_0) es siempre positiva. Así, la conclusión *cualitativa* es que un incremento (decremento) en el nivel de ingreso siempre dará como resultado un incremento (decremento) en el precio de equilibrio. Si conocemos los valores que las derivadas de las funciones de oferta y demanda toman en el equilibrio inicial, (8.36) producirá también una conclusión *cuantitativa*.

Esta explicación del ajuste de mercado tiene que ver con el efecto de un cambio en Y_0 sobre P^*. ¿Es posible hallar también el efecto sobre la cantidad de equilibrio $Q^*(= Q_d^* = Q_s^*)$? La respuesta es afirmativa. Puesto que en el estado de equilibrio se tiene $Q^* = S(P^*)$, y como $P^* = P^*(Y_0)$, podemos aplicar la regla de la cadena para obtener la derivada

$$\left(\frac{dQ^*}{dY_0}\right) = \frac{dS}{dP^*}\left(\frac{dP^*}{dY_0}\right) > 0 \left[\text{puesto que } \frac{dS}{dP^*} > 0\right] \qquad \textbf{(8.37)}$$

Por lo tanto, la cantidad de equilibrio también tiene una relación positiva con Y_0 en este modelo. De nuevo, (8.37) puede proveernos una conclusión cuantitativa si conocemos los valores que toman las distintas derivadas en el equilibrio.

Los resultados de (8.36) y (8.37), que agotan el contenido estático comparativo del modelo (puesto que el último contiene sólo una variable exógena y dos endógenas), no son sorprendentes. De hecho, no hacen más que transmitir el principio de que un cambio hacia arriba de la curva de demanda dará como resultado un mayor precio de equilibrio, como también una mayor cantidad de equilibrio. Esta misma proposición, al parecer, podría haberse obtenido en un instante a partir de un análisis gráfico simple. Esto parece correcto, pero no debemos olvidar el carácter mucho más general del procedimiento analítico que hemos usado aquí. Por su propia naturaleza, el análisis gráfico está limitado a un conjunto específico de curvas (la contraparte geométrica de un conjunto específico de funciones); por lo tanto, sus conclusiones son, en términos estrictos, pertinentes y aplicables a sólo ese conjunto de curvas. Por el contrario, la formulación de (8.32), simplificada como está, abarca el conjunto completo de combinaciones posibles de curvas de demanda con pendiente negativa y curvas de oferta con pendiente positiva. De forma que es mucho muy general. Asimismo, el procedimiento analítico que usamos aquí puede manejar muchos problemas de mayor complejidad que probarían estar más allá de las capacidades del método gráfico.

Método de ecuaciones simultáneas

El análisis del modelo (8.32) lo llevamos a cabo con base en una sola ecuación, la (8.35). Puesto que sólo una variable endógena se puede incorporar de manera fructífera en una ecuación, la inclusión de P^* significa la exclusión de Q^*. Como resultado, tuvimos primero que hallar (dP^*/dY_0) y luego inferir (dQ^*/dY_0) en un paso posterior. Ahora mostraremos cómo se pueden estudiar al mismo tiempo P^* y Q^*. Como hay dos variables endógenas, estableremos un sistema de dos ecuaciones. Primero, si $Q = Q_d = Q_s$ en (8.32) y reordenando, podemos expresar el modelo de mercado como

$$\begin{aligned} F^1(P, Q; Y_0) &= D(P, Y_0) - Q = 0 \\ F^2(P, Q; Y_0) &= S(P) - Q = 0 \end{aligned} \tag{8.38}$$

que está en la forma de (8.24), con $n = 2$ y $m = 1$. De nuevo adquiere interés la comprobación de las condiciones del teorema de la función implícita. Primero, puesto que se supone que las funciones de oferta y demanda poseen derivadas continuas, así debe ser para las funciones F^1 y F^2. Segundo, el jacobiano de variables endógenas (en el que intervienen P y Q) de hecho resulta ser no cero, sin importar dónde se evalúe, porque

$$|J| = \begin{vmatrix} \dfrac{\partial F^1}{\partial P} & \dfrac{\partial F^1}{\partial Q} \\ \dfrac{\partial F^2}{\partial P} & \dfrac{\partial F^2}{\partial Q} \end{vmatrix} = \begin{vmatrix} \dfrac{\partial D}{\partial P} & -1 \\ \dfrac{dS}{dP} & -1 \end{vmatrix} = \frac{dS}{dP} - \frac{\partial D}{\partial P} > 0 \tag{8.39}$$

Por consiguiente, si existe una solución de equilibrio (P^*, Q^*) (como es de suponer a fin de que tenga sentido hablar de estática comparativa), el teorema de la función implícita indica que podemos escribir las funciones implícitas

$$P^* = P^*(Y_0) \qquad \text{y} \qquad Q^* = Q^*(Y_0) \tag{8.40}$$

aun cuando no podamos resolver de forma explícita para P^* y Q^*. Sabemos que estas funciones tienen derivadas continuas. Además, (8.38) tendrá el estatus de un par de identidades en alguna vecindad del estado de equilibrio, de modo que también podemos escribir

$$\begin{aligned} D(P^*, Y_0) - Q^* &\equiv 0 \quad [\text{es decir, } F^1(P^*, Q^*; Y_0) \equiv 0] \\ S(P^*) - Q^* &\equiv 0 \quad [\text{es decir, } F^2(P^*, Q^*; Y_0) \equiv 0] \end{aligned} \tag{8.41}$$

De éstas, (dP^*/dY_0) y (dQ^*/dY_0) podemos determinar en forma simultánea mediante la regla de la función implícita (8.28′).

En el presente contexto, con F^1 y F^2 como se definen en (8.41), y con dos variables endógenas P^* y Q^* y una sola variable exógena Y_0, la regla de la función implícita toma la forma específica

$$\begin{bmatrix} \dfrac{\partial F^1}{\partial P^*} & \dfrac{\partial F^1}{\partial Q^*} \\ \dfrac{\partial F^2}{\partial P^*} & \dfrac{\partial F^2}{\partial Q^*} \end{bmatrix} \begin{bmatrix} \left(\dfrac{dP^*}{dY_0}\right) \\ \left(\dfrac{dQ^*}{dY_0}\right) \end{bmatrix} = \begin{bmatrix} -\dfrac{\partial F^1}{\partial Y_0} \\ -\dfrac{\partial F^2}{\partial Y_0} \end{bmatrix}$$

Note que las derivadas estáticas comparativas las escribimos aquí con el símbolo d en vez de ∂, porque sólo hay una variable exógena en el presente problema. De manera más específica, la última ecuación la podemos expresar como

$$\begin{bmatrix} \dfrac{\partial D}{\partial P^*} & -1 \\ \dfrac{dS}{dP^*} & -1 \end{bmatrix} \begin{bmatrix} \left(\dfrac{dP^*}{dY_0}\right) \\ \left(\dfrac{dQ^*}{dY_0}\right) \end{bmatrix} = \begin{bmatrix} -\dfrac{\partial D}{\partial Y_0} \\ 0 \end{bmatrix}$$

Por la regla de Cramer, y usando (8.39), vemos que la solución es

$$\left(\frac{dP^*}{dY_0}\right) = \frac{\begin{vmatrix} -\dfrac{\partial D}{\partial Y_0} & -1 \\ 0 & -1 \end{vmatrix}}{|J|} = \frac{\dfrac{\partial D}{\partial Y_0}}{|J|}$$

$$\left(\frac{dQ^*}{dY_0}\right) = \frac{\begin{vmatrix} \dfrac{\partial D}{\partial P^*} & -\dfrac{\partial D}{\partial Y_0} \\ \dfrac{dS}{dP^*} & 0 \end{vmatrix}}{|J|} = \frac{\dfrac{dS}{dP^*}\dfrac{\partial D}{\partial Y_0}}{|J|} \tag{8.42}$$

donde todas las derivadas de las funciones de oferta y demanda (incluso las que aparecen en el jacobiano) se evaluarán en el equilibrio inicial. Podemos comprobar que los resultados obtenidos son idénticos a los obtenidos antes en (8.36) y (8.37), por medio del método de una sola ecuación.

En vez de aplicar directamente la regla de la función implícita, podemos llegar al mismo resultado diferenciando primero por completo cada identidad en (8.41) a su vez, para obtener un sistema lineal de ecuaciones en las variables dP^* y dQ^*:

$$\frac{\partial D}{\partial P^*}dP^* - dQ^* = -\frac{\partial D}{\partial Y_0}dY_0$$

$$\frac{dS}{dP^*}dP^* - dQ^* = 0$$

y dividiendo después entre $dY_0 \neq 0$, e interpretando cada cociente de dos diferenciales como una derivada.

Uso de derivadas totales

En ambos métodos, el de una sola ecuación y el de ecuaciones simultáneas, ilustrados antes, hemos tomado las *diferenciales totales* de ambos lados de una identidad de equilibrio e igualado después los dos resultados para llegar a la regla de la función implícita. Sin embargo, en vez de tomar las diferenciales totales, podemos tomar, e igualar, las *derivadas totales* de los dos lados de la identidad de equilibrio respecto a una variable endógena particular o parámetro.

En el método de una sola ecuación, por ejemplo, la identidad de equilibrio es

$$D(P^*, Y_0) - S(P^*) \equiv 0 \qquad \text{[de (8.35)]}$$

donde

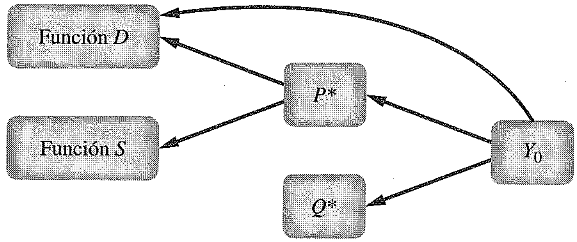
$$P^* = P^*(Y_0) \qquad \text{[de (8.34)]}$$

Tomando la derivada total de la identidad de equilibrio respecto a Y_0, que toma en cuenta los efectos indirectos así como los directos de un cambio en Y_0, obtendremos la ecuación

$$\underset{\begin{pmatrix}\text{efecto indirecto}\\ \text{de } Y_0 \text{ en } D\end{pmatrix}}{\frac{\partial D}{\partial P^*}\left(\frac{dP^*}{dY_0}\right)} + \underset{\begin{pmatrix}\text{efecto directo}\\ \text{de } Y_0 \text{ en } D\end{pmatrix}}{\frac{\partial D}{\partial Y_0}} - \underset{\begin{pmatrix}\text{efecto indirecto}\\ \text{de } Y_0 \text{ en } S\end{pmatrix}}{\frac{dS}{dP^*}\left(\frac{dP^*}{dY_0}\right)} = 0$$

Cuando de esta ecuación se determina (dP^*/dY_0), el resultado es idéntico al de (8.36).

FIGURA 8.7



En el método de ecuaciones simultáneas, por otro lado, hay un par de identidades de equilibrio:

$$D(P^*, Y_0) - Q^* \equiv 0$$
$$S(P^*) - Q^* \equiv 0 \qquad \text{[de (8.41)]}$$

donde $$P^* = P^*(Y_0) \qquad Q^* = Q^*(Y_0) \qquad \text{[de (8.40)]}$$

Ahora es más difícil distinguir los diversos efectos de Y_0, pero con la ayuda del mapa de canales de la figura 8.7, el patrón debe volverse claro. Este mapa de canales nos indica, por ejemplo, que al diferenciar la función D respecto a Y_0, debemos tomar en cuenta el efecto indirecto de Y_0 sobre D por P^*, así como el efecto directo de Y_0 (flecha curvada). Por otro lado, al diferenciar la función S respecto a Y_0, sólo debemos tomar en cuenta el efecto indirecto (por P^*). El resultado de la diferenciación total de las dos identidades respecto a Y_0 es, después de reacomodar términos, el siguiente par de ecuaciones:

$$\frac{\partial D}{\partial P^*}\left(\frac{dP^*}{dY_0}\right) - \left(\frac{dQ^*}{dY_0}\right) = -\frac{\partial D}{\partial Y_0}$$
$$\frac{dS}{dP^*}\left(\frac{dP^*}{dY_0}\right) - \left(\frac{dQ^*}{dY_0}\right) = 0$$

Éstas son, por supuesto, idénticas a las ecuaciones obtenidas por el método de diferencial total, y de nuevo llevan a las derivadas estáticas comparativas de (8.42).

Modelo de ingreso nacional (IS-LM)

Una aplicación característica del teorema de la función implícita es una forma funcional general del modelo IS-LM.[8] El equilibrio en este modelo macroeconómico se caracteriza por un nivel de ingreso y tasas de interés que producen de forma simultánea equilibrio tanto en el mercado de bienes como en el monetario.

Un mercado de bienes se describe mediante el siguiente conjunto de ecuaciones:

$$Y = C + I + G \qquad C = C(Y - T) \qquad G = G_0$$
$$I = I(r) \qquad T = T(Y)$$

Y es el nivel de producto interno bruto (PIB), o ingreso nacional. En esta forma del modelo, Y se puede considerar también como una oferta agregada. C, I, G y T son el consumo, la inversión, el gasto público y los impuestos, respectivamente.

[8] IS significa "la inversión es igual a los ahorros" y LM significa "la preferencia de liquidez es igual a la oferta monetaria".

1. Se supone que el consumo es una función estrictamente creciente de ingreso disponible $(Y - T)$. Si se denota el ingreso disponible como $Y^d = Y - T$, entonces la función de consumo se puede expresar como

$$C = C(Y^d)$$

donde $dC/dY^d = C'(Y^d)$ es la propensión marginal al consumo $(0 < C'(Y^d) < 1)$.

2. Se supone que el gasto de inversión es una función estrictamente decreciente de la tasa de interés, r:

$$\frac{dI}{dr} = I'(r) < 0$$

3. El sector público se describe mediante dos variables: gasto público (G) e impuestos (T). Comúnmente, se supone que el gasto público es exógeno (establecido por la política) mientras que se supone que los impuestos son una función creciente de ingreso. $\frac{dT}{dY} = T'(Y)$ es la tasa marginal de impuesto $(0 < T'(Y) < 1)$.

Si sustituimos estas funciones para C, I, G en la primera ecuación $Y = C + I + G$, obtenemos

$$Y = C(Y - T(Y)) + I(r) + G_0 \qquad \text{(curva IS)}$$

que produce una sola ecuación con dos variables endógenas: Y y r. Esta ecuación da todas las combinaciones de Y y r que producen equilibrio en el mercado de bienes. Esta ecuación define de manera implícita la curva IS.

Pendiente de la curva IS

Si expresamos la ecuación IS, como una identidad de equilibrio,

$$Y - C(Y^d) - I(r) - G_0 \equiv 0$$

entonces la diferencial total respecto a Y y r es

$$dY - C'(Y^d)[1 - T'(Y)]\, dY - I'(r)\, dr = 0$$

Nota: $$\frac{dY^d}{dY} = 1 - T'(Y)$$

Podemos reordenar los términos dY y dr para obtener una expresión para la pendiente de la curva IS:

$$\frac{dr}{dY} = \frac{1 - C'(Y^d)[1 - T'(Y)]}{I'(r)} < 0$$

Dadas las restricciones puestas en las derivadas de C, I y T, podemos comprobar fácilmente que la pendiente de la curva IS es negativa.

El mercado monetario podemos describirlo mediante las siguientes tres ecuaciones:

$$M^d = L(Y, r) \quad \text{[demanda de dinero]} \quad \text{donde} \quad L_Y > 0 \quad \text{y} \quad L_r < 0$$

$$M^s = M_0^s \qquad \text{[oferta de dinero]}$$

donde se supone que la oferta de dinero la determina de manera exógena la autoridad monetaria central,

$$M^d = M^s \qquad \text{[condición de equilibrio]}$$

Al sustituir las dos primeras ecuaciones en la tercera, obtenemos una expresión que define de forma implícita la curva LM, que de nuevo es de la naturaleza de una identidad de equilibrio.

$$L(Y, r) \equiv M_0^s$$

Pendiente de la curva LM

Puesto que ésta es una identidad de equilibrio, podemos tomar la diferencial total respecto a las dos variables endógenas, Y y r:

$$L_Y\, dY + L_r\, dr = 0$$

la cual podemos reordenarla a fin de obtener una expresión para la pendiente de la curva LM

$$\frac{dr}{dY} = -\frac{L_Y}{L_r} > 0$$

Puesto que $L_Y > 0$ y $L_r < 0$, podemos determinar que la pendiente de la curva LM es positiva.

El estado de equilibrio macroeconómico simultáneo de los mercados de bienes y monetario se describe mediante el siguiente sistema de ecuaciones:

$$\begin{aligned} Y &\equiv C(Y^d) + I(r) + G_0 \\ L(Y, r) &\equiv M_0^s \end{aligned}$$

que de manera implícita definen las dos variables endógenas, Y y r, como funciones de las variables exógenas, G_0 y M_0^s. Tomando la diferencial total del sistema, obtenemos

$$\begin{aligned} dY - C'(Y^d)[1 - T'(Y)]\, dY - I'(r)\, dr &= dG_0 \\ L_Y\, dY + L_r\, dr &= dM_0^s \end{aligned}$$

o, en forma matricial,

$$\begin{bmatrix} 1 - C'(Y^d)[1 - T'(Y)] & -I'(r) \\ L_Y & L_r \end{bmatrix} \begin{bmatrix} dY \\ dr \end{bmatrix} = \begin{bmatrix} dG_0 \\ dM_0^s \end{bmatrix}$$

El determinante jacobiano es

$$\begin{aligned} |J| &= \begin{vmatrix} 1 - C'(Y^d)[1 - T'(Y)] & -I'(r) \\ L_Y & L_r \end{vmatrix} \\ &= \{1 - C'(Y^d)[1 - T'(Y)]\}L_r + L_Y I'(r) < 0 \end{aligned}$$

Puesto que $|J| \neq 0$, este sistema satisface las condiciones del teorema de la función implícita y es posible escribir las funciones implícitas

$$Y^* = Y^*\big(G_0, M_0^s\big)$$

y

$$r^* = r^*\big(G_0, M_0^s\big)$$

aunque el sistema no se pueda resolver de manera explícita para Y^* y r^*. Si no es posible despejar Y^* y r^*, aún se pueden llevar a cabo ejercicios de estática comparativa para determinar los efectos de un cambio de una de las variables exógenas $\big(G_0, M_0^s\big)$ en los valores de equilibrio de Y^* y r^*. Considera las derivadas estáticas comparativas $\partial Y^*/\partial G_0$ y $\partial r^*/\partial G_0$ que se deducirán al aplicar el teorema de la función implícita al sistema de diferenciales totales en forma matricial

$$\begin{bmatrix} 1 - C'(Y^d)[1 - T'(Y)] & -I'(r) \\ L_Y & L_r \end{bmatrix} \begin{bmatrix} dY \\ dr \end{bmatrix} = \begin{bmatrix} dG_0 \\ dM_0^s \end{bmatrix}$$

Primero establecemos $dM_0^s = 0$ y dividimos ambos lados entre dG_0.

$$\begin{bmatrix} 1 - C' \cdot (1 - T') & -I'(r) \\ L_Y & L_r \end{bmatrix} \begin{bmatrix} \dfrac{dY^*}{dG_0} \\ \dfrac{dr^*}{dG_0} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Por medio de la regla de Cramer, obtenemos

$$\frac{dY^*}{dG_0} = \frac{\begin{vmatrix} 1 & -I' \\ 0 & L_r \end{vmatrix}}{|J|} = \frac{L_r}{|J|} = \frac{\ominus}{\ominus} > 0$$

y

$$\frac{dr^*}{dG_0} = \frac{\begin{vmatrix} 1 - C' \cdot (1 - T') & 1 \\ L_Y & 0 \end{vmatrix}}{|J|} = \frac{-L_Y}{|J|} = \frac{\ominus}{\ominus} > 0$$

Del teorema de la función implícita, estos cocientes de diferenciales, dY^*/dG_0 y dr^*/dG_0, podemos interpretarlos como derivadas parciales,

$$\frac{\partial Y^*(G_0, M_0^s)}{\partial G_0} \qquad \text{y} \qquad \frac{\partial r^*(G_0, M_0^s)}{\partial G_0}$$

que son las derivadas estáticas comparativas deseadas.

Ampliación del modelo: economía abierta

Una propiedad de un modelo que los economistas buscan es su robustez; la capacidad del modelo para aplicarse a entornos distintos. En este punto se extenderá el modelo básico para incorporar el sector externo.

1. *Exportaciones netas*. Sea que X denote las exportaciones, M las importaciones y E la tasa de intercambio (medida como el precio en nuestro país de la moneda extranjera). Las exportaciones son una función creciente del tipo de cambio.

$$X = X(E) \qquad \text{donde} \qquad X'(E) > 0$$

Las importaciones son una función decreciente del tipo de cambio, pero una función creciente del ingreso.

$$M = M(Y, E) \qquad \text{donde} \qquad M_Y > 0, \ M_E < 0$$

2. *Flujos de capital*. El flujo neto de capital hacia un país es una función de la tasa de interés nacional r y la tasa de interés mundial r_w. Sea K el influjo de capital neto tal que

$$K = K(r, r_w) \qquad \text{donde} \qquad K_r > 0, K_{r_w} < 0$$

3. *Balanza de pagos*. Las entradas y salidas de moneda extranjera para un país están separadas, por lo común, en dos cuentas: cuenta *corriente* (exportaciones netas de bienes y servicios) y la cuenta de *capital* (la compra de bonos nacionales y extranjeros). Juntas, las dos cuentas constituyen la balanza de pagos.

$$\begin{aligned} \text{BP} &= \text{cuenta corriente} + \text{cuenta de capital} \\ &= [X(E) - M(Y, E)] + K(r, r_w) \end{aligned}$$

Bajo un tipo de cambio flexible, éste se ajusta para mantener la balanza de pagos igual a cero. Tener la balanza de pagos igual a cero equivale a decir que la oferta de moneda extranjera es igual a la demanda de moneda extranjera que tiene un país.[9]

Equilibrio de economía abierta

El equilibrio en una economía abierta se caracteriza por tres condiciones: la demanda agregada es igual a la oferta agregada; la demanda de dinero es igual a la oferta de dinero; la balanza de pagos es igual a cero. Al añadir el sector externo al modelo básico se obtiene el siguiente sistema de ecuaciones

$$\begin{aligned} Y &= C(Y^d) + I(r) + G_0 + X(E) - M(Y, E) \\ L(Y, r) &= M_0^s \\ X(E) - M(Y, E) + K(r, r_w) &= 0 \end{aligned}$$

Puesto que hay tres ecuaciones, se requieren tres variables endógenas, que son Y, r y E. Las variables exógenas ahora son G_0, M_0^s y r_w. Expresar al sistema como identidades de equilibrio $F^1 \equiv 0$, $F^2 \equiv 0$, $F^3 \equiv 0$ nos permite hallar el jacobiano:

$$\begin{aligned} Y - C(Y^d) - I(r) - G_0 - X(E) + M(Y, E) &\equiv 0 \\ L(Y, r) - M_0^s &\equiv 0 \\ X(E) - M(Y, E) + K(r, r_w) &\equiv 0 \end{aligned}$$

$$|J| = \begin{vmatrix} 1 - C' \cdot (1 - T') + M_Y & -I' & M_E - X' \\ L_Y & L_r & 0 \\ -M_Y & K_r & X' - M_E \end{vmatrix}$$

Usando el desarrollo de Laplace de arriba abajo en la tercera columna, obtenemos

$$\begin{aligned} |J| &= (M_E - X') \begin{vmatrix} L_Y & L_r \\ -M_Y & K_r \end{vmatrix} + (X' - M_E) \begin{vmatrix} 1 - C' \cdot (1 - T') + M_Y & -I' \\ L_Y & L_r \end{vmatrix} \\ &= (M_E - X')(L_Y K_r + L_r M_Y) + (X' - M_E)\{[1 - C' \cdot (1 - T') + M_Y]L_r + I' L_Y\} \\ &= (M_E - X')\{L_Y(K_r - I') + L_r[C'(1 - T') - 1]\} \end{aligned}$$

Dadas las suposiciones respecto a los signos de las derivadas parciales y la restricción de que $0 < C' \cdot (1 - T') < 1$, podemos determinar que $|J| < 0$. Por lo tanto, podemos escribir las funciones implícitas

$$\begin{aligned} Y^* &= Y^*\left(G_0, M_0^s, r_w\right) \\ r^* &= r^*\left(G_0, M_0^s, r_w\right) \\ E^* &= E^*\left(G_0, M_0^s, r_w\right) \end{aligned}$$

[9] Bajo un tipo de cambio fijo, la balanza de pagos no necesariamente es cero. En ese caso, cualquier excedente o déficit se registra como *cambio de acuerdos oficiales*.

Tomar la diferencial total del sistema y escribirlo en forma matricial

$$\begin{bmatrix} 1 - C' \cdot (1 - T') + M_Y & -I' & M_E - X' \\ L_Y & L_r & 0 \\ -M_Y & K_r & X' - M_E \end{bmatrix} \begin{bmatrix} dY^* \\ dr^* \\ dE^* \end{bmatrix} = \begin{bmatrix} dG_0 \\ dM_0^s \\ -K_{r_w}\, dr_w \end{bmatrix}$$

nos permite llevar a cabo una serie de ejercicios estáticos comparativos. Consideremos el impacto de un cambio en las tasas de interés en el mundo r_w en los valores de equilibrio de Y, r y E. Si se establece $dG_0 = dM_0^s = 0$ y dividimos ambos lados entre dr_w, obtenemos

$$\begin{bmatrix} 1 - C' \cdot (1 - T') + M_Y & -I' & M_E - X' \\ L_Y & L_r & 0 \\ -M_Y & K_r & X' - M_E \end{bmatrix} \begin{bmatrix} \dfrac{dY^*}{dr_w} \\ \dfrac{dr^*}{dr_w} \\ \dfrac{dE^*}{dr_w} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -K_{r_w} \end{bmatrix}$$

Al usar la regla de Cramer obtenemos las derivadas estáticas comparativas

$$\frac{\partial Y^*}{\partial r_w} = \frac{\begin{vmatrix} 0 & -I' & M_E - X' \\ 0 & L_r & 0 \\ -K_{r_w} & K_r & X' - M_E \end{vmatrix}}{|J|} = \frac{(-K_{r_w})(-L_r)(M_E - X')}{|J|} > 0$$

y

$$\frac{\partial r^*}{\partial r_w} = \frac{\begin{vmatrix} 1 - C' \cdot (1 - T') + M_Y & 0 & M_E - X' \\ L_Y & 0 & 0 \\ -M_Y & -K_{r_w} & X' - M_E \end{vmatrix}}{|J|} = \frac{K_{r_w}(-L_Y)(M_E - X')}{|J|} > 0$$

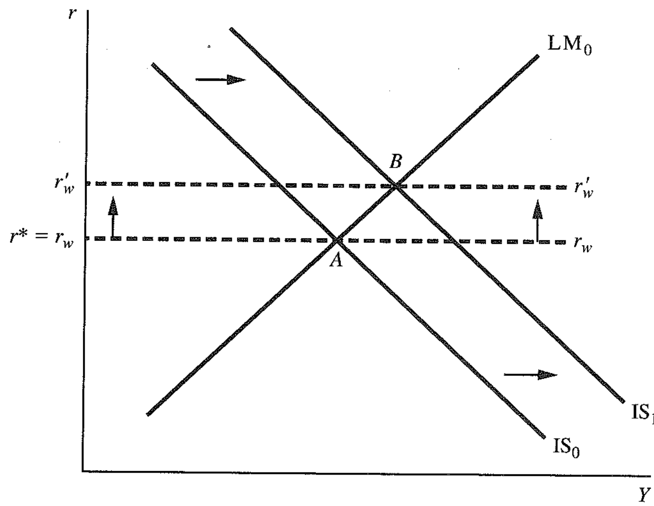
y

$$\frac{\partial E^*}{\partial r_w} = \frac{\begin{vmatrix} 1 - C' \cdot (1 - T') + M_Y & -I' & 0 \\ L_Y & L_r & 0 \\ -M_Y & K_r & -K_{r_w} \end{vmatrix}}{|J|}$$

$$= \frac{-K_{r_w}\{[1 - C' \cdot (1 - T') + M_Y]L_r + L_Y I'\}}{|J|} > 0$$

En este punto debemos comparar los resultados obtenidos con los principios macroeconómicos. De manera intuitiva, un aumento en la tasa de interés mundial debe conducir a un incremento de salidas de capital y una depreciación de la moneda nacional. Esto, a su vez, conduce a un incremento en las exportaciones netas y el ingreso. El aumento en el ingreso nacional causará un incremento en la demanda de dinero, lo cual aumenta la presión en las tasas de interés nacionales. Este resultado se ilustra de forma gráfica en la figura 8.8, donde un aumento en las tasas de interés mundiales da lugar al desplazamiento de la curva IS a la derecha.

FIGURA 8.8



Resumen del procedimiento

En el análisis del modelo de mercado de función general y el modelo de ingreso nacional no es posible obtener valores solución explícitos de las variables endógenas. En cambio, se confía en que el teorema de la función implícita permita escribir soluciones implícitas tales como

$$P^* = P^*(Y_0) \qquad \text{y} \qquad r^* = r^*(G_0, M_0^s)$$

La investigación posterior para las derivadas estáticas comparativas como (dP^*/dY_0) y $(\partial r^*/\partial G_0)$ se basa, entonces, por su importancia en el hecho conocido, gracias de nuevo al teorema de la función implícita, de que las funciones P^* y r^* poseen derivadas continuas.

Para facilitar la aplicación de ese teorema, es práctica común escribir la condición o condiciones de equilibrio del modelo en la forma de (8.19) u (8.24). Después, se comprueba si (1) la función o las funciones, F, tienen derivadas continuas y (2) el valor de F_y o el determinante jacobiano de variables endógenas (como podría ser el caso) es no cero en el equilibrio inicial del modelo. Sin embargo, cuando cada una de las funciones del modelo tiene derivadas continuas, suposición que suele adoptarse como algo normal en los modelos de función general, la primera condición se satisface de manera automática. Por lo tanto, como un asunto práctico, sólo es necesario comprobar el valor de F_y o el jacobiano de la variable endógena. Y si es no cero en el equilibrio, se puede proceder de inmediato a la tarea de hallar las derivadas estáticas comparativas.

Para ese fin, nos sirve de ayuda la regla de la función implícita. Para el caso de una sola ecuación, simplemente debemos igualar las variables endógenas a su valor de equilibrio (por ejemplo, $P = P^*$) en la condición de equilibrio, y entonces aplicar la regla establecida en (8.23) que dé como resultado la identidad de equilibrio. Para el caso de ecuaciones simultáneas, también se deben igualar todas las variables endógenas a sus respectivos valores de equilibrio en las condiciones de equilibrio. Entonces, podemos aplicar la regla de la función implícita, como se ilustra en (8.29), a las identidades de equilibrio resultantes, o bien, ejecutar los pasos siguientes para llegar al mismo resultado.

1. Tome la diferencial total de cada identidad de equilibrio.
2. Seleccione una, y sólo una, variable exógena (por ejemplo, X_0) como el único factor de desequilibrio e iguale a cero las diferenciales de *las otras* variables exógenas. Luego, divida

los términos restantes de cada identidad entre dX_0, e interprete cada cociente de dos diferenciales como una derivada estática comparativa, una derivada *parcial* si el modelo contiene dos o más variables exógenas.[10]

3. Resuelva el sistema de ecuaciones resultante para las derivadas estáticas comparativas que aparecen ahí, e interprete sus implicaciones económicas. En este paso, si usa la regla de Cramer, puede aprovechar el hecho de que, antes, al comprobar la condición $|J| \neq 0$, ya se calculó el determinante de la matriz de coeficientes del sistema de ecuaciones que ahora se resuelve.
4. Para el análisis de otro factor de desequilibrio (otra variable exógena), si existe, repita los pasos 2 y 3. Aunque surgirá un grupo diferente de derivadas estáticas comparativas en el nuevo sistema de ecuaciones, la matriz de coeficientes será la misma que antes y, por lo tanto, puede usar de nuevo el valor conocido de $|J|$.

Dado un modelo con m variables exógenas, se requieren exactamente m aplicaciones de los pasos 1, 2 y 3 para tener todas las derivadas estáticas comparativas.

[10] En lugar tomar los pasos 1 y 2, podríamos recurrir al método de derivada total diferenciando (ambos lados) cada identidad de equilibrio por completo con respecto a la variable exógena seleccionada. Para esto, resultará útil un mapa de canal.

EJERCICIO 8.6

1. Sea la condición de equilibrio para el ingreso nacional
   $$S(Y) + T(Y) = I(Y) + G_0 \qquad (S', T', I' > 0; \quad S' + T' > I')$$
   donde S, Y, T, I y G significan ahorro, ingreso nacional, impuestos, inversión y gasto público, respectivamente. Todas las derivadas son continuas.
   (*a*) Interprete los significados económicos de las derivadas S', T' e I'.
   (*b*) Compruebe si se satisfacen las condiciones del teorema de la función implícita. En caso afirmativo, escriba la identidad de equilibrio.
   (*c*) Encuentre (dY^*/dG_0) y explique sus implicaciones económicas.
2. Sean las funciones de oferta y demanda para un artículo
   $$Q_d = D(P, Y_0) \qquad (D_p < 0; \quad D_{Y_0} > 0)$$
   $$Q_s = S(P, T_0) \qquad (S_p > 0; \quad S_{T_0} < 0)$$
   donde Y_0 es el ingreso y T_0 es el impuesto sobre el artículo. Todas las derivadas son continuas.
   (*a*) Escriba la condición de equilibrio en una sola ecuación.
   (*b*) Compruebe si es aplicable el teorema de la función implícita. Si la respuesta es afirmativa, escriba la identidad de equilibrio.
   (*c*) Determine $(\partial P^*/\partial Y_0)$ y $(\partial P^*/\partial T_0)$, y explique sus implicaciones económicas.
   (*d*) Por medio de un procedimiento similar a (8.37), encuentre $(\partial Q^*/\partial Y_0)$ a partir de la función de la oferta y $(\partial Q^*/\partial T_0)$ de la función de la demanda. (¿Por qué no usar la función de la demanda para la primera y la función de la oferta para la última?)
3. Resuelva el problema 2 mediante el método de ecuaciones simultáneas.
4. Sean las funciones de la oferta y la demanda para un artículo
   $$Q_d = D(\underset{-}{P}, \underset{+}{t_0}) \qquad \text{y} \qquad Q_s = Q_{s0}$$

donde t_0 es la preferencia del consumidor por el artículo, y donde ambas derivadas parciales son continuas.

(*a*) ¿Cuál es el significado de los signos – y + debajo de las variables independientes P y t_0?
(*b*) Escriba la condición de equilibrio como una sola ecuación.
(*c*) ¿Es aplicable el teorema de la función implícita?
(*d*) ¿Cómo variaría el precio de equilibrio con la preferencia del consumidor?

5. Considere el siguiente modelo de ingreso nacional (sin considerar los impuestos):

$$Y - C(Y) - I(i) - G_0 = 0 \qquad (0 < C' < 1; \quad I' < 0)$$
$$kY + L(i) - M_{s0} = 0 \qquad (k = \text{constante positiva}; \quad L' < 0)$$

(*a*) ¿La primera ecuación es de la naturaleza de una condición de equilibrio?
(*b*) ¿Cuál es la cantidad demandada de dinero total en este modelo?
(*c*) Analice la estática comparativa del modelo cuando cambia la oferta de dinero (política monetaria) y cuando cambia el gasto público (política fiscal).

6. En el problema 5, suponga que mientras la demanda de dinero depende aún de Y como se especifica, ya no es afectada por la tasa de interés.

(*a*) ¿Cómo se debe revisar el enunciado del modelo?
(*b*) Escriba el nuevo jacobiano, llámelo $|J|'$. ¿$|J|'$ es numéricamente mayor o menor que $|J|$?
(*c*) ¿Aún se aplicaría la regla de la función implícita?
(*d*) Encuentre las nuevas derivadas estáticas comparativas.
(*e*) Al comparar la nueva $(\partial Y^*/\partial G_0)$ con la del problema 5, ¿qué se puede concluir acerca de la eficacia de la política fiscal en el nuevo modelo donde Y es independiente de i?
(*f*) Comparando la nueva $(\partial Y^*/\partial M_{s0})$ con la del problema 5, ¿qué se puede decir acerca de la eficacia de la política monetaria en el nuevo modelo?

8.7 Limitaciones de la estática comparativa

La estática comparativa es un área útil de estudio, porque en economía con frecuencia se está interesado en investigar la forma en que un cambio en un parámetro que causa desequilibrio afectará el estado de equilibrio de un modelo. Sin embargo, es importante entender que por su naturaleza particular la estática comparativa ignora el proceso de ajuste del equilibrio anterior al nuevo y también ignora la longitud del tiempo requerido en ese proceso de ajuste. Como una consecuencia, por necesidad debe ignorar también la posibilidad de que, como resultado de una inestabilidad inherente del modelo, es posible que no se logre nunca el nuevo equilibrio. El estudio del proceso de ajuste *per se* pertenece al campo de la *dinámica económica.* Cuando lleguemos a eso, prestaremos atención especial a la manera en que una variable cambia con el tiempo, y consideraremos explícitamente la cuestión de estabilidad de equilibrio.

Sin embargo, el tema importante de la dinámica debe esperar su turno. Mientras tanto, en la parte 4, emprenderemos el estudio del problema de *optimización*, una variedad especial sumamente importante del análisis de equilibrio con implicaciones relacionadas de estática comparativa (y complicaciones) propias.

Parte

4

Problemas de optimización

Capítulo 9

Optimización: una variedad especial de análisis de equilibrio

Cuando introdujimos el término equilibrio en el capítulo 3, hicimos una amplia distinción entre equilibrio final y equilibrio intermedio. En este último tipo, ejemplificado por el estudio de modelos de mercado e ingreso nacional, la interacción de ciertas fuerzas de oposición en el modelo, por ejemplo, las fuerzas de la oferta y la demanda en los modelos de mercado y las fuerzas de salidas y entradas en los modelos de ingreso, dicta un estado de equilibrio, si existe, en el que estas fuerzas de oposición se equilibran entre sí y, por lo tanto, se evita cualquier tendencia adicional a cambiar. El logro de este tipo de equilibrio es el resultado del equilibrio neutro de estas fuerzas y no requiere el esfuerzo consciente de alguien para lograr un objetivo específico. Los hogares consumidores que están detrás de las fuerzas de la demanda y las empresas que están detrás de las fuerzas de la oferta se esfuerzan por lograr una posición óptima bajo las circunstancias dadas; sin embargo, en lo que respecta al mercado, nadie intenta lograr un determinado precio de equilibrio o cantidad de equilibrio (a menos que, por supuesto, el gobierno pretenda controlar el precio). De manera similar, en la determinación del ingreso nacional, el equilibrio impersonal de salidas y entradas es lo que origina un estado de equilibrio, y no se requiere ningún esfuerzo consciente para lograr algún objetivo particular (p. ej., un intento para modificar un nivel de ingreso indeseable por medio de políticas monetarias o fiscales).

En esta parte del libro centramos la atención en el estudio del *equilibrio final*, donde definimos el estado de equilibrio como la posición óptima para una determinada unidad económica (un hogar, una empresa comercial o incluso una economía completa), en la cual la mencionada unidad económica estará luchando de manera deliberada por conseguir ese equilibrio. Como resultado, en este contexto, pero sólo en él, la advertencia anterior de que el equilibrio no implica conveniencia se vuelve irrelevante e inmaterial. En esta parte del libro, centramos la atención en las técnicas clásicas para localizar posiciones óptimas, las que hacen uso del cálculo diferencial. En el capítulo 13 describiremos los desarrollos más actuales, conocidos como programación matemática.

9.1 Valores óptimos y valores extremos

La economía es en esencia una ciencia de elección. Cuando se va a llevar a cabo un proyecto económico, como la producción de un nivel específico de producción, suele haber varias formas de lograrlo. Sin embargo, una (o más) de estas alternativas es más atractiva que otras desde el punto de vista de algún criterio, y la esencia del problema de optimización es elegir, con base en ese criterio especificado, la mejor alternativa posible.

El criterio de elección más común entre alternativas en economía es el objetivo de *maximizar* algo (por ejemplo, la ganancia de una empresa, la utilidad de un consumidor o la tasa de crecimiento de una empresa o de la economía de un país) o de *minimizar* algo (por ejemplo, el costo total de producción conocido). Económicamente, los problemas de maximización y minimización podríamos categorizarlos bajo el encabezado general de *optimización*, lo que significa "la búsqueda de lo mejor". Desde un punto de vista matemático puro, los términos *máximo* y *mínimo* no conllevan ninguna connotación de optimización. Por lo tanto, el término colectivo para máximo y mínimo, como conceptos matemáticos, es la designación más directa *extremo*, que significa un valor extremo.

En la formulación de un problema de optimización, lo primero que debemos hacer es delinear una *función objetivo* en la que la variable dependiente representa el objeto de maximización o minimización, y en la que el conjunto de variables independientes indica los objetos cuyas magnitudes económicas puede tomar y elegir la unidad económica en cuestión, con una visión a optimizar. Por lo tanto, se hará referencia a las variables independientes como *variables de elección*.[1] La esencia del proceso de optimización es hallar el conjunto de valores de las variables de elección que conducirán al extremo deseado de la función objetivo.

Por ejemplo, una empresa comercial podría buscar maximizar la ganancia π, es decir, maximizar la diferencia entre el ingreso total R y el costo total C. Puesto que, dentro del marco de un determinado estado de tecnología y una demanda de mercado particular para el producto de la empresa, R y C son funciones del nivel de producción Q, se deduce que π se puede expresar también como una función de Q:

$$\pi(Q) = R(Q) - C(Q)$$

Esta ecuación constituye la función objetivo pertinente, con π como el objeto de maximización y Q como la (única) variable de elección. El problema de optimización es, entonces, el de elegir el nivel de Q que maximiza a π. Note que si bien el nivel *óptimo* de π es por definición su nivel *máximo*, no se requiere que el nivel óptimo de la variable de elección Q por sí mismo sea máximo o mínimo.

A fin de presentar el problema en una forma más general para un análisis posterior (aunque sin limitarse aún a funciones objetivo de una variable), consideraremos la función general

$$y = f(x)$$

e intentaremos elaborar un procedimiento para hallar el nivel de x que maximizará o minimizará el valor de y. Supondremos que la función f es diferenciable continuamente.

[1] También se pueden llamar *variables de decisión*, o *variables de política*.

9.2 Máximo y mínimo relativo: criterio de la primera derivada

Puesto que la función objetivo $y = f(x)$ se expresa en la forma general, no hay restricción en cuanto a si es lineal o no lineal, o si es monótona o contiene partes crecientes o decrecientes. De entre los muchos tipos posibles de función compatibles con la forma de la función objetivo analizada en la sección 9.1, hemos seleccionado tres casos específicos que se ilustran en la figura 9.1. A pesar de su sencillez, las gráficas de la figura 9.1 nos deben dar una idea valiosa del problema de localizar el valor máximo o mínimo de la función $y = f(x)$.

Extremo relativo en relación con extremo absoluto

Si la función objetivo es una función constante, como en la figura 9.1*a*, todos los valores de la variable de elección x producirán el mismo valor de y, y la altura de cada punto de la gráfica de la función (como A o B o C) se podría considerar un máximo o, para el caso, un mínimo, o, de hecho, ninguno de los dos. En este caso, no hay elección importante que hacer en relación con el valor de x para la maximización o minimización de y.

En la figura 9.1*b*, la función es estrictamente creciente, y no hay máximo finito si el conjunto de números reales no negativos se toma como su dominio. Sin embargo, podríamos considerar que el punto final D a la izquierda (el intersecto y) representa un mínimo; de hecho, en este caso es el mínimo *absoluto* (o *global*) en la imagen de la función.

Los puntos E y F de la figura 9.1*c*, por otro lado, son ejemplos de un extremo *relativo* (o *local*), en el sentido de que cada uno representa un extremo en la vecindad inmediata del punto. El hecho de que el punto F es un mínimo relativo no es garantía de que también sea el mínimo global de la función, aunque podría suceder que éste fuera el caso. De manera similar, un punto máximo relativo como E podría ser o no ser un máximo global. Toma en cuenta también que una función puede tener muy bien varios extremos relativos, algunos de los cuales podrían ser máximos mientras que otros son mínimos.

En la mayor parte de los problemas económicos que trataremos, el interés principal, si no exclusivo, tendrá que ver con valores extremos distintos a los valores de una función en la orilla de un intervalo, porque en la mayoría de esta clase de problemas el dominio de la función objetivo está restringido al conjunto de números reales no negativos y, por lo tanto, un punto en el borde (a la izquierda) representará el nivel cero de la variable de elección, que con frecuencia no es de interés práctico. En realidad, el tipo de función que encontramos más a menudo en el análisis económico es el que se ilustra en la figura 9.1*c*, o alguna variante de ésta que contiene una sola inflexión en la curva. Por lo tanto, la explicación continúa sobre todo con referencia a la búsqueda de extremos *relativos* como los puntos E y F. Sin embargo, esto

FIGURA 9.1

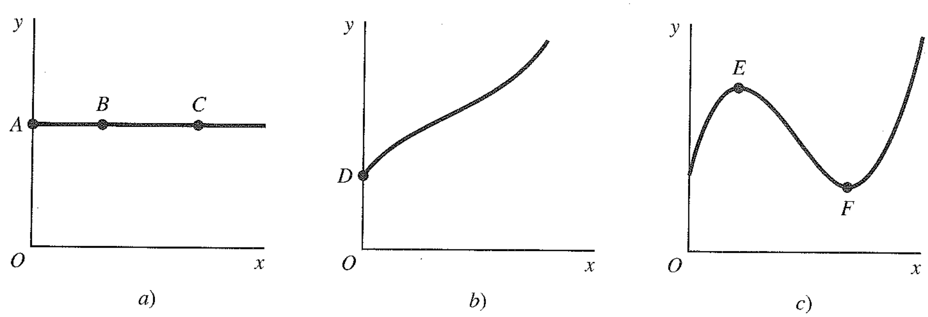
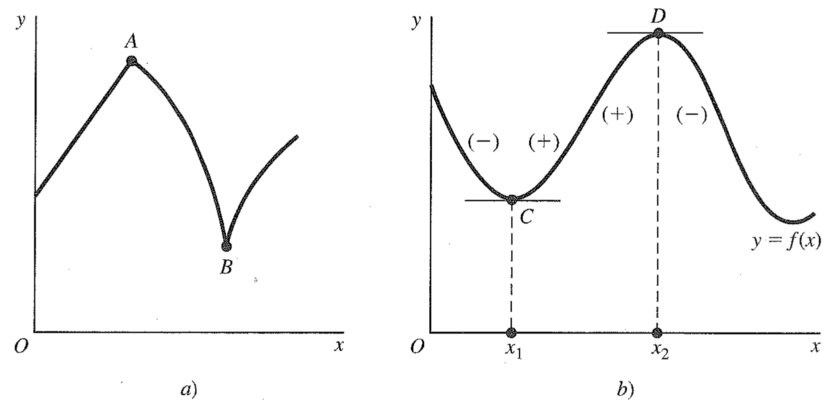


FIGURA 9.2



de ninguna manera nos impedirá conocer un máximo absoluto si lo deseamos, porque un máximo absoluto debe ser un máximo relativo o uno de los puntos en las orillas del dominio de la función. Por lo tanto, si conocemos todos los máximos relativos, sólo es necesario seleccionar el mayor y compararlo con los valores de la función en las orillas del dominio a fin de determinar el máximo absoluto. El mínimo absoluto de una función podemos hallarlo de manera análoga. En lo sucesivo, los valores extremos considerados serán *relativos* o *locales*, a menos que se indique lo contrario.

Criterio de la primera derivada

Como terminología alternativa, de ahora en adelante haremos referencia a la derivada de una función como su *primera* derivada (en lugar de derivada de *primer orden*). La razón de esto pronto será evidente.

Dada una función $y = f(x)$, la primera derivada $f'(x)$ desempeña un papel importante en la búsqueda de valores extremos. Esto se debe al hecho de que, si en $x = x_0$ ocurre un extremo relativo de la función, entonces (1) $f'(x_0)$ no existe, o (2) $f'(x_0) = 0$. La primera posibilidad se ilustra en la figura 9.2*a*, donde los puntos A y B muestran valores extremos relativos de y y, sin embargo, la derivada no está definida en ninguno de estos puntos picudos. Puesto que en la presente explicación estamos suponiendo que $y = f(x)$ es continua y posee una derivada continua, estamos descartando de hecho puntos picudos. Para funciones suaves, los valores extremos relativos pueden ocurrir sólo donde la primera derivada y tiene un valor cero. Esto se ilustra mediante los puntos C y D en la figura 9.2*b*, los cuales representan valores extremos, y se caracterizan por una pendiente cero, $f'(x_1) = 0$ y $f'(x_2) = 0$. También es fácil ver que cuando la pendiente no es cero es imposible que se tenga un mínimo relativo (el fondo de un valle) o un máximo relativo (la cúspide de una colina). Por esta razón, en el contexto de funciones suaves podemos tomar la condición $f'(x) = 0$ como una condición *necesaria* para un extremo relativo (ya sea máximo o mínimo).

No obstante, debemos agregar que una pendiente cero, aunque *necesaria*, *no* es *suficiente* para establecer un extremo relativo. En breve presentamos un ejemplo del caso donde una pendiente cero no tiene relación con un extremo. Sin embargo, al anexar cierta estipulación a la condición de pendiente cero, podemos obtener un criterio decisivo para un extremo relativo. Esto podemos expresarlo como sigue:

Criterio de la primera derivada para el extremo relativo Si la primera derivada de una función $f(x)$ en $x = x_0$ es $f'(x_0) = 0$, entonces el valor de la función en x_0, $f(x_0)$, será *a*) Un *máximo* relativo si el signo de la derivada $f'(x)$ es positivo a la izquierda inmediata del punto x_0 y es negativo a su derecha inmediata.

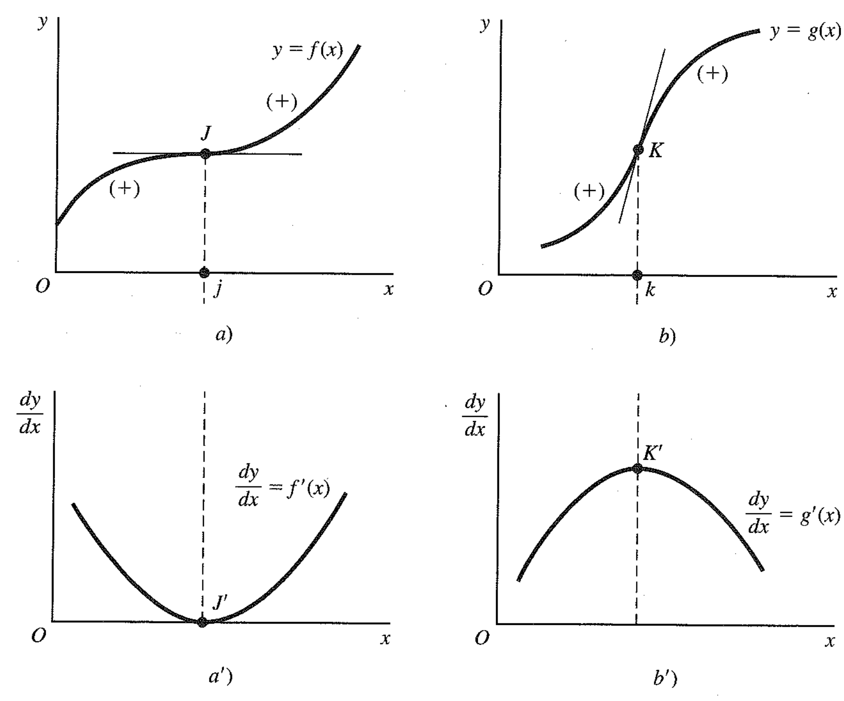
b) Un *mínimo* relativo si el signo de $f'(x)$ es negativo a la izquierda inmediata de x_0 a su derecha inmediata es positivo).

c) Ni un máximo relativo ni un mínimo relativo si $f'(x)$ tiene el mismo signo en la izquierda inmediata y la derecha inmediata del punto x_0.

El valor x_0 se llamará *valor crítico* de x si $f'(x_0) = 0$ y $f(x_0)$ se denominará *valor estacionario* de y (o de la función f). El punto con coordenadas x_0 y $f(x_0)$ se puede llamar, en consecuencia, *punto estacionario*. (La razón fundamental para la palabra *estacionario* debe ser obvia, siempre que la pendiente sea cero, el punto en cuestión nunca se sitúa en una cuesta ascendente o descendente, sino más bien está en una posición de reposo.) Entonces, de manera gráfica, la primera posibilidad listada en esta prueba establecerá el punto estacionario como la cúspide de una colina, como el punto D en la figura 9.2*b*, mientras que la segunda posibilidad establecerá el punto estacionario como el fondo de un valle, por ejemplo el punto C en el mismo diagrama. Sin embargo, ante la existencia de una tercera posibilidad, aún no analizada, no podemos considerar la condición $f'(x) = 0$ como una *condición suficiente* para un extremo relativo. Pero ahora se ve que, si satisfacemos la condición necesaria $f'(x) = 0$, *entonces* la estipulación de cambio de signo de la derivada nos puede servir como una *condición suficiente* para un máximo o mínimo relativo, dependiendo de la dirección del cambio de signo.

A continuación explicamos la tercera posibilidad. En la figura 9.3*a* mostramos que la función f logra una pendiente cero en el punto J (cuando $x = j$). Aun cuando $f'(j)$ es cero, lo que hace a $f(j)$ un valor estacionario, la derivada no cambia su signo de un lado de $x = j$ al otro; por lo tanto, de acuerdo con el criterio de la primera derivada, el punto J no produce un máxi-

FIGURA 9.3



mo ni un mínimo, como se confirma debidamente mediante la gráfica de la función. Además, ejemplifica lo que se conoce como un *punto de inflexión*.

El rasgo característico de un punto de inflexión es que, en ese punto, la función derivada (en vez de la primitiva) alcanza un valor extremo. Puesto que su valor extremo puede ser un máximo o un mínimo, tenemos dos tipos de puntos de inflexión. En la figura 9.3a', donde hemos graficado la derivada $f'(x)$, vemos que su valor es cero cuando $x = j$ (véase el punto J'), pero es positivo en ambos lados del punto J'; esto hace que J' sea un punto *mínimo* de la función derivada $f'(x)$.

El otro tipo de punto de inflexión se representa en la figura 9.3b, donde la pendiente de la función $g(x)$ aumenta hasta que alcanza el punto k y disminuye después. En consecuencia, la gráfica de la función derivada $g'(x)$ asume la forma que muestra la figura 9.3b' en la que el punto K' da un valor *máximo* de la función derivada $g'(x)$.[2]

En resumen, un extremo relativo debe ser un valor estacionario, pero un valor estacionario se puede relacionar con un extremo relativo o un punto de inflexión. Para hallar un máximo o mínimo relativo de una función particular, el procedimiento debe ser hallar primero los valores estacionarios de la función donde se satisface la condición $f'(x) = 0$ y, después, aplicar el criterio de la primera derivada para determinar si cada uno de los valores estacionarios es un máximo relativo, un mínimo relativo o ninguno.

Ejemplo 1

Halle los extremos relativos de la función

$$y = f(x) = x^3 - 12x^2 + 36x + 8$$

Primero, encontramos que la función derivada es

$$f'(x) = 3x^2 - 24x + 36$$

Para obtener los valores críticos, es decir, los valores de x que satisfacen la condición $f'(x) = 0$, igualamos a cero la función derivada y obtenemos la ecuación cuadrática

$$3x^2 - 24x + 36 = 0$$

Mediante factorización, o al aplicar la fórmula general de la ecuación cuadrática, obtenemos el siguiente par de raíces (soluciones):

$$x_1^* = 6 \quad [\text{en la cual se tiene } f'(6) = 0 \text{ y } f(6) = 8]$$
$$x_2^* = 2 \quad [\text{en la cual se tiene } f'(2) = 0 \text{ y } f(2) = 40]$$

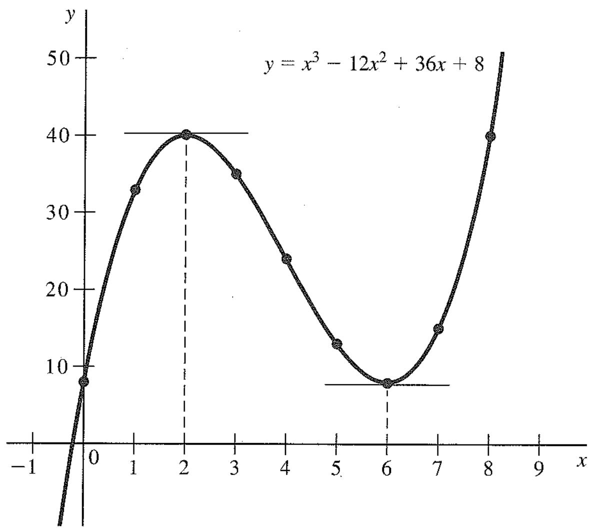
Puesto que $f'(6) = f'(2) = 0$, estos dos valores de x son los valores críticos deseados.

Es fácil comprobar que, en la vecindad inmediata de $x = 6$, tenemos $f'(x) < 0$ para $x < 6$ y $f'(x) > 0$ para $x > 6$; así, el valor de la función $f(6) = 8$ es un mínimo relativo. De manera similar, puesto que en la vecindad inmediata de $x = 2$, encontramos que $f'(x) > 0$ para $x < 2$ y $f'(x) < 0$ para $x > 2$, el valor de la función $f(2) = 40$ es un máximo relativo.

La figura 9.4 ilustra la gráfica de la función de este ejemplo. Esta gráfica podemos usarla para comprobar la ubicación de los valores extremos obtenidos mediante el criterio de la

[2] Note que un valor cero en la primera derivada, si bien es una condición necesaria para un extremo relativo, *no* se requiere para un punto de inflexión, porque la derivada $g'(x)$ tiene un valor positivo en $x = k$, y el punto K aún es un punto de inflexión.

FIGURA 9.4



primera derivada. Sin embargo, en realidad, en la mayor parte de los casos la " utilidad" fluye en la dirección opuesta: los valores extremos derivados matemáticamente ayudarán a trazar la gráfica. Desde el punto de vista ideal, el trazo preciso de una gráfica requiere conocer el valor de la función en cada punto del dominio; pero en realidad, en la práctica, sólo se seleccionan algunos puntos del dominio para fines de trazo, y el resto de los puntos se suelen completar por interpolación. La falla de esta práctica es que, a menos que por coincidencia se dé con el(los) punto(s) estacionario(s), se perderá la localización exacta del punto o puntos de retorno en la curva. Ahora, con el criterio de la primera derivada a disposición, podemos determinar con precisión estos puntos de retorno.

Ejemplo 2

Encuentre el extremo relativo de la función de costo promedio

$$AC = f(Q) = Q^2 - 5Q + 8$$

La derivada en este caso es $f'(Q) = 2Q - 5$, una función lineal. Al igualar a cero $f'(Q)$, obtenemos la ecuación lineal $2Q - 5 = 0$, que tiene la única raíz $Q^* = 2.5$. En este caso, éste es el único valor crítico. Para aplicar el criterio de la primera derivada, procedemos a hallar los valores de la derivada en, por ejemplo, $Q = 2.4$ y $Q = 2.6$, respectivamente. Puesto que $f'(2.4) = -0.2 < 0$ mientras que $f'(2.6) = 0.2 > 0$, concluimos que el valor estacionario $AC = f(2.5) = 1.75$ representa un mínimo relativo. La gráfica de la función de este ejemplo es en realidad una curva en forma de U, de modo que el mínimo relativo ya localizado también será el mínimo absoluto. Conocer la localización exacta de este punto debe ser de gran ayuda para graficar la curva AC.

EJERCICIO 9.2

1. Halle los valores estacionarios de las siguientes funciones (compruebe si son máximos o mínimos relativos o puntos de inflexión), suponiendo que el dominio es el conjunto de los números reales:

   (*a*) $y = -2x^2 + 8x + 7$ (*b*) $y = 5x^2 + x$ (*c*) $y = 3x^2 + 3$ (*d*) $y = 3x^2 - 6x + 2$

2. Encuentre los valores estacionarios de las siguientes funciones (compruebe si son máximos o mínimos relativos o puntos de inflexión), suponiendo que el dominio es el intervalo $0, \infty$):
   (*a*) $y = x^3 - 3x + 5$
   (*b*) $y = \frac{1}{3}x^3 - x^2 + x + 10$
   (*c*) $y = -x^3 + 4.5x^2 - 6x + 6$
3. Demuestre que la función $y = x + 1/x$ (con $x \neq 0$) tiene dos extremos relativos, uno un máximo y el otro un mínimo. ¿El "mínimo" es mayor o menor que el "máximo"? ¿Cómo es posible este resultado paradójico?
4. Sea $T = \phi(x)$ una función que representa un *total* (por ejemplo, producto total o costo total):
   (*a*) Escriba las expresiones para la función *marginal M* y la función *promedio A*.
   (*b*) Demuestre que, cuando A alcanza un extremo relativo, M y A deben tener el mismo valor.
   (*c*) ¿Qué principio general sugiere esto para el trazo de una curva marginal y una curva promedio en el mismo diagrama?
   (*d*) ¿Qué se puede concluir acerca de la elasticidad de la función total T en el punto donde A alcanza un valor extremo?

9.3 Derivada segunda y derivadas de orden superior

Hasta el momento hemos considerado sólo la primera derivada $f'(x)$ de una función $y = f(x)$. Introducimos ahora el concepto de *segunda derivada* (abreviatura para *derivada de segundo orden*), y derivadas, incluso, de orden superior. Éstas nos permitirán desarrollar criterios alternativos para localizar los extremos relativos de una función.

Derivada de una derivada

Puesto que la primera derivada $f'(x)$ es por sí misma una función de x, ésta también debe ser diferenciable respecto a x, siempre que sea continua y suave. El resultado de esta diferenciación, conocida como la segunda derivada de la función f, se denota mediante

$f''(x)$ donde el apóstrofo doble indica que $f(x)$ ha sido diferenciada respecto a x dos veces, y donde la expresión (x) que sigue al apóstrofo doble indica que la segunda derivada es de nuevo una función de x

o bien,

$\dfrac{d^2y}{dx^2}$ donde la notación proviene de la consideración de que la segunda derivada significa, de hecho, $\dfrac{d}{dx}\left(\dfrac{dy}{dx}\right)$; por ello, la d^2 (léase: "d dos") en el numerador y dx^2 (léase: "dx cuadrada") en el denominador de este símbolo.

Si la segunda derivada $f''(x)$ existe para todos los valores de x del dominio, se dice que la función $f(x)$ es *diferenciable dos veces*; si, además, $f''(x)$ es continua, se dice que la función $f(x)$ es *diferenciable continuamente dos veces*. Así como la notación $f \in C^{(1)}$ o $f \in C'$ se usa con frecuencia para indicar que la función f es diferenciable de manera continua, podemos emplear una notación análoga

$$f \in C^{(2)} \qquad \text{o} \qquad f \in C''$$

para indicar que f es diferenciable continuamente dos veces.

Como una función de x de la segunda derivada se puede diferenciar respecto a x de nuevo para producir una *tercera* derivada, la cual a su vez puede ser la fuente de una *cuarta* derivada, y así sucesivamente hasta el infinito, siempre y cuando se satisfaga la condición de diferenciabilidad. Estas derivadas de orden superior se simbolizan siguiendo la misma línea que la segunda derivada:

$$f'''(x),\ f^{(4)}(x),\ \ldots,\ f^{(n)}(x) \qquad \text{[con los superíndices encerrados en ( )]}$$

o bien,

$$\frac{d^3y}{dx^3},\ \frac{d^4y}{dx^4},\ \ldots,\ \frac{d^ny}{dx^n}$$

La última podemos escribirla también como $\frac{d^n}{dx^n}y$, donde la parte $\frac{d^n}{dx^n}$ sirve como un símbolo operador que instruye a tomar la n-ésima derivada de (alguna función) respecto a x.

Casi todas las funciones *específicas* con las que trabajaremos poseen derivadas continuas hasta cualquier orden que se desee; es decir, son continuamente diferenciables cualquier número de veces. Siempre que usemos una función *general*, como $f(x)$, suponemos siempre que tiene derivadas hasta cualquier orden que se requiera.

Ejemplo 1

Halle las derivadas primera a quinta de la función

$$y = f(x) = 4x^4 - x^3 + 17x^2 + 3x - 1$$

Las derivadas deseadas son:

$$\begin{aligned} f'(x) &= 16x^3 - 3x^2 + 34x + 3 \\ f''(x) &= 48x^2 - 6x + 34 \\ f'''(x) &= 96x - 6 \\ f^{(4)}(x) &= 96 \\ f^{(5)}(x) &= 0 \end{aligned}$$

En este ejemplo (polinomial) particular, se nota que cada función derivada sucesiva surge como un polinomio de orden menor, de cúbico a cuadrático, a lineal, a constante. Se nota también que la quinta derivada, siendo la derivada de una constante, es igual a cero para todos los valores de x; por lo tanto, podríamos haber escrito como $f^{(5)}(x) \equiv 0$ también. La ecuación $f^{(5)}(x) = 0$ debemos distinguirla de la ecuación $f^{(5)}(x_0) = 0$ (cero en x_0 solamente), así como entender que la expresión $f^{(5)}(x) \equiv 0$ no significa que la quinta derivada no exista; en realidad existe, y tiene el valor de cero.

Ejemplo 2

Determine las primeras cuatro derivadas de la función racional

$$y = g(x) = \frac{x}{1+x} \qquad (x \neq -1)$$

Estas derivadas las encontramos por medio de la regla del cociente o después de reescribir la función como $y = x(1+x)^{-1}$, por la regla del producto:

$$\left.\begin{aligned} g'(x) &= (1+x)^{-2} \\ g''(x) &= -2(1+x)^{-3} \\ g'''(x) &= 6(1+x)^{-4} \\ g^{(4)}(x) &= -24(1+x)^{-5} \end{aligned}\right\} \qquad (x \neq -1)$$

En este caso, la derivación repetida no tiende a simplificar las expresiones derivadas posteriores.

Tome en cuenta que, al igual que la función primitiva $g(x)$, las derivadas sucesivas obtenidas son por sí mismas funciones de x. Sin embargo, dados valores específicos de x, estas funciones derivadas tomarán entonces valores específicos. Cuando $x = 2$, por ejemplo, la segunda derivada del ejemplo 2 podemos evaluarla como

$$g''(2) = -2(3)^{-3} = \frac{-2}{27}$$

y de manera similar para otros valores de x. Es de suma importancia entender que para evaluar esta segunda derivada $g''(x)$ en $x = 2$, debemos obtener primero $g''(x)$ de $g'(x)$ y sustituir después $x = 2$ en la ecuación para $g''(x)$. Es *incorrecto* sustituir $x = 2$ en $g(x)$ o $g'(x)$ *antes* del proceso de diferenciación que conduce a $g''(x)$.

Interpretación de la segunda derivada

La función derivada $f'(x)$ mide la razón de cambio de la función f. De la misma manera, la función segunda derivada f'' es la medida de la razón de cambio de la primera derivada f'; en otras palabras, la segunda derivada mide la *razón de cambio* de la *razón de cambio* de la función original f. Dicho de otro modo, con un determinado incremento infinitesimal en la variable independiente x desde un punto $x = x_0$,

$$\left.\begin{matrix} f'(x_0) > 0 \\ f'(x_0) < 0 \end{matrix}\right\} \text{ significa que el } \textit{valor de la función} \text{ tiende a } \left\{\begin{matrix} \text{aumentar} \\ \text{disminuir} \end{matrix}\right.$$

mientras que, en relación con la segunda derivada,

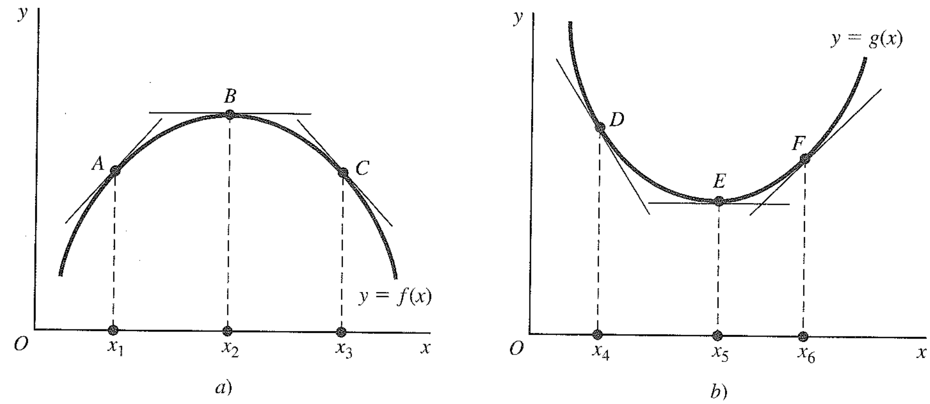
$$\left.\begin{matrix} f''(x_0) > 0 \\ f''(x_0) < 0 \end{matrix}\right\} \begin{matrix} \text{significa que } \textit{la pendiente de la curva} \\ \text{tiende a} \end{matrix} \left\{\begin{matrix} \text{aumentar} \\ \text{disminuir} \end{matrix}\right.$$

Por lo tanto, una primera derivada positiva acoplada con una segunda derivada positiva en $x = x_0$ implica que la pendiente de la curva en ese punto es *positiva y creciente*; en otras palabras, el valor de la función es creciente a una tasa creciente. De igual manera, una primera derivada positiva con una segunda derivada negativa indica que la pendiente de la curva es *positiva pero decreciente*, el valor de la función es creciente a una tasa decreciente. Una primera derivada negativa se puede interpretar de manera análoga, pero en este caso conviene hacer una advertencia: cuando $f'(x_0) < 0$ y $f''(x_0) > 0$, la pendiente de la curva es *negativa y creciente*, pero esto *no* significa que la pendiente esté cambiando, por ejemplo, de (–10) a (–11); por el contrario, el cambio debe ser desde (–11), hasta (–10), porque (–11) es un número más pequeño que (–10); en otras palabras, la pendiente negativa tiende a ser *menos* inclinada a medida que x aumenta. Por último, cuando $f'(x_0) < 0$ y $f''(x_0) < 0$, la pendiente de la curva debe ser *negativa y decreciente*. Esto se refiere a una pendiente negativa que tiende a ser *cada vez más inclinada* cuando x aumenta.

Todo esto se aclara con una explicación gráfica. En la figura 9.5*a* se ilustra una función con $f''(x) < 0$ en todas partes. Puesto que la pendiente debe disminuir de manera permanente cuando x aumenta sobre la gráfica, se pasará, cuando se va de izquierda a derecha, por un punto A con una pendiente positiva, luego un punto B con pendiente cero, y después un punto C con una pendiente negativa. Podría suceder también que una función con $f''(x) < 0$ se caracterice por $f'(x) > 0$ en todas partes y, por lo tanto, su gráfica sea sólo la porción creciente de una curva en forma de U invertida, o bien, con $f'(x) < 0$ en todo lugar, la porción decreciente de esa curva.

En la figura 9.5*b* se ilustra el caso opuesto de una función con $f''(x) > 0$ en todas partes. Aquí, a medida que se pasa por los puntos D a E a F, la pendiente crece de modo permanente,

FIGURA 9.5



pasa de negativa a cero y luego a positiva. De nuevo, se agrega que una función caracterizada por $f''(x) > 0$ en todas partes se puede representar, dependiendo de la especificación de la primera derivada, como la porción decreciente o creciente de una curva en forma de U.

En la figura 9.5 es evidente que la segunda derivada $f''(x)$ se relaciona con la *curvatura* de una gráfica y determina la forma en que la curva tiende a doblarse por sí misma. Para describir los dos tipos de curvaturas analizadas, se hace referencia a la de la figura 9.5*a* como *estrictamente cóncava,* y a la de la figura 9.5*b* como *estrictamente convexa*. Una función cuya gráfica es estrictamente cóncava (convexa) se llama *función estrictamente cóncava* (*convexa*). La representación geométrica de una función estrictamente cóncava es de la siguiente manera. Si tomamos cualquier par de puntos M y N sobre su curva y se unen mediante una recta, el segmento MN debe quedar por completo *debajo* de la curva, excepto en los puntos M y N. La caracterización de una función estrictamente convexa la obtenemos al sustituir la palabra *arriba* por la palabra *debajo* en el último enunciado. Pruebe lo anterior en la figura 9.5. Si relajamos un poco la condición de caracterización, de modo que permitamos que el segmento de recta MN quede *ya sea* debajo *o* a lo largo de (coincidiendo con) la curva, entonces estaremos describiendo una *función cóncava*, sin el adverbio *estrictamente*. De manera similar, si el segmento de recta MN se ubica *ya sea* arriba *o* a lo largo de la curva, entonces la función es *convexa*, de nuevo sin el adverbio *estrictamente*. Note que, puesto que el segmento de recta MN puede coincidir con una curva (no estrictamente) cóncava o convexa, la última podría contener un segmento lineal. En contraste, una curva *estrictamente* cóncava o convexa nunca contiene un segmento lineal en ninguna parte. Se deduce que si bien una función estrictamente cóncava (convexa) es de manera automática una función cóncava (convexa), lo contrario no es cierto.[3]

De la explicación anterior sobre la segunda derivada podemos inferir que si la segunda derivada $f''(x)$ es negativa para *toda* x, entonces la función primitiva $f(x)$ debe ser estrictamente cóncava. De manera similar, $f(x)$ debe ser estrictamente convexa si $f''(x)$ es positiva para *toda* x. A pesar de esto, *no* es válido invertir esta inferencia y decir que, si $f(x)$ es estrictamente cóncava (convexa), entonces $f''(x)$ debe ser negativa (positiva) para toda x. Esto es porque, en ciertos casos excepcionales, la segunda derivada podría tener un valor *cero* en un punto estacionario sobre tal curva. Un ejemplo de esto lo encontramos en la función $y = f(x) = x^4$ que describe una curva estrictamente convexa, pero cuyas derivadas

$$f'(x) = 4x^3 \qquad f''(x) = 12x^2$$

[3] Estos conceptos se analizan con más detalle en la sección 11.5.

indican que, en el punto estacionario donde $x = 0$, el valor de la segunda derivada es $f''(0) = 0$. Sin embargo, advierta que en cualquier otro punto, con $x \neq 0$, la segunda derivada de esta función tiene el signo positivo (esperado). Por lo tanto, aparte de la posibilidad de un valor cero en un punto estacionario, podemos esperar en general que la segunda derivada de una función estrictamente cóncava o convexa se adhiera a un solo signo algebraico.

Para otros tipos de función, la segunda derivada podría tomar tanto valores positivos como negativos, dependiendo del valor de x. En la figura 9.3a y b, por ejemplo, tanto $f(x)$ como $g(x)$ experimentan un cambio de signo en la segunda derivada en sus respectivos puntos de inflexión J y K. De acuerdo con la figura 9.3a', la pendiente de $f'(x)$, es decir, el valor de $f''(x)$, cambia de negativa a positiva en $x = j$; con la pendiente de $g'(x)$, es decir, el valor de $g''(x)$, ocurre lo contrario, como se observa en la figura 9.3b'. Traducido en términos de curvatura, esto significa que la gráfica de $f(x)$ cambia de estrictamente cóncava a estrictamente convexa en el punto J, mientras que la gráfica de $g(x)$ tiene el cambio inverso en el punto K. En consecuencia, en lugar de caracterizar un punto de inflexión como un punto donde la primera derivada alcanza un valor extremo, es posible caracterizarlo opcionalmente como un punto donde la función experimenta un cambio de curvatura o un cambio en el signo de su segunda derivada.

Una aplicación

Las dos curvas de la figura 9.5 ejemplifican las gráficas de funciones cuadráticas, las cuales podemos expresar por lo común en la forma

$$y = ax^2 + bx + c \qquad (a \neq 0)$$

Del análisis de la segunda derivada, podemos ahora obtener una forma conveniente de determinar si una función cuadrática particular tendrá una gráfica estrictamente convexa (en forma de U) o estrictamente cóncava (en forma de U invertida).

Puesto que la segunda derivada de la función cuadrática citada es $d^2y/dx^2 = 2a$, esta derivada tendrá siempre el mismo signo algebraico que el coeficiente a. Teniendo en cuenta que una segunda derivada positiva implica una función estrictamente convexa, se infiere que un coeficiente positivo a en la función cuadrática anterior da lugar a una gráfica en forma de U. Por el contrario, un coeficiente negativo a da lugar a una curva estrictamente cóncava, en forma de una U invertida.

Como sugerimos al final de la sección 9.2, el extremo relativo de esta función probará ser también su extremo absoluto, porque en una función cuadrática podemos encontrar sólo un valle o una cúspide, evidente en una U o U invertida, respectivamente.

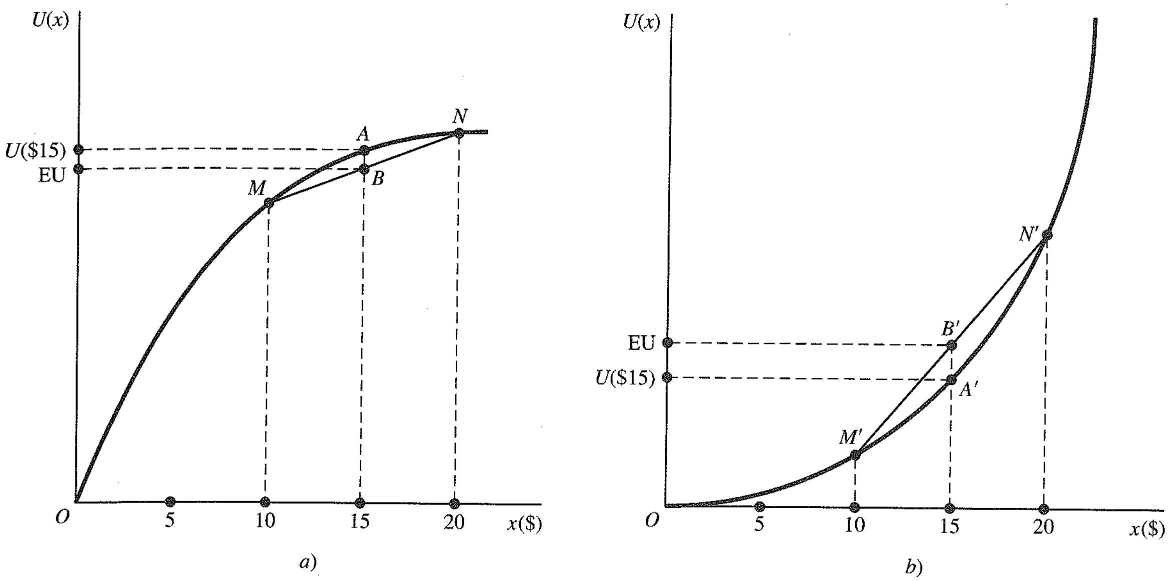
Actitudes hacia el riesgo

La aplicación más común del concepto de utilidad marginal se relaciona con el consumo de bienes. En otra aplicación se considera la utilidad marginal del *ingreso*, y en la presente explicación analizamos el *pago* de un juego de apuesta, y usamos este concepto para distinguir entre diferentes actitudes de los individuos hacia el riesgo.

Considere el juego donde, por una suma fija de dinero pagada por adelantado (el *costo* del juego), se puede lanzar un dado y obtener \$10 si aparece un número impar, o \$20 si el número es par. En vista de la probabilidad igual de los dos resultados, el *valor esperado de pago* matemáticamente es

$$\text{VE} = 0.5 \times \$10 + 0.5 \times \$20 = \$15$$

FIGURA 9.6



Se considera que se trata de un *juego justo*, o *apuesta justa*, si el costo del juego es exactamente \$15. A pesar de su carácter de justo, el juego aún implica un riesgo, porque aunque conocemos la distribución de probabilidad de los dos posibles resultados, desconocemos el resultado real de un juego particular. Por lo tanto, las personas que tienen "aversión al riego" declinarían de manera congruente jugar tal juego. Por otro lado, hay personas que "aman" o "prefieren" el riesgo y estarían dispuestas a jugar juegos justos, o incluso juegos con posibilidades en contra (es decir, con el costo del juego superior al valor esperado de pago).

La explicación para tales actitudes hacia el riesgo las encontramos fácilmente en las diferentes funciones de utilidad que posee la gente. Suponga que un jugador tiene la función de utilidad estrictamente cóncava $U = U(x)$, ilustrada en la figura 9.6*a*, donde x denota el pago, con $U(0) = 0$, $U'(x) > 0$ (utilidad marginal positiva de ingreso o pago) y $U''(x) < 0$ (utilidad marginal decreciente) para toda x. La decisión económica que enfrenta esta persona tiene que ver con la elección entre dos cursos de acción. Primero, al no jugar, la persona ahorra el costo de \$15 del juego (= VE) y, por lo tanto, disfruta el nivel de utilidad U(\$15), medido por la altura del punto A sobre la curva. Segundo, al jugar, la persona tiene 0.5 de probabilidades de recibir \$10 y, por lo tanto, disfrutar de U(\$10) (véase el punto M), más 0.5 probabilidad de recibir \$20 y, por consiguiente, disfrutar de U(\$20) (véase el punto N). La *utilidad esperada de jugar* es, por lo tanto, igual a

$$\text{UE} = 0.5 \times U(\$10) + 0.5 \times U(\$20)$$

la cual, siendo el promedio de la altura de M y de N, se mide por la altura del punto B, el punto medio del segmento de recta MN. Puesto que una función de utilidad estrictamente cóncava tiene la propiedad de que el segmento de recta MN debe quedar debajo del arco MN, el punto B debe ser menor que el punto A; es decir, UE, la utilidad esperada por jugar, no satisface la utilidad del costo del juego, y éste se debe evitar. Por esta razón, una función de utilidad estrictamente cóncava se relaciona con el comportamiento de aversión al riesgo.

Para una persona que gusta del riesgo el proceso de decisión es análogo, pero se hará la elección opuesta, porque ahora la función de utilidad pertinente es una estrictamente convexa.

En la figura 9.6*b*, $U(\$15)$, la utilidad de conservar los \$15 al no jugar el juego se muestra por medio del punto A' en la curva, y UE, la utilidad esperada de jugar, está dada por B', el punto medio en el segmento de recta $M'N'$. Pero esta vez el segmento de recta $M'N'$ está arriba del arco $M'N'$, y el punto B' está arriba del punto A'. Así, sin duda, hay un incentivo positivo para jugar el juego. En oposición a la situación de la figura 9.6*a*, podemos relacionar una función de utilidad estrictamente convexa con el comportamiento de preferir el riesgo.

EJERCICIO 9.3

1. Encuentre las derivadas segunda y tercera de las siguientes funciones:
   (*a*) $ax^2 + bx + c$ (*c*) $\dfrac{3x}{1-x}$ $(x \neq 1)$
   (*b*) $7x^4 - 3x - 4$ (*d*) $\dfrac{1+x}{1-x}$ $(x \neq 1)$
2. ¿Cuál de las siguientes funciones cuadráticas es estrictamente convexa?
   (*a*) $y = 9x^2 - 4x + 8$ (*c*) $u = 9 - 2x^2$
   (*b*) $w = -3x^2 + 39$ (*d*) $v = 8 - 5x + x^2$
3. Dibuje (*a*) una curva cóncava que *no* es estrictamente cóncava y (*b*) una curva que califica al mismo tiempo como curva cóncava y convexa.
4. Dada la función $y = a - \dfrac{b}{c+x}$ $(a, b, c > 0;\ x \geq 0)$, determine la forma general de su gráfica al examinar (*a*) sus derivadas primera y segunda, (*b*) su intersección con la vertical y (*c*) el límite de y cuando x tiende a infinito. Si esta función se va a usar como función de consumo, ¿cómo se deben restringir los parámetros a fin de que tenga sentido económico?
5. Dibuje la gráfica de una función $f(x)$ tal que $f'(x) \equiv 0$, y la gráfica de una función $g(x)$ tal que $g'(3) = 0$. Resuma en un enunciado la diferencia esencial entre $f(x)$ y $g(x)$ en términos del concepto de punto estacionario.
6. Se dice que una persona que no es contraria ni favorable al riesgo (indiferente hacia un juego justo) es "neutral al riesgo".
   (*a*) ¿Qué clase de función de utilidad usaría para caracterizar a esta persona?
   (*b*) Usando el juego de lanzar un dado detallado en el texto, describa la relación entre $U(\$15)$ y UE para la persona neutral al riesgo.

9.4 Criterio de la segunda derivada

Volviendo al par de puntos extremos B y E de la figura 9.5 y recordando la relación recién establecida entre la segunda derivada y la curvatura de una curva, debemos poder ver la validez del segundo criterio para un extremo relativo:

Criterio de la segunda derivada para el extremo relativo Si el valor de la primera derivada de una función f en $x = x_0$ es $f'(x_0) = 0$, entonces el valor de la función en x_0, $f(x_0)$, será

a) Un *máximo* relativo si el valor de la segunda derivada en x_0 es $f''(x_0) < 0$.
b) Un *mínimo* relativo si el valor de la segunda derivada en x_0 es $f''(x_0) > 0$.

En general, es más conveniente usar esta prueba que el criterio de la primera derivada, porque no requiere comprobar el signo de la derivada a la izquierda y a la derecha de x_0; sin embargo, tiene la desventaja de que no podemos sacar una conclusión inequívoca en el caso de que $f''(x_0) = 0$, porque el valor estacionario $f(x_0)$ puede ser *ya sea* un máximo relativo, *o bien*, un mínimo relativo, *o* incluso un valor de inflexión.[4] Cuando encontremos la situación de $f''(x_0) = 0$, debemos volver al criterio de la primera derivada, o recurrir a otro criterio, que se desarrollará en la sección 9.6, la cual tiene que ver con la tercera derivada o incluso derivadas superiores. Sin embargo, para la mayor parte de los problemas de economía, la prueba de la segunda derivada por lo común es adecuada para determinar un máximo o mínimo relativo.

Ejemplo 1

Encuentre el extremo relativo de la función

$$y = f(x) = 4x^2 - x$$

Las derivadas primera y segunda son

$$f'(x) = 8x - 1 \qquad \text{y} \qquad f''(x) = 8$$

Al igualar $f'(x)$ a cero y resolver la ecuación resultante, encontramos que el (único) valor crítico es $x^* = \frac{1}{8}$, que produce el (único) valor estacionario $f\left(\frac{1}{8}\right) = -\frac{1}{16}$. Debido a que la segunda derivada es positiva (en este caso es de hecho positiva para cualquier valor de x), el extremo se establece como un mínimo. Además, puesto que la función dada se representa como una curva en U, el mínimo relativo es también un mínimo absoluto.

Ejemplo 2

Encuentre los extremos relativos de la función

$$y = g(x) = x^3 - 3x^2 + 2$$

Las dos primeras derivadas de esta función son

$$g'(x) = 3x^2 - 6x \qquad \text{y} \qquad g''(x) = 6x - 6$$

Si igualamos $g'(x)$ a cero y resolvemos la ecuación cuadrática, $3x^2 - 6x = 0$, obtenemos los valores críticos $x_1^* = 2$ y $x_2^* = 0$, que a su vez producen los dos valores estacionarios:

$$g(2) = -2 \qquad [\text{un mínimo porque} \quad g''(2) = 6 > 0]$$
$$g(0) = 2 \qquad [\text{un máximo porque} \quad g''(0) = -6 < 0]$$

Condiciones necesarias en relación con suficientes

Como en el criterio de la primera derivada, la condición de pendiente cero $f'(x) = 0$ desempeña el papel de una condición *necesaria* en el criterio de la segunda derivada. Puesto que esta condición se basa en la derivada de primer orden, suele llamarse *condición de primer orden*. Una vez que vemos que se satisface la condición de primer orden en $x = x_0$, el signo negativo (positivo) de $f''(x_0)$ es *suficiente* para establecer el valor estacionario en cuestión como un máximo relativo (mínimo). Estas condiciones suficientes, que se basan en la derivada de segundo orden, por lo común se denominan *condiciones de segundo orden*.

[4] Para ver que un punto de inflexión es posible cuando $f''(x_0) = 0$, tome como referencia la figura 9.3*a* y 9.3*a'*. El punto J del diagrama superior es un punto de inflexión, con $x = j$ como su valor crítico. Puesto que la curva $f'(x)$ en el diagrama inferior alcanza un mínimo en $x = j$, la pendiente de $f'(x)$ [es decir, $f''(x)$] debe ser cero en el valor crítico $x = j$. Así, el punto J ilustra un punto de inflexión que ocurre cuando $f''(x_0) = 0$.

Para ver que un extremo relativo es congruente también con $f''(x_0) = 0$, considere la función $y = x^4$. Esta función se grafica como una curva en forma de U y tiene un mínimo, $y = 0$, obtenido en el valor crítico $x = 0$. Puesto que la segunda derivada de esta función es $f''(x) = 12x^2$, obtenemos de nuevo un valor cero para esta derivada en el valor crítico $x = 0$. Por lo tanto, esta función ilustra un extremo relativo que aparece cuando $f''(x_0) = 0$.

TABLA 9.1 Condiciones para un extremo relativo: $y = f(x)$

Condición	Máximo	Mínimo
Necesaria de primer orden	$f'(x) = 0$	$f'(x) = 0$
Necesaria de segundo orden†	$f''(x) \le 0$	$f''(x) \ge 0$
Suficiente de segundo orden†	$f''(x) < 0$	$f''(x) > 0$

†Aplicable sólo después de que cumplió la condición necesaria de primer orden.

Vale la pena mencionar de nuevo que la condición de primer orden es *necesaria*, pero *no suficiente*, para un máximo o mínimo relativo (¿recuerda los puntos de inflexión?). Por otro lado, la condición de segundo orden de que $f''(x)$ es negativa (positiva) en el valor crítico x_0 es *suficiente* para un máximo relativo (mínimo), pero *no es necesaria*. [¿Recuerda el extremo relativo que ocurre cuando $f''(x_0) = 0$?] Por esta razón, debemos evitar la siguiente forma de razonar: "puesto que ya sabemos que el valor estacionario $f(x_0)$ es un mínimo, debemos tener $f''(x_0) > 0$". Este razonamiento es imperfecto porque trata incorrectamente al signo positivo de $f''(x_0)$ como una condición necesaria para que $f(x_0)$ sea un mínimo.

Esto no quiere decir que nunca usemos las derivadas de segundo orden para expresar las condiciones *necesarias* de los extremos relativos. De hecho sí podemos usarlas. Sin embargo, debemos tener cuidado para considerar el hecho de que puede ocurrir un máximo (mínimo) relativo no sólo cuando $f''(x_0)$ es negativa (positiva), sino también cuando $f''(x_0)$ es cero. En consecuencia, las *condiciones necesarias de segundo orden* debemos expresarlas en términos de desigualdades débiles: para que un valor estacionario $f(x_0)$ sea un

$\left\{\begin{matrix}\text{máximo}\\ \text{mínimo}\end{matrix}\right\}$, relativo, es necesario que $\quad f''(x_0)\left\{\begin{matrix}\le\\ \ge\end{matrix}\right\}0.$

La explicación anterior se resume en la tabla 9.1. Todas las ecuaciones y desigualdades de la tabla son de la naturaleza de las condiciones (requerimientos) por satisfacer, y no especificaciones descriptivas de una determinada función. En particular, la ecuación $f'(x) = 0$ no significa que la función f tenga una pendiente cero en todas partes; más bien, expresa la condición de que sólo los valores de x que cumplen este requerimiento pueden calificar como valores críticos.

Condiciones para la maximización de la ganancia

A continuación presentaremos un ejemplo económico de problemas de valores extremos, es decir, problemas de optimización.

Una de las primeras cosas que aprende un estudiante de economía es que, a fin de maximizar la ganancia, una empresa debe igualar el costo marginal y el ingreso marginal. Mostraremos cómo se deduce en forma matemática esta condición. Para mantener el análisis en un nivel general, trabajaremos con la función de ingreso total $R = R(Q)$ y la función de costo total $C = C(Q)$, las cuales son funciones de una sola variable Q. De éstas se deduce que una función de ganancia (la función objetivo) podemos formularla también en términos de Q (la variable de elección):

$$\pi = \pi(Q) = R(Q) - C(Q) \qquad \textbf{(9.1)}$$

Para hallar el nivel de producción de máxima ganancia debemos satisfacer la condición necesaria de primer orden para un máximo: $d\pi/dQ = 0$. En consecuencia, se diferencia (9.1) respecto a Q e igualar a cero la derivada resultante: el resultado es

$$\begin{aligned}\frac{d\pi}{dQ} &\equiv \pi'(Q) = R'(Q) - C'(Q)\\ &= 0 \text{ si y sólo si } R'(Q) = C'(Q)\end{aligned} \qquad \textbf{(9.2)}$$

Así, la producción *óptima* (producción de *equilibrio*) Q^* debe satisfacer la ecuación $R'(Q^*) = C'(Q^*)$, o bien, MR = MC. Esta condición constituye la condición de primer orden para maximización de la ganancia.

Sin embargo, la condición de primer orden puede dar lugar a un mínimo y no a un máximo; así, debemos comprobar la condición de segundo orden siguiente. La segunda derivada se obtiene diferenciando la primera derivada en (9.2) respecto a Q:

$$\begin{aligned}\frac{d^2\pi}{dQ^2} &\equiv \pi''(Q) = R''(Q) - C''(Q) \\ &\leq 0 \text{ si y sólo si } R''(Q) \leq C''(Q)\end{aligned}$$

Esta última desigualdad es la condición necesaria de segundo orden para maximización. Si no se satisface, entonces es posible que Q^* no maximice la ganancia; de hecho, la minimiza. Si $R''(Q^*) = C''(Q^*)$, no se puede llegar a una conclusión definitiva. El mejor escenario es hallar $R''(Q^*) < C''(Q^*)$, que satisface la condición suficiente de segundo orden para un máximo. En ese caso, con absoluta certeza se puede tomar a Q^* como una producción que maximiza la ganancia. Desde el punto de vista económico esto significaría que, si la tasa de cambio de MR es menor que la tasa de cambio de MC en el nivel de producción donde MC = MR, entonces esa producción maximizará la ganancia.

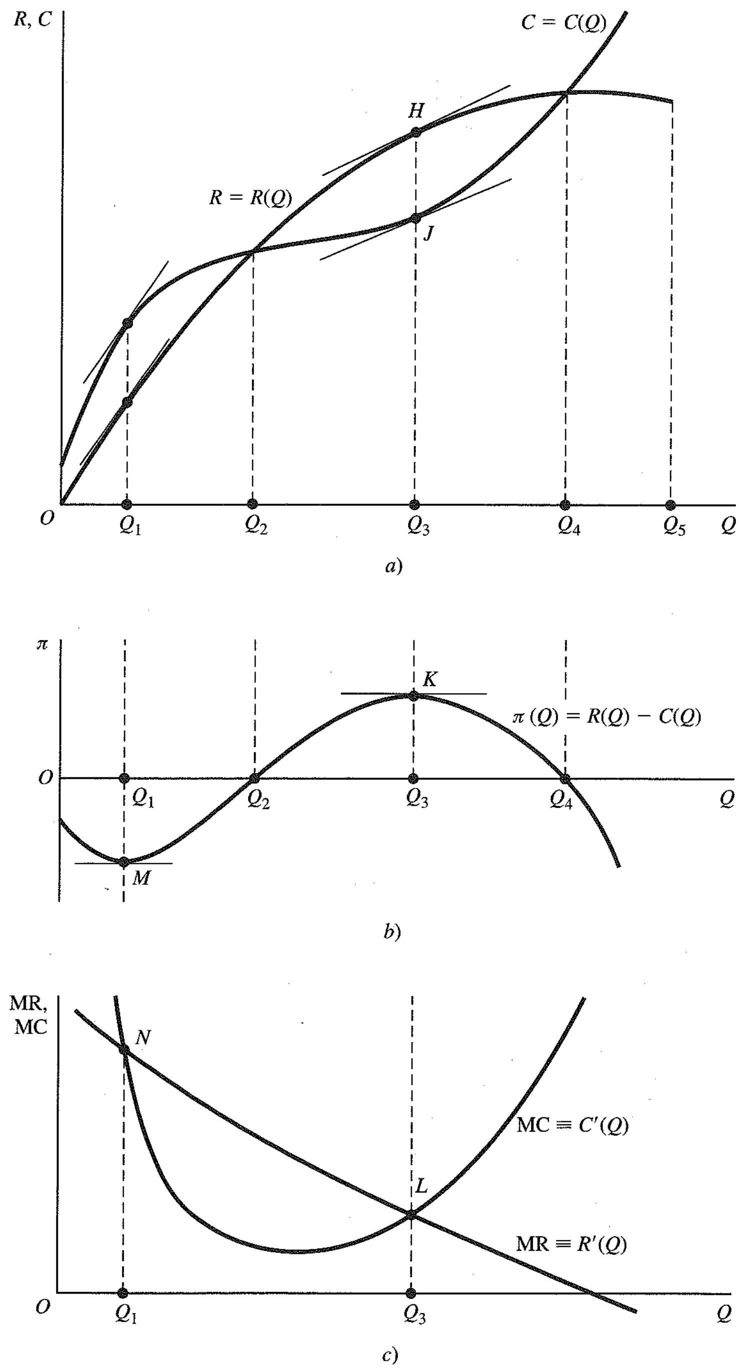
Estas condiciones se ilustran en la figura 9.7. La figura 9.7*a* muestra una curva de ingreso total y una de costo total, las cuales se cruzan dos veces, en los niveles de producción Q_2 y Q_4. En el intervalo abierto (Q_2, Q_4), el ingreso total R excede al costo total C y, por lo tanto, p es positiva. Pero en los intervalos $[0, Q_2)$ y $(Q_4, Q_5]$, donde Q_5 representa el límite superior de la capacidad productiva de la empresa, π es negativa. Este hecho se refleja en la figura 9.7*b*, donde la curva de ganancia, obtenida al graficar la distancia vertical entre las curvas R y C para cada nivel de producción, queda arriba del eje horizontal sólo en el intervalo (Q_2, Q_4).

Cuando se fija $d\pi/dQ = 0$, de acuerdo con la condición de primer orden, la intención es localizar el punto máximo K sobre la curva, en la producción Q_3, donde la pendiente de la curva es cero. Sin embargo, el punto mínimo relativo M (producción Q_1) se ofrecerá por sí mismo como candidato, porque también satisface el requerimiento de pendiente cero. A continuación recurrimos a la condición de segundo orden para eliminar la clase "equivocada" de extremo.

La condición de primer orden $d\pi/dQ = 0$ es equivalente a la condición $R'(Q) = C'(Q)$. En la figura 9.7*a* el nivel de producción Q_3 satisface esto, porque las curvas R y C tienen la misma pendiente en Q_3 (las rectas tangentes trazadas para las dos curvas en H y J son paralelas entre sí). Lo mismo se cumple para la producción Q_1. Puesto que la igualdad de las pendientes de R y C significa la igualdad de MR y MC, las producciones Q_3 y Q_1 deben estar donde se cruzan las curvas MR y MC, como lo ilustra la figura 9.7*c*.

¿Cómo entra en acción la condición de segundo orden? Consideremos primero la figura 9.7*b*. En el punto K, la segunda derivada de la función π (excepto el caso excepcional de valor cero) tiene un valor negativo, $\pi''(Q_3) < 0$, porque la curva tiene forma de U invertida alrededor de K. Esto significa que Q_3 maximizará la ganancia. Por otro lado, en el punto M se esperaría que $\pi''(Q_1) > 0$; por lo tanto, Q_1 proporciona un mínimo relativo para π. La condición suficiente de segundo orden para un máximo podemos expresarla de otra manera como $R''(Q) < C''(Q)$, es decir, la pendiente de la curva MR es menor que la pendiente de la curva MC. De la figura 9.7*c*, resulta evidente que la producción Q_3 satisface esta condición, puesto que la pendiente de MR es negativa, mientras que la de MC es positiva en el punto L. Pero la producción Q_1 viola esta condición porque tanto MC como MR tienen pendientes ne-

FIGURA 9.7



gativas, y la de MR es *numéricamente menor* que la de MC en el punto N, lo cual significa que $R''(Q_1)$ es *mayor* que $C''(Q_1)$. De hecho, la producción Q_1 también viola la condición *necesaria* de segundo orden para un máximo relativo, pero satisface la condición *suficiente* de segundo orden para un mínimo relativo.

Ejemplo 3 Sean las funciones $R(Q)$ y $C(Q)$

$$R(Q) = 1\,200Q - 2Q^2$$
$$C(Q) = Q^3 - 61.25Q^2 + 1\,528.5Q + 2\,000$$

Entonces, la función de ganancia es

$$\pi(Q) = -Q^3 + 59.25Q^2 - 328.5Q - 2\,000$$

donde R, C y π están expresados en dólares y Q está en unidades de (por ejemplo) toneladas por semana. Esta función de ganancia tiene dos valores críticos, $Q = 3$ y $Q = 36.5$, porque

$$\frac{d\pi}{dQ} = -3Q^2 + 118.5Q - 328.5 = 0 \quad \text{cuando } Q = \begin{cases} 3 \\ 36.5 \end{cases}$$

Sin embargo, puesto que la segunda derivada es

$$\frac{d^2\pi}{dQ^2} = -6Q + 118.5 \qquad \begin{cases} > 0 & \text{cuando } Q = 3 \\ < 0 & \text{cuando } Q = 36.5 \end{cases}$$

la producción que maximiza la ganancia es $Q^* = 36.5$ (toneladas por semana). (La otra producción minimiza la ganancia.) Al sustituir Q^* en la función de ganancia, podemos hallar la ganancia maximizada como $\pi^* = \pi(36.5) = 16\,318.44$ (dólares por semana).

Como método alternativo al precedente, podemos hallar primero las funciones MR y MC y luego igualar las dos, es decir, encontrar su intersección. Puesto que

$$R'(Q) = 1{,}200 - 4Q$$
$$C'(Q) = 3Q^2 - 122.5Q + 1\,528.5$$

igualar las dos funciones dará como resultado una ecuación cuadrática idéntica con $d\pi/dQ = 0$ que ha producido los dos valores críticos de Q citados antes.

Coeficientes de una función de costo total cúbica

En el ejemplo 3 usamos una función cúbica para representar la función de costo total. La curva de costo total tradicional $C = C(Q)$, como lo ilustra la figura 9.7*a*, se supone que contiene dos curvaturas, una forma un segmento cóncavo (costo marginal decreciente) y la posterior uno convexo (costo marginal creciente). Puesto que la gráfica de una función cúbica siempre contiene dos curvaturas, como lo ilustra la figura 9.4, se debe ajustar bien a ese papel. Sin embargo, la figura 9.4 revela de inmediato un problema: es posible que la función cúbica produzca un segmento con pendiente descendente en su gráfica, mientras que la función de costo total, para que tenga sentido económico, debe tener pendiente ascendente en todas partes (una producción más grande conlleva siempre un costo total mayor). Si deseamos usar una función de costo total, como

$$C = C(Q) = aQ^3 + bQ^2 + cQ + d \qquad \textbf{(9.3)}$$

es esencial que asignemos restricciones apropiadas a los parámetros para evitar que alguna vez la curva C se doble hacia abajo.

Una forma equivalente de expresar este requerimiento es que la función MC debe ser positiva en todas partes, y esto se asegura sólo si el *mínimo absoluto* de la función MC resulta ser positivo. Al diferenciar (9.3) respecto a Q, obtenemos la función MC

$$MC = C'(Q) = 3aQ^2 + 2bQ + c \qquad \textbf{(9.4)}$$

la cual, debido a que es cuadrática, representa una parábola como en la figura 9.7c. A fin de que la curva MC permanezca positiva en todas partes (arriba del eje horizontal), es necesario que la parábola tenga forma de U (de lo contrario, con una U invertida, la curva se limita a extenderse hacia el segundo cuadrante). Por consiguiente, el coeficiente del término Q^2 de (9.4) tiene que ser positivo, es decir, se debe imponer la restricción $a > 0$. Sin embargo, esta restricción de ningún modo es suficiente, porque el valor mínimo de una curva MC en forma de U, llámelo $\mathrm{MC}_{\text{mín}}$ (un mínimo relativo que también resulta ser un mínimo absoluto), aún puede aparecer debajo del eje horizontal. De esta forma, a continuación debemos hallar $\mathrm{MC}_{\text{mín}}$ y determinar las restricciones paramétricas que lo harían positivo.

De acuerdo con lo que sabemos de un extremo relativo, el mínimo de MC ocurrirá donde

$$\frac{d}{dQ}\mathrm{MC} = 6aQ + 2b = 0$$

El nivel de producción que satisface esta condición de primer orden es

$$Q^* = \frac{-2b}{6a} = \frac{-b}{3a}$$

Esto minimiza (en vez de maximizar) a MC porque de seguro la segunda derivada $d^2(\mathrm{MC})/dQ^2 = 6a$ es positiva en vista de la restricción $a > 0$. Conocer Q^* nos permite ahora calcular $\mathrm{MC}_{\text{mín}}$, pero primero podríamos inferir el signo del coeficiente b. Como están descartados los niveles de producción negativos, vemos que b nunca puede ser positiva (dada $a > 0$). Además, puesto que suponemos que la ley de ingresos decrecientes se establece en un nivel de producción positivo (es decir, suponemos que MC tiene un segmento decreciente inicial), Q^* debe ser positiva (en lugar de cero). En consecuencia, se debe imponer la restricción $b < 0$.

Ahora, es una cuestión simple sustituir la producción Q^* que minimiza a MC en (9.4) para encontrar que

$$\mathrm{MC}_{\text{mín}} = 3a\left(\frac{-b}{3a}\right)^2 + 2b\frac{-b}{3a} + c = \frac{3ac - b^2}{3a}$$

Así, para garantizar el carácter positivo de $\mathrm{MC}_{\text{mín}}$, se debe imponer la restricción[5] $b^2 < 3ac$. Esta última restricción de hecho implica también la restricción $c > 0$. (¿Por qué?)

En la explicación anterior se requirieron los parámetros a, b y c. ¿Qué pasa con el otro parámetro d? La respuesta es que también es necesaria una restricción para d, pero ésa no tiene nada que ver con el problema de mantener positiva a MC. Si permitimos que $Q = 0$ en (9.3), encontraremos que $C(0) = d$. Así, el papel de d es determinar solamente la intersección ver-

[5] Esta restricción también podríamos haberla obtenido por el método de *completar el cuadrado*. La función MC podemos transformarla de manera sucesiva como sigue:

$$\begin{aligned}\mathrm{MC} &= 3aQ^2 + 2bQ + c \\ &= \left(3aQ^2 + 2bQ + \frac{b^2}{3a}\right) - \frac{b^2}{3a} + c \\ &= \left(\sqrt{3a}Q + \sqrt{\frac{b^2}{3a}}\right)^2 + \frac{-b^2 + 3ac}{3a}\end{aligned}$$

Puesto que es posible que la expresión cuadrada sea cero, debemos requerir, a fin de asegurar el carácter positivo de MC, que $b^2 < 3ac$ a sabiendas de que $a > 0$.

tical de la curva C, sin que tenga nada que ver con su pendiente. Puesto que el significado económico de d es el costo fijo de una empresa, la restricción apropiada (en el contexto de corto plazo) sería $d > 0$.

En resumen, los coeficientes de la función de costo total (9.3) debemos restringirlos como sigue (suponiendo el contexto de corto plazo):

$$a, c, d > 0 \qquad b < 0 \qquad b^2 < 3ac \tag{9.5}$$

Como podemos comprobar fácilmente, la función $C(Q)$ del ejemplo 3 satisface a (9.5).

Curva de ingreso marginal con pendiente ascendente

La figura 9.7c muestra que la curva de ingreso marginal tiene pendiente descendente en todas partes. Por supuesto, ésta es la manera como se dibuja de ordinario la curva MR para una empresa bajo competencia imperfecta. Sin embargo, no podemos descartar *a priori* la posibilidad de que la pendiente de la curva MR sea ascendente en parte o por completo.[6]

Dada una función de ingreso promedio $\text{AR} = f(Q)$, la función de ingreso marginal podemos expresarla mediante

$$\text{MR} = f(Q) + Qf'(Q) \qquad \text{[de (7.7)]}$$

Así que la pendiente de la curva MR se determina a partir de la derivada

$$\frac{d}{dQ}\text{MR} = f'(Q) + f'(Q) + Qf''(Q) = 2f'(Q) + Qf''(Q)$$

Siempre y cuando la curva AR tenga pendiente descendente (como sucedería en competencia imperfecta), con seguridad el término $2f'(Q)$ es negativo. Pero el término $Qf''(Q)$ puede ser negativo, cero o positivo, dependiendo del signo de la segunda derivada de la función AR, es decir, en función de si la curva AR es estrictamente cóncava, lineal, o convexa. Si la curva AR es estrictamente convexa, ya sea en su totalidad (como lo ilustra la figura 7.2) o a lo largo de un segmento específico, existirá la posibilidad de que el término (positivo) $Qf''(Q)$ pudiera dominar al término (negativo) $2f'(Q)$; así, la curva MR tendría una pendiente ascendente en forma parcial o completa.

Ejemplo 4

Sea la función de ingreso promedio

$$\text{AR} = f(Q) = 8\,000 - 23Q + 1.1Q^2 - 0.018Q^3$$

Según se comprueba (ver el ejercicio 9.4-7), esta función da lugar a una curva AR de pendiente descendente, como es apropiado para una empresa bajo competencia imperfecta. Puesto que

$$\text{MR} = f(Q) + Qf'(Q) = 8\,000 - 46Q + 3.3Q^2 - 0.072Q^3$$

se deduce que la pendiente de MR es

$$\frac{d}{dQ}\text{MR} = -46 + 6.6Q - 0.216Q^2$$

Debido a que se trata de una función cuadrática y puesto que el coeficiente de Q^2 es negativo, $d\text{MR}/dQ$ debemos graficarla como una curva U invertida en función de Q, como lo ilustra la figura 9.5a. Si un segmento de esta curva está arriba del eje horizontal, la pendiente de MR tomará valores positivos.

[6] Este punto se resalta de manera enfática en John P. Formby, Stephen Layson y W. James Smith, "The Law of Demand, Positive Sloping Marginal Revenue, and Multiple Profit Equilibria", en *Economic Inquiry*, abril de 1982, pp. 303 a 311.

Con $dMR/dQ = 0$, y aplicando la fórmula cuadrática, encontramos que los dos ceros de la función cuadrática son $Q_1 = 10.76$ y $Q_2 = 19.79$ (aproximadamente). Esto significa que, para valores de Q en el intervalo abierto (Q_1, Q_2), la curva dMR/dQ está arriba del eje horizontal. Por lo tanto, la curva de ingreso marginal tiene pendiente positiva para niveles de producción entre Q_1 y Q_2.

La presencia de un segmento de la curva MR con pendiente positiva tiene implicaciones interesantes. Esta clase de curva MR puede producir más de una intersección con la curva MC que satisface la condición suficiente de segundo orden para maximización de la ganancia. Sin embargo, aunque tales intersecciones constituyen óptimos locales, sólo uno de ellos es el óptimo global que está buscando la empresa.

EJERCICIO 9.4

1. Halle los máximos y mínimos relativos de y mediante el criterio de la segunda derivada:

   (*a*) $y = -2x^2 + 8x + 25$

   (*b*) $y = x^3 + 6x^2 + 9$

   (*c*) $y = \frac{1}{3}x^3 - 3x^2 + 5x + 3$

   (*d*) $y = \dfrac{2x}{1-2x} \quad \left(x \neq \dfrac{1}{2}\right)$

2. El señor Greenthumb desea cercar un campo de flores rectangular, usando una pared de su casa como un lado del rectángulo. Los otros tres lados se encerrarán con malla de alambre, de la cual tiene sólo 64 pies disponibles. ¿Cuáles son la longitud L y el ancho W del rectángulo con el cual obtendría el área de plantación más grande posible? ¿Cómo se asegura de que su respuesta sea el área más grande y no la más pequeña?

3. Una empresa tiene las siguientes funciones de costo total y demanda:

   $$C = \tfrac{1}{3}Q^3 - 7Q^2 + 111Q + 50$$

   $$Q = 100 - P$$

   (*a*) ¿La función de costo total satisface las restricciones de coeficientes de (9.5)?

   (*b*) Escriba la función de ingreso total R en términos de Q.

   (*c*) Formule la función de ganancia total π en términos de Q.

   (*d*) Encuentre el nivel de producción Q^* de maximización de ganancia.

   (*e*) ¿Cuál es la ganancia máxima?

4. Si el coeficiente b de (9.3) fuera cero, ¿qué sucedería con las curvas de costo marginal y costo total?

5. Para expresar las siguientes suposiciones usaremos una función de ganancia cuadrática $\pi(Q) = hQ^2 + jQ + k$:

   (*a*) Si la producción es nula, la ganancia será negativa (como resultado de los costos fijos).

   (*b*) La función de producción es estrictamente cóncava.

   (*c*) La ganancia máxima ocurre en un nivel de producción positivo Q^*.

   ¿Qué restricciones de parámetro se requieren?

6. Una empresa en un mercado competitivo puro tiene una sola variable de insumo L (mano de obra), y la tasa de salario es W_0 por periodo. Sus costos fijos le cuestan un total de F dólares por periodo. El precio del producto es P_0.

   (*a*) Escriba la función de producción, la función de ingreso, la función de costo y la función de ganancia de la empresa.

(*b*) ¿Cuál es la condición de primer orden para la maximización de ganancia? Dé a esta condición una interpretación económica.
(*c*) ¿Qué circunstancias económicas asegurarían que se maximizara la ganancia en vez de minimizarse?

7. Use el siguiente procedimiento para comprobar que la curva AR del ejemplo 4 tiene pendiente negativa:
(*a*) Denote la pendiente de AR mediante S. Escriba una expresión para S.
(*b*) Encuentre el valor máximo de S, $S_{máx}$, mediante el criterio de la segunda derivada.
(*c*) Deduzca del valor de $S_{máx}$ que la curva AR tiene pendiente negativa en todas partes.

9.5 Series de Maclaurin y series de Taylor

Llegó el momento de elaborar un criterio para los extremos relativos que podamos aplicar aun cuando la segunda derivada resulte tener un valor cero en el punto estacionario. Sin embargo, antes de que se pueda hacer eso, es necesario analizar primero el llamado desarrollo de una función $y = f(x)$ en lo que se conoce, respectivamente, como una *serie de Maclaurin* (expansión alrededor del punto $x = 0$) y una *serie de Taylor* (desarrollo respecto a cualquier punto $x = x_0$).

Expandir una función $y = f(x)$ alrededor de un punto x_0 significa, en el presente contexto, transformar esa función en una forma *polinomial*, en la que los coeficientes de los distintos términos se expresan en términos de los valores de las derivadas $f'(x_0)$, $f''(x_0)$, etcétera, evaluadas en el punto de expansión x_0. En la serie de Maclaurin, éstas se desarrollarán en $x = 0$; por lo tanto, se tiene $f'(0)$, $f''(0)$, etc., en los coeficientes. El resultado del desarrollo es una serie de potencias porque, siendo un polinomio, consta de una suma de función de potencia.

Serie de Maclaurin de una función polinomial

Consideremos primero la expansión de una función *polinomial* de n-ésimo grado,

$$f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 + \cdots + a_nx^n \tag{9.6}$$

en un polinomio equivalente de n-ésimo grado donde los coeficientes (a_0, a_1, etc.) los expresamos en términos de los valores de derivadas $f'(0)$, $f''(0)$, etc. Puesto que esto tiene que ver con la transformación de un polinomio en otro del mismo grado, éste podría parecer un ejercicio estéril y sin objetivo, pero en realidad servirá para que esclarezcamos la idea completa de expansión.

Puesto que la serie de potencias después del desarrollo tendrá que ver con las derivadas de varios órdenes de la función f, primero determinamos éstas. Mediante diferenciación sucesiva de (9.6), obtenemos las derivadas como sigue:

$$\begin{aligned}
f'(x) &= a_1 + 2a_2x + 3a_3x^2 + 4a_4x^3 + \cdots + na_nx^{n-1} \\
f''(x) &= 2a_2 + 3(2)a_3x + 4(3)a_4x^2 + \cdots + n(n-1)a_nx^{n-2} \\
f'''(x) &= 3(2)a_3 + 4(3)(2)a_4x + \cdots + n(n-1)(n-2)a_nx^{n-3} \\
f^{(4)}(x) &= 4(3)(2)a_4 + 5(4)(3)(2)a_5x + \cdots + n(n-1)(n-2)(n-3)a_nx^{n-4} \\
&\vdots \\
f^{(n)}(x) &= n(n-1)(n-2)(n-3)\cdots(3)(2)(1)a_n
\end{aligned}$$

Tome en cuenta que cada diferenciación sucesiva reduce el número de términos en uno, se elimina la constante aditiva de enfrente, hasta que, en la n-ésima derivada, se tiene un solo término de producto (un término constante). Estas derivadas podemos evaluarlas en varios valores de x; aquí, evaluaremos en $x = 0$, con el resultado de que eliminaremos todos los términos con x. Así, sólo quedan los siguientes valores de derivadas excepcionalmente nítidos:

$$f'(0) = a_1 \qquad f''(0) = 2a_2 \qquad f'''(0) = 3(2)a_3 \qquad f^{(4)}(0) = 4(3)(2)a_4$$
$$\cdots \qquad f^{(n)}(0) = n(n-1)(n-2)(n-3)\cdots(3)(2)(1)a_n \tag{9.7}$$

Si adoptamos un símbolo de abreviatura $n!$ (léase: "n factorial"), definido como

$$n! \equiv n(n-1)(n-2)\cdots(3)(2)(1) \qquad (n = \text{un entero positivo})$$

de tal manera que, por ejemplo, $2! = 2 \times 1 = 2$ y $3! = 3 \times 2 \times 1 = 6$, etc. (Con $0!$ definido como igual a 1), entonces el resultado en (9.7) podemos reescribirlo como

$$a_1 = \frac{f'(0)}{1!} \qquad a_2 = \frac{f''(0)}{2!} \qquad a_3 = \frac{f'''(0)}{3!} \qquad a_4 = \frac{f^{(4)}(0)}{4!} \qquad \cdots \qquad a_n = \frac{f^{(n)}(0)}{n!}$$

Al sustituir éstas en (9.6) y utilizar el hecho obvio de que $f(0) = a_0$, la función $f(x)$ podemos expresarla ahora como un nuevo polinomio del mismo grado, pero equivalente, en el que los coeficientes se expresan en términos de derivadas evaluadas en $x = 0$:[7]

$$f(x) = \frac{f(0)}{0!} + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \cdots + \frac{f^{(n)}(0)}{n!}x^n \qquad \text{[fórmula de Maclaurin]} \tag{9.8}$$

Este nuevo polinomio, conocido como serie de Maclaurin de la función polinomial $f(x)$, representa la expansión de la función $f(x)$ respecto a cero ($x = 0$). Note que el punto de expansión (aquí, 0) es simplemente el valor de x que usaremos para evaluar $f(x)$ y todas sus derivadas.

Ejemplo 1

Encuentre la serie de Maclaurin para la función

$$f(x) = 2 + 4x + 3x^2 \tag{9.9}$$

Esta función tiene las derivadas

$$\begin{aligned} f'(x) &= 4 + 6x \\ f''(x) &= 6 \end{aligned} \quad \text{de modo que} \left\{ \begin{aligned} f'(0) &= 4 \\ f''(0) &= 6 \end{aligned} \right.$$

Así, la serie de Maclaurin es

$$\begin{aligned} f(x) &= f(0) + f'(0)x + \frac{f''(0)}{2}x^2 \\ &= 2 + 4x + 3x^2 \end{aligned}$$

Con el renglón anterior se comprueba que la serie de Maclaurin representa correctamente la función dada.

[7] Puesto que $0! = 1$ y $1! = 1$, los dos primeros términos de la derecha de los signos de igualdad en (9.8) se pueden escribir en forma simplificada como $f(0)$ y $f'(0)x$, respectivamente. Aquí incluimos los denominadores $0!$ y $1!$ para llamar la atención en relación con la simetría entre los distintos términos de la expansión.

Serie de Taylor de una función polinomial

En términos generales, la función polinomial de (9.6) podemos desarrollarla alrededor de algún punto x_0, no necesariamente cero. En aras de la simplicidad, explicaremos lo anterior por medio de la función cuadrática específica en (9.9) y después generalizaremos el resultado.

Con el fin de llevar a cabo el desarrollo en torno a un punto específico x_0, podemos interpretar primero cualquier valor de x como una *desviación* de x_0. De manera más específica, sea $x = x_0 + \delta$, donde δ representa la desviación del valor de x_0. Con esta interpretación, la función dada (9.9) y sus derivadas se convierten en

$$\begin{aligned} f(x) &= 2 + 4(x_0 + \delta) + 3(x_0 + \delta)^2 \\ f'(x) &= 4 + 6(x_0 + \delta) \\ f''(x) &= 6 \end{aligned} \qquad \textbf{(9.10)}$$

Sabemos que la expresión $(x_0 + \delta) = x$ es una variable en la función, pero como x_0 en el presente contexto es un número *fijo* (elegido), sólo δ se puede considerar de modo apropiado como una variable en (9.10). En consecuencia, $f(x)$ es de hecho una función de δ, por ejemplo, $g(\delta)$:

$$g(\delta) = 2 + 4(x_0 + \delta) + 3(x_0 + \delta)^2 \qquad [\equiv f(x)]$$

con derivadas

$$\begin{aligned} g'(\delta) &= 4 + 6(x_0 + \delta) \qquad [\equiv f'(x)] \\ g''(\delta) &= 6 \qquad\qquad\qquad\quad\; [\equiv f''(x)] \end{aligned}$$

Ya sabemos cómo desarrollar $g(\delta)$ respecto a cero ($\delta = 0$). De acuerdo con (9.8), tal desarrollo produce la siguiente serie de Maclaurin:

$$g(\delta) = \frac{g(0)}{0!} + \frac{g'(0)}{1!}\delta + \frac{g''(0)}{2!}\delta^2 \qquad \textbf{(9.11)}$$

Sin embargo, puesto que permitimos que $x = x_0 + \delta$, el hecho de que $\delta = 0$ significa que $x = x_0$; por consiguiente, con base en la identidad $g(\delta) \equiv f(x)$, podemos escribir para el caso de $\delta = 0$:

$$g(0) = f(x_0) \qquad g'(0) = f'(x_0) \qquad g''(0) = f''(x_0)$$

Al escribir éstas en (9.11), descubrimos el resultado para representar el desarrollo de $f(x)$ respecto al punto x_0, porque los coeficientes ahora tienen que ver con las derivadas $f'(x_0)$, $f''(x_0)$, etc., todas evaluadas en $x = x_0$:

$$f(x)[=g(\delta)] = \frac{f(x_0)}{0!} + \frac{f'(x_0)}{1!}(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 \qquad \textbf{(9.12)}$$

Debemos comparar este resultado, el polinomio de Taylor de $f(x)$, con el polinomio de Maclaurin de $g(\delta)$ de (9.11).

Puesto que para la función específica en consideración, (9.9), tenemos

$$f(x_0) = 2 + 4x_0 + 3x_0^2 \qquad f'(x_0) = 4 + 6x_0 \qquad f''(x_0) = 6$$

el polinomio de Taylor en (9.12) se convierte en

$$\begin{aligned} f(x) &= 2 + 4x_0 + 3x_0^2 + (4 + 6x_0)(x - x_0) + \tfrac{6}{2}(x - x_0)^2 \\ &= 2 + 4x + 3x^2 \end{aligned}$$

Con esto se comprueba que el polinomio de Taylor representa correctamente la función dada.

Podemos generalizar la fórmula de expansión de (9.12) para aplicarla al polinomio de n-ésimo grado de (9.6). La fórmula generalizada es

$$f(x) = \frac{f(x_0)}{0!} + \frac{f'(x_0)}{1!}(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n \qquad \text{[fórmula de Taylor]} \tag{9.13}$$

Esto difiere de la fórmula de Maclaurin de (9.8) sólo en la sustitución de cero por x_0 como el punto de expansión, y en la sustitución de x por la expresión $(x - x_0)$. Lo que expresamos en (9.13) es que, dado un polinomio de n-ésimo grado $f(x)$, si permitimos que $x = 7$ (por ejemplo) en los términos de la derecha de (9.13), seleccionamos un número arbitrario x_0 y evaluamos y sumamos estos términos, al final tenemos $f(7)$, el valor de $f(x)$ en $x = 7$.

Ejemplo 2

Tomando $x_0 = 3$ como el punto de expansión, podemos escribir (9.6) en forma equivalente como

$$f(x) = f(3) + f'(3)(x - 3) + \frac{f''(3)}{2}(x - 3)^2 + \cdots + \frac{f^{(n)}(3)}{n!}(x - 3)^n$$

Expansión de una función arbitraria

Hasta ahora, hemos mostrado cómo una función polinomial de n-ésimo grado puede expresarse en otra forma polinomial equivalente de n-ésimo grado. Resulta que también es posible expresar alguna función *arbitraria* $\phi(x)$, una que no necesariamente es un polinomio, en una forma polinomial similar a (9.13), siempre que $\phi(x)$ tenga derivadas continuas finitas hasta el orden deseado en el punto de expansión x_0.

De acuerdo con la proposición matemática conocida como *teorema de Taylor*, dada una función arbitraria $\phi(x)$, si conocemos el valor de la función en $x = x_0$ [es decir, $\phi(x_0)$] y los valores de sus derivadas en x_0 [es decir, $\phi'(x_0)$, $\phi''(x_0)$, etc.], entonces esta función se puede desarrollar en torno al punto x_0 como sigue (n = a un entero positivo fijo elegido de manera arbitraria):

$$\begin{aligned}\phi(x) &= \left[\frac{\phi(x_0)}{0!} + \frac{\phi'(x_0)}{1!}(x - x_0) + \frac{\phi''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{\phi^{(n)}(x_0)}{n!}(x - x_0)^n\right] + R_n \\ &\equiv P_n + R_n \qquad \text{[fórmula de Taylor con residuo]}\end{aligned} \tag{9.14}$$

donde P_n representa el polinomio de n-ésimo grado (entre corchetes) [los primeros $(n + 1)$ términos de la derecha], y R_n denota un *residuo*, que explicamos en la página 248.[8] La presencia de R_n es lo que distingue a (9.14) de la fórmula de Taylor (9.13) y, por esta razón, (9.14) se llama *fórmula de Taylor con residuo*. La forma del polinomio P_n y el tamaño del residuo R_n dependerán del valor de n que se elija. Mientras más grande sea n, más términos habrá en P_n; en consecuencia, R_n en general tomará un valor diferente para cada n distinta. Este hecho explica la necesidad de usar un índice n en estos dos símbolos. Como ayuda para recordar, podemos identificar a n como el orden de la derivada superior en P_n. (En el caso especial de $n = 0$, no aparecerá ninguna derivada en P_n.)

[8] No debemos confundir el símbolo R_n (residuo) con el símbolo R^n (espacio de dimensión n).

La presencia de R_n en (9.14) se debe al hecho de que aquí tratamos con una función arbitraria ϕ que no siempre se puede transformar *exactamente* en la forma polinomial mostrada en (9.13), sino sólo se puede aproximar. Por lo tanto, se incluye un término residual como complemento para la parte P_n, a fin de representar la discrepancia entre $\phi(x)$ y P_n. Así, P_n constituye una aproximación polinomial a $\phi(x)$, con el término R_n como una medida del error de aproximación. Si elegimos $n = 1$, por ejemplo, tenemos

$$\phi(x) = [\phi(x_0) + \phi'(x_0)(x - x_0)] + R_1 = P_1 + R_1$$

donde P_1 consta de $n + 1 = 2$ términos y constituye una aproximación *lineal* a $\phi(x)$. Si elegimos $n = 2$, aparecerá un término de segunda potencia, de tal manera que

$$\phi(x) = \left[\phi(x_0) + \phi'(x_0)(x - x_0) + \frac{\phi''(x_0)}{2!}(x - x_0)^2\right] + R_2 = P_2 + R_2$$

donde P_2, que consta de $n + 1 = 3$ términos, es una aproximación *cuadrática* a $\phi(x)$. Y así sucesivamente. El hecho de que podamos crear aproximaciones polinomiales a alguna función arbitraria (siempre y cuando tenga derivadas continuas, finitas) es de gran importancia práctica. Las funciones polinomiales, incluso las de grado superior, son relativamente fáciles de resolver, y si sirven como buenas aproximaciones para algunas funciones difíciles, nos facilitarán el camino, según lo ilustran los dos ejemplos siguientes.

Debemos señalar que la función arbitraria $\phi(x)$ podría abarcar el polinomio de n-ésimo grado de (9.6) como un caso especial. Para este último caso, si el desarrollo es hacia otro polinomio de n-ésimo grado, el resultado de (9.13) se aplicará de manera exacta; en otras palabras, podemos usar el resultado de (9.14), con $R_n \equiv 0$. Sin embargo, si el polinomio de n-ésimo grado se va a desarrollar hacia un polinomio de *menor* grado, entonces este último podemos considerarlo sólo como una aproximación a $f(x)$, y debe aparecer un residuo; en ese caso, el resultado de (9.14) se puede aplicar con un residuo no cero. Por lo tanto, la fórmula de Taylor en la forma de (9.14) es perfectamente general.

Ejemplo 3

Desarrolla la función no polinomial

$$\phi(x) = \frac{1}{1 + x}$$

alrededor del punto $x_0 = 1$, con $n = 4$. Requeriremos las primeras cuatro derivadas de $\phi(x)$, las cuales son

$$\phi'(x) = -(1 + x)^{-2} \quad \text{de tal manera que} \quad \phi'(1) = -(2)^{-2} = \frac{-1}{4}$$

$$\phi''(x) = 2(1 + x)^{-3} \qquad \phi''(1) = 2(2)^{-3} = \frac{1}{4}$$

$$\phi'''(x) = -6(1 + x)^{-4} \qquad \phi'''(1) = -6(2)^{-4} = \frac{-3}{8}$$

$$\phi^{(4)}(x) = 24(1 + x)^{-5} \qquad \phi^{(4)}(1) = 24(2)^{-5} = \frac{3}{4}$$

Asimismo, vemos que $\phi(1) = \frac{1}{2}$. Por lo tanto, con $x_0 = 1$ de (9.14) y utilizando las derivadas obtenidas, llegamos a la siguiente serie de Taylor con residuo:

$$\begin{aligned}\phi(x) &= \frac{1}{2} - \frac{1}{4}(x - 1) + \frac{1}{8}(x - 1)^2 - \frac{1}{16}(x - 1)^3 + \frac{1}{32}(x - 1)^4 + R_4 \\ &= \frac{31}{32} - \frac{13}{16}x + \frac{1}{2}x^2 - \frac{3}{16}x^3 + \frac{1}{32}x^4 + R_4\end{aligned}$$

Aquí también es posible elegir $x_0 = 0$ como el punto de expansión. En ese caso, con x_0 igual a cero de (9.14), el desarrollo dará como resultado una *serie de Maclaurin con residuo.*

Ejemplo 4

Expanda la función cuadrática

$$\phi(x) = 5 + 2x + x^2$$

alrededor de $x_0 = 1$, con $n = 1$. Esta función es, como (9.9) en el ejemplo 1, un polinomio de segundo grado. Pero como $n = 1$, la tarea asignada es expandir en un polinomio de *primer* grado, es decir, hallar una aproximación lineal para la función cuadrática dada; así, es seguro que aparezca un término residual. Por esta razón, debemos considerar a $\phi(x)$ como una función "arbitraria" para el fin de este desarrollo de Taylor.

Para llevar a cabo este desarrollo sólo requerimos la primera derivada $\phi'(x) = 2 + 2x$. Evaluada en $x_0 = 1$, la función y su derivada producen

$$\phi(x_0) = \phi(1) = 8 \qquad \phi'(x_0) = \phi'(1) = 4$$

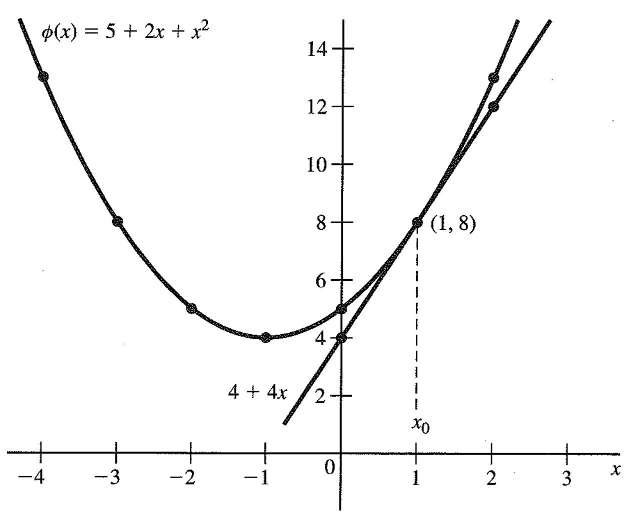
Así, la fórmula de Taylor con residuo produce

$$\begin{aligned}\phi(x) &= \phi(x_0) + \phi'(x_0)(x - x_0) + R_1 \\ &= 8 + 4(x - 1) + R_1 = 4 + 4x + R_1\end{aligned}$$

donde el término $(4 + 4x)$ es una aproximación lineal y el término R_1 representa el error de aproximación.

En la figura 9.8, $\phi(x)$ se grafica como una parábola, y su aproximación lineal como una recta tangente a la curva $\phi(x)$ en el punto (1, 8). La presencia del punto de tangencia en $x = 1$ no es un asunto de coincidencia, sino la consecuencia directa del hecho de que el punto de expansión se establece en ese valor particular de x. Esto hace pensar que, cuando una función arbitraria $\phi(x)$ se aproxima mediante un polinomio, este último produce el valor exacto de $\phi(x)$ *en* (y sólo *en*) el punto de desarrollo, con error de aproximación cero ($R_1 = 0$). En otro lugar, R_1 es estrictamente no cero y, de hecho, muestra errores de aproximación cada vez más

FIGURA 9.8



grandes cuando se intenta aproximar $\phi(x)$ para valores de x más y más alejados del punto de desarrollo x_0. Así, al intentar aproximar alguna función $\phi(x)$ mediante un polinomio, si el interés se centra en obtener una aproximación precisa en la vecindad de un valor específico de x, por ejemplo x_0, entonces se debe elegir x_0 como el punto de expansión.

La construcción de la figura 9.8 recuerda mucho la figura 8.1. De hecho, ambas figuras tienen que ver con "aproximaciones". Pero hay una diferencia en el alcance de la aproximación. En la figura 8.1 se intenta aproximar Δy mediante la diferencial dy con la ayuda de una recta tangente dibujada en x_0, un determinado valor inicial de x. Por otro lado, la figura 9.8 tiene como finalidad aproximar de manera más general una curva completa mediante una recta particular, es decir, aproximar la altura de la curva en algún valor de x, por ejemplo x_1, mediante la altura correspondiente de la recta en x_1. Note que en ambos casos el error de aproximación varía con el valor de x. En la figura 8.1, el error (la diferencia entre dy y Δy) se hace más pequeño a medida que Δx disminuye, o a medida que x se aproxima a x_0, en el cual se dibuja la recta tangente. En la figura 9.8, el error (la discrepancia vertical entre la recta y la curva) se hace pequeño cuando x se aproxima a x_0, el punto de desarrollo elegido.

Forma de Lagrange del residuo

Ahora debemos comentar más acerca del residuo. De acuerdo con la *forma de Lagrange del residuo*, R_n se puede expresar como

$$R_n = \frac{\phi^{(n+1)}(p)}{(n+1)!}(x - x_0)^{n+1} \tag{9.15}$$

donde p es algún número entre x (el punto donde se desea evaluar la función arbitraria ϕ) y x_0 (el punto donde se expande la función ϕ). Tome en cuenta que esta expresión se asemeja mucho al término que debe seguir al último término de P_n de (9.14), excepto que aquí la derivada se evalúa en un punto p en lugar de x_0. Puesto que el punto p no se especifica de otro modo, esta fórmula no permite calcular R_n; sin embargo, tiene gran importancia analítica, por lo que procedemos a ilustrar en forma gráfica su significado, aunque sólo lo haremos para el caso simple de $n = 0$.

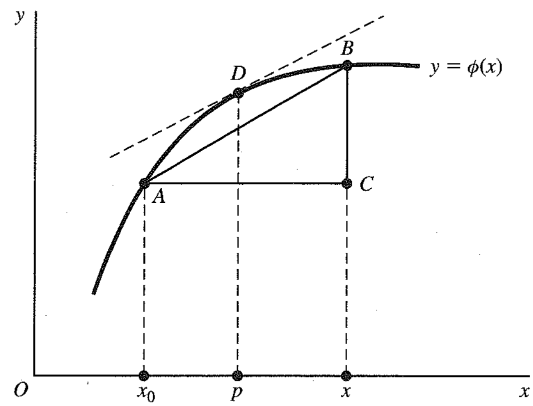
Cuando $n = 0$, no aparecerá ninguna derivada en la parte P_0 del polinomio; por lo tanto, (9.14) se reduce a

$$\phi(x) = P_0 + R_0 = \phi(x_0) + \phi'(p)(x - x_0)$$

o bien,
$$\phi(x) - \phi(x_0) = \phi'(p)(x - x_0)$$

Este resultado, una versión simple del *teorema del valor medio*, expresa que la diferencia entre el valor de la función ϕ en x_0 y en algún otro valor x podemos expresarla como el producto de la diferencia $(x - x_0)$ y la derivada ϕ' evaluada en p (con p como un punto entre x y x_0). Consideremos la figura 9.9, donde la función $\phi(x)$ se muestra como una curva continua con valores de derivada definidos en todos los puntos. Sea x_0 el punto de expansión elegido, y sea x *cualquier* punto sobre el eje horizontal. Si intentamos aproximar $\phi(x)$, o la distancia xB, mediante $\phi(x_0)$, o la distancia x_0A, tendremos un error igual a $\phi(x) - \phi(x_0)$, o la distancia CB. Lo que dice el teorema del valor medio es que el error CB, que constituye el valor del residuo R_0 en el desarrollo, se puede expresar como $\phi'(p)(x - x_0)$, donde p es algún punto entre x y x_0. Primero, en la curva entre los puntos A y B se localiza un punto D tal que la recta

FIGURA 9.9



tangente en D es paralela a la recta AB; tal punto D debe existir, puesto que la curva pasa de A a B de una manera continua y suave. Entonces, el residuo será

$$\begin{aligned} R_0 = CB = \frac{CB}{AC}AC &= (\text{pendiente de } AB) \cdot AC \\ &= (\text{pendiente de la tangente en } D) \cdot AC \\ &= (\text{pendiente de la curva en } x = p) \cdot AC \\ &= \phi'(p)(x - x_0) \end{aligned}$$

donde el punto p está entre x y x_0, como se requiere. Esto demuestra la razón fundamental de la forma de Lagrange del residuo para el caso $n = 0$. R_0 se expresa siempre como $\phi'(p)(x - x_0)$ porque, aunque no es posible asignar un valor específico a p, podemos estar seguros de que tal punto existe.

La ecuación (9.15) proporciona una forma de expresar el término del residuo R_n, pero no elimina a R_n como una fuente de discrepancia entre $\phi(x)$ y el polinomio P_n. Sin embargo, si incrementamos n (y, por lo tanto, aumentamos el grado del polinomio) de manera indefinida, vemos que

$$R_n \to 0 \text{ cuando } n \to \infty \quad \text{de tal manera que} \quad P_n \to \phi(x) \text{ cuando } n \to \infty$$

entonces se dice que la serie de Taylor converge a $\phi(x)$ en el punto de desarrollo, y esta serie podemos escribirla como una *serie infinita convergente* como sigue:

$$\phi(x) = \frac{\phi(x_0)}{0!} + \frac{\phi'(x_0)}{1!}(x - x_0) + \frac{\phi''(x_0)}{2!}(x - x_0)^2 + \cdots \qquad \textbf{(9.16)}$$

Note que ya no se muestra el término R_n; en su lugar están unos puntos suspensivos que significan que el polinomio contiene un número finito de términos subsecuentes cuyas estructuras matemáticas siguen el patrón que indican los términos previos. En este caso (conveniente), será posible hacer que P_n sea una aproximación a $\phi(x)$ tan precisa como se desee al elegir un valor suficientemente grande para n, es decir, incluyendo una cantidad de términos bastante grande en el polinomio P_n. Un ejemplo importante de esto lo analizamos en la sección 10.2.

EJERCICIO 9.5

1. Encuentre el valor de las siguientes expresiones factoriales:
   (*a*) $5!$ (*b*) $8!$ (*c*) $\frac{4!}{3!}$ (*d*) $\frac{6!}{4!}$ (*e*) $\frac{(n+2)!}{n!}$
2. Halle los primeros cinco términos de la serie de Maclaurin (es decir, elija $n = 4$ y sea $x_0 = 0$) para:
   (*a*) $\phi(x) = \frac{1}{1-x}$ (*b*) $\phi(x) = \frac{1-x}{1+x}$
3. Determine la serie de Taylor con $n = 4$ y $x_0 = -2$, para las dos funciones del problema 2.
4. Con base en la fórmula de Taylor con la forma de Lagrange del residuo [véase (9.14) y (9.15)], demuestre que en el punto de expansión ($x = x_0$) la serie de Taylor siempre da *exactamente* el valor de la función en ese punto, $\phi(x_0)$, no solamente una aproximación.

9.6 Criterio de la *N*-ésima derivada para el extremo relativo de una función de una variable

El desarrollo de una función en una serie de Taylor (o Maclaurin) es útil como un mecanismo de aproximación en la circunstancia que $R_n \to 0$ cuando $n \to \infty$, pero por el momento nuestro interés se centra en su aplicación en la obtención de un criterio general para un extremo relativo.

Expansión de Taylor y extremo relativo

Como elemento preparatorio para esa tarea, definimos de nuevo un extremo relativo como:

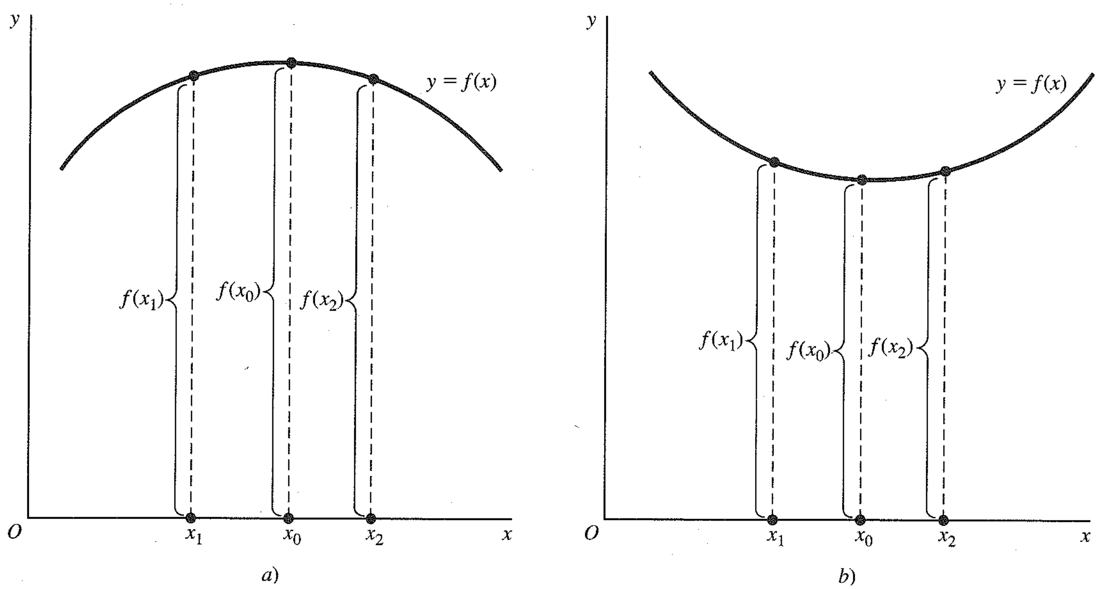
> Una función $f(x)$ alcanza un valor máximo (mínimo) relativo en x_0 si $f(x) - f(x_0)$ es negativa (positiva) para valores de x en la vecindad inmediata de x_0, tanto a su izquierda como a su derecha.

Esto se aclara al referirse a la figura 9.10, donde x_1 es un valor de x a la izquierda de x_0, y x_2 es un valor de x a la derecha de x_0. En la figura 9.10*a*, $f(x_0)$ es un máximo relativo; por lo tanto, $f(x_0)$ excede a $f(x_1)$ y $f(x_2)$. En resumen, $f(x) - f(x_0)$ es negativa para cualquier valor de x en la vecindad inmediata de x_0. Lo contrario es cierto para la figura 9.10*b*, donde $f(x_0)$ es un mínimo relativo y, por consiguiente, $f(x) - f(x_0) > 0$.

Si suponemos que $f(x)$ tiene derivadas continuas, finitas, hasta el orden deseado en el punto $x = x_0$, la función $f(x)$, que no es necesariamente un polinomio, podemos expandirla en torno al punto x_0 como una serie de Taylor. Con base en (9.14) (después de cambiar como es debido ϕ a f), y usar la forma de Lagrange del residuo, podemos escribir

$$f(x) - f(x_0) = f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + \frac{f^{(n+1)}(p)}{(n+1)!}(x - x_0)^{n+1} \qquad \textbf{(9.17)}$$

FIGURA 9.10



Si podemos determinar el signo de la expresión $f(x) - f(x_0)$ para valores de x a la izquierda y derecha inmediatas de x_0, llegamos a una conclusión en cuanto a si $f(x_0)$ es un extremo, y en caso afirmativo, si es un máximo o un mínimo. Para esto, es necesario que examinemos la suma del lado derecho de (9.17). En total, hay $(n + 1)$ términos en esta suma, n términos de P_n, más el residuo que es de grado $(n + 1)$, y por lo tanto el número real de términos es indefinido; así que depende del valor elegido de n. Sin embargo, al elegir a n de manera apropiada, podemos estar seguros de que existirá siempre sólo un término simple a la derecha. Esto nos simplifica muchísimo la tarea de evaluar el signo de $f(x) - f(x_0)$ y determinar si $f(x_0)$ es un extremo, y si es así, de qué clase.

Algunos casos específicos

Lo anterior se aclara con algunos ejemplos específicos

Caso 1 $\qquad f'(x_0) \neq 0$

Si la primera derivada de x_0 es no cero, se elige $n = 0$, de modo que el residuo será de primer grado. Entonces sólo habrá $n + 1 = 1$ término del lado derecho, lo cual significa que sólo estará presente el residuo R_0. Es decir, tenemos

$$f(x) - f(x_0) = \frac{f'(p)}{1!}(x - x_0) = f'(p)(x - x_0)$$

donde p es algún número entre x_0 y un valor de x en la vecindad inmediata de x_0. Note que p debe estar muy cerca de x_0.

¿Cuál es el signo de la expresión de la derecha? Como resultado de la continuidad de la derivada, $f'(p)$ tendrá el mismo signo que $f'(x_0)$ puesto que, como se mencionó antes, p es muy, muy cercana a x_0. En el caso presente, $f'(p)$ debe ser no cero; de hecho, debe ser un número positivo o negativo específico. ¿Pero qué pasa con la parte $(x - x_0)$? Cuando se va de la izquierda de x_0 a su derecha, x cambia de una magnitud $x_1 < x_0$ a una magnitud $x_2 > x_0$ (véase la figura 9.10). En consecuencia, la expresión $(x - x_0)$ debe cambiar de negativa a positiva a medida que se avanza, y $f(x) - f(x_0) = f'(p)(x - x_0)$ también debe cambiar de signo de la izquierda de x_0 a su derecha. Sin embargo, esto viola la nueva definición de un extremo relativo; por lo tanto, no puede existir un máximo relativo en $f(x_0)$ cuando $f'(x_0) \neq 0$, un hecho que ya es bien conocido.

Caso 2 $\qquad f'(x_0) = 0;\ f''(x_0) \neq 0$

En este caso, elija $n = 1$ para que el residuo sea de segundo grado. Entonces, al inicio habrá $n + 1 = 2$ términos a la derecha. Pero uno de estos términos se anula porque $f'(x_0) = 0$, y de nuevo queda un solo término que evaluar:

$$\begin{aligned} f(x) - f(x_0) &= f'(x_0)(x - x_0) + \frac{f''(p)}{2!}(x - x_0)^2 \\ &= \tfrac{1}{2} f''(p)(x - x_0)^2 \qquad [\text{porque} \quad f'(x_0) = 0] \end{aligned}$$

Como antes, $f''(p)$ tendrá el mismo signo que $f''(x_0)$, signo que es específico e invariante, mientras que la parte $(x - x_0)^2$, por ser un cuadrado, siempre es positiva. Así, la expresión $f(x) - f(x_0)$ debe tomar el mismo signo que $f''(x_0)$ y, según la definición anterior de extremo relativo, especificará

$$\begin{matrix} \text{Un máximo relativo de } f(x) \text{ si } f''(x_0) < 0 \\ \text{Un mínimo relativo de } f(x) \text{ si } f''(x_0) > 0 \end{matrix} \qquad [\text{con} \quad f'(x_0) = 0]$$

Esto se conoce como el criterio de la segunda derivada introducida antes.

Caso 3 $\qquad f'(x_0) = f''(x_0) = 0$, pero $f'''(x_0) \neq 0$

Aquí encontramos una situación imposible resolver con el criterio de la segunda derivada, porque $f''(x_0)$ es cero. Sin embargo, con la ayuda de la serie de Taylor establecemos sin dificultad un resultado definitivo.

Eligimos $n = 2$; entonces, aparecerán al principio tres términos a la derecha, pero eliminaremos dos porque $f'(x_0) = f''(x_0) = 0$, de manera que de nuevo tenemos un solo término que evaluar:

$$\begin{aligned} f(x) - f(x_0) &= f'(x_0)(x - x_0) + \frac{1}{2!} f''(x_0)(x - x_0)^2 + \frac{1}{3!} f'''(p)(x - x_0)^3 \\ &= \frac{1}{6} f'''(p)(x - x_0)^3 \qquad [\text{porque } f'(x_0) = 0,\ f''(x_0) = 0] \end{aligned}$$

Como antes, el signo de $f'''(p)$ es idéntico al de $f'''(x_0)$ como resultado de la continuidad de la derivada y porque p es muy cercana a x_0. Pero la parte $(x - x_0)^3$ tiene un signo cambiante. En particular, puesto que $(x - x_0)$ es negativa a la izquierda de x_0, también lo será $(x - x_0)^3$; sin embargo, a la derecha de x_0, la parte $(x - x_0)^3$ será positiva. Así que hay un cambio en el signo de $f(x) - f(x_0)$ cuando se pasa por x_0, lo cual va en contra de la definición de un extremo relativo. No obstante, sabemos que x_0 es un valor crítico $[f'(x_0) = 0]$ y, por lo tanto, debe dar un punto de inflexión, en vista de que no da un extremo relativo.

Caso 4 $\qquad f'(x_0) = f''(x_0) = \cdots = f^{(N-1)}(x_0) = 0$, pero $f^{(N)}(x_0) \neq 0$

Éste es un caso muy general y, por lo tanto, podemos deducir de él un resultado general. Note que los valores de las derivadas son cero hasta que llegamos a la N-ésima.

De manera análoga a los tres casos anteriores, la serie de Taylor para el caso 4 se reduce a

$$f(x) - f(x_0) = \frac{1}{N!} f^{(N)}(p)(x - x_0)^N$$

De nuevo, $f^{(N)}(p)$ toma el mismo signo que $f^{(N)}(x_0)$, que es invariante. El signo de la parte $(x - x_0)^N$, por otro lado, *variará* si N es *impar* (ver casos 1 y 3) y *permanecerá sin cambio* (positiva) si N es *par* (ver caso 2). Cuando N es impar, en consecuencia, $f(x) - f(x_0)$ cambiará de signo cuando se pasa por el punto x_0, así que se viola la definición de un extremo relativo (lo cual significa que x_0 debe producir un punto de inflexión sobre la curva). Pero cuando N es par, $f(x) - f(x_0)$ no cambia de signo de la izquierda de x_0 a su derecha, y esto establece el valor estacionario $f(x_0)$ como un máximo o mínimo relativo, dependiendo de si $f^{(N)}(x_0)$ es negativa o positiva.

Criterio de la *N*-ésima derivada

Por fin, entonces, podemos expresar el siguiente criterio general.

Criterio de la *N*-ésima derivada para el extremo relativo de una función de una variable Si la primera derivada de una función $f(x)$ en x_0 es $f'(x_0) = 0$ y si el primer valor *no cero* de derivada en x_0 encontrado en la derivación sucesiva es el de la N-ésima derivada, $f^{(N)}(x_0) \neq 0$, entonces el valor estacionario $f(x_0)$ será

a) Un *máximo* relativo si N es un número par y $f^{(N)}(x_0) < 0$
b) Un *mínimo* relativo si N es un número par y $f^{(N)}(x_0) > 0$
c) Un *punto de inflexión* si N es impar.

En la afirmación anterior debe quedar claro que el criterio de la N-ésima derivada funciona si y sólo si la función $f(x)$ es capaz de producir, tarde o temprano, un valor de derivada no cero en el valor crítico x_0. Aunque existen funciones excepcionales que no satisfacen esta condición, la mayor parte de las funciones que es probable que encontremos producirán alguna $f^{(N)}(x_0)$ diferente de cero en la diferenciación sucesiva.[9] Por lo tanto, el criterio es útil en la mayoría de los casos.

[9] Si $f(x)$ es una función constante, entonces es evidente que $f'(x) = f''(x) = \cdots = 0$, de manera que no se encuentra ningún valor de derivada diferente de cero. Sin embargo, éste es un caso trivial, ya que es una función constante no requiere prueba para el extremo. Como un ejemplo no trivial, considera la función

$$y = \begin{cases} e^{-1/x^2} & \text{(para } \neq 0) \\ 0 & \text{(para } = 0) \end{cases}$$

donde la función $y = e^{-1/x^2}$ es una función exponencial, que introduciremos en el capítulo 10. Por sí misma, $y = e^{-1/x^2}$ es discontinua en $x = 0$, porque $x = 0$ no está en el dominio (la división entre cero no está definida). Sin embargo, puesto que $\lim_{x \to 0} y = 0$, podemos, al anexar la condición de que $y = 0$ para $x = 0$, llenar el vacío en el dominio y, por consiguiente, obtener una función continua. La gráfica de esta función muestra que logra un mínimo en $x = 0$. Pero resulta que, en $x = 0$, todas las derivadas (hasta cualquier orden) tienen valores cero. Por lo tanto, no podemos aplicar el criterio de la N-ésima derivada para confirmar el hecho determinable de manera gráfica de que la función tiene un mínimo en $x = 0$. Para una explicación más detallada de este caso excepcional, consulta R. Courant, *Differential and Integral Calculus* (traducido por E. J. McShane), Interscience, Nueva York, vol. 1, 2a. ed., 1937, pp. 196, 197 y 336.

Ejemplo 1 Examine la función $y = (7 - x)^4$ para su extremo relativo. Puesto que $f'(x) = -4(7 - x)^3$ es cero cuando $x = 7$, tomamos $x = 7$ como el valor crítico para la prueba, con $y = 0$ como el valor estacionario de la función. Por derivación sucesiva (continua hasta que encontramos un valor de derivada no cero en el punto $x = 7$), obtenemos

$$f''(x) = 12(7 - x)^2 \quad \text{de modo que} \quad f''(7) = 0$$

$$f'''(x) = -24(7 - x) \qquad f'''(7) = 0$$

$$f^{(4)}(x) = 24 \qquad f^{(4)}(7) = 24$$

Puesto que 4 es un número par y como $f^{(4)}(7) = 24$ es positivo, concluimos que el punto (7, 0) representa un mínimo relativo.

Según se comprueba fácilmente, esta función representa una curva estrictamente convexa. En vista de que la segunda derivada en $x = 7$ es cero (en lugar de positiva), este ejemplo nos sirve para ilustrar la afirmación anterior en relación con la segunda derivada y la curvatura de una curva (sección 9.3) para el efecto de que, si bien una $f''(x)$ positiva para toda x implica una $f(x)$ estrictamente convexa, una $f(x)$ estrictamente convexa *no* implica una $f''(x)$ positiva para toda x. Y lo más importante es que también sirve para ilustrar el hecho de que, dada una curva estrictamente convexa (cóncava), el extremo encontrado en esa curva debe ser un mínimo (máximo), porque tal extremo cumplirá la condición suficiente de segundo orden, *o bien*, si no es así, cumplirá otra condición suficiente (de orden superior) para un mínimo (máximo).

EJERCICIO 9.6

1. Encuentre los valores estacionarios de las siguientes funciones:
   (*a*) $y = x^3$ (*b*) $y = -x^4$ (*c*) $y = x^6 + 5$
   Determine por medio del criterio de la *N*-ésima derivada si representan máximos relativos, mínimos relativos o puntos de inflexión.
2. Determine los valores estacionarios de las siguientes funciones:
   (*a*) $y = (x - 1)^3 + 16$ (*c*) $y = (3 - x)^6 + 7$
   (*b*) $y = (x - 2)^4$ (*d*) $y = (5 - 2x)^4 + 8$
   Use el criterio de la *N*-ésima derivada para determinar la naturaleza exacta de estos valores estacionarios.

Capítulo 10

Funciones exponenciales y logarítmicas

La prueba de la N-ésima derivada desarrollada en el capítulo 9 nos proporciona un medio para realizar la tarea de localizar los valores extremos de alguna función objetivo, siempre y cuando tenga que ver sólo con una variable de elección, posea derivadas del orden deseado, y se produzca un valor de derivada diferente a cero en el valor crítico x_0 al derivar sucesivamente. Sin embargo, en los ejemplos citados en el capítulo 9, empleamos sólo funciones polinomiales y racionales, para las que se sabe cómo obtener las derivadas necesarias. Suponga que la función objetivo es una *exponencial*, como

$$y = 8^{x-\sqrt{x}}$$

Entonces aún no podemos aplicar el criterio de la derivada, porque todavía no sabemos cómo diferenciar tal función. Esto es lo que haremos en el presente capítulo.

Las funciones exponenciales, así como las funciones logarítmicas estrechamente relacionadas, tienen aplicaciones importantes en economía, en particular en relación con problemas de crecimiento y en general en la dinámica económica. Sin embargo, la aplicación relevante en esta parte del libro, tiene que ver con una clase de problemas de optimización en los que la variable de elección es el *tiempo*. Por ejemplo, cierto comerciante de vinos podría tener una reserva, de la cual se sabe que el valor de mercado aumenta con el tiempo en algún modo prescrito. El problema es determinar el mejor momento para vender esas existencias con base en la función del valor del vino, después de tomar en consideración el costo de interés relacionado con tener el capital monetario inmovilizado en esa reserva. Las funciones exponenciales podrían ayudar a solucionar esta clase de problemas de dos maneras. Primero, el valor del vino podría aumentar con el tiempo de acuerdo con una *ley de crecimiento exponencial*; en ese caso, tendríamos una función exponencial para el valor del vino. Segundo, cuando se considera que se pagan intereses, la presencia de capitalización del interés con seguridad introducirá una función exponencial. Así, debemos estudiar la naturaleza de las funciones exponenciales antes de poder analizar este tipo de problema de optimización.

Puesto que el objetivo primario es tratar el tiempo como variable de elección, cambiamos ahora al símbolo t, en lugar de x, para indicar la variable independiente en la discusión posterior. (Este símbolo t puede representar también otras variables distintas al tiempo.)

10.1 Naturaleza de las funciones exponenciales

El término *exponente* significa un indicador de la potencia a la cual se va a elevar una variable. En expresiones de potencia, como x^3 o x^5, los exponentes son *constantes*; pero no hay razón por la que no se pueda tener un exponente *variable*, como en 3^x o 3^t, donde el número 3 se va a elevar a distintas potencias (varios valores de x o t). Una función cuya variable *independiente* aparece en el papel de un exponente se llama *función exponencial*.

Función exponencial simple

En su versión más simple, la función exponencial se puede representar en la forma

$$y = f(t) = b^t \qquad (b > 1) \tag{10.1}$$

donde y y t son las variables dependiente e independiente, respectivamente, y b denota una *base* fija del exponente. El dominio de tal función es el conjunto de todos los números reales. Así, a diferencia de los exponentes en una función polinomial, el exponente variable t de (10.1) no está limitado a enteros positivos, a menos que deseemos imponer tal restricción.

Entonces, ¿por qué la restricción de $b > 1$? La explicación es la siguiente: puesto que el dominio de la función de (10.1) consta del conjunto de los números reales, es posible que t tome el valor $\frac{1}{2}$. Si se permite que b sea negativa, la potencia un medio de b indicaría tomar la raíz cuadrada de un número negativo. Si bien ésta no es una tarea imposible, en realidad se preferiría optar por lo fácil restringiendo b a valores positivos. Sin embargo, una vez que adoptamos la restricción $b > 0$ podríamos ir también hasta la restricción $b > 1$: la restricción $b > 1$ difiere de $b > 0$ sólo en que se excluyen los casos de (1) $0 < b < 1$ y (2) $b = 1$; pero como se mostrará, el primer caso puede ser incluido en la restricción $b > 1$, mientras que el segundo se puede descartar de inmediato. Considere el primer caso. Si $b = \frac{1}{5}$, entonces se tiene

$$y = \left(\frac{1}{5}\right)^t = \frac{1}{5^t} = 5^{-t}$$

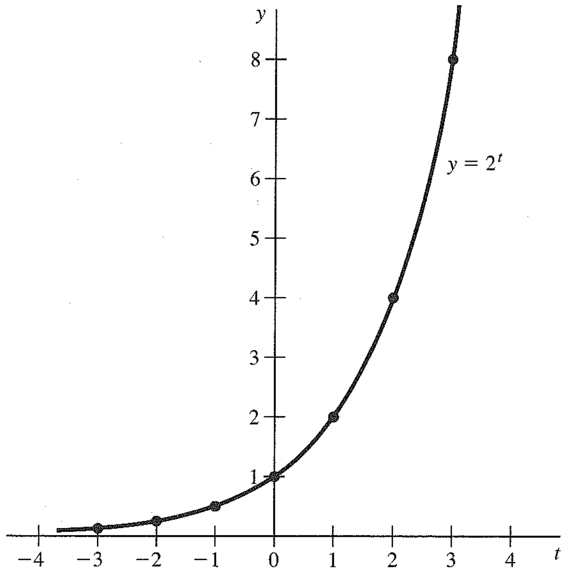
Esto muestra que una función con una base fraccionaria se puede reescribir fácilmente en una con una base mayor que 1. En cuanto al segundo caso, el hecho de que $b = 1$ producirá la función $y = 1^t = 1$, así que la función exponencial degenera en una función constante; por lo tanto, ésta se podría calificar como miembro de la familia exponencial.

Forma gráfica

La gráfica de la función exponencial de (10.1) toma la forma general de la curva de la figura 10.1. La curva trazada se basa en el valor $b = 2$; pero incluso para otros valores de b prevalecerá la misma configuración general

Se pueden observar varias características sobresalientes de este tipo de curva exponencial. Primero, es continua y uniforme en todas partes; por lo tanto, la función debe ser diferenciable en todo lugar (en realidad, es continuamente diferenciable cualquier número de veces). Segundo, es estrictamente creciente y, de hecho, y aumenta a una tasa creciente en todas partes; en consecuencia, las derivadas primera y segunda de la función $y = b^t$ deben ser positivas, un hecho que podemos confirmar después de haber obtenido las fórmulas de diferenciación per-

FIGURA 10.1



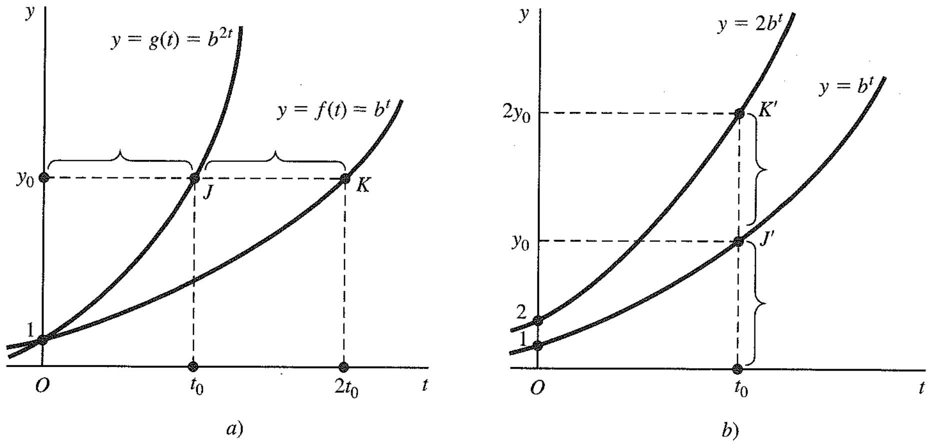
tinentes. Tercero, se nota que, aunque el dominio de la función contiene números negativos y positivos, la imagen de la función se limita al intervalo abierto $(0, \infty)$; es decir, la variable dependiente y es siempre *positiva*, sin importar el signo de la variable independiente t.

La monotonía estricta de la función exponencial tiene por lo menos dos implicaciones interesantes e importantes. Primera, inferimos que la función exponencial debe tener una función inversa, la cual por sí misma es estrictamente monótona; esta función inversa resulta ser una función *logarítmica*. Segunda, puesto que monotonía estricta significa que hay un valor único de t para un determinado valor de y, y como la imagen de la función exponencial está en el intervalo $(0, \infty)$, se deduce que se debe poder expresar *cualquier número positivo* como una potencia única de una base $b > 1$. Esto se puede ver en la figura 10.1, donde la curva de $y = 2^t$ abarca todos los valores positivos de y en su imagen; por lo tanto, cualquier valor positivo de y se debe poder expresar como alguna potencia única del número 2. Incluso si se cambia la base a algún otro número real mayor que 1, se cumple la misma imagen, así que es posible expresar cualquier número positivo y como una potencia de cualquier base $b > 1$.

Función exponencial generalizada

Este último punto merece examinarse más de cerca. Si una y positiva puede expresarse como potencia de varias bases alternativas, entonces debe existir un procedimiento general de *conversión de base*. En el caso de la función $y = 9^t$, por ejemplo, se puede transformar fácilmente en $y = (3^2)^t = 3^{2t}$, de modo que la base pasó de 9 a 3, siempre y cuando se modifique el exponente como es debido de t a $2t$. Este cambio de exponente, requerido por la conversión de base, no crea ningún tipo nuevo de función, porque, si se permite que $w = 2t$, entonces $y = 3^{2t} = 3^w$ aún está en la forma de (10.1). Sin embargo, desde el punto de vista de la base 3, el exponente ahora es $2t$ en vez de t. ¿Cuál es el efecto de agregar un coeficiente numérico al exponente t (en este caso 2)?

FIGURA 10.2



La respuesta se encuentra en la figura 10.2*a*, donde se trazan las curvas, una para la función $y = f(t) = b^t$ y otra para la función $y = g(t) = b^{2t}$. Puesto que el exponente de la última expresión es exactamente el doble de la primera, y puesto que se adopta la misma base para las dos funciones, la asignación de un valor arbitrario $t = t_0$ en la función g y $t = 2t_0$ en la función f debe producir el mismo valor:

$$f(2t_0) = g(t_0) = b^{2t_0} = y_0$$

Así que la distancia y_0J será la mitad de y_0K. Mediante un razonamiento similar, para cualquier valor de y la función g debe estar a la mitad entre la función f y el eje vertical. Por lo tanto, se podría concluir que *duplicar* el exponente tiene el efecto de comprimir a la *mitad* la curva exponencial hacia el eje y, mientras que *reducir a la mitad* el exponente ampliará la curva lejos del eje y al *doble* de la distancia horizontal.

Es interesante que ambas funciones compartan la misma intersección vertical

$$f(0) = g(0) = b^0 = 1$$

El cambio de exponente t a $2t$, o cualquier otro múltiplo de t, deja intacta la intersección vertical. En términos de *compresión*, esto se debe a que comprimir una distancia horizontal cero producirá una distancia cero.

El cambio de exponente es una forma de modificar, y generalizar, la función exponencial de (10.1); otra forma es anexar un coeficiente a b^t, como $2b^t$. [*Cuidado*: $2b^t \neq (2b)^t$.] El efecto de tal coeficiente es también comprimir o extender la curva, excepto que esta vez la dirección es vertical. En la figura 10.2*b*, la curva superior representa $y = 2b^t$, y la inferior es $y = b^t$. Para todo valor de t, la primera debe ser el doble de alta, porque tiene un valor y que es dos veces el primero. Así, se tiene $t_0J' = J'K'$. Note que la intersección vertical también se modifica en el presente caso. Podemos concluir que *duplicar* el coeficiente (aquí, de 1 a 2) sirve para extender la curva lejos del eje horizontal al *doble* de la distancia vertical, mientras que *reducir a la mitad* el coeficiente comprime la curva a la *mitad* hacia el eje t.

Conociendo ya las dos modificaciones recién explicadas, la función exponencial $y = b^t$ se puede generalizar ahora a la forma

$$y = ab^{ct} \qquad \textbf{(10.2)}$$

donde a y c son agentes de "compresión" o "extensión". Cuando se les asignan varios valores, modifican la posición de la curva exponencial, de modo que generan una familia completa de curvas exponenciales (funciones). Si a y c son positivas, prevalecerá la configuración general mostrada en la figura 10.2; sin embargo, si a o c o ambas son *negativas*, entonces las modificaciones fundamentales ocurrirán en la configuración de la curva (véase el ejercicio 10.1-5).

Una base preferida

Lo que suscitó el análisis del cambio de exponente de t a ct fue la pregunta de conversión de base. Pero, concediendo la factibilidad de la conversión de base, ¿por qué se querría realizar? Una respuesta es que algunas bases son más convenientes que otras en lo que respecta a manejos matemáticos.

Es bastante curioso que, en el cálculo, la base preferida sea cierto número irracional denotado por el símbolo e:

$$e = 2.71828\ldots$$

Cuando esta base e se usa en una función exponencial, se denomina *función exponencial natural*, de la cual algunos ejemplos son

$$y = e^t \qquad y = e^{3t} \qquad y = Ae^{rt}$$

Estas funciones ilustrativas se expresan también con otras notaciones

$$y = \exp(t) \qquad y = \exp(3t) \qquad y = A\exp(rt)$$

donde la abreviatura exp (para exponencial) indica que e tendrá como exponente la expresión entre paréntesis.

La elección de tal número insólito como $e = 2.71828\ldots$ como la base preferida sin duda parece desconcertante. Pero hay una razón excelente para esta elección, ¡porque la función e^t posee la notable propiedad de ser su propia derivada! Es decir,

$$\frac{d}{dt}e^t = e^t$$

un hecho que reduce el trabajo de diferenciación a nada. Dotados de esta regla de diferenciación, que se probará en la sección 10.5, será fácil hallar también la derivada de una función exponencial natural más complicada como $y = Ae^{rt}$. Para esto, primero permita que $w = rt$, de manera que la función se convierte en

$$y = Ae^w \qquad \text{donde } w = rt, \text{ y } A, r \text{ son constantes}$$

Entonces, por la regla de la cadena, se puede escribir

$$\frac{dy}{dt} = \frac{dy}{dw}\frac{dw}{dt} = Ae^w(r) = rAe^{rt}$$

Es decir,

$$\frac{d}{dt}Ae^{rt} = rAe^{rt} \tag{10.3}$$

De este modo, debe quedar del todo clara la conveniencia matemática de la base e.

EJERCICIO 10.1

1. Trace en un solo diagrama las gráficas de las funciones exponenciales $y = 3^t$ y $y = 3^{2t}$.
   (*a*) ¿Muestran las gráficas la misma relación posicional general que se observa en la figura 10.2*a*?
   (*b*) ¿Comparten estas curvas la misma intersección vertical? ¿Por qué?
   (*c*) Bosqueje la gráfica de la función $y = 3^{3t}$ en el mismo diagrama.
2. Represente en un solo diagrama las gráficas de las funciones exponenciales $y = 4^t$ y $y = 3(4^t)$.
   (*a*) ¿Muestran las gráficas la relación posicional general sugerida en la figura 10.2*b*?
   (*b*) ¿Tienen las dos curvas la misma intersección *y*? ¿Por qué?
   (*c*) Bosqueje la gráfica de la función $y = \frac{3}{2}(4^t)$ en el mismo diagrama.
3. Si se da por sentado que e^t es su propia derivada, use la regla de la cadena para hallar dy/dt para las siguientes funciones:
   (*a*) $y = e^{5t}$ (*b*) $y = 4e^{3t}$ (*c*) $y = 6e^{-2t}$
4. Según la explicación acerca de (10.1), ¿espera que la función $y = e^t$ sea estrictamente creciente a una tasa creciente? Compruebe su respuesta determinando los signos de las derivadas primera y segunda de esta función; recuerde que el dominio de esta función es el conjunto de los números reales, es decir, el intervalo $(-\infty, \infty)$.
5. Si se asignan valores negativos para *a* y *c* en (10.2), ya no prevalecerá la forma general de las curvas en la figura 10.2. Examine el cambio en la configuración de curva al contrastar (*a*) el caso de $a = -1$ con el caso de $a = 1$ y (*b*) el caso de $c = -1$ con el caso de $c = 1$.

10.2 Funciones exponenciales naturales y el problema de crecimiento

Las preguntas pertinentes aún sin contestar son: ¿cómo está definido el número *e*?, ¿tiene algún significado económico además de su importancia matemática como una base conveniente? y ¿en qué formas se aplican las funciones exponenciales naturales al análisis económico?

El número *e*

Consideremos la siguiente función:

$$f(m) = \left(1 + \frac{1}{m}\right)^m \tag{10.4}$$

Si a *m* le asignamos valores cada vez mayores, entonces $f(m)$ tomará también valores mayores; en particular, encontramos que

$$\begin{aligned}
f(1) &= \left(1 + \tfrac{1}{1}\right)^1 = 2 \\
f(2) &= \left(1 + \tfrac{1}{2}\right)^2 = 2.25 \\
f(3) &= \left(1 + \tfrac{1}{3}\right)^3 = 2.37037\ldots \\
f(4) &= \left(1 + \tfrac{1}{4}\right)^4 = 2.44141\ldots \\
&\;\vdots
\end{aligned}$$

Además, si m se incrementa de forma indefinida, entonces $f(m)$ convergirá al número $2.71828\ldots \equiv e$; así que e se puede definir como el límite de (10.4) cuando $m \to \infty$:

$$e \equiv \lim_{m\to\infty} f(m) = \lim_{m\to\infty}\left(1+\frac{1}{m}\right)^m \tag{10.5}$$

Que el valor aproximado de e sea 2.71828 se comprueba al hallar la serie de Maclaurin de la función $\phi(x) = e^x$, con x usada aquí para facilitar la aplicación directa de la fórmula de desarrollo (9.14). Esta serie produce una aproximación polinomial a e^x y, por lo tanto, el valor de $e\ (= e^1)$ se podría aproximar si se fija $x = 1$ en este polinomio. Si el término del residuo R_n se aproxima a cero cuando el número de términos en la serie se incrementa de forma indefinida, es decir, si la serie es convergente a $\phi(x)$, entonces se puede aproximar el valor de e a cualquier grado deseado de precisión al hacer suficientemente grande el número de términos incluidos.

Para lograr esto, se necesita tener derivadas de varios órdenes para la función. Si se acepta el hecho de que la primera derivada de e^x es la misma e^x, se ve que la derivada de $\phi(x)$ es simplemente e^x y, de manera similar, que la segunda, tercera, o cualquier derivada de orden superior debe ser e^x también. Por consiguiente, cuando evaluamos todas las derivadas en el punto de desarrollo ($x_0 = 0$), tenemos el resultado nítido

$$\phi'(0) = \phi''(0) = \cdots = \phi^{(n)}(0) = e^0 = 1$$

En consecuencia, si se establece que $x_0 = 0$ en (9.14), la serie de Maclaurin de e^x es

$$\begin{aligned} e^x = \phi(x) &= \phi(0) + \phi'(0)x + \frac{\phi''(0)}{2!}x^2 + \frac{\phi'''(0)}{3!}x^3 + \cdots + \frac{\phi^{(n)}(0)}{n!}x^n + R_n \\ &= 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots + \frac{1}{n!}x^n + R_n \end{aligned}$$

El término del residuo, de acuerdo con (9.15), se puede escribir como

$$R_n = \frac{\phi^{(n+1)}(p)}{(n+1)!}x^{n+1} = \frac{e^p}{(n+1)!}x^{n+1} \qquad [\phi^{(n+1)}(x) = e^x;\ \therefore \phi^{(n+1)}(p) = e^p]$$

En vista de que la expresión factorial $(n+1)!$ crece con más rapidez que la expresión de potencia x^{n+1} (para una x finita) cuando n aumenta, se deduce que $R_n \to 0$ cuando $n \to \infty$. Así, la serie de Maclaurin converge y, como resultado, el valor de e^x se puede expresar como una serie infinita convergente como sigue:

$$e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + \frac{1}{5!}x^5 + \cdots \tag{10.6}$$

Como caso especial, para $x = 1$ encontramos que

$$\begin{aligned} e &= 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \frac{1}{5!} + \cdots \\ &= 2 + 0.5 + 0.1666667 + 0.0416667 + 0.0083333 + 0.0013889 \\ &\quad + 0.0001984 + 0.0000248 + 0.0000028 + 0.0000003 + \cdots \\ &= 2.7182819 \end{aligned}$$

Si deseamos una cifra precisa hasta cinco decimales, podemos escribir $e = 2.71828$. Note que no es necesario preocuparnos acerca de términos posteriores en la serie infinita, porque serán de magnitud insignificante si sólo nos interesan cinco decimales.

Una interpretación económica de *e*

Desde el punto de vista matemático, el número e es la expresión del límite en (10.5). Pero, ¿posee también algún significado económico? La respuesta es que se puede interpretar como el resultado de un modo especial de capitalización del interés.

Suponga que, empezando con un principal (o capital) de \$1, se encuentra un banquero hipotético que ofrece la tasa de interés inusual de 100 por ciento anual (\$1 de interés por año). Si el interés se capitaliza una vez al año, el valor según los libros al final del año será \$2; este valor se denota mediante $V(1)$, donde el número entre paréntesis indica la frecuencia de capitalización dentro de un año:

$$\begin{aligned} V(1) &= \text{principal inicial } (1 + \text{tasa de interés}) \\ &= 1(1 + 100\%) = \left(1 + \tfrac{1}{1}\right)^1 = 2 \end{aligned}$$

Sin embargo, si se tiene un interés compuesto cada medio año, se acumulará un interés que equivale a 50% (la mitad de 100%) del principal al final de 6 meses. Por consiguiente, tendremos \$1.50 como el nuevo principal durante el segundo periodo de 6 meses, en el cual el interés se calculará en 50% de \$1.50. Así, el valor según los libros al final del año será $1.50(1 + 50\%)$; es decir,

$$V(2) = (1 + 50\%)(1 + 50\%) = \left(1 + \tfrac{1}{2}\right)^2$$

Mediante un razonamiento análogo, podemos escribir $V(3) = (1 + \frac{1}{3})^3$, $V(4) = (1 + \frac{1}{4})^4$, etc.; o bien, en general

$$V(m) = \left(1 + \frac{1}{m}\right)^m \qquad \textbf{(10.7)}$$

donde m representa la frecuencia de capitalización en un año.

En el caso límite, cuando el interés se capitaliza *de manera continua* durante el año, es decir, cuando m se vuelve infinita, el valor según los libros crecerá como una “bola de nieve”, que al final de un año será

$$\lim_{m\to\infty} V(m) = \lim_{m\to\infty} \left(1 + \frac{1}{m}\right)^m = e \text{ (dólares)} \qquad \text{[por (10.5)]}$$

Así, el número $e = 2.71828$ se interpreta como el valor al final del año al que crecerá un principal de \$1 si el interés a la tasa de 100 por ciento por año se capitaliza de forma continua.

Note que la tasa de interés de 100 por ciento es sólo una *tasa de interés nominal*, porque si \$1 se convierte en \$$e$ = \$2.718 después de un año, la *tasa de interés efectiva* es en este caso de alrededor de 172 por ciento anual.

Interés compuesto y la función *Ae^rt*

El proceso continuo de capitalización del interés recién descrito se puede generalizar en tres direcciones para tomar en cuenta: (1) más años de capitalización, (2) un principal distinto de \$1 y (3) una tasa de interés nominal diferente de 100 por ciento.

Si un principal de \$1 se convierte en \$$e$ después de un año de capitalización del interés y si se permite que \$$e$ sea el nuevo principal en el segundo año (durante el cual cada peso crecerá de nuevo en \$$e$), el valor según los libros al final de 2 años se convertirá en \$$e(e)$ = \$$e^2$. De la misma manera, se convierte en \$$e^3$ al final de 3 años o, en términos generales, será \$$e^t$ después de t años.

A continuación, se procede a cambiar el principal de \$1 a una cantidad no especificada, \$$A$. Este cambio es fácil: si \$1 se convierte en \$$e^t$ después de t años de capitalización continua a la tasa nominal de 100 por ciento anual, es lógico que \$$A$ aumentará a \$$Ae^t$.

¿Qué pasa con una tasa de interés nominal distinta a 100%, por ejemplo, $r = 0.05$ (= 5 por ciento)? El efecto de este cambio de tasa es alterar la expresión Ae^t a Ae^{rt}, según se comprueba a partir de lo siguiente. Con un principal inicial de \$$A$, que se invertirá durante t años a una tasa de interés nominal r, la fórmula de interés compuesto (10.7) se debe modificar a la forma

$$V(m) = A\left(1+\frac{r}{m}\right)^{mt} \tag{10.8}$$

La inserción del coeficiente A refleja el cambio de principal del nivel previo de \$1. La expresión del cociente r/m significa que, en cada uno de los m periodos de capitalización en un año, sólo será aplicable en realidad $1/m$ de la tasa nominal r. Por último, el exponente mt indica que, como el interés se va a capitalizar m veces al año, debe haber un total de mt capitalizaciones en t años.

La fórmula (10.8) se puede transformar en una forma alternativa

$$\begin{aligned} V(m) &= A\left[\left(1+\frac{r}{m}\right)^{m/r}\right]^{rt} \\ &= A\left[\left(1+\frac{1}{w}\right)^{w}\right]^{rt} \qquad \text{donde } w \equiv \frac{m}{r} \end{aligned} \tag{10.8'}$$

A medida que se incrementa la frecuencia de capitalización m, la variable w recién creada debe crecer a una tasa igual; así, cuando $m \to \infty$, se tiene $w \to \infty$ y la expresión entre corchetes de (10.8′), en virtud de (10.5), tiende al número e. En consecuencia, se ve que el valor según los libros en el proceso generalizado de capitalización continua es

$$V \equiv \lim_{m\to\infty} V(m) = Ae^{rt} \tag{10.8''}$$

como anticipamos.

Note que en (10.8) t es una variable *discreta* (lo contrario de una *variable continua*): ésta sólo puede tomar valores que son múltiplos enteros de $1/m$. Por ejemplo, si $m = 4$ (capitalización trimestral), entonces t sólo puede tomar los valores de $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$, 1, etcétera, lo cual indica que $V(m)$ tomará un nuevo valor sólo al final de cada nuevo trimestre. Sin embargo, cuando $m \to \infty$, como en (10.8″), $1/m$ se vuelve infinitesimal, por lo que t será continua. En ese caso, es legítimo hablar de fracciones de un año y permitir que t sea, por ejemplo, 1.2 o 2.35.

El resultado es que las expresiones e, e^t, Ae^t y Ae^{rt} se pueden interpretar desde el punto de vista económico en relación con el interés compuesto continuo, como se resume en la tabla 10.1.

Tasa de crecimiento instantánea

Se debe señalar, sin embargo, que la capitalización del interés es una interpretación ilustrativa, pero no exclusiva, de la función exponencial natural Ae^{rt}. La capitalización del interés sola-

TABLA 10.1
Capitalización continua de intereses

Principal, $	Tasa de interés nominal	Años de capitalización continua	Valor en libros, al final del proceso de capitalización, $
1	100% (= 1)	1	e
1	100%	t	e^t
A	100%	t	Ae^t
A	r	t	Ae^{rt}

mente ejemplifica el proceso general del *crecimiento exponencial* (aquí, el crecimiento de una suma de capital monetario con el tiempo), y se puede aplicar también al crecimiento de población, riqueza o capital real.

Aplicado a cierto contexto distinto al de capitalización del interés, el coeficiente r en Ae^{rt} ya no denota la tasa de interés nominal. Entonces, ¿qué significado económico toma? La respuesta es que r se puede reinterpretar como la *tasa de crecimiento instantánea* de la función Ae^{rt}. (De hecho, ésta es la razón de que se adopte el símbolo r, para la tasa de crecimiento, en primer lugar.) Dada la función $V = Ae^{rt}$, que da el valor de V en cada punto del tiempo t, la tasa de cambio de V se encuentra en la derivada

$$\frac{dV}{dt} = rAe^{rt} = rV \qquad \text{[véase (10.3)]}$$

Pero la *tasa de crecimiento* de V es simplemente la *tasa de cambio* en V expresada en términos (de porcentaje) relativos, es decir, formulada como una relación al valor de V misma. Así, para cualquier punto de tiempo, se tiene

$$\text{Tasa de crecimiento de } V \equiv \frac{dV/dt}{V} = \frac{rV}{V} = r \qquad \textbf{(10.9)}$$

como se afirmó antes.

Respecto a esta tasa de crecimiento se deben hacer varias observaciones. Primero, se aclara un punto fundamental en relación con el concepto de tiempo, a saber, la distinción entre un *punto* de tiempo y un *periodo*. La variable V (que denota una suma de dinero, o el tamaño de la población, etc.) es un concepto de *existencias*, que se refiere a la pregunta: ¿cuánto de esto *existe* en un determinado momento? Como tal, V se relaciona con el concepto *puntual* de tiempo; en cada punto de tiempo, V toma un valor único. El cambio en V, por otro lado, representa un *flujo*, que conlleva la pregunta: ¿cuánto de esto *tiene lugar* durante un determinado lapso? Por consiguiente, un cambio en V y, de la misma manera, la tasa de cambio de V debe tener referencia con algún periodo especificado, por ejemplo, un año.

Con esto en mente, retomemos (10.9) para hacer algunos comentarios:

1. La tasa de crecimiento definida en (10.9) es una tasa de crecimiento *instantánea*. Como la derivada $dV/dt = rAe^{rt}$ toma un valor distinto en un punto diferente de t, lo mismo que $V = Ae^{rt}$, su relación debe tener referencia también con un punto específico (o *instante*) de t. En este sentido, la tasa de crecimiento es instantánea.
2. En el presente caso la tasa instantánea de crecimiento es una constante r y, por lo tanto, la tasa de crecimiento es uniforme en todos los puntos de tiempo. Sin embargo, es posible que esto no resulte cierto para todas las situaciones de crecimiento.

3. Aunque la tasa de crecimiento r se mide en un punto particular del tiempo, su magnitud tiene la connotación de tanto por ciento *por unidad de tiempo*, por ejemplo, por año (si t se mide en unidades de año). El crecimiento, por su propia naturaleza, puede ocurrir sólo en un intervalo de tiempo. Ésta es la razón por la que una sola imagen fija (que registra la situación en un instante) nunca podría describir, por ejemplo, el desarrollo de un niño, mientras que dos imágenes fijas tomadas en diferentes tiempos, por ejemplo, a un año de distancia, pueden lograr esto. Decir que V tiene una tasa de crecimiento de r en el instante $t = t_0$ significa en realidad que, si se permite que la tasa de cambio $dV/dt(= rV)$ que prevalece en $t = t_0$ continúe inamovible durante una unidad completa de tiempo (un año), entonces V habrá crecido en la cantidad rV al final de año.
4. Para la función exponencial $V = Ae^{rt}$, la *tasa* de crecimiento *porcentual* es constante en todos los valores de t, pero la *cantidad absoluta* de incremento de V aumenta a media que pasa el tiempo, porque la tasa porcentual se calculará sobre bases cada vez más grandes.

Al interpretar r como la tasa de crecimiento instantánea, es claro que en lo sucesivo se requerirá poco esfuerzo para hallar la tasa de crecimiento de una función exponencial natural de la forma $y = Ae^{rt}$, siempre y cuando r sea constante. Dada una función $y = 75e^{0.02t}$, por ejemplo, se puede leer de inmediato la tasa de crecimiento de y como 0.02 o 2 por ciento por periodo.

Crecimiento continuo en relación con crecimiento discreto

El análisis anterior, aunque interesante desde el punto de vista analítico, aún es debatible por lo que respecta a la relevancia económica, porque en realidad el crecimiento no siempre toma lugar sobre una base *continua*, ni siquiera en la capitalización del interés. Por fortuna, aun en casos de crecimiento *discreto*, donde los cambios ocurren sólo una vez por periodo en vez de un instante a otro, se puede justificar el uso de la función de crecimiento exponencial continuo.

Entre otros, en los casos donde la frecuencia de capitalización es relativamente alta, aunque no infinita, el patrón continuo de crecimiento se puede considerar como una aproximación al patrón de crecimiento real. Pero, lo que es más importante, se puede mostrar que un problema de crecimiento discreto o continuo siempre se transforma en una versión continua equivalente.

Suponga que se tiene un patrón de crecimiento geométrico (por ejemplo, la capitalización *discreta* del interés), como se ilustra mediante la siguiente sucesión:

$$A,\ A(1+i),\ A(1+i)^2,\ A(1+i)^3, \ldots$$

donde la tasa de interés efectiva por periodo se denota mediante i y donde el exponente de la expresión $(1 + i)$ denota el número de periodos cubiertos en la capitalización. Si se considera que $(1 + i)$ es la base b en una expresión exponencial, entonces la sucesión dada se puede resumir mediante la función exponencial Ab^t, excepto que, como resultado de la naturaleza discreta del problema, t se restringe sólo a valores enteros. Además, $b = 1 + i$ es un número positivo (positivo incluso si i es una tasa de interés *negativa*, por ejemplo –0.04), de modo que siempre se puede expresar como una potencia de cualquier número real mayor que 1, incluso e. Esto significa que debe existir un número r tal que[1]

$$1 + i = b = e^r$$

[1] El método de hallar el número t, dado un valor específico de b, se analizará en la sección 10.4.

Por lo tanto, se puede transformar Ab^t en una función exponencial natural:

$$A(1+i)^t = Ab^t = Ae^{rt}$$

Para cualquier valor de t, en este contexto valores enteros de t, la función Ae^{rt} producirá exactamente el mismo valor que $A(1+i)^t$, tal que $A(1+i) = Ae^r$ y $A(1+i)^2 = Ae^{2r}$. En consecuencia, aunque se considera un caso *discreto* $A(1+i)^t$, aún podemos trabajar con la función exponencial natural *continua* Ae^{rt}. Esto explica por qué las funciones exponenciales naturales se aplican en gran medida en el análisis económico a pesar del hecho de que no todos los patrones de crecimiento son en realidad continuos.

Descuento y crecimiento negativo

Pasamos ahora, por un momento, de la capitalización de interés al concepto estrechamente relacionado del *descuento*. En un problema de interés compuesto, buscamos calcular el *valor futuro* V (principal más interés) a partir de un determinado *valor presente* A (principal inicial). El problema del *descuento* es lo contrario de hallar el valor presente A de una suma particular V, la cual estará disponible t años a partir de ahora.

Tomamos primero el caso discreto. Si la cantidad de principal A se convertirá en el valor futuro de $A(1+i)^t$ después de t años de capitalización anual a la tasa de interés i por año, es decir, si

$$V = A(1+i)^t$$

entonces, al dividir ambos lados de la ecuación entre la expresión no cero $(1+i)^t$, se obtiene la fórmula de descuento:

$$A = \frac{V}{(1+i)^t} = V(1+i)^{-t} \tag{10.10}$$

la cual lleva un exponente negativo. Se debe entender que en esta fórmula se han invertido los papeles de V y A: V ahora es un dato, mientras que A es la incógnita, que se calculará a partir de i (la tasa de descuento) y t (el número de años), así como V.

De manera similar, para el caso continuo, si el principal A se convertirá en Ae^{rt} después de t años de capitalización continua a la tasa r de acuerdo con la fórmula

$$V = Ae^{rt}$$

entonces podemos deducir la fórmula correspondiente de descuento continuo dividiendo ambos lados de la última ecuación entre e^{rt}:

$$A = \frac{V}{e^{rt}} = Ve^{-rt} \tag{10.11}$$

Aquí, A es de nuevo la incógnita (en vez de V), que calcularemos a partir del valor futuro V especificado, la tasa nominal de descuento r, y el número de años t. La expresión e^{-rt} se denomina *factor de descuento*.

Si tomamos a (10.11) como una función de crecimiento exponencial, leemos de inmediato a $-r$ como la tasa instantánea de crecimiento de A. Como es negativa, esta tasa es de hecho una *tasa de disminución*. Así como la capitalización del interés ejemplifica el proceso de crecimiento, el descuento ilustra crecimiento negativo.

EJERCICIO 10.2

1. Use la forma de serie infinita de e^x de (10.6) para hallar el valor aproximado de:
   (*a*) e^2 (*b*) $\sqrt{e}\ (= e^{1/2})$
   (Redondee el cálculo de cada término a tres decimales y continúe con la serie hasta que obtenga un término 0.000.)
2. Dada la función $\phi(x) = e^{2x}$:
   (*a*) Escriba la parte polinomial P_n de su serie de Maclaurin.
   (*b*) Escriba la forma de Lagrange del residuo R_n. Determine si $R_n \to 0$ cuando $n \to \infty$, es decir, si la serie converge a $\phi(x)$.
   (*c*) Si es convergente, de tal manera que $\phi(x)$ se pueda expresar como una serie infinita, escriba esta serie.
3. Escriba una expresión exponencial para el valor futuro de:
   (*a*) \$70, compuesto de forma continua a la tasa de interés de 4% durante 3 años;
   (*b*) \$690, compuesto de manera continua a la tasa de interés de 5% durante 2 años.
   (Estas tasas de interés son tasas nominales por año.)
4. ¿Cuál es la tasa de crecimiento instantánea de y en cada una de las siguientes expresiones?
   (*a*) $y = e^{0.07t}$ (*c*) $y = Ae^{0.4t}$
   (*b*) $y = 15e^{0.03t}$ (*d*) $y = 0.03e^t$
5. Muestre que las dos funciones $y_1 = Ae^{rt}$ (capitalización de interés) y $y_2 = Ae^{-rt}$ (descuento) son imágenes especulares entre sí en relación con el eje y [cf. ejercicio 10.1-5, inciso (*b*)].

10.3 Logaritmos

Las funciones exponenciales tienen relación estrecha con las *funciones logarítmicas* (*funciones log*, para abreviar). Antes de poder analizar las funciones log, se debe entender primero el significado del término *logaritmo*.

Significado de logaritmo

Cuando tenemos dos números, como 4 y 16, los cuales pueden estar relacionados entre sí mediante la ecuación $4^2 = 16$, definimos el *exponente* 2 como el *logaritmo* de 16 de base 4, y escribimos

$$\log_4 16 = 2$$

De este ejemplo debe quedar claro que el logaritmo no es sino la *potencia* a la que la base (4) se debe elevar para obtener un número particular (16). En general, se puede afirmar que

$$y = b^t \quad \Leftrightarrow \quad t = \log_b y \tag{10.12}$$

lo cual indica que el log de y de base b (denotado por $\log_b y$) es la potencia a la que se debe elevar la base b a fin de obtener el valor de y. Por esta razón, es correcto, aunque repetitivo, escribir,

$$b^{\log_b y} = y$$

Dada y, el proceso de hallar su logaritmo $\log_b y$ se conoce como *obtener* el log de *y de base b*. El proceso inverso, el de hallar y a partir de un valor conocido de su logaritmo $\log_b$, se designa como *tomar el antilog de* $\log_b y$.

En el análisis de funciones exponenciales remarcamos que la función $y = b^t$ (con $b > 1$) es estrictamente creciente. Esto significa que, para cualquier valor positivo de y, hay un exponente *único* t (no necesariamente positivo) tal que $y = b^t$; además, mientras más grande sea el valor de y, más grande debe ser t, como se ve en la figura 10.2. Traducido en logaritmos, la monotonía estricta de la función exponencial indica que cualquier número positivo y debe poseer un logaritmo *único* t de base $b > 1$ tal que mientras más grande sea y, más grande es su logaritmo. Según se ilustra en las figuras 10.1 y 10.2, y es necesariamente positiva en la función exponencial $y = b^t$; en consecuencia, un número negativo o cero no posee logaritmo.

Logaritmo común y logaritmo natural

La base del logaritmo, $b > 1$, no tiene que estar restringida a ningún número particular, pero en las aplicaciones reales de logaritmos dos números son ampliamente elegidos como bases, el número 10 y el número e. Cuando 10 es la base, el logaritmo se conoce como el *logaritmo común*, simbolizado por $\log_{10}$ (o si el contexto es claro, simplemente por log). Con e como la base, el logaritmo se conoce como *logaritmo natural* y se denota ya sea por $\log_e$ o por ln (para logaritmo natural). También se puede usar el símbolo log (sin el subíndice e) si no es ambiguo en el contexto particular.

Los logaritmos comunes, utilizados con frecuencia en el trabajo *computacional*, se ejemplifican mediante lo siguiente:

$$
\begin{aligned}
\log_{10} 1\,000 &= 3 \qquad && [\text{porque } 10^3 = 1\,000] \\
\log_{10} 100 &= 2 && [\text{porque } 10^2 = 100] \\
\log_{10} 10 &= 1 && [\text{porque } 10^1 = 10] \\
\log_{10} 1 &= 0 && [\text{porque } 10^0 = 1] \\
\log_{10} 0.1 &= -1 && [\text{porque } 10^{-1} = 0.1] \\
\log_{10} 0.01 &= -2 && [\text{porque } 10^{-2} = 0.01]
\end{aligned}
$$

Observe la estrecha relación que hay entre el conjunto de números justo a la izquierda de los signos de igualdad y el conjunto de los números que están inmediatamente a la derecha. De éstos, debe ser evidente que el logaritmo común de un número entre 10 y 100 debe estar entre 1 y 2, y que el logaritmo común de un número entre 1 y 10 debe ser una fracción positiva, etc. Los logaritmos exactos se obtienen de una tabla de logaritmos comunes o calculadoras electrónicas con la capacidad de calcular logaritmos.[2]

Sin embargo, en el trabajo *analítico* los logaritmos naturales demuestran ser mucho más convenientes que los logaritmos comunes. Puesto que por la definición de logaritmo se tiene la relación

$$y = e^t \quad \Leftrightarrow \quad t = \log_e y \quad (\text{o bien } t = \ln y) \qquad \textbf{(10.13)}$$

es fácil ver que la conveniencia analítica de e en las funciones exponenciales se amplía de forma automática al ámbito de los logaritmos con e como base.

[2] Más fundamentalmente, el valor de un logaritmo, como el valor de e, se puede calcular (o aproximar) recurriendo al desarrollo de Maclaurin de una función log, de una manera similar a la descrita en (10.6). Sin embargo, esta deducción no se trata aquí.

Los siguientes ejemplos sirven para ilustrar los logaritmos naturales:

$$\begin{aligned} \ln e^3 &= \log_e e^3 = 3 \\ \ln e^2 &= \log_e e^2 = 2 \\ \ln e^1 &= \log_e e^1 = 1 \\ \ln 1 &= \log_e e^0 = 0 \\ \ln \frac{1}{e} &= \log_e e^{-1} = -1 \end{aligned}$$

El principio general que surge de estos ejemplos es que, dada una expresión e^k, donde k es algún número real, podemos leer automáticamente el exponente k como el logaritmo natural de e^k. Por lo tanto, en general se tiene el resultado de que $\ln e^k = k$.[3]

El logaritmo común y el logaritmo natural son convertibles entre sí; es decir, se puede cambiar la base de un logaritmo, lo mismo que la base de una expresión exponencial. Después de haber estudiado las reglas básicas de los logaritmos se procede a obtener un par de fórmulas de conversión.

Reglas de los logaritmos

Los logaritmos son de la naturaleza de los exponentes; por lo tanto, siguen ciertas reglas que tienen relación estrecha con las de los exponentes presentadas en la sección 2.5. Éstas pueden ser de gran ayuda en la simplificación de operaciones matemáticas. Las primeras tres reglas se expresan sólo en términos del logaritmo natural, pero también son válidas cuando se reemplaza el símbolo ln por $\log_b$.

Regla I (log de un producto) $\qquad \ln(uv) = \ln u + \ln v \qquad (u, v > 0)$

Ejemplo 1

$$\ln(e^6 e^4) = \ln e^6 + \ln e^4 = 6 + 4 = 10$$

Ejemplo 2

$$\ln(Ae^7) = \ln A + \ln e^7 = \ln A + 7$$

DEMOSTRACIÓN Por definición, $\ln u$ es la potencia a la cual se debe elevar e para obtener el valor de u; así, $e^{\ln u} = u$.[4] De manera similar, se tiene $e^{\ln v} = v$ y $e^{\ln(uv)} = uv$. Esta última es una expresión exponencial para uv. Sin embargo, otra expresión de uv se obtiene mediante multiplicación directa de u y v:

$$uv = e^{\ln u} e^{\ln v} = e^{\ln u + \ln v}$$

Así, al igualar las dos expresiones para uv anteriores, se encuentra

$$e^{\ln(uv)} = e^{\ln u + \ln v} \quad \text{y, por consiguiente, } \ln(uv) = \ln u + \ln v$$

Regla II (log de un cociente) $\qquad \ln(u/v) = \ln u - \ln v \qquad (u, v > 0)$

[3] Como ayuda nemotécnica, observe que cuando al símbolo ln (o $\log_e$) se le coloca al final la expresión e^k, el símbolo ln parece cancelar al símbolo e, y deja a k como respuesta.

[4] Note que cuando e se eleva a la potencia $\ln u$, el símbolo e y el símbolo ln de nuevo parecen cancelarse, y queda u como respuesta.

Ejemplo 3

$$\ln(e^2/c) = \ln e^2 - \ln c = 2 - \ln c$$

Ejemplo 4

$$\ln(e^2/e^5) = \ln e^2 - \ln e^5 = 2 - 5 = -3$$

La demostración de esta regla es muy similar a la de la regla I y, por lo tanto, se deja como ejercicio.

Regla III (log de una potencia) $\ln u^a = a \ln u \qquad (u > 0)$

Ejemplo 5

$$\ln e^{15} = 15 \ln e = 15$$

Ejemplo 6

$$\ln A^3 = 3 \ln A$$

DEMOSTRACIÓN Por definición, $e^{\ln u} = u$; y de manera similar, $e^{\ln u^a} = u^a$. Sin embargo, otra expresión para u^a se forma como sigue:

$$u^a = (e^{\ln u})^a = e^{a \ln u}$$

Al igualar los exponentes en las dos expresiones para u^a, se obtiene el resultado deseado, $\ln u^a = a \ln u$.

Estas tres reglas son mecanismos útiles para simplificar las operaciones matemáticas en cierto tipo de problemas. La regla I sirve para convertir, vía logaritmos, una operación multiplicativa (uv) en una aditiva $(\ln u + \ln v)$; la regla II convierte una división (u/v) en una resta $(\ln u - \ln v)$; y la regla III permite reducir una potencia a una constante multiplicativa. Además, estas reglas se pueden combinar entre sí. Asimismo, se pueden leer hacia atrás, y aplicarse a la inversa.

Ejemplo 7

$$\ln(uv^a) = \ln u + \ln v^a = \ln u + a \ln v$$

Ejemplo 8

$$\ln u + a \ln v = \ln u + \ln v^a = \ln(uv^a) \qquad \text{[Ejemplo 7 a la inversa]}$$

Sin embargo, advierta que cuando se tienen de inicio expresiones *aditivas*, los logaritmos no son útiles. En particular, se debe recordar que

$$\ln(u \pm v) \neq \ln u \pm \ln v$$

A continuación introducimos otras dos reglas que tienen que ver con cambios en la base de un logaritmo.

Regla IV (conversión de base log) $\log_b u = (\log_b e)(\log_e u) \qquad (u > 0)$

Esta regla, que se asemeja a la regla de la cadena (tenga en cuenta la "cadena" $b \nearrow^{e} \searrow_{e} \nearrow^{u}$, nos permite obtener un logaritmo $\log_e u$ (de base e) del logaritmo $\log_b u$ (de base b), o viceversa.

DEMOSTRACIÓN Sea $u = e^p$, de modo que $p = \log_e u$. Entonces se deduce que

$$\log_b u = \log_b e^p = p \log_b e = (\log_e u)(\log_b e)$$

La regla IV se puede generalizar fácilmente a

$$\log_b u = (\log_b c)(\log_c u)$$

donde c es alguna base distinta de b.

Regla V (inversión de la base log) $$\log_b e = \frac{1}{\log_e b}$$

Esta regla, que se asemeja a la regla de diferenciación de la función inversa, permite obtener de inmediato el log de b de base e con el log de e de base b, y viceversa (esta regla se puede generalizar también en la forma $\log_b c = 1/\log_c b$).

DEMOSTRACIÓN Como una aplicación de la regla IV, si $u = b$; entonces tenemos

$$\log_b b = (\log_b e)(\log_e b)$$

Pero la expresión del lado izquierdo es $\log_b b = 1$; por lo tanto, $\log_b e$ y $\log_e b$ deben ser recíprocas, como se afirma en la regla V.

De las dos últimas reglas es fácil deducir el siguiente par de fórmulas de conversión entre log común y log natural:

$$\begin{aligned} \log_{10} N &= (\log_{10} e)(\log_e N) = 0.4343 \log_e N \\ \log_e N &= (\log_e 10)(\log_{10} N) = 2.3026 \log_{10} N \end{aligned} \qquad \textbf{(10.14)}$$

para un número real positivo N. El primer signo igual en cada fórmula se justifica fácilmente mediante la regla IV. En la primera fórmula, el valor 0.4343 (el log común de 2.71828) se obtiene de una tabla de logaritmos comunes o con una calculadora electrónica; en la segunda, el valor 2.3026 (el log natural de 10) es solamente el recíproco de 0.4343, calculado de este modo como resultado de la regla V.

Ejemplo 9 $\log_e 100 = 2.3026(\log_{10} 100) = 2.3026(2) = 4.6052$. Por el contrartio, se tiene $\log_{10} 100 = 0.4343(\log_e 100) = 0.4343(4.6052) = 2$.

Una aplicación

Las reglas anteriores de logaritmos nos permiten resolver con facilidad ciertas *ecuaciones exponenciales* simples (*funciones* exponenciales que se igualan a cero). Por ejemplo, si intentamos hallar el valor de x que satisface la ecuación

$$ab^x - c = 0 \qquad (a, b, c > 0)$$

podemos intentar primero transformar esta ecuación exponencial, mediante el uso de logaritmos, en una ecuación *lineal* y luego resolverla como tal. Para este propósito, el término c se debe pasar al lado derecho:

$$ab^x = c$$

Esto es porque no hay expresión log simple para la expresión aditiva $(ab^x - c)$, pero existen expresiones logarítmicas convenientes para el término multiplicativo ab^x y para c de manera individual. Así, después de pasar c al segundo miembro y tomar el log (por ejemplo, base 10) de ambos lados, se tiene

$$\log a + x \log b = \log c$$

que es una ecuación lineal en la variable x, con la solución

$$x = \frac{\log c - \log a}{\log b}$$

EJERCICIO 10.3

1. ¿Cuáles son los valores de los siguientes logaritmos?
   (*a*) $\log_{10} 10\,000$ (*c*) $\log_3 81$
   (*b*) $\log_{10} 0.0001$ (*d*) $\log_5 3\,125$
2. Evalúe lo siguiente:
   (*a*) $\ln e^7$ (*c*) $\ln(1/e^3)$ (*e*) $(e^{\ln 3})!$
   (*b*) $\log_e e^{-4}$ (*d*) $\log_e(1/e^2)$ (*f*) $\ln e^x - e^{\ln x}$
3. Evalúe lo siguiente mediante la aplicación de las reglas de los logaritmos:
   (*a*) $\log_{10}(100)^{13}$ (*c*) $\ln(3/B)$ (*e*) $\ln ABe^{-4}$
   (*b*) $\log_{10} \frac{1}{100}$ (*d*) $\ln Ae^2$ (*f*) $(\log_4 e)(\log_e 64)$
4. De las siguientes expresiones, ¿cuáles son válidas?
   (*a*) $\ln u - 2 = \ln \dfrac{u}{e^2}$ (*c*) $\ln u + \ln v - \ln w = \ln \dfrac{uv}{w}$
   (*b*) $3 + \ln v = \ln \dfrac{e^3}{v}$ (*d*) $\ln 3 + \ln 5 = \ln 8$
5. Pruebe que $\ln (u/v) = \ln u - \ln v$.

10.4 Funciones logarítmicas

Cuando una variable se expresa como una función del logaritmo de otra variable, la función se conoce como *función logarítmica*. Ya se han visto dos versiones de este tipo de función en (10.12) y (10.13), a saber,

$$t = \log_b y \qquad \text{y} \qquad t = \log_e y\ (= \ln y)$$

que difieren entre sí sólo respecto al logaritmo.

Funciones logarítmica y exponencial

Como expresamos antes, las funciones log son funciones inversas de ciertas funciones exponenciales. Un examen de las dos funciones log previas nos confirma que son las funciones inversas respectivas de las funciones exponenciales

$$y = b^t \qquad \text{y} \qquad y = e^t$$

porque las funciones log citadas son los resultados de invertir los papeles de las variables dependiente e independiente de las funciones exponenciales correspondientes. Se debe entender que el símbolo t se usa aquí como símbolo general, y no necesariamente significa *tiempo*. Aun cuando sea el caso, su presencia como variable *dependiente* no significa que el tiempo se determine mediante alguna variable y; significa sólo que un determinado valor de y se relaciona con un único punto de tiempo.

Como inversas de funciones (exponenciales) estrictamente crecientes, las funciones logarítmicas también deben ser estrictamente crecientes, lo cual es congruente con la afirmación anterior de que mientras más grande sea un número, mayor es su logaritmo para cualquier base particular. Esta propiedad se puede expresar en forma simbólica en términos de las dos proposiciones siguientes: para dos valores positivos de y (y_1 y y_2),

$$
\begin{aligned}
\ln y_1 = \ln y_2 \quad &\Leftrightarrow \quad y_1 = y_2 \\
\ln y_1 > \ln y_2 \quad &\Leftrightarrow \quad y_1 > y_2
\end{aligned}
\tag{10.15}
$$

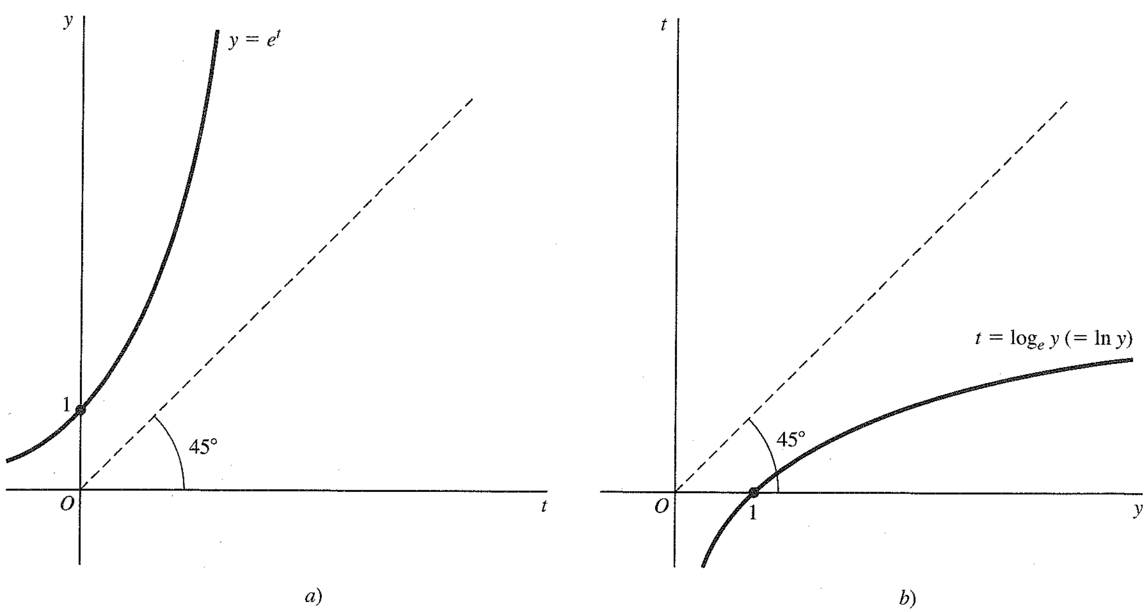
Estas proposiciones son válidas también si se reemplaza ln por $\log_b$.

Forma gráfica

La monotonía y otras propiedades generales de las funciones logarítmicas se ve con claridad en sus gráficas. Dada la gráfica de la función exponencial $y = e^t$, podemos obtener la gráfica de la función log correspondiente al trazar de nuevo la gráfica original con los dos ejes cambiados de posición. Este resultado se ilustra en la figura 10.3. Note que si la gráfica de la figura 10.3*b* se colocara sobre la gráfica de la figura 10.3*a*, con el eje y sobre el eje y y el eje t sobre el eje t, las dos curvas deben coincidir con exactitud. Como aparecen en la figura 10.3, con los ejes intercambiados, las dos curvas son imágenes especulares (como deben ser las gráficas de cualquier par de funciones inversas) respecto a la recta de 45° que pasa por el origen.

Esta relación de imagen especular tiene varias implicaciones notables. Para empezar, aunque ambas son estrictamente crecientes, en la curva log observamos que y aumenta a una *tasa decreciente* (segunda derivada negativa), en contraposición a la curva exponencial, en la que observamos que y aumenta a una tasa creciente. Otro contraste interesante es que, si bien la función exponencial tiene una *imagen* positiva, la función log tiene un *dominio* positivo. (Esta

FIGURA 10.3



última restricción en el dominio de la función log es otra forma de expresar que sólo los números positivos poseen logaritmos.) Una tercera consecuencia de la relación de imagen especular es que, así como $y = e^t$ tiene una intersección vertical en 1, la función log $t = \log_e y$ debe cruzar el eje horizontal en $y = 1$, indicando que $\log_e 1 = 0$. En vista de que la base del logaritmo no afecta a esta intersección horizontal, por ejemplo, $\log_{10} 1 = 0$, se puede inferir de la forma general de la curva log de la figura 10.3*b* que, para *cualquier* base,

$$\left.\begin{array}{r} 0 < y < 1 \\ y = 1 \\ y > 1 \end{array}\right\} \quad \Leftrightarrow \quad \left\{\begin{array}{l} \log y < 0 \\ \log y = 0 \\ \log y > 0 \end{array}\right. \tag{10.16}$$

Como comprobación, se pueden verificar los dos conjuntos de ejemplos de logaritmos comunes y naturales de la sección 10.3. Además, se puede notar que

$$\log y \to \left\{\begin{array}{c} \infty \\ -\infty \end{array}\right\} \text{cuando } y \to \left\{\begin{array}{l} \infty \\ 0^+ \end{array}\right. \tag{10.16'}$$

La comparación gráfica de la función logarítmica y la función exponencial en la figura 10.3 se basa en las funciones simples $y = e^t$ y $t = \ln y$. El mismo resultado general prevalece si comparamos la función exponencial generalizada $y = Ae^{rt}$ con su función log correspondiente. Con las constantes A y r (positivas) para *comprimir* o *extender* la curva exponencial, se asemejará a la forma general de la figura 10.3*a*, excepto por que su intersección vertical estará en $y = A$ en vez de en $y = 1$ (cuando $t = 0$, se tiene $y = Ae^0 = A$). En consecuencia, su función inversa debe tener una intersección *horizontal* en $y = A$. En general, respecto a la recta de 45°, la curva log correspondiente será una imagen especular de la curva exponencial.

Si lo que deseamos es la expresión algebraica específica de la inversa de $y = Ae^{rt}$, obtenemos ésta tomando el logaritmo natural de ambos lados de esta función exponencial [lo cual, de acuerdo con la primera proposición de (10.15), no modifica la ecuación] y luego se despeja t:

$$\ln y = \ln(Ae^{rt}) = \ln A + rt \ln e = \ln A + rt$$

Por consiguiente,

$$t = \frac{\ln y - \ln A}{r} \qquad (r \neq 0) \tag{10.17}$$

Este resultado, una función log, constituye la inversa de la función exponencial $y = Ae^{rt}$. Como se afirmó, la función de (10.17) tiene una intersección horizontal en $y = A$, porque cuando $y = A$, se tiene $\ln y = \ln A$ y, por lo tanto, $t = 0$.

Conversión de base

En la sección 10.2 se afirmó que la función exponencial $y = Ab^t$ siempre se puede convertir en una función exponencial *natural* $y = Ae^{rt}$. Ahora ya se puede obtener una fórmula de conversión. Sin embargo, en lugar de Ab^t considere la conversión de la expresión más general Ab^{ct} en Ae^{rt}. Puesto que la esencia del problema es hallar una r a partir de valores de b y c tal que

$$e^r = b^c$$

todo lo que se requiere es expresar r como una función de b y c. Esta tarea es fácil de llevar a cabo si tomamos el logaritmo natural de ambos lados de la última ecuación:

$$\ln e^r = \ln b^c$$

El lado izquierdo se puede leer de inmediato como igual a r, de modo que la función deseada (fórmula de conversión) surge como

$$r = \ln b^c = c \ln b \tag{10.18}$$

Esto indica que la función $y = Ab^{ct}$ se puede reescribir siempre en la forma de base natural, $y = Ae^{(c \ln b)t}$.

Ejemplo 1

Convierta $y = 2^t$ a una función exponencial natural. Aquí se tiene $A = 1$, $b = 2$ y $c = 1$. Por lo tanto, $r = c \ln b = \ln 2$, y la función exponencial deseada es

$$y = Ae^{rt} = e^{(\ln 2)t}$$

Si queremos, podemos calcular también el valor numérico de ln 2 por medio de (10.14) y una tabla de logaritmos comunes como sigue:

$$\ln 2 = 2.3026 \log_{10} 2 = 2.3026(0.3010) = 0.6931 \tag{10.19}$$

Entonces el resultado anterior se podría expresar de otra manera como $y = e^{0.6931t}$.

Ejemplo 2

Convierta $y = 3(5)^{2t}$ a una función exponencial natural. En este ejemplo, $A = 3$, $b = 5$ y $c = 2$, y con la fórmula (10.18) se obtiene $r = 2 \ln 5$. Por lo tanto la función deseada es

$$y = Ae^{rt} = 3e^{(2 \ln 5)t}$$

De nuevo, si queremos, podemos calcular que

$$2 \ln 5 = \ln 25 = 2.3026 \log_{10} 25 = 2.3026(1.3979) = 3.2188$$

así que el resultado anterior se puede expresar de otra manera como $y = 3e^{3.2188t}$.

También podemos, por supuesto, convertir las funciones log de la forma $t = \log_b y$ en funciones log naturales equivalentes. Para lograrlo, es suficiente aplicar la regla IV de los logaritmos, que se puede expresar como

$$\log_b y = (\log_b e)(\log_e y)$$

La sustitución directa de este resultado en la función log proporcionada produce la función log natural deseada:

$$\begin{aligned} t &= \log_b y = (\log_b e)(\log_e y) \\ &= \frac{1}{\log_e b} \log_e y \qquad \text{[por la regla V de los logaritmos]} \\ &= \frac{\ln y}{\ln b} \end{aligned}$$

Por el mismo procedimiento, se puede transformar la función log más general $t = a \log_b(cy)$ en la forma equivalente

$$t = a(\log_b e)(\log_e cy) = \frac{a}{\log_e b} \log_e(cy) = \frac{a}{\ln b} \ln(cy)$$

Ejemplo 3 Convierta la función $t = \log_2 y$ y en log natural. Puesto que en este ejemplo se tiene $b = 2$ y $a = c = 1$, la función deseada es

$$t = \frac{1}{\ln 2} \ln y$$

Sin embargo, por (10.19) podemos expresarla también como $t = (1/0.6931) \ln y$.

Ejemplo 4 Convierta la función $t = 7 \log_{10}(2y)$ en una función logarítmica natural. Los valores de las constantes son en este caso $a = 7$, $b = 10$ y $c = 2$; en consecuencia, la función deseada es

$$t = \frac{7}{\ln 10} \ln(2y)$$

Pero puesto que $\ln 10 = 2.3026$, como indica (10.14), esta función se puede reescribir como $t = (7/2.3026) \ln(2y) = 3.0400 \ln(2y)$

En la explicación precedente hemos seguido la práctica de expresar t como una función de y cuando la función es logarítmica. La única razón para proceder de esto modo es nuestro deseo de remarcar la relación de función inversa entre las funciones exponencial y logarítmica. Cuando se estudia *por sí sola* una función log, se escribirá $y = \ln t$ (en vez de $t = \ln y$), como es habitual. Naturalmente, nada del aspecto analítico de la explicación se verá afectado por un intercambio de símbolos.

EJERCICIO 10.4

1. La forma de la función inversa de $y = Ae^{rt}$ de (10.17) requiere que r sea distinta de cero. ¿Cuál es el significado de este requerimiento cuando se considera en relación con la función exponencial original $y = Ae^{rt}$?
2. (*a*) Bosqueje una gráfica de la función exponencial $y = Ae^{rt}$; indique el valor de la intersección vertical.
   (*b*) Luego trace la gráfica de la función logarítmica $t = \dfrac{\ln y - \ln A}{r}$ e indique el valor de la intersección horizontal.
3. Encuentre la función inversa de $y = ab^{ct}$.
4. Transforme las siguientes funciones a sus formas exponenciales naturales:
   (*a*) $y = 8^{3t}$ (*c*) $y = 5(5)^t$
   (*b*) $y = 2(7)^{2t}$ (*d*) $y = 2(15)^{4t}$
5. Transforme las funciones siguientes a sus formas logarítmicas naturales:
   (*a*) $t = \log_7 y$ (*c*) $t = 3 \log_{15}(9y)$
   (*b*) $t = \log_8(3y)$ (*d*) $t = 2 \log_{10} y$
6. Determine la tasa de interés nominal de capitalización continua por año (r) que sea equivalente a una tasa de interés de capitalización discreta (i) de
   (*a*) Cinco por ciento anual, capitalizado anualmente.
   (*b*) Cinco por ciento anual, capitalizado cada medio año.
   (*c*) Seis por ciento anual, capitalizado cada medio año.
   (*d*) Seis por ciento anual, capitalizado trimestralmente.

7. (*a*) Al describir la figura 10.3, en el texto se expresa que, si las dos curvas se colocan una sobre otra, presentan una relación de imagen especular. ¿Dónde se localiza el "espejo"?
   (*b*) Si trazamos una función $f(x)$ y su negativa, $-f(x)$, en el mismo diagrama, ¿mostrarán también las dos curvas una relación de imagen especular? En caso afirmativo, ¿dónde se localiza el espejo en este caso?
   (*c*) Si bosquejamos las gráficas de Ae^{rt} y Ae^{-rt} en el mismo diagrama, ¿serán imágenes especulares entre sí las dos curvas? Si es así, ¿dónde se localiza el "espejo"?

10.5 Derivadas de funciones exponenciales y logarítmicas

Antes afirmamos que la función e^t es su propia derivada. Resulta que la función log natural, ln t, posee una derivada bastante conveniente también, a saber, $d(\ln t)/dt = 1/t$. Este hecho refuerza la preferencia por la base e. Enseguida probamos la validez de estas dos fórmulas de derivadas, y luego deducimos las fórmulas para ciertas variantes de las expresiones exponencial y logarítmica e^t y ln t.

Regla de la función log

La derivada de la función log $y = \ln t$ es

$$\frac{d}{dt}\ln t = \frac{1}{t}$$

Para probar esto, recuerde que, por definición, la derivada de $y = \psi(t) = \ln t$ tiene el siguiente valor en $t = N$ (suponiendo que $t \to N^+$):

$$\begin{aligned}\psi'(N) &= \lim_{t\to N^+} \frac{\psi(t) - \psi(N)}{t - N} = \lim_{t\to N^+} \frac{\ln t - \ln N}{t - N} \\ &= \lim_{t\to N^+} \frac{\ln(t/N)}{t - N} \qquad \text{[por la regla II de los logaritmos]}\end{aligned}$$

Ahora introducimos un símbolo $m \equiv \dfrac{N}{t - N}$. Entonces se puede escribir $\dfrac{1}{t - N} = \dfrac{m}{N}$, y también $\dfrac{t}{N} = 1 + \dfrac{t - N}{N} = 1 + \dfrac{1}{m}$. Por lo tanto, la expresión que está a la derecha del signo de límite en la ecuación previa se puede convertir a la forma

$$\frac{1}{t - N}\ln\frac{t}{N} = \frac{m}{N}\ln\left(1 + \frac{1}{m}\right) = \frac{1}{N}\ln\left(1 + \frac{1}{m}\right)^m \quad \text{[por la regla III de los logaritmos]}$$

Note que, cuando $t \to N^+$, m tiende a infinito. Así, para hallar el valor de derivada deseado, podemos tomar el límite de la última expresión de la ecuación precedente cuando $m \to \infty$:

$$\psi'(N) = \lim_{m\to\infty} \frac{1}{N}\ln\left(1 + \frac{1}{m}\right)^m = \frac{1}{N}\ln e = \frac{1}{N} \qquad \text{[por (10.5)]}$$

Sin embargo, puesto que N puede ser cualquier número para el cual se define un logaritmo, podemos generalizar este resultado y escribir $\psi'(t) = d(\ln t)/dt = 1/t$. Esto prueba la regla de la función log para $t \to N^+$.

El caso de $t \to N^-$ necesita algunas modificaciones, pero en esencia la prueba es similar. Ahora la derivada de $y = \ln t$ tiene el valor

$$\psi'(N) = \lim_{t\to N^-} \frac{\psi(t)-\psi(N)}{t-N} = \lim_{t\to N^-} \frac{\psi(N)-\psi(t)}{N-t}$$
$$= \lim_{t\to N^-} \frac{\ln N - \ln t}{N-t} = \lim_{t\to N^-} \frac{\ln(N/t)}{N-t}$$

Sea $\mu = t/(N-t)$ entonces $1/(N-t) = \mu/t$ y $N/t = 1+(N-t)/t = 1+1/\mu$. Estas ecuaciones permiten escribir la expresión a la derecha del último signo de límite en la ecuación anterior para $\psi'(N)$ como

$$\frac{1}{N-t}\ln\frac{N}{t} = \frac{\mu}{t}\ln\left(1+\frac{1}{\mu}\right) = \frac{1}{t}\ln\left(1+\frac{1}{\mu}\right)^{\mu}$$

Cuando $t \to N^-$, $\mu \to \infty$. Así, el valor deseado de la derivada es

$$\psi'(N) = \lim_{t\to N^-} \frac{1}{N}\ln e = \frac{1}{N}$$

el mismo resultado que para el caso de $t \to N^+$. Con esto se completa la demostración de la regla de la función log. Observe, una vez más, que en el proceso de demostración no se emplea ningún valor numérico específico y, por lo tanto, el resultado es aplicable en general.

Regla de la función exponencial

La derivada de la función $y = e^t$ es

$$\frac{d}{dt}e^t = e^t$$

Este resultado se deduce fácilmente de la regla de la función log. Se sabe que la función inversa de la función $y = e^t$ es $t = \ln y$, con derivada $dt/dy = 1/y$. Así, por la regla de la función inversa, se puede escribir de inmediato

$$\frac{d}{dt}e^t = \frac{dy}{dt} = \frac{1}{dt/dy} = \frac{1}{1/y} = y = e^t$$

Reglas generalizadas

Las reglas de las funciones logaritmo y exponencial se pueden generalizar a casos donde la variable t en la expresión e^t y $\ln t$ se reemplaza por alguna *función* de t, por ejemplo $f(t)$. Las versiones generalizadas de las dos reglas son

$$\begin{aligned} &\frac{d}{dt}e^{f(t)} = f'(t)e^{f(t)} \qquad \left[\text{o} \quad \frac{d}{dt}e^u = e^u\frac{du}{dt}\right] \\ &\frac{d}{dt}\ln f(t) = \frac{f'(t)}{f(t)} \qquad \left[\text{o} \quad \frac{d}{dt}\ln v = \frac{1}{v}\frac{dv}{dt}\right] \end{aligned} \tag{10.20}$$

Las demostraciones para (10.20) no requieren más que la aplicación directa de la regla de la cadena. Dada una función $y = e^{f(t)}$ se puede establecer primero $u = f(t)$, de manera que $y = e^u$. Entonces, por la regla de la cadena, la derivada surge como

$$\frac{d}{dt}e^{f(t)} = \frac{d}{dt}e^u = \frac{de^u}{du}\frac{du}{dt} = e^u\frac{du}{dt} = e^{f(t)}f'(t)$$

De manera similar, dada una función $y = \ln f(t)$, se puede permitir primero que $v = f(t)$, a fin de formar una cadena: $y = \ln v$, donde $v = f(t)$. Entonces, por la regla de la cadena, se tiene

$$\frac{d}{dt} \ln f(t) = \frac{d}{dt} \ln v = \frac{d \ln v}{dv} \frac{dv}{dt} = \frac{1}{v} \frac{dv}{dt} = \frac{1}{f(t)} f'(t)$$

Note que sólo la modificación real introducida en (10.20) más allá de las reglas simples $de^t/dt = e^t$ y $d(\ln t)/dt = 1/t$ es el factor multiplicativo $f'(t)$.

Ejemplo 1 Encuentre la derivada de la función $y = e^{rt}$. Aquí, el exponente es $rt = f(t)$, con $f'(t) = r$; así

$$\frac{dy}{dt} = \frac{d}{dt} e^{rt} = re^{rt}$$

Ejemplo 2 Halle dy/dt de la función $y = e^{-t}$. Puesto que en este caso $f(t) = -t$, de tal manera que $f'(t) = -1$. Como resultado,

$$\frac{dy}{dt} = \frac{d}{dt} e^{-t} = -e^{-t}$$

Ejemplo 3 Encuentre dy/dt de la función $y = \ln(at)$. Puesto que en este caso $f(t) = at$, con $f'(t) = a$, la derivada es

$$\frac{d}{dt} \ln(at) = \frac{a}{at} = \frac{1}{t}$$

que es, cuestión bastante interesante, idéntica a la derivada de $y = \ln t$.

Este ejemplo ilustra el hecho de que una constante multiplicativa para t *dentro de* una expresión logarítmica desaparece en el proceso de derivación. Pero note que, para una constante k *multiplicada por fuera,* se tiene

$$\frac{d}{dt} k \ln t = k \frac{d}{dt} \ln t = \frac{k}{t}$$

así, una constante multiplicativa *fuera* de la expresión logarítmica no desaparece con la derivación.

Ejemplo 4 Halle la derivada de la función $y = \ln t^c$. Con $f(t) = t^c$ y $f'(t) = ct^{c-1}$, la fórmula de (10.20) produce

$$\frac{d}{dt} \ln t^c = \frac{ct^{c-1}}{t^c} = \frac{c}{t}$$

Ejemplo 5 Encuentre dy/dt a partir de $y = t^3 \ln t^2$. Debido a que esta función es un producto de dos términos t^3 y $\ln t^2$, se debe usar la regla del producto:

$$\begin{aligned} \frac{dy}{dt} &= t^3 \frac{d}{dt} \ln t^2 + (\ln t^2) \frac{d}{dt} t^3 \\ &= t^3 \left(\frac{2t}{t^2}\right) + (\ln t^2)(3t^2) \\ &= 2t^2 + 3t^2(2 \ln t) \qquad \text{[Regla III de los logaritmos]} \\ &= 2t^2(1 + 3 \ln t) \end{aligned}$$

Caso de base b

Para funciones exponenciales y logarítmicas con base b, las derivadas son

$$\frac{d}{dt}b^t = b^t \ln b \quad \left[\textit{Advertencia: } \frac{d}{dt}b^t \neq tb^{t-1}\right]$$
$$\frac{d}{dt}\log_b t = \frac{1}{t \ln b} \tag{10.21}$$

Note que en el caso especial de base e (cuando $b = e$), se tiene $\ln b = \ln e = 1$, para que estas dos derivadas se reduzcan a la regla básica de función exponencial $(d/dt)e^t = e^t$ y la regla básica de función logarítmica $(d/dt)\ln t = 1/t$, respectivamente.

Las demostraciones para (10.21) no son difíciles. Para el caso de b^t, la prueba se basa en la identidad $b \equiv e^{\ln b}$, que nos permite escribir

$$b^t = e^{(\ln b)t} = e^{t \ln b}$$

(Escriba $t \ln b$, en lugar de $\ln bt$, a fin de remarcar que t no es parte del argumento de la expresión logarítmica.) Por lo tanto,

$$\begin{aligned}\frac{d}{dt}b^t &= \frac{d}{dt}e^{t\ln b} = (\ln b)(e^{t \ln b}) \qquad \text{[por (10.20)]}\\ &= (\ln b)(b^t) = b^t \ln b\end{aligned}$$

Por otro lado, para probar la segunda parte de (10.21) se cuenta con la propiedad logarítmica básica de que

$$\log_b t = (\log_b e)(\log_e t) = \frac{1}{\ln b}\ln t$$

que da lugar a la derivada

$$\frac{d}{dt}\log_b t = \frac{d}{dt}\left(\frac{1}{\ln b}\ln t\right) = \frac{1}{\ln b}\frac{d}{dt}\ln t = \frac{1}{\ln b}\left(\frac{1}{t}\right)$$

Las versiones más generales de estas dos fórmulas son

$$\frac{d}{dt}b^{f(t)} = f'(t)b^{f(t)} \ln b$$
$$\frac{d}{dt}\log_b f(t) = \frac{f'(t)}{f(t)}\frac{1}{\ln b} \tag{10.21'}$$

De nuevo, se observa que si $b = e$, entonces $\ln b = 1$, y estas fórmulas se reducen a (10.20).

Ejemplo 6

Encuentre la derivada de la función $y = 12^{1-t}$. Aquí, $b = 12$, $f(t) = 1 - t$ y $f'(t) = -1$; por lo tanto,

$$\frac{dy}{dt} = -(12)^{1-t} \ln 12$$

Derivadas superiores

Las derivadas superiores de funciones exponenciales y logarítmicas, al igual que las de otros tipos de funciones, son solamente los resultados de diferenciación repetida.

Ejemplo 7 Encuentre la *segunda* derivada de $y = b^t$ (con $b > 1$). La primera derivada, por (10.21), es $y'(t) = b^t \ln b$ (donde $\ln b$ es, por supuesto, una constante); así, diferenciando una vez más respecto a t, se tiene

$$y''(t) = \frac{d}{dt}y'(t) = \left(\frac{d}{dt}b^t\right) \ln b = (b^t \ln b) \ln b = b^t(\ln b)^2$$

Note que $y = b^t$ es siempre positiva y $\ln b$ (para $b > 1$) también es positivo [por (10.16)]; por lo tanto, $y'(t) = b^t \ln b$ debe ser positiva. Y $y''(t)$, siendo un producto de b^t y un número cuadrado, también es positiva. Estos hechos confirman la afirmación previa de que la función exponencial $y = b^t$ se incremente monotónicamente a una tasa creciente.

Ejemplo 8 Obtenga la *segunda* derivada de $y = \ln t$. La primera derivada es $y' = 1/t = t^{-1}$; por consiguiente, la segunda derivada es

$$y'' = -t^{-2} = \frac{-1}{t^2}$$

En vista de que el dominio de esta función consta del intervalo abierto $(0, \infty)$, $y' = 1/t$ debe ser un número positivo. Por otro lado, y'' siempre es negativa. Juntas, estas conclusiones sirven para confirmar la aseveración anterior de que la función log $y = \ln t$ se incrementa monotónicamente a una tasa creciente.

Aplicación

Una de las virtudes principales del logaritmo es su capacidad para convertir una multiplicación en una suma, y una división en una resta. Esta propiedad se aprovecha cuando se está diferenciando un *producto* o *cociente* complicado de cualquier tipo de funciones (no necesariamente exponencial o logarítmica).

Ejemplo 9 Encuentre dy/dx a partir de

$$y = \frac{x^2}{(x+3)(2x+1)}$$

En lugar de aplicar las reglas del producto y del cociente, se podría tomar primero el logaritmo natural de ambos lados de la ecuación para reducir la función a la forma

$$\ln y = \ln x^2 - \ln(x+3) - \ln(2x+1)$$

De acuerdo con (10.20), la derivada del lado izquierdo respecto a x es

$$\frac{d}{dx}\begin{pmatrix}\text{lado}\\ \text{izquierdo}\end{pmatrix} \quad \frac{1}{y}\frac{dy}{dx}$$

mientras que el lado derecho da

$$\frac{d}{dx}\begin{pmatrix}\text{lado}\\ \text{derecho}\end{pmatrix} = \frac{2x}{x^2} - \frac{1}{x+3} - \frac{2}{2x+1} = \frac{7x+6}{x(x+3)(2x+1)}$$

Cuando se igualan los dos resultados y ambos lados se multiplican por y, se obtiene la derivada deseada, como sigue:

$$\begin{aligned}\frac{dy}{dx} &= \frac{7x+6}{x(x+3)(2x+1)}y \\ &= \frac{7x+6}{x(x+3)(2x+1)}\frac{x^2}{(x+3)(2x+1)} = \frac{x(7x+6)}{(x+3)^2(2x+1)^2}\end{aligned}$$

Ejemplo 10 Encuentre dy/dx de $y = x^a e^{kx-c}$. Tomando el log natural de ambos lados, se tiene

$$\ln y = a \ln x + \ln e^{kx-c} = a \ln x + kx - c$$

Al diferenciar ambos lados respecto a x y usar (10.20) obtenemos

$$\frac{1}{y}\frac{dy}{dx} = \frac{a}{x} + k$$

y

$$\frac{dy}{dx} = \left(\frac{a}{x} + k\right) y = \left(\frac{a}{x} + k\right) x^a e^{kx-c}$$

Sin embargo, note que si la función dada contiene términos aditivos, entonces podría *no* ser deseable convertir la función en la forma log.

EJERCICIO 10.5

1. Encuentre las derivadas de:
   (*a*) $y = e^{2t+4}$
   (*b*) $y = e^{1-9t}$
   (*c*) $y = e^{t^2+1}$
   (*d*) $y = 5e^{2-t^2}$
   (*e*) $y = e^{ax^2+bx+c}$
   (*f*) $y = xe^x$
   (*g*) $y = x^2 e^{2x}$
   (*h*) $y = axe^{bx+c}$
2. (*a*) Compruebe la derivada del ejemplo 3 utilizando la ecuación $\ln(at) = \ln a + \ln t$.
   (*b*) Compruebe el resultado del ejemplo 4 por medio de la ecuación $\ln t^c = c \ln t$.
3. Encuentre las derivadas de:
   (*a*) $y = \ln(7t^5)$
   (*b*) $y = \ln(at^c)$
   (*c*) $y = \ln(t+19)$
   (*d*) $y = 5\ln(t+1)^2$
   (*e*) $y = \ln x - \ln(1+x)$
   (*f*) $y = \ln[x(1-x)^8]$
   (*g*) $y = \ln\left(\frac{2x}{1+x}\right)$
   (*h*) $y = 5x^4 \ln x^2$
4. Encuentre las derivadas de:
   (*a*) $y = 5^t$
   (*b*) $y = \log_2(t+1)$
   (*c*) $y = 13^{2t+3}$
   (*d*) $y = \log_7 7x^2$
   (*e*) $y = \log_2(8x^2+3)$
   (*f*) $y = x^2 \log_3 x$
5. Compruebe las dos fórmulas de (10.21′).
6. Demuestre que la función $V = Ae^{rt}$ (con $A, r > 0$) y la función $A = Ve^{-rt}$ (con $V, r > 0$) son estrictamente monótonas, pero en direcciones opuestas, y que son estrictamente convexas en forma (véase ejercicio 10.2-5).
7. Encuentre las derivadas de las siguientes funciones tomando primero el logaritmo natural de ambos lados:
   (*a*) $y = \dfrac{3x}{(x+2)(x+4)}$
   (*b*) $y = (x^2+3)e^{x^2+1}$

10.6 Fecha óptima

Lo que hemos aprendido acerca de las funciones exponenciales y logarítmicas se puede aplicar ahora a algunos problemas sencillos de fecha óptima.

Problema de almacenaje de vino

Suponga que un comerciante de vino posee cierta cantidad de esta bebida (por ejemplo, una caja), la cual puede vender al tiempo presente ($t = 0$) por una suma de $\$K$ o almacenar durante algún tiempo y luego vender a un valor más alto. Se sabe que el valor creciente (V) del vino es la siguiente función de tiempo:

$$V = Ke^{\sqrt{t}} \quad [= K \exp(t^{1/2})] \tag{10.22}$$

de modo que si $t = 0$ (vender ahora), entonces $V = K$. El problema es averiguar cuándo debe vender a fin de maximizar la ganancia, suponiendo que el costo de almacenaje es cero.[5]

Como el costo del vino es un gasto realizado, el comerciante ya lo pagó, y como se supone que no existe costo de almacenaje, maximizar la ganancia es lo mismo que maximizar el ingreso de ventas, o el valor V. Sin embargo, hay un problema. Cada valor de V correspondiente a un punto específico de t representa una suma de un peso a cobrar en una fecha distinta y, como resultado del elemento de interés en cuestión, no es directamente comparable con el valor V de otra fecha. La forma de salir de esta dificultad es *descontar* cada cifra V a su equivalente *valor presente* (el valor en el instante $t = 0$), porque entonces todos los valores V estarán en una base comparable.

Suponga que la tasa de interés sobre la base de capitalización continua es al nivel de r. Entonces, de acuerdo con (10.11), el valor presente de V se puede expresar como

$$A(t) = Ve^{-rt} = Ke^{\sqrt{t}}e^{-rt} = Ke^{\sqrt{t}-rt} \tag{10.22'}$$

donde A, que denota el valor presente de V, es por sí misma una función de t. Por lo tanto, el problema equivale a hallar el valor de t que maximiza a A.

Condiciones de maximización

La condición de primer orden para maximizar A es tener $dA/dt = 0$. Para hallar esta derivada, podemos diferenciar directamente (10.22′) respecto a t, o hacerlo de modo indirecto tomando primero el log natural de ambos lados de (10.22′) y diferenciando después respecto a t. A continuación de ilustra este último procedimiento.

Primero obtenemos de (10.22′) la ecuación

$$\ln A(t) = \ln K + \ln e^{\sqrt{t}-rt} = \ln K + (t^{1/2} - rt)$$

Al diferenciar ambos lados respecto a t, obtenemos

$$\frac{1}{A}\frac{dA}{dt} = \frac{1}{2}t^{-1/2} - r$$

o bien,

$$\frac{dA}{dt} = A\left(\frac{1}{2}t^{-1/2} - r\right)$$

Puesto que $A \neq 0$, la condición $dA/dt = 0$ se satisface si y sólo si

$$\frac{1}{2}t^{-1/2} = r \qquad \text{o} \qquad \frac{1}{2\sqrt{t}} = r \qquad \text{o} \qquad \frac{1}{2r} = \sqrt{t}$$

Esto significa que la duración óptima de tiempo de almacenaje es

$$t^* = \left(\frac{1}{2r}\right)^2 = \frac{1}{4r^2}$$

[5] La consideración de costo de almacenamiento conlleva una dificultad la cual aún no podemos resolver. Después, en el capítulo 14, se volverá a este problema.

Si $r = 0.10$, por ejemplo, entonces $t^* = 25$, y el comerciante debe almacenar la caja de vino durante 25 años. Note que mientras más alta es la tasa de interés (tasa de descuento), más corto será el periodo de almacenaje óptimo.

La condición de primer orden, $1/(2\sqrt{t}) = r$, admite una interpretación económica fácil. La expresión del lado izquierdo solamente representa la tasa de crecimiento del valor del vino V, porque de (10.22)

$$\begin{aligned}\frac{dV}{dt} &= \frac{d}{dt}K\exp(t^{1/2}) = K\frac{d}{dt}\exp(t^{1/2}) && [K \text{ constante}]\\ &= K\left(\frac{1}{2}t^{-1/2}\right)\exp(t^{1/2}) && [\text{por } (10.20)]\\ &= \left(\frac{1}{2}t^{-1/2}\right)V && [\text{por } (10.22)]\end{aligned}$$

de modo que la tasa de crecimiento de V es de hecho la expresión del lado izquierdo de la condición de primer orden:

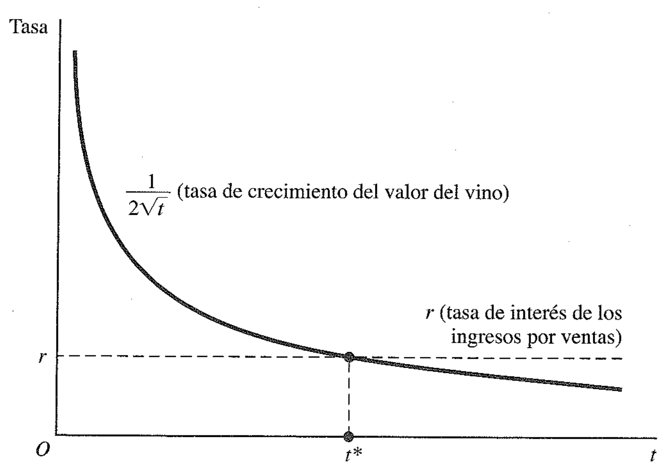
$$r_V \equiv \frac{dV/dt}{V} = \frac{1}{2}t^{-1/2} = \frac{1}{2\sqrt{t}}$$

Por el contrario, la expresión r del lado derecho es la tasa de interés o la tasa de crecimiento de interés compuesto de fondo de caja por cobrar *si* el vino se vende de inmediato, un aspecto de *costo de oportunidad* de almacenar el vino. Así, la igualación de las dos tasas instantáneas, como se ilustra en la figura 10.4, es un intento por tratar de conservar el vino hasta que se anule por completo la ventaja del almacenaje, es decir, esperar el momento cuando la tasa de crecimiento (decreciente) del valor del vino coincide con la tasa de interés (constante) de los ingresos por ventas al contado.

Lo siguiente es comprobar si el valor de t^* satisface la condición de segundo orden para la maximización de A. La segunda derivada de A es

$$\frac{d^2A}{dt^2} = \frac{d}{dt}A\left(\frac{1}{2}t^{-1/2} - r\right) = A\frac{d}{dt}\left(\frac{1}{2}t^{-1/2} - r\right) + \left(\frac{1}{2}t^{-1/2} - r\right)\frac{dA}{dt}$$

FIGURA 10.4



Pero, puesto que el término final sale cuando se evalúa en el punto de equilibrio (óptimo), donde $dA/dt = 0$, el resultado es

$$\frac{d^2A}{dt^2} = A\frac{d}{dt}\left(\frac{1}{2}t^{-1/2} - r\right) = A\left(-\frac{1}{4}t^{-3/2}\right) = \frac{-A}{4\sqrt{t^3}}$$

En vista de que $A > 0$, esta segunda derivada es negativa cuando se evalúa en $t^* > 0$; así, se tiene la seguridad de que el valor solución t^* en realidad maximiza la ganancia.

Problema del corte de madera

Un problema similar, que tiene que ver con una elección del mejor tiempo para actuar, es el del corte de madera.

Suponga que el valor de los árboles maderables (ya plantados en algún terreno) es la siguiente función de tiempo creciente:

$$V = 2^{\sqrt{t}}$$

expresada en unidades de \$1 000. Suponiendo una tasa de descuento de r (sobre la base continua) y suponiendo también costo cero de conservación durante el periodo de crecimiento de los árboles, ¿cuál es la fecha para cortar la madera para venta?

Como en el problema del vino, primero se debe convertir V a su valor presente:

$$A(t) = Ve^{-rt} = 2^{\sqrt{t}}e^{-rt}$$

así,

$$\ln A = \ln 2^{\sqrt{t}} + \ln e^{-rt} = \sqrt{t}\ln 2 - rt = t^{1/2}\ln 2 - rt$$

Para maximizar A establecemos $dA/dt = 0$. Obtenemos la primera derivada al diferenciar ln A respecto a t y multiplicar después por A:

$$\frac{1}{A}\frac{dA}{dt} = \frac{1}{2}t^{-1/2}\ln 2 - r$$

así,

$$\frac{dA}{dt} = A\left(\frac{\ln 2}{2\sqrt{t}} - r\right)$$

Puesto que $A \neq 0$, la condición $dA/dt = 0$ se satisface si y sólo si

$$\frac{\ln 2}{2\sqrt{t}} = r \quad \text{o bien,} \quad \sqrt{t} = \frac{\ln 2}{2r}$$

En consecuencia, el número óptimo de años de crecimiento es

$$t^* = \left(\frac{\ln 2}{2r}\right)^2$$

Partiendo de esta solución, es evidente que mientras más alta sea la tasa de descuento, más pronto se debe cortar el árbol.

Para asegurarse de que t^* es una solución de maximización (en vez de minimización), se debe comprobar la condición de segundo orden. Pero esto se deja como ejercicio.

En este ejemplo hemos separado el costo de plantación al suponer que los árboles ya estaban plantados, en cuyo caso es legítimo no considerar el costo de plantación (suprimido) en la decisión de optimización. Si la decisión no es la de cuándo cosechar sino la de plantar o no, entonces el costo de plantación (en el que se incurre en el *presente*) debe ser comparado como es debido con el valor *presente* de la producción de madera, calculado con t fijo en el valor óptimo t^*. Por ejemplo, si $r = 0.05$, tenemos

$$t^* = \left(\frac{0.6931}{0.10}\right)^2 = (6.931)^2 = 48.0 \text{ años}$$

y

$$A^* = 2^{6.931}e^{-0.05(48.0)} = (122.0222)e^{-2.40}$$
$$= 122.0222(0.0907) = \$11.0674 \text{ (en miles)}$$

Así que sólo un costo de plantación menor que A^* hará rentable el proyecto, de nuevo, siempre y cuando sea nulo el costo de conservación.

EJERCICIO 10.6

1. Si el valor del vino crece de acuerdo con la función $V = Ke^{2\sqrt{t}}$, en vez de como en (10.22), ¿cuánto tiempo debe almacenar el vino el comerciante?
2. Compruebe la condición de segundo orden para el problema del corte de madera.
3. Como una generalización del problema de optimización ilustrado en la presente sección, demuestre que:
   (*a*) Con cualquier función de valor $V = f(t)$ y dada una tasa continua de descuento r, la condición de primer orden para que el valor presente A de V alcance un máximo es que la tasa de crecimiento V sea igual a r.
   (*b*) La condición suficiente de segundo orden para un máximo equivale en realidad a la estipulación de que la tasa de crecimiento de V sea estrictamente decreciente con el tiempo.
4. Analice la estática comparativa del problema de almacenaje de vino.

10.7 Más aplicaciones de derivadas exponenciales y logarítmicas

Aparte de su uso en problemas de optimización, las fórmulas de derivadas de la sección 10.5 tienen más aplicaciones económicas útiles.

Determinación de la tasa de crecimiento

Cuando una variable y es una función del tiempo, $y = f(t)$, su tasa instantánea de crecimiento se define como[6]

$$r_y \equiv \frac{dy/dt}{y} = \frac{f'(t)}{f(t)} = \frac{\text{función marginal}}{\text{función total}} \qquad \textbf{(10.23)}$$

Pero de (10.20) vemos que esta relación es precisamente la derivada de $\ln f(t) = \ln y$. Así, para hallar la tasa instantánea de crecimiento de una función de tiempo $f(t)$, en lugar de diferenciar respecto a t, y luego dividir entre $f(t)$, simplemente tome su log natural y luego derive ln

[6] Si la variable t *no* denota tiempo, la expresión $(dy/dt)/y$ se denomina *tasa proporcional de cambio* de y respecto a t.

$f(t)$ respecto al tiempo.[7] Este método alternativo podría ser un enfoque más simple, si $f(t)$ fuera una expresión multiplicativa o divisional que, al tomar el logaritmo, se redujera a una suma o diferencia de términos aditivos.

Ejemplo 1 Halle la tasa de crecimiento de $V = Ae^{rt}$, donde t denota tiempo. Ya se sabe que la tasa de crecimiento de V es r, pero se comprobará al hallar la derivada de ln V:

$$\ln V = \ln A + rt \ln e = \ln A + rt \qquad [\text{una constante}]$$

Por lo tanto,

$$r_V = \frac{d}{dt}\ln V = 0 + \frac{d}{dt}rt = r$$

como se quería demostrar.

Ejemplo 2 Encuentre la tasa de crecimiento de $y = 4^t$. En este caso, se tiene

$$\ln y = \ln 4^t = t \ln 4$$

Por consiguiente

$$r_y = \frac{d}{dt}\ln y = \ln 4$$

Así es como debe ser, porque $e^{\ln 4} \equiv 4$ y, en consecuencia, $y = 4^t$ se puede reescribir como $y = e^{(\ln 4)t}$, que de inmediato permitiría leer (ln 4) como la tasa de crecimiento de y.

Tasa de crecimiento de una combinación de funciones

Para llevar esta explicación más lejos, examine la tasa de crecimiento instantánea de un *producto* de dos funciones de tiempo:

$$y = uv \quad \text{donde} \begin{cases} u = f(t) \\ v = g(t) \end{cases}$$

Tomando el log natural de y, se obtiene

$$\ln y = \ln u + \ln v$$

Así, la tasa de crecimiento deseada es

$$r_y = \frac{d}{dt}\ln y = \frac{d}{dt}\ln u + \frac{d}{dt}\ln v$$

Pero los dos términos del lado derecho son las tasas de crecimiento de u y v, respectivamente. Por lo tanto, se tiene la regla

$$r_{(uv)} = r_u + r_v \tag{10.24}$$

Expresada en palabras, la tasa instantánea de crecimiento de un *producto* es la *suma* de las tasas instantáneas de crecimiento de los componentes.

Mediante un procedimiento similar, podemos demostrar que la tasa de crecimiento de un *cociente* es la *diferencia* entre las tasas de crecimiento de los componentes (véase el ejercicio 10.7-4):

$$r_{(u/v)} = r_u - r_v \tag{10.25}$$

[7] Si se traza el log natural de una función $f(t)$ contra t en un diagrama bidimensional, la pendiente de la curva, en consecuencia, indica la tasa de crecimiento de $f(t)$. Esto prueba el fundamento de las denominadas gráficas de escala semilog, que se emplean para comparar las tasas de crecimiento de variables diferentes, o las tasas de crecimiento de la misma variable en distintos países.

Ejemplo 3 Si el consumo C crece a la tasa α, y si la población H (en número de individuos) crece a la tasa β, ¿cuál es la tasa de crecimiento del consumo *per cápita*? Puesto que el consumo *per cápita* es igual a C/H, su tasa de crecimiento debe ser

$$r_{(C/H)} = r_C - r_H = \alpha - \beta$$

Considere ahora la tasa instantánea de crecimiento de una *suma* de dos funciones de tiempo:

$$z = u + v \qquad \text{donde} \begin{cases} u = f(t) \\ v = g(t) \end{cases}$$

Esta vez, el log natural será

$$\ln z = \ln(u+v) \qquad [\neq \ln u + \ln v]$$

Así,

$$\begin{aligned} r_z = \frac{d}{dt}\ln z &= \frac{d}{dt}\ln(u+v) \\ &= \frac{1}{u+v}\frac{d}{dt}(u+v) \qquad [\text{por } (10.20)] \\ &= \frac{1}{u+v}[f'(t) + g'(t)] \end{aligned}$$

Pero de (10.23) se tiene $r_u = f'(t)/f(t)$, de modo que $f'(t) = f(t)r_u = ur_u$. De manera similar, se tiene $g'(t) = vr_v$. Como resultado, podemos escribir la regla

$$r_{(u+v)} = \frac{u}{u+v}r_u + \frac{v}{u+v}r_v \tag{10.26}$$

que demuestra que la tasa de crecimiento de una *suma* es un *promedio ponderado* de las tasas de crecimiento de los componentes.

Del mismo modo, se tiene (véase el ejercicio 10.7-5)

$$r_{(u-v)} = \frac{u}{u-v}r_u - \frac{v}{u-v}r_v \tag{10.27}$$

Ejemplo 4 Las exportaciones de bienes de un país, $G = G(t)$, tienen la tasa de crecimiento de a/t, y sus exportaciones de servicios, $S = S(t)$, tienen la tasa de crecimiento de b/t. ¿Cuál es la tasa de crecimiento de sus exportaciones totales? Puesto que las exportaciones totales son $X(t) = G(t) + S(t)$, una suma, su tasa de crecimiento debe ser

$$\begin{aligned} r_X &= \frac{G}{X}r_G + \frac{S}{X}r_S \\ &= \frac{G}{X}\left(\frac{a}{t}\right) + \frac{S}{X}\left(\frac{b}{t}\right) = \frac{Ga + Sb}{Xt} \end{aligned}$$

Determinación de la elasticidad puntual

Se ha visto que, dada $y = f(t)$, la derivada de $\ln y$ mide la tasa instantánea de crecimiento de y. Ahora se verá qué sucede cuando, dada una función $y = f(x)$, se deriva $\ln y$ respecto a $\ln x$, y no respecto a x.

Para empezar, definimos $u \equiv \ln y$ y $v \equiv \ln x$. Entonces se puede observar una cadena de relación que une a u con y, y por lo tanto a x y v como sigue:

$$u \equiv \ln y \qquad y = f(x) \qquad x \equiv e^{\ln x} \equiv e^{v}$$

En consecuencia, la derivada de $\ln y$ respecto a $\ln x$ es

$$\begin{aligned}\frac{d(\ln y)}{d(\ln x)} = \frac{du}{dv} &= \frac{du}{dy}\frac{dy}{dx}\frac{dx}{dv} \\ &= \left(\frac{d}{dy}\ln y\right)\left(\frac{dy}{dx}\right)\left(\frac{d}{dv}e^{v}\right) = \frac{1}{y}\frac{dy}{dx}e^{v} = \frac{1}{y}\frac{dy}{dx}x = \frac{dy}{dx}\frac{x}{y}\end{aligned}$$

Pero esta expresión es precisamente la de la elasticidad puntual de la función. Por lo tanto, se ha establecido el principio general que, para una función $y = f(x)$, la elasticidad puntual de y respecto a x es

$$\varepsilon_{yx} = \frac{d(\ln y)}{d(\ln x)} \tag{10.28}$$

Se debe observar que el subíndice yx en este símbolo es un indicador de que y y x son las dos variables en cuestión y no implica la multiplicación de y y x. Esto es diferente del caso de $r_{(uv)}$, donde el subíndice *denota* un producto. De nuevo, ahora se tiene otra forma de hallar la elasticidad puntual de una función mediante el uso de logaritmos, lo cual prueba a menudo ser un método más fácil, *si* la función dada viene en la forma de una expresión multiplicativa o divisional.

Ejemplo 5

Encuentre la elasticidad puntual de demanda, dado que $Q = k/P$, donde k es una constante positiva. Ésta es la ecuación de una hipérbola rectangular (véase la figura 2.8*d*); y, como es bien sabido, una función de demanda de esta forma tiene una elasticidad puntual unitaria en todos los puntos. Para demostrar esto, se aplicará (10.28). Puesto que el log natural de la función de demanda es

$$\ln Q = \ln k - \ln P$$

es necesaria la elasticidad de demanda (Q respecto a P)

$$\varepsilon_d = \frac{d(\ln Q)}{d(\ln P)} = -1 \quad \text{o bien} \quad |\varepsilon_d| = 1$$

El resultado de (10.28) se obtuvo por medio de la regla de la cadena de derivadas. Es interesante observar que una regla de la cadena similar se cumpla para elasticidades; es decir, dada una función $y = g(w)$, donde $w = h(x)$, se tiene

$$\varepsilon_{yx} = \varepsilon_{yw}\varepsilon_{wx} \tag{10.29}$$

La demostración es:

$$\varepsilon_{yw}\varepsilon_{wx} = \left(\frac{dy}{dw}\frac{w}{y}\right)\left(\frac{dw}{dx}\frac{x}{w}\right) = \frac{dy}{dw}\frac{dw}{dx}\frac{w}{y}\frac{x}{w} = \frac{dy}{dx}\frac{x}{y} = \varepsilon_{yx}$$

EJERCICIO 10.7

1. Encuentre la tasa instantánea de crecimiento:
   (*a*) $y = 5t^2$ (*c*) $y = ab^t$ (*e*) $y = t/3^t$
   (*b*) $y = at^c$ (*d*) $y = 2^t(t^2)$
2. Si la población crece de acuerdo con la función $H = H_0(2)^{bt}$ y el consumo mediante la función $C = C_0e^{at}$, encuentre las tasas de crecimiento de población, consumo y consumo per cápita usando el log natural.
3. Si y se relaciona con x mediante $y = x^k$, ¿cómo se relacionarán las tasas de crecimiento r_y y r_x?
4. Compruebe que si $y = u/v$, donde $u = f(t)$ y $v = g(t)$, entonces la tasa de crecimiento de y será $r_y = r_u - r_v$, como se muestra en (10.25).
5. El ingreso real y se define como el ingreso nominal Y deflactado[8] por el nivel de precio P. ¿Cómo se relaciona r_y (para el ingreso real) con r_Y (para el ingreso nominal)?
6. Compruebe la regla de la tasa de crecimiento (10.27).
7. Dada la función de demanda $Q_d = k/P^n$, donde k y n son constantes positivas, encuentre la elasticidad puntual de demanda ε_d por medio de (10.28) (cf. ejercicio 8.1-4).
8. (*a*) Dada $y = wz$, donde $w = g(x)$ y $z = h(x)$, establezca que $\varepsilon_{yx} = \varepsilon_{wx} + \varepsilon_{zx}$.
   (*b*) Dada $y = u/v$, donde $u = G(x)$ y $v = H(x)$, establezca que $\varepsilon_{yx} = \varepsilon_{ux} - \varepsilon_{vx}$.
9. Dada $y = f(x)$, demuestre que la derivada $d(\log_b y)/d(\log_b x)$, log base b en vez de e, mide también la elasticidad puntual ε_{yx}.
10. Demuestre que, si la demanda de dinero M_d es una función del ingreso nacional $Y = Y(t)$ y la tasa de interés $i = i(t)$, la tasa de crecimiento de M_d se puede expresar como una suma ponderada de r_Y y r_i,

    $$r_{M_d} = \varepsilon_{M_dY} r_Y + \varepsilon_{M_di} r_i$$

    donde las ponderaciones son las elasticidades de M_d respecto a Y e i, respectivamente.
11. Dada la función de producción $Q = F(K, L)$, encuentre una expresión general para la tasa de crecimiento de Q en términos de las tasas de crecimiento de K y L.

[8] [Nota del revisor] Se usa deflactar o deflatar como traducción de "deflate", que significa quitar el efecto de la inflación de precios: "desinflar". El ingreso nominal y se deflacta dividiéndolo entre el índice de precios P.

Capítulo 11

El caso de más de una variable de elección

En el capítulo 9 analizamos el problema de la optimización dentro del marco de una función objetivo con una sola variable de elección. En el capítulo 10 el análisis se extendió a las funciones objetivo exponenciales, pero todavía se trató con una sola variable de elección. Ahora desarrollaremos una manera de hallar los valores extremos de una función objetivo que tiene que ver con más de dos variables. Sólo entonces podremos tratar el tipo de problema que confronta, por ejemplo, una empresa multiproducto, donde la decisión de maximizar la ganancia está basada en la elección de niveles de producción óptimos para varios artículos y la combinación óptima de varios insumos distintos.

Analizaremos primero el caso de una función objetivo de dos variables de elección $z = f(x, y)$, a fin de aprovechar su capacidad de graficación. Después, los resultados analíticos pueden generalizarse al caso de n variables no graficables; sin embargo, respecto al número de variables, se supondrá en general que, cuando se escribe en forma general, la función objetivo posee derivadas parciales continuas hasta cualquier orden deseado. Esto asegura la suavidad y diferenciabilidad de la función objetivo, así como de sus derivadas parciales.

Para funciones de algunas variables, los valores extremos son de dos clases: (1) absolutos o globales y (2) relativos o locales. Como anteriormente, centramos la atención sobre todo en los extremos relativos, y por esta razón con frecuencia abandonamos el adjetivo "relativo", a sabiendas de que, a menos que se especifique lo contrario, se hace referencia a los extremos *relativos*. No obstante, en la sección 11.5 daremos la debida consideración a las condiciones para extremos *absolutos*.

11.1 Versión diferencial de condiciones de optimización

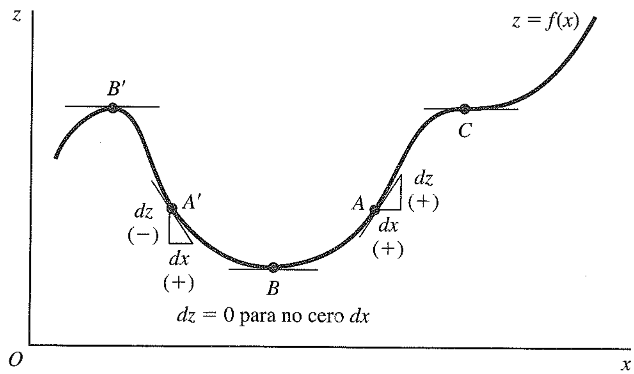
La explicación del capítulo 9 sobre condiciones de optimización para problemas con una sola variable de elección se basó por completo en términos de *derivadas*, en oposición a las diferenciales. A fin de prepararse para el estudio de problemas con dos o más variables de elección, será útil también conocer cómo podemos expresar de modo equivalente esas condiciones en términos de *diferenciales*.

Condición de primer orden

Dada una función $z = f(x)$, se puede, como se explicó en la sección 8.1, escribir la diferencial

$$dz = f'(x)\,dx \tag{11.1}$$

FIGURA 11.1



y usar dz como una aproximación al cambio real, Δz, inducido por el cambio de x de x_0 a $x_0 + \Delta x$; mientras más pequeña sea Δx, mejor es la aproximación. De (11.1) es claro que si $f'(x) > 0$, dz y dx deben tomar el mismo signo algebraico; esto se ilustra mediante el punto A en la figura 11.1 (ver fig. 8.1*b*). En el caso opuesto, donde $f'(x) < 0$, ejemplificado por el punto A', dz y dx toman signos algebraicos opuestos. Puesto que los puntos como A y A', donde $f'(x) \neq 0$ y por consiguiente $dz \neq 0$, no se pueden calificar como puntos estacionarios, tiene sentido que una condición necesaria para que z alcance un extremo (un valor estacionario) sea $dz = 0$. Dicho con más precisión, la condición se debe expresar como "$dz = 0$ para una dx arbitraria no cero", puesto que una dx cero (ningún cambio en x) carece de importancia en el contexto presente. En el punto B de la figura 11.1, ocurre un mínimo de z, y en el punto B' se presenta un máximo de z. En ambos casos, con la recta tangente horizontal, es decir, con$f'(x) = 0$, dz (el lado vertical del triángulo formado con la recta tangente como hipotenusa) de hecho se reduce a cero. Así, la condición de derivada de primer orden "$f'(x) = 0$" se puede traducir en la condición diferencial de primer orden "$dz = 0$ para una dx arbitraria distinta de cero". Sin embargo, recuerde que si bien esta condición diferencial es necesaria para un extremo, de ningún modo es suficiente, porque un punto de inflexión como C en la figura 11.1 también puede satisfacer la condición de que $dz = 0$ para una dx arbitraria y diferente de cero.

Condición de segundo orden

Las condiciones suficientes de segundo orden para extremos de z son, en términos de derivadas, $f''(x) < 0$ (para un máximo) y $f''(x) > 0$ (para un mínimo) en el punto estacionario. Para traducir estas condiciones en diferenciales equivalentes se necesita la noción de *diferencial de segundo orden*, definida como la diferencial de una diferencial, es decir, $d(dz)$, que por lo común se denota mediante d^2z (léase: "d dos z").

Dado que $dz = f'(x)\,dx$, se puede obtener d^2z sólo mediante diferenciación adicional de dz. Sin embargo, debe tener presente que dx, en este contexto representa un cambio arbitrario o un determinado cambio diferente de cero en x, se tratará como una constante durante la diferenciación. En consecuencia, dz puede variar sólo con $f'(x)$, pero puesto que $f'(x)$ es a su vez una función de x, en el análisis final dz varía sólo con x. En vista de esto, tenemos

$$\begin{aligned} d^2z \equiv d(dz) &= d[f'(x)\,dx] && \text{[por (11.1)]} \\ &= [df'(x)]\,dx && [dx \text{ es constante}] \\ &= [f''(x)\,dx]\,dx = f''(x)\,dx^2 && \textbf{(11.2)} \end{aligned}$$

Tome en cuenta que el exponente 2 aparece en (11.2) en dos formas diferentes. En el símbolo d^2z, el exponente 2 (léase: "dos") indica la diferencial de *segundo orden* de z; pero en el símbolo $dx^2 \equiv (dx)^2$, el exponente 2 (léase: "cuadrada") denota el *cuadrado* de la diferencial de primer orden dx. El resultado de (11.2) proporciona un vínculo directo entre d^2z y $f''(x)$. En vista de que sólo estamos considerando valores diferentes de cero para dx, el término dx^2 es siempre positivo; así, d^2z y $f''(x)$ deben tener el mismo signo algebraico. Del mismo modo que una $f''(x)$ positiva (negativa) en un punto estacionario delinea un valle (cúspide), también debe ser así con una d^2z positiva (negativa) en ese punto.

Se deduce que la condición de *derivada* "$f''(x) < 0$ es suficiente para que un máximo de z" se pueda expresar de modo equivalente como la condición *diferencial* "$d^2z < 0$ para que una dx diferente de cero arbitraria sea suficiente para un máximo de z". La traslación de la condición para un mínimo de z es análoga; sólo se requiere invertir el sentido de la desigualdad en el enunciado anterior. Al ir más allá, podríamos concluir con base en (11.2) que las condiciones *necesarias* de segundo orden son

$$\begin{aligned} &\text{Para el máximo de } z\text{:} \quad f''(x) \leq 0 \\ &\text{Para el mínimo de } z\text{:} \quad f''(x) \geq 0 \end{aligned}$$

que se pueden traducir, respectivamente, en

$$\left.\begin{aligned} &\text{Para el máximo de } z\text{:} \quad d^2z \leq 0 \\ &\text{Para el mínimo de } z\text{:} \quad d^2z \geq 0 \end{aligned}\right\} \text{ para valores arbitrarios de } dx \text{ distintos de cero}$$

Condiciones diferenciales contra condiciones de derivadas

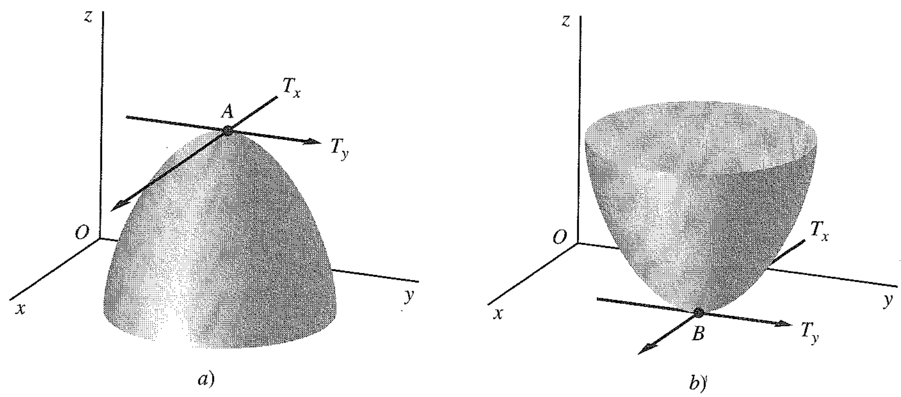
Ahora que hemos demostrado la posibilidad de expresar la versión de derivada de las condiciones de primer y segundo orden en términos de dz y d^2z, tal vez le gustaría preguntar el por qué del interés en desarrollar un nuevo conjunto de condiciones diferenciales cuando ya se disponía de las condiciones de derivada. La respuesta es que las condiciones diferenciales, pero no las de derivada, se expresan en formas que pueden generalizarse de modo directo del caso de una variable a casos con dos o más variables de elección. Para ser más específicos, la condición de primer orden (valor cero para dz) y la condición de segundo orden (negatividad o positividad para d^2z) son aplicables con igual validez a todos los casos, siempre que la frase "para valores arbitrarios de dx distintos de cero" se modifique como es debido para reflejar el cambio en el número de variables de elección.

Esto no significa que las condiciones de derivada no tendrán nada más que ver. Por el contrario, puesto que las condiciones de derivada son, desde el punto de vista operacional, más convenientes en cuanto a la aplicación —después de llevar a cabo el proceso de generalización por medio de las condiciones diferenciales a casos con más de dos variables de elección— intentaremos desarrollar y hacer uso de condiciones de derivada apropiadas a esos casos.

11.2 Valores extremos de una función de dos variables

Para una función de una variable de elección, el valor extremo se representa de modo gráfico mediante la cúspide de una colina o el fondo de un valle en una gráfica bidimensional. Con *dos* variables de elección, la gráfica de la función, $z = f(x, y)$, se convierte en una superficie en un espacio de tres dimensiones, y aunque los valores extremos aún se tienen que relacionar con cúspides y fondos, estas "colinas" y "valles" por sí mismos toman entonces un carácter

FIGURA 11.2



tridimensional; en el nuevo contexto se conformarán como domos y tazones, respectivamente. Los dos diagramas de la figura 11.2 sirven como ilustraciones. El punto A del diagrama a, la cúspide de un domo, constituye un máximo; el valor de z en este punto es más grande que cualquier otro punto en su vecindad inmediata. De manera similar, el punto B del diagrama b, el fondo de un tazón, representa un mínimo; en cualquier parte de su vecindad inmediata el valor de la función excede al del punto B.

Condición de primer orden

Para la función

$$z = f(x, y)$$

la condición necesaria de primer orden para un extremo (ya sea máximo o mínimo) de nuevo tiene que ver con $dz = 0$. Pero como aquí hay dos variables independientes, dz es ahora una diferencial *total*; así, la condición de primer orden debe modificarse a la forma

$$dz = 0 \text{ para valores arbitrarios de } dx \text{ y } dy\text{, al menos uno diferente de cero} \qquad \textbf{(11.3)}$$

El fundamento de (11.3) es similar a la explicación de la condición $dz = 0$ para el caso de una variable: un punto extremo debe ser un punto estacionario, y en un punto estacionario, dz como una aproximación al cambio real Δz debe ser cero para dx y dy arbitrarias, al menos uno diferente de cero.

En este caso de dos variables, la diferencial total es

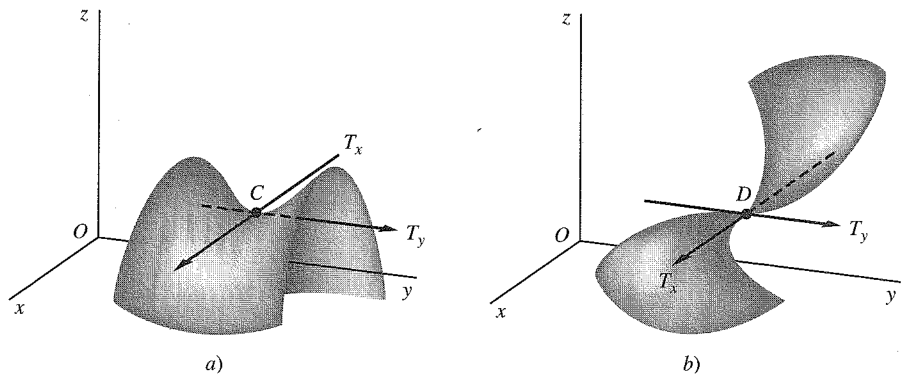
$$dz = f_x\, dx + f_y\, dy \qquad \textbf{(11.4)}$$

A fin de satisfacer la condición (11.3), es necesario y suficiente que las dos derivadas parciales f_x y f_y sean al mismo tiempo iguales a cero. Así, la versión de derivada equivalente de la condición de primer orden (11.3) es

$$f_x = f_y = 0 \qquad \left[\begin{matrix}\text{o} \\ \text{bien,}\end{matrix}\ \frac{\partial z}{\partial x} = \frac{\partial z}{\partial y} = 0\right] \qquad \textbf{(11.5)}$$

Hay una interpretación gráfica sencilla de esta condición. Respecto al punto A de la figura 11.2a, $f_x = 0$ en ese punto significa que la recta tangente T_x, trazada por A y paralela al plano

FIGURA 11.3



xz (manteniendo a y constante), debe tener una pendiente cero. De la misma manera, $f_y = 0$ en el punto A significa que la recta tangente T_y, trazada por A y paralela al plano yz (manteniendo a x constante), también debe tener una pendiente cero. Puede comprobar fácilmente que estos requerimientos de recta tangente en realidad se aplican también al punto mínimo B en la figura 11.2*b*. Esto se debe a que tanto la condición (11.5) como la (11.3) son condiciones necesarias tanto para un máximo como para un mínimo.

Como en la explicación anterior, la condición de primer orden es *necesaria*, pero no *suficiente*. En los dos diagramas de la figura 11.3 se puede ver que no es suficiente para establecer un extremo. En el punto C del diagrama a, tanto T_x como T_y tienen pendientes cero, pero este punto no califica como un extremo: mientras que es un *mínimo* cuando se considera el plano yz, ¡resulta ser un *máximo* cuando se considera contra el plano xz! Un punto con tal "personalidad dual" se conoce, por razones gráficas, como un *punto silla*. De manera similar, el punto D de la figura 11.3*b*, si bien se caracteriza porque T_x y T_y no tienen inclinación respecto a z, no es tampoco un extremo; su localización en la superficie torcida hace que sea un *punto de inflexión*, ya sea que se vea contra el plano xz o el yz. Estos contraejemplos descartan de manera definitiva la condición de primer orden como una condición suficiente para un extremo.

Para desarrollar una condición suficiente, debemos examinar la diferencial total de segundo orden, que se relaciona con derivadas parciales de segundo orden.

Derivadas parciales de segundo orden

La función $z = f(x, y)$ puede dar lugar a *dos* derivadas parciales de primer orden,

$$f_x \equiv \frac{\partial z}{\partial x} \qquad \text{y} \qquad f_y \equiv \frac{\partial z}{\partial y}$$

Puesto que f_x es por sí misma una función de x (así como de y), podemos medir la tasa de cambio de f_x respecto a x, mientras que y permanece fija, mediante una derivada parcial de segundo orden denotada por f_{xx}, o bien, $\partial^2 z/\partial x^2$:

$$f_{xx} \equiv \frac{\partial}{\partial x}(f_x) \quad \text{o bien,} \quad \frac{\partial^2 z}{\partial x^2} \equiv \frac{\partial}{\partial x}\left(\frac{\partial z}{\partial x}\right)$$

La notación f_{xx} tiene un subíndice doble que significa que la función primitiva f ha sido diferenciada parcialmente respecto a x dos veces, mientras que la notación $\partial^2 z/\partial x^2$ se asemeja a la de d^2z/dx^2 excepto por el uso del símbolo parcial. De manera análoga, podemos usar la segunda derivada parcial

$$f_{yy} \equiv \frac{\partial}{\partial y}(f_y) \quad \text{o bien,} \quad \frac{\partial^2 z}{\partial y^2} = \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial y}\right)$$

para denotar la tasa de cambio de f_y respecto a y, mientras x se mantiene constante.

Recuerde que f_x es también una función de y y que f_y es también una función de x. Por lo tanto, podemos escribir dos segundas derivadas parciales:

$$f_{xy} \equiv \frac{\partial^2 z}{\partial x\,\partial y} \equiv \frac{\partial}{\partial x}\left(\frac{\partial z}{\partial y}\right) \qquad \text{y} \qquad f_{yx} \equiv \frac{\partial^2 z}{\partial y\,\partial x} \equiv \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial x}\right)$$

Éstas se llaman *derivadas parciales cruzadas* (o *mixtas*) porque cada una mide la tasa de cambio de una derivada parcial de primer orden respecto a la "otra" variable.

Conviene repetir que las derivadas parciales de segundo orden de $z = f(x, y)$, al igual que z y las primeras derivadas f_x y f_y, son también funciones de las variables x y y. Cuando el hecho requiere énfasis, podemos escribir f_{xx} como $f_{xx}(x, y)$ y f_{xy} como $f_{xy}(x, y)$, etc. En este mismo sentido, podemos usar la notación $f_{yx}(1, 2)$ para denotar el valor de f_{yx} evaluada en $x = 1$ y $y = 2$, etcétera.

Aunque f_{xy} y f_{yx} se han definido por separado, según una proposición conocida como *teorema de Young*, tendrán valores idénticos siempre y cuando las dos derivadas parciales cruzadas sean continuas. En ese caso, el orden secuencial en el que se emprende la diferenciación es irrelevante, porque $f_{xy} = f_{yx}$. Para los tipos ordinarios de funciones *específicas* con las que se trabaja, por lo común se cumple esta condición de continuidad; para la *mayoría* de las funciones, como ya se mencionó, se supone que siempre se cumple la condición de continuidad. Por consiguiente, podemos esperar hallar derivadas parciales cruzadas idénticas. De hecho, el teorema se aplica también a funciones de tres o más variables. Dada $z = g(u, v, w)$, por ejemplo, las derivadas parciales mixtas se caracterizarán por $g_{uv} = g_{vu}$, $g_{vw} = g_{wv}$, etc., siempre que estas derivadas parciales sean continuas.

Ejemplo 1

Encuentre las cuatro derivadas parciales de segundo orden de

$$z = x^3 + 5xy - y^2$$

Las primeras derivadas parciales de esta función son

$$f_x = 3x^2 + 5y \qquad \text{y} \qquad f_y = 5x - 2y$$

Al proseguir con la diferenciación, obtenemos

$$f_{xx} = 6x \qquad f_{yx} = 5 \qquad f_{xy} = 5 \qquad f_{yy} = -2$$

Como se esperaba, f_{yx} y f_{xy} son idénticas.

Ejemplo 2

Encuentre las derivadas parciales de $z = x^2e^{-y}$. En este caso, las derivadas parciales son

$$f_x = 2xe^{-y} \qquad \text{y} \qquad f_y = -x^2e^{-y}$$

Así, tenemos

$$f_{xx} = 2e^{-y} \qquad f_{yx} = -2xe^{-y} \qquad f_{xy} = -2xe^{-y} \qquad f_{yy} = x^2e^{-y}$$

De nuevo, vemos que $f_{yx} = f_{xy}$.

Tenga en cuenta que las segundas derivadas parciales son funciones de las variables originales x y y. Este hecho es suficientemente claro en el ejemplo 2, pero es cierto incluso para el ejemplo 1, aunque en ese caso sucede que algunas segundas derivadas parciales son funciones *constantes*.

Diferencial total de segundo orden

Dada la diferencial total dz en (11.4), y el concepto de derivadas parciales de segundo orden a disposición, podemos obtener una expresión para la diferencial total de segundo orden d^2z mediante la diferenciación posterior de dz. Es necesario recordar que en la ecuación $dz = f_x\,dx + f_y\,dy$, los símbolos dx y dy representan cambios arbitrarios o específicos en x y y; así que debemos tratarlos como constantes durante la diferenciación. Como resultado, dz depende sólo de f_x y f_y, y puesto que f_x y f_y son por sí mismas funciones de x y y, dz, al igual que z, es una función de x y y.

Para obtener d^2z, solamente se aplica la definición de una diferencial, como se muestra en (11.4), a dz. Así,

$$\begin{aligned}
d^2z \equiv d(dz) &= \frac{\partial(dz)}{\partial x}\,dx + \frac{\partial(dz)}{\partial y}\,dy \qquad \text{[cf. (11.4)]}\\
&= \frac{\partial}{\partial x}(f_x\,dx + f_y\,dy)\,dx + \frac{\partial}{\partial y}(f_x\,dx + f_y\,dy)\,dy\\
&= (f_{xx}\,dx + f_{xy}\,dy)\,dx + (f_{yx}\,dx + f_{yy}\,dy)\,dy\\
&= f_{xx}\,dx^2 + f_{xy}\,dy\,dx + f_{yx}\,dx\,dy + f_{yy}\,dy^2\\
&= f_{xx}\,dx^2 + 2f_{xy}\,dx\,dy + f_{yy}\,dy^2 \qquad [f_{xy} = f_{yx}]
\end{aligned} \tag{11.6}$$

Note una vez más que el exponente 2 aparece en (11.6) en dos formas distintas. En el símbolo d^2z, el exponente 2 indica la diferencial total de *segundo orden* de z; pero en el símbolo $dx^2 \equiv (dx)^2$, el exponente denota el *cuadrado* de la diferencial de primer orden dx.

El resultado de (11.6) muestra la magnitud de d^2z en términos de valores dados de dx y dy, medidos desde algún punto (x_0, y_0) en el dominio. A fin de calcular d^2z, también es necesario conocer las derivadas parciales de segundo orden f_{xx}, f_{xy} y f_{yy}, evaluadas en (x_0, y_0), así como conocer las derivadas parciales de primer orden para calcular dz a partir de (11.4).

Ejemplo 3

Dada $z = x^3 + 5xy - y^2$, encuentre dz y d^2z. Esta función es la misma que la del ejemplo 1. Sustituyendo las distintas derivadas ya obtenidas en (11.4) y (11.6), hallamos[1]

$$dz = (3x^2 + 5y)\,dx + (5x - 2y)\,dy$$

[1] Otra forma de llegar a estos resultados es mediante diferenciación directa de la función:

$$\begin{aligned}
dz &= d(x^3) + d(5xy) - d(y^2)\\
&= 3x^2\,dx + 5y\,dx + 5x\,dy - 2y\,dy
\end{aligned}$$

La diferenciación posterior de dz (teniendo presente que dx y dy son constantes) produce

$$\begin{aligned}
d^2z &= d(3x^2)\,dx + d(5y)\,dx + d(5x)\,dy - d(2y)\,dy\\
&= (6x\,dx)\,dx + (5\,dy)\,dx + (5\,dx)\,dy - (2\,dy)\,dy\\
&= 6x\,dx^2 + 10\,dx\,dy - 2\,dy^2
\end{aligned}$$

y

$$d^2z = 6x\,dx^2 + 10\,dx\,dy - 2\,dy^2$$

Podemos calcular también dz y d^2z en puntos específicos en el dominio. En el punto $x = 1$ y $y = 2$, por ejemplo, tenemos

$$dz = 13\,dx + dy \qquad \text{y} \qquad d^2z = 6\,dx^2 + 10\,dx\,dy - 2\,dy^2$$

Condición de segundo orden

En el caso de una variable, $d^2z < 0$ en un punto estacionario identifica al punto como el máximo de una colina en un espacio bidimensional. De manera similar, en el caso de dos variables, $d^2z < 0$ en un punto estacionario identificaría al punto como la cúspide de un domo en un espacio tridimensional. Una vez que satisfacemos la condición necesaria de primer orden, la condición suficiente de segundo orden para un máximo de $z = f(x, y)$ es

$$d^2z < 0 \text{ para valores arbitrarios de } dx \text{ y } dy\text{, al menos uno diferente de cero} \qquad \textbf{(11.7)}$$

Por otro lado, un valor d^2z positivo en un punto estacionario se relaciona con el fondo de un tazón. La condición suficiente de segundo orden para un mínimo de $z = f(x, y)$ es

$$d^2z > 0 \text{ para valores arbitrarios de } dx \text{ y } dy\text{, al menos uno diferente de cero} \qquad \textbf{(11.8)}$$

La razón de por qué (11.7) y (11.8) son sólo condiciones suficientes, pero no necesarias, es que de nuevo es posible que d^2z tome un valor cero en un máximo o un mínimo. Por esta razón, las condiciones *necesarias* de segundo orden debemos expresarlas con desigualdades débiles como:

$$\begin{array}{ll} \text{Para el máximo de } z\text{:} & d^2z \le 0 \\ \text{Para el mínimo de } z\text{:} & d^2z \ge 0 \end{array} \left.\vphantom{\begin{array}{l}a\\b\end{array}}\right\} \begin{array}{l}\text{para valores arbitrarios de } dx \text{ y } dy\text{, al menos} \\ \text{uno diferente de cero}\end{array} \qquad \textbf{(11.9)}$$

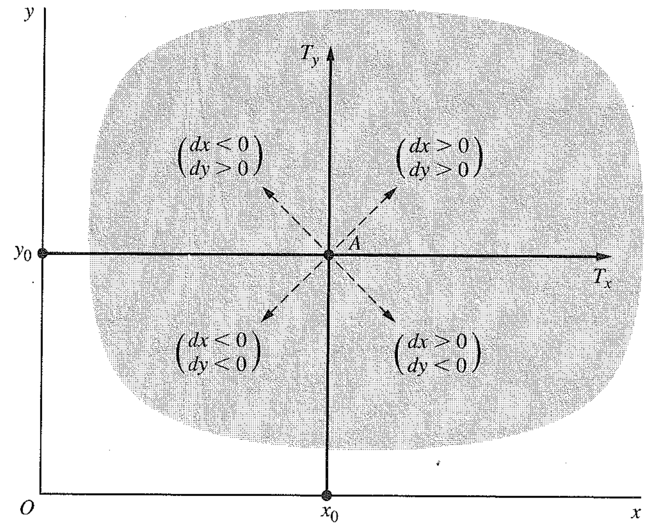
En lo que sigue, centraremos más la atención en las condiciones suficientes de segundo orden.

Por conveniencia operacional, las condiciones diferenciales de segundo orden podemos traducirlas en condiciones equivalentes en derivadas de segundo orden. En el caso de dos variables, (11.6) muestra que esto conllevaría restricciones en los signos de las derivadas parciales de segundo orden f_{xx}, f_{xy} y f_{yy}. La traducción real requeriría conocer las formas cuadráticas, que analizaremos en la sección 11.3. Pero primero podríamos introducir el resultado principal aquí: para cualquier valor de dx y dy, al menos uno diferente de cero,

$$d^2z \begin{cases} < 0 & \text{si y sólo si} \quad f_{xx} < 0; \quad f_{yy} < 0; \quad \text{y} \quad f_{xx}f_{yy} > f_{xy}^2 \\ > 0 & \text{si y sólo si} \quad f_{xx} > 0; \quad f_{yy} > 0; \quad \text{y} \quad f_{xx}f_{yy} > f_{xy}^2 \end{cases}$$

Advierta que el signo de d^2z depende no sólo de f_{xx} y f_{yy}, que tiene que ver con la configuración superficial alrededor del punto A (figura 11.4) en las dos direcciones básicas mostradas por T_x (este-oeste) y T_y (norte-sur), sino también de la derivada parcial cruzada f_{xy}. El papel que desempeña esta última derivada parcial es asegurar que la superficie en cuestión produzca secciones transversales (bidimensionales) con el mismo tipo de configuración (colina o valle, según el caso) no sólo en las dos direcciones básicas (este-oeste y norte-sur), sino también en todas las direcciones posibles (por ejemplo, noreste-sudeste).

FIGURA 11.4



Este resultado, junto con la condición de primer orden (11.5), nos permite construir la tabla 11.1. Debemos entender que las segundas derivadas parciales se evaluarán en el punto estacionario donde $f_x = f_y = 0$. También debemos remarcar que la condición suficiente de segundo orden *no es necesaria* para un extremo; en particular, si un valor estacionario se caracteriza por $f_{xx}f_{yy} = f_{xy}^2$ en violación de esa condición, ese valor estacionario podría resultar un extremo. Por otro lado, en el caso de otro tipo de violación, con un punto estacionario caracterizado por $f_{xx}f_{yy} < f_{xy}^2$, podemos identificar ese punto como un punto silla, porque el signo de d^2z en ese caso será indefinido (positivo para algunos valores de dx y dy, pero negativo para otros).

Ejemplo 4

Encuentre el o los valores extremos de $z = 8x^3 + 2xy - 3x^2 + y^2 + 1$. Primero determinamos las primeras y segundas derivadas parciales:

$$f_x = 24x^2 + 2y - 6x \qquad f_y = 2x + 2y$$
$$f_{xx} = 48x - 6 \qquad f_{yy} = 2 \qquad f_{xy} = 2$$

La condición de primer orden requiere que hagamos que se satisfagan las ecuaciones simultáneas $f_x = 0$ y $f_y = 0$; es decir,

$$24x^2 + 2y - 6x = 0$$
$$2y + 2x = 0$$

La segunda ecuación implica que $y = -x$, y cuando sustituimos esta información en la primera ecuación obtenemos $24x^2 - 8x = 0$, que produce el par de soluciones

$$x_1^* = 0 \quad [\text{que implica } y_1^* = -x_1^* = 0]$$
$$x_2^* = \tfrac{1}{3} \quad \left[\text{que implica } y_2^* = -\tfrac{1}{3}\right]$$

Para aplicar la condición de segundo orden, se nota que cuando

$$x_1^* = y_1^* = 0$$

TABLA 11.1 Condiciones para el extremo relativo: $z = f(x, y)$

Condición	Máximo	Mínimo
Condición necesaria de primer orden	$f_x = f_y = 0$	$f_x = f_y = 0$
Condición suficiente de segundo orden[1]	$f_{xx}, f_{yy} < 0$ y $f_{xx} f_{yy} > f_{xy}^2$	$f_{xx}, f_{yy} > 0$ y $f_{xx} f_{yy} > f_{xy}^2$

[1] Aplicable sólo después de que se satisfizo la condición necesaria de primer orden.

f_{xx} resulta ser -6, mientras que f_{yy} es 2, de modo que $f_{xx} f_{yy}$ es negativa y necesariamente menor que un valor cuadrado f_{xy}^2. Esto no cumple la condición de segundo orden. El hecho de que f_{xx} y f_{yy} tengan signos opuestos indica que la superficie en cuestión se curvará hacia arriba en una dirección y hacia abajo en otra, por lo que dará lugar a un punto silla.

¿Y la otra solución? Cuando evaluamos en $x_2^* = \frac{1}{3}$, se encuentra que $f_{xx} = 10$, lo cual, junto con el hecho de que $f_{yy} = f_{xy} = 2$, satisface las tres partes de la condición suficiente de segundo orden para un mínimo. Por lo tanto, si establecemos que $x = \frac{1}{3}$ y $y = -\frac{1}{3}$ en la función dada, podemos obtener como un mínimo de z el valor $z^* = \frac{23}{27}$. En este ejemplo existe sólo un extremo relativo (un mínimo), que podemos representar mediante la terna ordenada

$$(x^*, y^*, z^*) = \left(\frac{1}{3}, \frac{-1}{3}, \frac{23}{27}\right)$$

Ejemplo 5 Encuentre el o los valores extremos de $z = x + 2ey - e^x - e^{2y}$. Las derivadas pertinentes de esta función son

$$f_x = 1 - e^x \qquad f_y = 2e - 2e^{2y}$$
$$f_{xx} = \quad -e^x \qquad f_{yy} = -4e^{2y} \qquad f_{xy} = 0$$

Para satisfacer la condición necesaria, debemos tener

$$1 - e^x = 0$$
$$2e - 2e^{2y} = 0$$

la cual tiene sólo una solución, a saber, $x^* = 0$ y $y^* = \frac{1}{2}$. Para confirmar la naturaleza del valor de z correspondiente a esta solución (el valor estacionario), evaluamos las derivadas de segundo orden en $x = 0$ y $y = \frac{1}{2}$, y encontramos que $f_{xx} = -1$, $f_{yy} = -4e$ y $f_{xy} = 0$. Puesto que f_{xx} y f_{yy} son negativas y como, además, $(-1)(-4e) > 0$, podemos concluir que el valor de z en cuestión,

$$z^* = 0 + e - e^0 - e^1 = -1$$

es un valor máximo de la función. Este punto máximo en la superficie dada podemos denotarla mediante la tercia ordenada $(x^*, y^*, z^*) = (0, \frac{1}{2}, -1)$.

Observe que para evaluar las segundas derivadas parciales en x^* y y^*, primero debemos emprender la diferenciación y luego, como paso final, sustituiremos los valores específicos de x^* y y^* en las derivadas.

EJERCICIO 11.2

Use la tabla 11.1 para hallar el o los valores extremos de cada una de las cuatro funciones siguientes, y determine si son máximos o mínimos:

1. $z = x^2 + xy + 2y^2 + 3$
2. $z = -x^2 - y^2 + 6x + 2y$

3. $z = ax^2 + by^2 + c$; considere cada uno de los tres casos:
   (*a*) $a > 0$, $b > 0$ (*b*) $a < 0$, $b < 0$ (*c*) a y b con signos opuestos
4. $z = e^{2x} - 2x + 2y^2 + 3$
5. Considere la función $z = (x - 2)^4 + (y - 3)^4$.
   (*a*) Establezca mediante razonamiento intuitivo que z alcanza un mínimo ($z^* = 0$) en $x^* = 2$ y $y^* = 3$.
   (*b*) ¿Se satisface la condición necesaria de primer orden de la tabla 11.1?
   (*c*) ¿Se satisface la condición suficiente de segundo orden de la tabla 11.1?
   (*d*) Encuentre el valor de d^2z. ¿Satisface la condición necesaria de segundo orden para un mínimo en (11.9)?

11.3 Formas cuadráticas, una incursión

La expresión para d^2z en la última línea de (11.6) ejemplifica lo que se conoce como *formas cuadráticas*, para las cuales existen criterios establecidos a fin de determinar si sus signos son siempre positivos, negativos, no positivos o no negativos, para valores arbitrarios de dx y dy, al menos uno diferente de cero. Puesto que la condición de segundo orden para extremo depende directamente del signo de d^2z, esos criterios son de interés directo.

Para empezar, se define una *forma* como expresión polinomial en la cual cada término componente tiene un grado uniforme. El encuentro anterior con polinomios se confinó al caso de una sola variable: $a_0 + a_1x + \cdots + a_nx^n$. Cuando intervienen más variables, cada término de un polinomio podría contener una o varias variables, elevada cada una a una potencia entera no negativa, por ejemplo $3x + 4x^2y^3 - 2yz$. En el caso especial donde cada término tiene un grado uniforme, es decir, donde la suma de exponentes en cada término es uniforme, el polinomio se conoce como una *forma*. Por ejemplo, $4x - 9y + z$ es una *forma lineal* en tres variables, porque cada uno de sus términos es de primer grado. Por otro lado, el polinomio $4x^2 - xy + 3y^2$, en el que cada término es de segundo grado (suma de exponentes enteros $= 2$), constituye una *forma cuadrática* en dos variables. Se podrían encontrar formas cuadráticas en tres variables, como $x^2 + 2xy - yw + 7w^2$, o de hecho en n variables.

Diferencial total de segundo orden como una forma cuadrática

Si consideramos las diferenciales dx y dy de (11.6) como variables y las derivadas parciales como coeficientes, es decir, si hacemos

$$\begin{array}{lll} u \equiv dx & v \equiv dy & \\ a \equiv f_{xx} & b \equiv f_{yy} & h \equiv f_{xy} [= f_{yx}] \end{array} \qquad \textbf{(11.10)}$$

entonces la diferencial total de segundo orden

$$d^2z = f_{xx}\,dx^2 + 2f_{xy}\,dx\,dy + f_{yy}\,dy^2$$

se puede identificar de modo fácil como una forma cuadrática q en las dos variables u y v:

$$q = au^2 + 2huv + bv^2 \qquad \textbf{(11.6')}$$

Tome en cuenta que en esta forma cuadrática $dx \equiv u$ y $dy \equiv v$ se expresan en el papel de variables, mientras que las segundas derivadas parciales se tratan como constantes, lo contrario de la situación cuando se estuvo diferenciando dz para obtener d^2z. La razón de esta inversión de papeles radica en la naturaleza modificada del problema que vamos a tratar ahora. La condición suficiente de segundo orden para el extremo estipula que d^2z sea positiva definida (para un mínimo) y negativa definida (para un máximo), sin importar los valores que dx y dy pudieran tomar (siempre y cuando no sean cero ambas). Por lo tanto, es obvio que en el presente contexto dx y dy deben ser consideradas como *variables*. Por otro lado, las segundas derivadas parciales asumirán valores específicos en los puntos que estamos examinando como puntos extremos posibles y, por lo tanto, las podemos considerar como *constantes*.

La pregunta principal es, entonces: ¿qué restricciones deben satisfacer a, b y h de (11.6′), mientras que u y v puede tomar cualquier valor, a fin de asegurar un signo definitivo para q?

Formas cuadráticas positivas definidas y negativas definidas

Por terminología, destaquemos que se dice que una forma cuadrática q es

$$\left.\begin{array}{l}\textit{Positiva definida}\\ \textit{Positiva semidefinida}\\ \textit{Negativa semidefinida}\\ \textit{Negativa definida}\end{array}\right\}\ \text{si } q \text{ es invariablemente}\ \left\{\begin{array}{ll}\text{positiva} & (>0)\\ \text{no negativa} & (\geq 0)\\ \text{no positiva} & (\leq 0)\\ \text{negativa} & (<0)\end{array}\right.$$

sin que importen los valores de las variables de la forma cuadrática, no todos cero. Por otro lado, si q cambia los signos cuando las variables asumen valores diferentes, se dice que q es *indefinida*. Los casos de certidumbre positiva y negativa de $q = d^2z$ se relacionan con las condiciones *suficientes* de segundo orden para un mínimo y un máximo, respectivamente. Los casos de *semi*definición se relacionan con las condiciones *necesarias* de segundo orden. Cuando $q = d^2z$ es indefinida, tenemos el síntoma de punto de silla.

Prueba de los determinantes para la definición de signo

Una prueba usada de manera extensa para la definición de signo de q requiere examinar los signos de ciertos determinantes. Esta prueba se aplica con más facilidad a la definición positiva y negativa (contrario a la semidefinición); es decir, se aplica más fácilmente a las condiciones suficientes de segundo orden (en oposición a las condiciones necesarias). Aquí, el análisis se confina sólo a las condiciones suficientes.[2]

Para el caso de dos variables, es relativamente fácil obtener las condiciones de los determinantes para definir el signo de q. En primer lugar, se ve que los signos de los términos primero y tercero de (11.6′) son independientes de los valores de las variables u y v, porque estas variables aparecen en cuadrados. Así, es fácil especificar la condición para la definición positiva o negativa de estos términos solamente, restringiendo los signos de a y b. El problema radica en el término intermedio; pero si se puede convertir todo el polinomio en una expresión tal que las variables u y v aparezcan sólo en algunos cuadrados, la definición del signo de q de nuevo será manejable.

[2] Para una explicación referente a una prueba de determinantes para las condiciones necesarias de segundo orden, véase Alpha C. Chiang, *Elements of Dynamic Optimization*, Waveland Press Inc., 1992, pp. 85-90.

El mecanismo que hará el truco es el de completar el cuadrado. Al sumar h^2v^2/a al lado derecho de (11.6′), y restar la misma cantidad, podemos reescribir la forma cuadrática como sigue:

$$\begin{aligned} q &= au^2 + 2huv + \frac{h^2}{a}v^2 + bv^2 - \frac{h^2}{a}v^2 \\ &= a\left(u^2 + \frac{2h}{a}uv + \frac{h^2}{a^2}v^2\right) + \left(b - \frac{h^2}{a}\right)v^2 \\ &= a\left(u + \frac{h}{a}v\right)^2 + \frac{ab - h^2}{a}(v^2) \end{aligned}$$

Ahora que las variables u y v aparecen sólo en cuadrados, podemos establecer el signo de q por completo en los valores de los coeficientes a, b y h como sigue:

$$q \text{ es} \begin{Bmatrix} \text{definida positiva} \\ \text{definida negativa} \end{Bmatrix} \text{ si y sólo si } \begin{Bmatrix} a > 0 \\ a < 0 \end{Bmatrix} \quad \text{y} \quad ab - h^2 > 0 \qquad \textbf{(11.11)}$$

Ahora que (1) $ab - h^2$ debe ser *positiva* en ambos casos y (2) como un prerrequisito para la positividad de $ab - h^2$, el producto ab debe ser positivo (puesto que debe exceder el término cuadrado h^2); por consiguiente, esta condición implica de modo automático que a y b deben tomar el signo algebraico idéntico.

Esta condición se puede expresar de manera más sucinta mediante el uso de determinantes. Observamos primero que la forma cuadrática de (11.6′) se puede reconfigurar en el siguiente formato simétrico, cuadrado:

$$\begin{aligned} q = \quad & a(u^2) + h(uv) \\ + & h(vu) + b(v^2) \end{aligned}$$

con los términos cuadrados colocados en la diagonal y con el término $2huv$ dividido en dos partes iguales y colocado fuera de la diagonal. Los coeficientes forman ahora una matriz simétrica, con a y b en la diagonal principal y h fuera de la diagonal. Vista desde esta perspectiva, se advierte que la forma cuadrática es la matriz de 1×1 (un escalar) que resulta de la multiplicación de matrices siguiente:

$$q = [u \quad v] \begin{bmatrix} a & h \\ h & b \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

Tome en cuenta que es un caso más generalizado del producto de matrices $x'Ax$ analizado en la sección 4.4, ejemplo 5. En ese ejemplo, con una matriz diagonal (una matriz simétrica con sólo ceros como sus elementos fuera de la diagonal) como A, el producto $x'Ax$ representa una suma de cuadrados ponderada. Aquí, con cualquier matriz simétrica como A (permitiendo que aparezcan elementos diferentes de cero fuera de la diagonal), el producto $x'Ax$ es una forma cuadrática.

El determinante de la matriz de coeficientes de 2×2, $\begin{vmatrix} a & h \\ h & b \end{vmatrix}$, al cual se le conoce como el *discriminante* de la forma cuadrática q, y que se denotará mediante $|D|$, proporciona la pista para el criterio en (11.11), porque este último se puede expresar también como:

$$q \text{ es} \begin{Bmatrix} \text{definida positiva} \\ \text{definida negativa} \end{Bmatrix} \text{ si y sólo si } \begin{Bmatrix} |a| > 0 \\ |a| < 0 \end{Bmatrix} \quad \text{y} \quad \begin{vmatrix} a & h \\ h & b \end{vmatrix} > 0 \qquad \textbf{(11.11')}$$

El determinante $|a| = a$ es simplemente el *primer menor principal director* de $|D|$. Por otro lado, el determinante $\begin{vmatrix} a & h \\ h & b \end{vmatrix}$ es el *segundo menor principal director* de $|D|$. En el presente caso, sólo hay dos menores principales directores disponibles, y sus signos servirán para determinar la definición positiva o negativa de q.

Cuando se traduce (11.11′), vía (11.10), en términos de la diferencial total de segundo orden d^2z, tenemos

$$d^2z \text{ es} \begin{Bmatrix} \text{definida positiva} \\ \text{definida negativa} \end{Bmatrix} \text{si y sólo si} \begin{Bmatrix} f_{xx} > 0 \\ f_{xx} < 0 \end{Bmatrix} \quad \text{y} \quad \begin{vmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{vmatrix} = f_{xx}f_{yy} - f_{xy}^2 > 0$$

Recordando que la última desigualdad implica que se requiere que f_{xx} y f_{yy} tomen el mismo *signo*, se ve que ésta es precisamente la condición suficiente de segundo orden presentada en la tabla 11.1.

En general, el discriminante de una forma cuadrática

$$q = au^2 + 2huv + bv^2$$

es el determinante simétrico $\begin{vmatrix} a & h \\ h & b \end{vmatrix}$. En el caso particular de la forma cuadrática

$$d^2z = f_{xx}\,dx^2 + 2f_{xy}\,dx\,dy + f_{yy}\,dy^2$$

el discriminante es un determinante con las derivadas parciales de segundo orden como sus elementos. Tal determinante se llama *determinante hessiano* (o sólo *hessiano*). En el caso de dos variables, el hessiano es

$$|H| = \begin{vmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{vmatrix}$$

que, en vista del teorema de Young ($f_{xy} = f_{yx}$), es simétrico, como debe ser un discriminante. Debemos distinguir con cuidado el determinante hessiano del determinante jacobiano descrito en la sección 7.6.

Ejemplo 1 ¿Es $q = 5u^2 + 3uv + 2v^2$ definida positiva o negativa? El discriminante de q es $\begin{vmatrix} 5 & 1.5 \\ 1.5 & 2 \end{vmatrix}$, con menores principales directores

$$5 > 0 \qquad \text{y} \qquad \begin{vmatrix} 5 & 1.5 \\ 1.5 & 2 \end{vmatrix} = 7.75 > 0$$

Por lo tanto q es definida positiva.

Ejemplo 2 Dada $f_{xx} = -2$, $f_{xy} = 1$ y $f_{yy} = -1$ en un cierto punto de una función $z = f(x, y)$, ¿d^2z tiene un signo definido en ese punto sin importar los valores de dx y dy? El discriminante de la forma cuadrática d^2z es en este caso $\begin{vmatrix} -2 & 1 \\ 1 & -1 \end{vmatrix}$, con los menores principales directores

$$-2 < 0 \qquad \text{y} \qquad \begin{vmatrix} -2 & 1 \\ 1 & -1 \end{vmatrix} = 1 > 0$$

Así, d^2z es definida negativa.

Formas cuadráticas de tres variables

¿Se pueden obtener condiciones similares para una forma cuadrática de *tres* variables?

Una forma cuadrática con tres variables u_1, u_2 y u_3 se puede representar de manera general como

$$\begin{aligned} q(u_1, u_2, u_3) = \quad & d_{11}(u_1^2) \quad + d_{12}(u_1 u_2) + d_{13}(u_1 u_3) \\ & + d_{21}(u_2 u_1) + d_{22}(u_2^2) \quad + d_{23}(u_2 u_3) \\ & + d_{31}(u_3 u_1) + d_{32}(u_3 u_2) + d_{33}(u_3^2) \\ = & \sum_{i=1}^{3} \sum_{j=1}^{3} d_{ij} u_i u_j \end{aligned} \qquad \textbf{(11.12)}$$

donde la notación doble $\sum$ (doble suma) significa que se permite que tanto el índice i como el índice j tomen los valores 1, 2 y 3; por lo tanto, la expresión de doble suma es equivalente al arreglo de 3×3 mostrado en la ecuación (11.12). Este arreglo cuadrado de la forma cuadrática, por cierto, se considera siempre simétrico, aun cuando se ha escrito el par de coeficientes (d_{12}, d_{21}) o (d_{23}, d_{32}) como si los dos miembros de cada par fueran diferentes. Porque, si sucede que el término de la forma cuadrática en el cual intervienen las variables u_1 y u_2 es, por ejemplo, $12u_1u_2$, podemos hacer que $d_{12} = d_{21} = 6$, así que $d_{12}u_1u_2 = d_{21}u_2u_1$, y aplicar un procedimiento similar para hacer simétricos los otros elementos que están fuera de la diagonal.

Esta forma cuadrática de tres variables se puede expresar de nuevo como un producto de tres matrices:

$$q(u_1, u_2, u_3) = [u_1 \quad u_2 \quad u_3] \begin{bmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \equiv u'Du \qquad \textbf{(11.12')}$$

Como en el caso de dos variables, la primera matriz (un vector renglón) y la tercera matriz (un vector columna) solamente listan las variables, y la intermedia (D) es una matriz de coeficientes simétrica de la versión de arreglo cuadrado de la forma cuadrática de (11.12). Sin embargo, esta vez se puede formar un total de *tres* menores principales directores a partir de su discriminante, a saber,

$$|D_1| \equiv d_{11} \qquad |D_2| \equiv \begin{vmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{vmatrix} \qquad |D_3| \equiv \begin{vmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{vmatrix}$$

donde $|D_i|$ denota el i-ésimo menor principal director del discriminante $|D|$.[3] Las condiciones para la definición positiva o negativa de nuevo se pueden establecer en términos de ciertas restricciones de signo en estos menores principales.

Mediante el ahora ya conocido mecanismo de completar el cuadrado, la forma cuadrática (11.12) se puede convertir en una expresión en la cual las tres variables aparecen sólo como

[3] Hasta aquí se ha considerado el i-ésimo menor principal director $|D_i|$ como un subdeterminante que se forma al retener los primeros i elementos de la diagonal principal de $|D|$. Puesto que la noción de *menor* implica la *eliminación* de algo del determinante original, es posible que se prefiera ver el i-ésimo menor principal director como un subdeterminante formado al eliminar los últimos $(n - i)$ renglones y columnas de $|D|$.

componentes de algunos cuadrados. En particular, recordando que $a_{12} = a_{21}$, etc., tenemos

$$q = d_{11}\left(u_1 + \frac{d_{12}}{d_{11}}u_2 + \frac{d_{13}}{d_{11}}u_3\right)^2 + \frac{d_{11}d_{22} - d_{12}^2}{d_{11}}\left(u_2 + \frac{d_{11}d_{23} - d_{12}d_{13}}{d_{11}d_{22} - d_{12}^2}u_3\right)^2$$
$$+ \frac{d_{11}d_{22}d_{33} - d_{11}d_{23}^2 - d_{22}d_{13}^2 - d_{33}d_{12}^2 + 2d_{12}d_{13}d_{23}}{d_{11}d_{22} - d_{12}^2}(u_3)^2$$

Esta suma de cuadrados será positiva (negativa) para cualquier valor de u_1, u_2 y u_3, todos no cero, si y sólo si los coeficientes de las tres expresiones cuadradas son todos positivos (negativos). Pero los tres coeficientes (en el orden dado) se pueden expresar en términos de los tres menores principales directores como sigue:

$$|D_1| \qquad \frac{|D_2|}{|D_1|} \qquad \frac{|D_3|}{|D_2|}$$

Por consiguiente, para una *definición positiva*, la condición necesaria y suficiente es triple:

$$\begin{aligned} &|D_1| > 0 \\ &|D_2| > 0 \qquad \text{[cuando previamente se tenga } |D_1| > 0] \\ &|D_3| > 0 \qquad \text{[cuando previamente se tenga } |D_2| > 0] \end{aligned}$$

En otras palabras, los tres menores principales directores deben ser positivos. Para una *definición negativa*, la condición necesaria y suficiente se convierte en:

$$\begin{aligned} &|D_1| < 0 \\ &|D_2| > 0 \qquad \text{[cuando previamente se tenga } |D_1| < 0] \\ &|D_3| < 0 \qquad \text{[cuando previamente se tenga } |D_2| > 0] \end{aligned}$$

Es decir, los tres menores principales directores deben alternar en signo en la manera especificada.

Ejemplo 3

Determine si $q = u_1^2 + 6u_2^2 + 3u_3^2 - 2u_1u_2 - 4u_2u_3$ es definida positiva o negativa. El discriminante de q es

$$\begin{vmatrix} 1 & -1 & 0 \\ -1 & 6 & -2 \\ 0 & -2 & 3 \end{vmatrix}$$

con tres menores principales directores como sigue:

$$1 > 0 \qquad \begin{vmatrix} 1 & -1 \\ -1 & 6 \end{vmatrix} = 5 > 0 \qquad \text{y} \qquad \begin{vmatrix} 1 & -1 & 0 \\ -1 & 6 & -2 \\ 0 & -2 & 3 \end{vmatrix} = 11 > 0$$

Por lo tanto, la forma cuadrática es positiva definida.

Ejemplo 4

Determine si $q = 2u^2 + 3v^2 - w^2 + 6uv - 8uw - 2vw$ es definida positiva o negativa. El discriminante se puede escribir como $\begin{vmatrix} 2 & 3 & -4 \\ 3 & 3 & -1 \\ -4 & -1 & -1 \end{vmatrix}$, y se encuentra que su primer menor principal director es $2 > 0$, pero el segundo menor principal director es $\begin{vmatrix} 2 & 3 \\ 3 & 3 \end{vmatrix} = -3 < 0$.

Esto viola la condición para definición positiva y negativa; así que q no es positiva definida ni negativa definida.

Formas cuadráticas de n variables

Como una extensión del resultado precedente para el caso de n variables, se establecerá sin demostración que para la forma cuadrática

$$q(u_1, u_2, \ldots, u_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij} u_i u_j \qquad [\text{donde } d_{ij} = d_{ji}]$$
$$= \underset{(1\times n)}{u'} \underset{(n\times n)}{D} \underset{(n\times 1)}{u} \qquad [\text{cf. } (11.12')]$$

la condición necesaria y suficiente para *definición positiva* es que los menores principales directores de $|D|$, a saber,

$$|D_1| \equiv d_{11} \qquad |D_2| \equiv \begin{vmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{vmatrix} \qquad \cdots \qquad |D_n| \equiv \begin{vmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ d_{n1} & d_{n2} & \cdots & d_{nn} \end{vmatrix}$$

todos sean positivos. La condición necesaria y suficiente correspondiente para *definición negativa* es que los menores principales directores alternan en signo como sigue:

$$|D_1| < 0 \qquad |D_2| > 0 \qquad |D_3| < 0 \qquad \text{(etc.)}$$

de modo que todos los de número *impar* sean negativos y los de número par sean positivos. El n-ésimo menor principal director, $|D_n| = |D|$, debe ser positivo si n es par, pero negativo si n es impar. Esto se puede expresar de manera breve mediante la desigualdad $(-1)^n|D_n| > 0$.

Prueba de la raíz característica para definición de signo de una forma cuadrática

Aparte de la prueba precedente de los determinantes para la definición de signo de una forma cuadrática $u'Du$, hay una prueba opcional que utiliza el concepto de las denominadas raíces características de la matriz D. Este concepto surge en problemas de la naturaleza siguiente. Dada una matriz D de $n \times n$, ¿podemos encontrar un escalar r y un vector $x \neq 0$ de $n \times 1$, tal que la ecuación matricial

$$\underset{(n\times n)}{D} \underset{(n\times 1)}{x} = r \underset{(n\times 1)}{x} \qquad \textbf{(11.13)}$$

se cumpla? En caso afirmativo, el escalar r se denomina *raíz característica* de la matriz D y x se llama *vector característico* de esa matriz.[4]

La ecuación matricial $Dx = rx$ se puede reescribir como $Dx - rIx = 0$, o bien

$$(D - rI)x = 0 \qquad \text{donde 0 es una matriz } n \times 1 \qquad \textbf{(11.13')}$$

[4] Las raíces características se conocen también como valores propios, autovalores o "eigenvalores". Los vectores característicos también se llaman *vectores propios, autovectores* o *"eigenvectores"*.

Esto, por supuesto, representa un sistema de n ecuaciones lineales homogéneas. Puesto que queremos una solución no trivial para x, se requiere que la matriz de coeficientes $(D - rI)$, llamada *matriz característica* de D, sea singular; en otras palabras, se debe hacer que su determinante se anule:

$$|D - rI| = \begin{vmatrix} d_{11} - r & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} - r & \cdots & d_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ d_{n1} & d_{n2} & \cdots & d_{nn} - r \end{vmatrix} = 0 \qquad \textbf{(11.14)}$$

La ecuación (11.14) se llama *ecuación característica* de la matriz D. Puesto que al aplicar el desarrollo de Laplace el determinante $|D - rI|$ producirá un polinomio de n-ésimo grado en la variable r, (11.14) es de hecho una ecuación polinomial de n-ésimo grado. Por lo tanto, habrá un total de n raíces, $(r_1, \ldots, r_n)$, cada una de las cuales califica como una raíz característica. Si D es simétrica, como es el caso del contexto de la forma cuadrática, las raíces características siempre serán números reales, pero pueden tomar cualquier signo algebraico, o ser cero.

En vista de que estos valores de r anularán el determinante $|D - rI|$, la sustitución de cualquiera de ellos (por ejemplo, r_i) en el sistema de ecuaciones (11.13′) producirá un vector correspondiente $x|_{r=r_i}$. Para ser precisos, si es homogéneo el sistema, producirá un número infinito de vectores correspondientes a la raíz r_i. Sin embargo, aplicaremos un proceso de *normalización* (que explicaremos en el ejemplo 5) y seleccionaremos un miembro particular del conjunto infinito cuyo vector característico corresponda a r_i; este vector se denotará por v_i. Con un total de n raíces características, debe haber un total de n vectores característicos correspondientes.

Ejemplo 5

Determine las raíces y vectores característicos de la matriz $\begin{bmatrix} 2 & 2 \\ 2 & -1 \end{bmatrix}$. Al sustituir D en (11.14) por la matriz dada, obtenemos la ecuación

$$\begin{vmatrix} 2-r & 2 \\ 2 & -1-r \end{vmatrix} = r^2 - r - 6 = 0$$

con raíces $r_1 = 3$ y $r_2 = -2$. Cuando se usa la primera raíz, la ecuación matricial (11.13′) toma la forma de

$$\begin{bmatrix} 2-3 & 2 \\ 2 & -1-3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 & 2 \\ 2 & -4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Debido a que los dos renglones de la matriz de coeficientes son linealmente dependientes, como se esperaría en vista de (11.14), hay un número infinito de soluciones, que se pueden expresar mediante la ecuación $x_1 = 2x_2$. Para producir una solución única, se *normaliza* la solución imponiendo la restricción $x_1^2 + x_2^2 = 1$.[5] Entonces, puesto que

$$x_1^2 + x_2^2 = (2x_2)^2 + x_2^2 = 5x_2^2 = 1$$

se puede obtener (al sacar la raíz cuadrada positiva) $x_2 = 1/\sqrt{5}$, y también $x_1 = 2x_2 = 2/\sqrt{5}$. Así que el primer vector característico es

$$v_1 = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}$$

[5] De modo más general, para el caso de n variables, se requiere que $\sum_{i=1}^{n} x_i^2 = 1$.

De manera similar, si usamos la segunda raíz $r_2 = -2$ en (11.13′), obtenemos la ecuación

$$\begin{bmatrix} 2-(-2) & 2 \\ 2 & -1-(-2) \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

que tiene la solución $x_1 = -\frac{1}{2}x_2$. Al llevar a cabo la normalización, hallamos

$$x_1^2 + x_2^2 = \left(-\tfrac{1}{2}x_2\right)^2 + x_2^2 = \tfrac{5}{4}x_2^2 = 1$$

que produce $x_2 = 2/\sqrt{5}$ y $x_1 = -1/\sqrt{5}$. Así, el segundo vector característico es

$$v_2 = \begin{bmatrix} -1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$$

El conjunto de vectores característicos que obtenemos de esta manera posee dos propiedades importantes: primera, el producto escalar $v_i'v_i$ $(i = 1, 2, \ldots, n)$ debe ser igual a la unidad, puesto que

$$v_i'v_i = [x_1 \quad x_2 \quad \cdots \quad x_n]\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \sum_{i=1}^{n} x_i^2 = 1 \qquad \text{[por normalización]}$$

Segunda, el producto escalar $v_i'v_j$ (donde $i \neq j$) siempre es cero.[6] En suma, podemos escribir que

$$v_i'v_i = 1 \qquad \text{y} \qquad v_i'v_j = 0 \qquad (i \neq j) \tag{11.15}$$

Estas propiedades demostrarán su utilidad después (véase el ejemplo 6). Por terminología, cuando dos vectores generan un producto escalar con valor cero, se dice que los vectores son *ortogonales* (perpendiculares) entre sí.[7] Por consiguiente, cada par de vectores característicos de la matriz D debe ser ortogonal. La otra propiedad, $v_i'v_i = 1$, es indicativa de normalización. Juntas, estas dos propiedades explican el hecho de que digamos que los vectores caracterís-

[6] Para demostrar esto, se observa que, mediante (11.13), se puede escribir $Dv_j = r_jv_j$, y $Dv_i = r_iv_i$. Al premultiplicar ambos lados de cada una de estas ecuaciones por un vector apropiado, se tiene

$$v_i'Dv_j = v_i'r_jv_j = r_jv_i'v_j \qquad [r_j \text{ es un escalar}]$$
$$v_j'Dv_i = v_j'r_iv_i = r_iv_j'v_i = r_iv_i'v_j \qquad [v_i'v_j = v_j'v_i]$$

Puesto que $v_i'Dv_j$ y $v_j'Dv_i$ son 1×1, y como se transponen entre sí (recuerde que $D' = D$ porque D es simétrica), deben representar el mismo escalar. Se deduce que las expresiones del extremo derecho en estas dos ecuaciones son iguales; por consiguiente, se tiene

$$(r_j - r_i)v_i'v_j = 0$$

Ahora si $r_j \neq r_i$ (raíces distintas), entonces $v_i'v_j$ tiene que ser cero a fin de que se cumpla la ecuación, y esto establece la afirmación. Si $r_j = r_i$ (raíces repetidas), además, siempre será posible hallar dos vectores normalizados linealmente independientes que satisfacen $v_i'v_j = 0$. Por lo tanto, se puede expresar en general que $v_i'v_j = 0$, siempre que $i \neq j$.

[7] Como una ilustración sencilla de esto, considere los dos vectores unitarios de un espacio bidimensional, $e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ y $e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Estos vectores quedan, respectivamente, en los dos ejes y, por lo tanto, son perpendiculares. Al mismo tiempo, se encuentra que $e_1'e_2 = e_2'e_1 = 0$. Véase también el ejercicio 4.3-4.

ticos ($v_1, \ldots, v_n$) son un conjunto de vectores *ortonormales*. Debe intentar comprobar la ortonormalidad de los dos vectores característicos del ejemplo 5.

Ahora estamos preparados para explicar la forma en que las raíces características y los vectores característicos de la matriz D pueden ser útiles para determinar la definición de signo de la forma cuadrática $u'Du$. En esencia, la idea es transformar de nuevo $u'Du$ (en el que intervienen no sólo los términos cuadrados $u_1^2, \ldots, u_n^2$, sino también los términos de producto cruzado como u_1u_2 y u_2u_3) en una forma que contiene sólo términos cuadrados. En teoría, el método es similar al proceso de completar el cuadrado usado antes para obtener la prueba de los determinantes. Sin embargo, en el caso presente, la transformación posee la característica adicional de que cada término cuadrado tiene como coeficiente una de las raíces característica, de modo que los signos de las n raíces proporcionarán información suficiente para determinar la definición de signo de la forma cuadrática.

La transformación que logra un buen resultado es la siguiente. Sean los vectores característicos $v_1, \ldots, v_n$ los que constituyen las columnas de una matriz T:

$$\underset{(n\times n)}{T} = [v_1 \quad v_2 \quad \cdots \quad v_n]$$

y luego se aplica la transformación $\underset{(n\times 1)}{u} = \underset{(n\times n)}{T}\ \underset{(n\times 1)}{y}$ a la forma cuadrática $u'Du$:

$$\begin{aligned} u'Du &= (Ty)'D(Ty) = y'T'DTy \qquad [\text{por } (4.11)] \\ &= y'Ry \qquad \text{donde} \qquad R \equiv T'DT \end{aligned}$$

Como resultado, la forma cuadrática original de las variables u_i se convierte en otra forma cuadrática en las variables y_i. Como las variables u_i y las variables y_i toman el mismo tipo de valores, la transformación no afecta la definición de signo de la forma cuadrática. De esta forma podemos considerar también ahora el signo de la forma cuadrática $y'Ry$. Lo que hace fascinante a esta última forma cuadrática es que la matriz R resultará ser una diagonal, con las raíces $r_1, \ldots, r_n$ de la matriz D mostrada a lo largo de su diagonal, y con ceros en cualquier otra parte, así que tenemos, de hecho,

$$\begin{aligned} u'Du = y'Ry &= [y_1 \quad y_2 \quad \cdots \quad y_n] \begin{bmatrix} r_1 & 0 & \cdots & 0 \\ 0 & r_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & r_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \\ &= r_1y_1^2 + r_2y_2^2 + \cdots + r_ny_n^2 \end{aligned} \qquad \textbf{(11.16)}$$

la cual es una expresión en la que intervienen sólo términos cuadrados. Por lo tanto, la transformación $R \equiv T'DT$ nos proporciona un procedimiento para *diagonalizar* la matriz simétrica D en la matriz diagonal especial R.

Ejemplo 6 Compruebe que la matriz $\begin{bmatrix} 2 & 2 \\ 2 & -1 \end{bmatrix}$ del ejemplo 5 se puede diagonalizar en la matriz $\begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & -2 \end{bmatrix}$. Con base en los vectores característicos encontrados en el ejemplo 5, la matriz de transformación T debe ser

$$T = [v_1 \quad v_2] = \begin{bmatrix} 2/\sqrt{5} & -1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}$$

Así, podemos escribir

$$R \equiv T'DT = \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} 2 & 2 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & -2 \end{bmatrix}$$

la cual comprueba como es debido el proceso de diagonalización.

Para probar el resultado de la diagonalización de (11.16), podemos escribir (en parte) la matriz R como sigue:

$$R \equiv T'DT = \begin{bmatrix} v_1' \\ v_2' \\ \vdots \\ v_n' \end{bmatrix} D[v_1 \quad v_2 \quad \cdots \quad v_n]$$

De esta manera, comprobamos con facilidad que $D[v_1 \quad v_2 \quad \cdots \quad v_n]$ se puede reescribir como $[Dv_1 \quad Dv_2 \quad \cdots \quad Dv_n]$. Además, por (11.13), también podemos escribir esto como $[r_1v_1 \quad r_2v_2 \quad \cdots \quad r_nv_n]$. Por consiguiente, vemos que

$$R = \begin{bmatrix} v_1' \\ v_2' \\ \vdots \\ v_n' \end{bmatrix} [r_1v_1 \quad r_2v_2 \quad \cdots \quad r_nv_n] = \begin{bmatrix} r_1v_1'v_1 & r_2v_1'v_2 & \cdots & r_nv_1'v_n \\ r_1v_2'v_1 & r_2v_2'v_2 & \cdots & r_nv_2'v_n \\ \cdots & \cdots & \cdots & \cdots \\ r_1v_n'v_1 & r_2v_n'v_2 & \cdots & r_nv_n'v_n \end{bmatrix}$$

$$= \begin{bmatrix} r_1 & 0 & \cdots & 0 \\ 0 & r_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & r_n \end{bmatrix} \qquad \text{[por (11.5)]}$$

que es precisamente lo que pretendemos demostrar.

En vista del resultado de (11.16), podemos expresar de modo formal la prueba de la raíz característica para la definición de signo de una forma cuadrática como sigue:

1. $q = u'Du$ es definida positiva (negativa), si y sólo si *cada* raíz característica de D es positiva (negativa).
2. $q = u'Du$ es semidefinida positiva (negativa), si y sólo si todas las raíces características de D son no negativas (no positivas).
3. $q = u'Du$ es indefinida, si y sólo si algunas de las raíces características de D son positivas y algunas son negativas.

Tome en cuenta que, al aplicar esta prueba, todo lo que necesitamos son las raíces características; los vectores característicos no se requieren a menos que deseemos encontrar la matriz de transformación T. Advierta también que esta prueba, a diferencia de la prueba de los determinantes descrita con anterioridad, permite comprobar las condiciones necesarias de segundo orden (parte 2 de la prueba) al mismo tiempo con las condiciones suficientes (parte 1 de la prueba). Sin embargo, ésta tiene una desventaja. Cuando la matriz D es de una dimensión alta, es posible que no sea fácil resolver la ecuación polinomial (11.14) a fin de determinar las raíces características necesarias para la prueba. En esos casos, tal vez sea preferible la prueba de los determinantes.

EJERCICIO 11.3

1. Por multiplicación matricial directa, exprese cada uno de los siguientes productos de matrices como una forma cuadrática:

   (*a*) $[u \quad v]\begin{bmatrix} 4 & 2 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} u \\ v \end{bmatrix}$ (*c*) $[x \quad y]\begin{bmatrix} 5 & 2 \\ 4 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$

   (*b*) $[u \quad v]\begin{bmatrix} -2 & 3 \\ 1 & -4 \end{bmatrix}\begin{bmatrix} u \\ v \end{bmatrix}$ (*d*) $[dx \quad dy]\begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}\begin{bmatrix} dx \\ dy \end{bmatrix}$

2. En el problema 1*b* y *c*, las matrices de coeficientes no son simétricas respecto a la diagonal principal. Compruebe que al promediar los elementos que están fuera de la diagonal

   y, por consiguiente, convertirlos respectivamente en $\begin{bmatrix} -2 & 2 \\ 2 & -4 \end{bmatrix}$ y $\begin{bmatrix} 5 & 3 \\ 3 & 0 \end{bmatrix}$ obtenemos las mismas formas cuadráticas que antes.

3. Con base en sus matrices de coeficientes (las versiones *simétricas*), establezca mediante la prueba de los determinantes si las formas cuadráticas del problema 1*a*, *b* y *c* son positivas definidas o negativas definidas.

4. Exprese cada una de las siguientes formas cuadráticas como un producto de matrices en el que interviene una matriz de coeficientes *simétrica*:

   (*a*) $q = 3u^2 - 4uv + 7v^2$ (*d*) $q = 6xy - 5y^2 - 2x^2$

   (*b*) $q = u^2 + 7uv + 3v^2$ (*e*) $q = 3u_1^2 - 2u_1u_2 + 4u_1u_3 + 5u_2^2 + 4u_3^2 - 2u_2u_3$

   (*c*) $q = 8uv - u^2 - 31v^2$ (*f*) $q = -u^2 + 4uv - 6uw - 4v^2 - 7w^2$

5. De los discriminantes obtenidos de las matrices de coeficientes simétricas del problema 4, constate mediante la prueba de los determinantes cuáles de las formas cuadráticas son positivas definidas y cuáles negativas definidas.

6. Encuentre las raíces características de cada una de las siguientes matrices:

   (*a*) $D = \begin{bmatrix} 4 & 2 \\ 2 & 3 \end{bmatrix}$ (*b*) $E = \begin{bmatrix} -2 & 2 \\ 2 & -4 \end{bmatrix}$ (*c*) $F = \begin{bmatrix} 5 & 3 \\ 3 & 0 \end{bmatrix}$

   ¿Qué se puede concluir acerca de los signos de las formas cuadráticas $u'Du$, $u'Eu$ y $u'Fu$? (compruebe sus resultados contra el problema 3).

7. Encuentre los vectores característicos de la matriz $\begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$.

8. Dada una forma cuadrática $u'Du$, donde D es 2×2, la ecuación característica de D se puede escribir como

   $$\begin{vmatrix} d_{11} - r & d_{12} \\ d_{21} & d_{22} - r \end{vmatrix} = 0 \qquad (d_{12} = d_{21})$$

   Desarrolle el determinante, expresa las raíces de esta ecuación mediante el uso de la fórmula cuadrática, y deduzca lo siguiente:

   (*a*) Ningún número imaginario (un número con $\sqrt{-1}$) puede ocurrir en r_1 y r_2.

   (*b*) Para tener raíces repetidas, la matriz D debe ser de la forma de $\begin{bmatrix} c & 0 \\ 0 & c \end{bmatrix}$.

   (*c*) Para tener semidefinición positiva o negativa, se podría anular el discriminante de la forma cuadrática, es decir, $|D| = 0$ es posible.

11.4 Funciones objetivo con más de dos variables

Cuando en una función objetivo aparecen $n > 2$ variables de elección, ya no es posible graficar la función, aunque podemos hablar de una *hipersuperficie* en un espacio de $(n + 1)$ dimensiones. En tal hipersuperficie (no graficable) podrían existir análogos de cúspide de domos o fondos de tazones de $(n + 1)$ dimensiones. ¿Cómo se identifican?

Condición de primer orden para el extremo

Consideremos específicamente una función de tres variables de elección,

$$z = f(x_1, x_2, x_3)$$

con primeras derivadas parciales f_1, f_2 y f_3 y segundas derivadas parciales f_{ij} $(\equiv \partial^2 z/\partial x_i \partial x_j)$, con $i, j = 1, 2,3$. En virtud del teorema de Young, tenemos $f_{ij} = f_{ji}$.

Esta explicación hace pensar que, para tener un máximo o un mínimo de z, es necesario que $dz = 0$ para valores arbitrarios de dx_1, dx_2 y dx_3, al menos uno diferente de cero. Puesto que ahora el valor de dz es

$$dz = f_1\, dx_1 + f_2\, dx_2 + f_3\, dx_3 \tag{11.17}$$

y puesto que dx_1, dx_2 y dx_3 son cambios arbitrarios en las variables independientes, al menos alguno diferente de cero, la única forma de garantizar una dz cero es tener $f_1 = f_2 = f_3 = 0$. Así, de nuevo la condición necesaria para el extremo es que todas las derivadas parciales de primer orden sean cero, lo mismo que para el caso de dos variables.[8]

Condición de segundo orden

El cumplimiento de la condición de primer orden marca ciertos valores de z como los valores estacionarios de la función objetivo. Si en un valor estacionario de z se encuentra que d^2z es positiva definida, esto bastará para establecer el valor de z como un mínimo. De manera análoga, la definición negativa de d^2z es una condición suficiente para que el valor estacionario sea un máximo. Esto da lugar a las preguntas de cómo expresar d^2z cuando hay tres variables en la función y cómo determinar su definición positiva o negativa.

La expresión para d^2z la obtenemos mediante la diferenciación de dz en (11.17). En tal proceso, como en (11.6), debemos tratar las derivadas f_i como variables y las diferenciales dx_i como constantes.

[8] Como un caso especial, tome en cuenta que si se trabaja con una función $z = f(x_1, x_2, x_3)$ definida de modo implícito por una ecuación $F(z, x_1, x_2, x_3) = 0$, donde

$$f_i \equiv \frac{\partial z}{\partial x_i} = \frac{-\partial F/\partial x_i}{\partial F/\partial z} \qquad (i = 1, 2, 3)$$

entonces la condición de primer orden $f_1 = f_2 = f_3 = 0$ equivale a la condición

$$\frac{\partial F}{\partial x_1} = \frac{\partial F}{\partial x_2} = \frac{\partial F}{\partial x_3} = 0$$

puesto que el valor del denominador $\partial F/\partial z \neq 0$ no tiene importancia.

Así, tenemos

$$\begin{aligned}
d^2z = d(dz) &= \frac{\partial(dz)}{\partial x_1}\,dx_1 + \frac{\partial(dz)}{\partial x_2}\,dx_2 + \frac{\partial(dz)}{\partial x_3}\,dx_3 \\
&= \frac{\partial}{\partial x_1}(f_1\,dx_1 + f_2\,dx_2 + f_3\,dx_3)\,dx_1 \\
&\quad + \frac{\partial}{\partial x_2}(f_1\,dx_1 + f_2\,dx_2 + f_3\,dx_3)\,dx_2 \\
&\quad + \frac{\partial}{\partial x_3}(f_1\,dx_1 + f_2\,dx_2 + f_3\,dx_3)\,dx_3 \\
&= f_{11}\,dx_1^2 + f_{12}\,dx_1\,dx_2 + f_{13}\,dx_1\,dx_3 \\
&\quad + f_{21}\,dx_2\,dx_1 + f_{22}\,dx_2^2 + f_{23}\,dx_2\,dx_3 \\
&\quad + f_{31}\,dx_3\,dx_1 + f_{32}\,dx_3\,dx_2 + f_{33}\,dx_3^2
\end{aligned} \tag{11.18}$$

que es una forma cuadrática similar a (11.12); en consecuencia, los criterios para definición positiva y negativa aprendidos antes se pueden aplicar aquí de manera directa.

En la determinación de la definición positiva o negativa de d^2z, debemos de nuevo, como lo hicimos en (11.6′), considerar a las dx_i como variables que pueden tomar cualquier valor (aunque al menos uno diferente de cero), mientras se considera a las derivadas f_{ij} como coeficientes a los que se pueden imponer ciertas restricciones. Los coeficientes de (11.18) dan lugar al determinante hessiano simétrico

$$|H| = \begin{vmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{vmatrix}$$

cuyos menores principales directores se pueden denotar por medio de

$$|H_1| = f_{11} \qquad |H_2| = \begin{vmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{vmatrix} \qquad |H_3| = |H|$$

Con base en los criterios de los determinantes para definición positiva y negativa, podemos enunciar la condición suficiente de segundo orden para un extremo de z como sigue:

$$z^* \text{ es un } \begin{Bmatrix} \text{máximo} \\ \text{mínimo} \end{Bmatrix}$$

$$\text{si } \begin{cases} |H_1| < 0; \quad |H_2| > 0; \quad |H_3| < 0 & (d^2z \text{ negativa definida}) \\ |H_1| > 0; \quad |H_2| > 0; \quad |H_3| > 0 & (d^2z \text{ positiva definida}) \end{cases} \tag{11.19}$$

Al usar esta condición, debemos evaluar todos los menores principales directores en el punto estacionario donde $f_1 = f_2 = f_3 = 0$.

Podríamos aplicar también la prueba de la raíz característica y relacionar la definición positiva (negativa) de d^2z con la positividad (negatividad) de las raíces características de la *matriz hessiana* $\begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}$. De hecho, en lugar de decir que la diferencial total de segundo orden d^2z es positiva (negativa) definida, también podemos expresar que la matriz hessiana H (para distinguirla del determinante hessiano $|H|$) es positiva (negativa) definida. Sin embargo, note que en este uso la definición de signo de H se refiere al signo de la forma cuadrática d^2z con la cual se relaciona H, *no* a los signos de los elementos de H *per se*.

Ejemplo 1

Encuentre el o los valores extremos de

$$z = 2x_1^2 + x_1x_2 + 4x_2^2 + x_1x_3 + x_3^2 + 2$$

La condición de primer orden para el extremo tiene que ver con la satisfacción simultánea de las tres ecuaciones siguientes:

$$\begin{aligned}
(f_1 =)\ 4x_1 + \ x_2 + \ x_3 &= 0 \\
(f_2 =)\ x_1 + 8x_2 \qquad &= 0 \\
(f_3 =)\ x_1 \qquad + 2x_3 &= 0
\end{aligned}$$

Debido a que se trata de un sistema lineal homogéneo, en el que las tres ecuaciones son independientes (el determinante de la matriz de coeficientes no se anula), sólo existe la solución simple $x_1^* = x_2^* = x_3^* = 0$. Esto significa que sólo hay un valor estacionario, $z^* = 2$.

El determinante hessiano de esta función es

$$|H| = \begin{vmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{vmatrix} = \begin{vmatrix} 4 & 1 & 1 \\ 1 & 8 & 0 \\ 1 & 0 & 2 \end{vmatrix}$$

cuyos menores principales directores son todos positivos:

$$|H_1| = 4 \qquad |H_2| = 31 \qquad |H_3| = 54$$

Así, mediante (11.9) podemos concluir que $z^* = 2$ es un mínimo.

Ejemplo 2

Encuentre el o los valores extremos de

$$z = -x_1^3 + 3x_1x_3 + 2x_2 - x_2^2 - 3x_3^2$$

Estas primeras derivadas parciales son

$$f_1 = -3x_1^2 + 3x_3 \qquad f_2 = 2 - 2x_2 \qquad f_3 = 3x_1 - 6x_3$$

Al igualar a cero todas las f_i, obtenemos tres ecuaciones simultáneas, una no lineal y dos lineales:

$$\begin{aligned}
-3x_1^2 \qquad + 3x_3 &= 0 \\
- 2x_2 \qquad &= -2 \\
3x_1 \qquad - 6x_3 &= 0
\end{aligned}$$

Puesto que la segunda ecuación da $x_2^* = 1$ y la tercera ecuación implica que $x_1^* = 2x_3^*$, la sustitución de éstas en la primera ecuación produce dos soluciones:

$$(x_1^*, x_2^*, x_3^*) = \begin{cases} (0, 1, 0), \text{ que implica } z^* = 1 \\ \left(\frac{1}{2}, 1, \frac{1}{4}\right), \text{ que implica } z^* = \frac{17}{16} \end{cases}$$

Las derivadas parciales de segundo orden, dispuestas de manera apropiada, producen el hessiano

$$|H| = \begin{vmatrix} -6x_1 & 0 & 3 \\ 0 & -2 & 0 \\ 3 & 0 & -6 \end{vmatrix}$$

en el que el primer elemento ($-6x_1$) se reduce a 0 bajo la primera solución (con $x_1^* = 0$) y a -3 bajo la segunda (con $x_1^* = \frac{1}{2}$). De inmediato resulta obvio que la primera solución no satisface la condición suficiente de segundo orden, puesto que $|H_1| = 0$. Sin embargo, se podría recurrir a la prueba de la raíz característica para más información. Para este fin, se aplica la ecuación característica (11.14). Puesto que la forma cuadrática que estamos probando es d^2z, cuyo discriminante es el determinante hessiano, debemos sustituir los elementos del hessiano por los elementos d_{ij} de esa ecuación. Por consiguiente, la ecuación característica es (para la primera solución)

$$\begin{vmatrix} -r & 0 & 3 \\ 0 & -2-r & 0 \\ 3 & 0 & -6-r \end{vmatrix} = 0$$

que al llevar a cabo el desarrollo se convierte en la ecuación cúbica

$$r^3 + 8r^2 + 3r - 18 = 0$$

Si empleamos el teorema I de la sección 3.3, encontraremos una raíz entera -2. La función cúbica debe ser divisible entre $(r+2)$, y podemos factorizar la ecuación cúbica y reescribir la ecuación precedente como

$$(r+2)(r^2 + 6r - 9) = 0$$

Del término $(r+2)$ se desprende con claridad que una de las raíces características es $r_1 = -2$. Las otras dos raíces se encuentran al aplicar la fórmula cuadrática al otro término; éstas son $r_2 = -3 + \frac{1}{2}\sqrt{72}$ y $r_3 = -3 - \frac{1}{2}\sqrt{72}$. En vista de que r_1 y r_3 son negativas pero r_2 es positiva, la forma cuadrática d^2z es indefinida, así que se violan las condiciones necesarias de segundo orden para una z máxima y para una z mínima. Por lo tanto, la primera solución ($z^* = 1$) no es un extremo.

En cuanto a la segunda solución, la situación es más sencilla. Puesto que los menores principales directores

$$|H_1| = -3 \qquad |H_2| = 6 \qquad \text{y} \qquad |H_3| = -18$$

alternan de signo como es debido, la prueba de los determinantes es concluyente. De acuerdo con (11.9), la solución $z^* = \frac{17}{16}$ es un máximo.

Caso de *n* variables

Cuando hay n variables de elección, la función objetivo se puede expresar como

$$z = f(x_1, x_2, \ldots, x_n)$$

La diferencial total será, entonces,

$$dz = f_1\,dx_1 + f_2\,dx_2 + \cdots + f_n\,dx_n$$

de modo que la condición necesaria para el extremo ($dz = 0$ para dx_i arbitrarias, aunque al menos una diferente de cero) significa que se requiere que las n derivadas parciales de primer orden sean cero.

La diferencial de segundo orden d^2z de nuevo será una forma cuadrática, derivable de manera análoga a (11.18) y expresable mediante un arreglo de $n \times n$. Los coeficientes de ese arreglo, dispuestos en forma apropiada, producen ahora el hessiano (simétrico)

$$|H| = \begin{vmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ f_{n1} & f_{n2} & \cdots & f_{nn} \end{vmatrix}$$

TABLA 11.2
Prueba de los determinantes para el extremo relativo: $z = f(x_1, x_2, \ldots, x_n)$

Condición	Máximo	Mínimo
Condición necesaria de primer orden	$f_1 = f_2 = \cdots = f_n = 0$	$f_1 = f_2 = \cdots = f_n = 0$
Condición suficiente de segundo orden[1]	$\|H_1\| < 0; \|H_2\| > 0;$ $\|H_3\| < 0; \ldots; (-1)^n\|H_n\| > 0$	$\|H_1\|, \|H_2\|, \ldots, \|H_n\| > 0$

[1] Aplicable sólo después de que se satisface la condición necesaria de primer orden.

con los menores principales directores $|H_1|, |H_2|, \ldots, |H_n|$, como se definió antes. La condición suficiente de segundo orden para el extremo es, como antes, que todos los menores principales sean positivos (para un mínimo en z) y que deben alternar de signo como es debido (para un mínimo en z), con el primero negativo.

En resumen, si centramos la atención en la prueba de los determinantes, tenemos los criterios como se listan en la tabla 11.2, que es válida para una función objetivo de cualquier número de variables de elección. Como casos especiales, podemos tener $n = 1$ o $n = 2$. Cuando $n = 1$, la función objetivo es $z = f(x)$, y las condiciones para maximización, $f_1 = 0$ y $|H_1| < 0$, se reducen a $f'(x) = 0$ y $f''(x) < 0$, exactamente como aprendimos en la sección 9.4. De manera similar, cuando $n = 2$, la función objetivo es $z = f(x_1, x_2)$, de modo que la condición de primer orden para el máximo es $f_1 = f_2 = 0$, mientras que la condición suficiente de segundo orden se convierte en

$$f_{11} < 0 \qquad \text{y} \qquad \begin{vmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{vmatrix} = f_{11}f_{22} - f_{12}^2 > 0$$

que es sólo un nuevo planteamiento de la información presentada en la tabla 11.1.

EJERCICIO 11.4

Encuentre los valores extremos, si los hay, de las siguientes cuatro funciones. Compruebe si son máximos o mínimos mediante la prueba de los determinantes.

1. $z = x_1^2 + 3x_2^2 - 3x_1x_2 + 4x_2x_3 + 6x_3^2$
2. $z = 29 - \left(x_1^2 + x_2^2 + x_3^2\right)$
3. $z = x_1x_3 + x_1^2 - x_2 + x_2x_3 + x_2^2 + 3x_3^2$
4. $z = e^{2x} + e^{-y} + e^{w^2} - (2x + 2e^w - y)$

Luego, conteste las siguientes preguntas en relación con matrices hessianas y sus raíces características.

5. (*a*) ¿Cuáles de los problemas del 1 al 4 producen matrices hessianas diagonales? En cada caso, ¿los elementos de la diagonal poseen un signo uniforme?
   (*b*) ¿Qué se puede concluir acerca de las raíces características de cada matriz hessiana diagonal hallada? ¿Y acerca de la definición de signo de d^2z?
   (*c*) ¿Concuerdan los resultados de la prueba de las raíces características con las de la prueba de los determinantes?
6. (*a*) Encuentre las raíces características de la matriz hessiana del problema 3.
   (*b*) ¿Qué puede concluir de sus resultados?
   (*c*) ¿Su respuesta en (*b*) es congruente con el resultado de la prueba de los determinantes del problema 3?

11.5 Condiciones de segundo orden en relación con la concavidad y la convexidad

Las condiciones de segundo orden, ya sea que se expresen en términos de los menores principales del determinante hessiano o las raíces características de la matriz hessiana, siempre tienen que ver con la cuestión de si un punto estacionario es la cúspide de una colina o el fondo de un valle. En otras palabras, se relacionan con la forma en que una curva, superficie o hipersuperficie (cualquiera que sea el caso) se curva alrededor de un punto estacionario. En el caso de una sola variable de elección, con $z = f(x)$, la configuración de colina (valle) se manifiesta en una curva en forma de U invertida (en forma de U). Para la función de dos variables $z = f(x, y)$, la configuración de colina (valle) toma la forma de una superficie con forma de domo (o de tazón), según se ilustra en la figura 11.2*a* (fig. 11.2*b*). Cuando se presentan tres o más variables de elección, las colinas y valles ya no son graficables; sin embargo, se puede considerar a las "colinas" y "valles" sobre superficies.

Una función que da lugar a una colina (valle) en todo el dominio es *cóncava* (*convexa*).[9] Para esta explicación, tomamos al dominio como R^n, donde n es el número de variables de elección. En vista de que las caracterizaciones de colinas y valles se refieren a todo el dominio, la concavidad y convexidad son, obviamente, conceptos globales. Para una clasificación más concreta podríamos también distinguir por un lado entre concavidad y convexidad, y por otro entre concavidad *estricta* y convexidad *estricta*. En el caso *no* estricto, se permite que la colina o valle contenga una o más porciones planas (en oposición a las curvadas), como segmentos de recta (sobre una curva) o segmentos planos (en una superficie). La presencia del término *estricto* descarta tales segmentos rectos o planos. Las dos superficies mostradas en la figura 11.2 ilustran funciones estrictamente cóncavas y estrictamente convexas, respectivamente. Por otro lado, la curva de la figura 6.5 es convexa (muestra un valle), pero no estrictamente convexa (contiene segmentos de recta). Una función estrictamente cóncava (estrictamente convexa) debe ser cóncava (convexa), pero lo contrario no es cierto.

En vista de la relación existente entre concavidad y concavidad estricta con una configuración de colina global, un extremo de una función cóncava debe ser una cúspide, un máximo (en oposición a un mínimo). Además, ese máximo debe ser un máximo absoluto (en oposición a un máximo relativo), puesto que la colina abarca todo el dominio. Sin embargo, tal vez ese máximo absoluto no sea único, porque podrían ocurrir varios máximos si la colina contiene una parte superior horizontal plana. La última posibilidad se puede descartar sólo cuando se especifica concavidad estricta,porque sólo entonces la cúspide constará de un solo punto y el máximo absoluto es *único*. Un máximo absoluto único (no único) se denomina también máximo absoluto *fuerte* (*débil*).

Mediante un razonamiento análogo, un extremo de una función *convexa* debe ser un mínimo absoluto (o global), el cual podría no ser único. Pero un extremo de una función *estrictamente convexa* debe ser un mínimo absoluto.

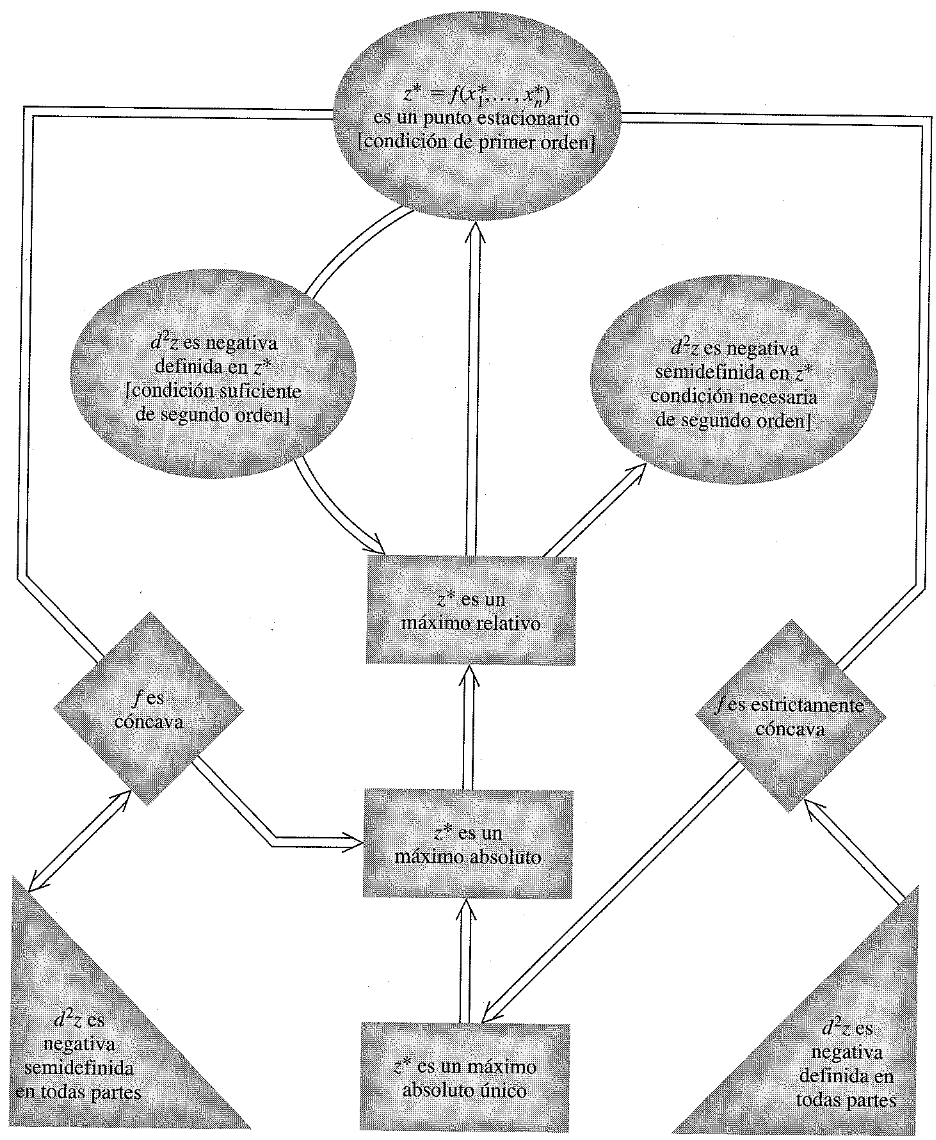
En los párrafos precedentes, las propiedades de concavidad y convexidad se toman como si fueran de ámbito global. Si son válidas sólo para una porción de la curva o superficie (sólo en un subconjunto S del dominio), entonces el máximo y el mínimo relacionados son relativos (o locales) para ese subconjunto del dominio, puesto que no se puede estar seguro de la situación que prevalece fuera del subconjunto S. En la explicación anterior de la definición de signo de d^2z (o de la matriz hessiana H), los menores principales directores del determinante hessiano los eva-

[9] Si la colina (valle) pertenece sólo a un subconjunto S del dominio, se dice que la función es *cóncava* (*convexa*) en S.

luamos sólo en el punto estacionario. Por lo tanto, al limitar la comprobación de la configuración de colina o valle a una vecindad pequeña del punto estacionario, podríamos analizar sólo los máximos y mínimos *relativos*. No obstante, podría suceder que d^2z tenga un signo definido en todas partes, sin importar dónde se evalúen los menores principales directores. En ese caso, la colina o valle deben abarcar todo el dominio, y el máximo o mínimo hallado sería de naturaleza absoluta. Para ser específicos, si d^2z es semidefinida negativa (positiva) *en todas partes*, la función $z = f(x_1, x_2, \ldots, x_n)$ debe ser cóncava (convexa), y si d^2z es definida negativa (positiva) en *todas partes*, la función f debe ser estrictamente cóncava (estrictamente convexa).

Este análisis se resume en la figura 11.5 para una función diferenciable dos veces de modo continuo $z = f(x_1, x_2, \ldots, x_n)$. Por razones de claridad, centramos la atención sólo en concavidad y máximo; sin embargo, las relaciones ilustradas seguirán siendo válidas si las palabras

FIGURA 11.5



cóncava, *negativa* y *máximo* se reemplazan, respectivamente, por *convexa*, *positiva* y *mínimo*. Para leer la figura 11.5, recuerde que el símbolo ⇒ (aquí alargado e incluso curvado) significa "implica". Cuando ese símbolo se extiende de un espacio cerrado (digamos, un rectángulo) a otro (como un óvalo), significa que el primero implica (es suficiente para) el segundo; asimismo, significa que el último es necesario al primero. Cuando el símbolo ⇒ se extiende de un espacio cerrado por un segundo a un tercero, significa que el primer espacio, cuando va acompañado por el segundo, implica el tercero.

Desde esta perspectiva, la columna media de la figura 11.5 (leyendo de arriba hacia abajo) expresa que la condición de primer orden es necesaria para que z^* sea un máximo relativo, y el carácter de máximo relativo de z^* es, a su vez, suficiente para que z^* sea un máximo absoluto, etc. Por otro lado, al leer esa columna de abajo hacia arriba, vemos que el hecho de que z^* sea un máximo absoluto único es suficiente para identificar a z^* como un máximo relativo, y así sucesivamente. Los tres óvalos de la parte superior tienen que ver con las condiciones de segundo orden en el punto estacionario z^*. Por lo que se relacionan sólo con un máximo relativo. Por otro lado, los diamantes y triángulos de la parte inferior describen propiedades globales que permiten sacar conclusiones acerca de un máximo absoluto. Tome en cuenta que si bien en la explicación anterior indicamos sólo que la semidefinición negativa de d^2z en todas partes es *suficiente* para la concavidad de la función f, hemos agregado en la figura 11.5 la información de que la condición también es *necesaria*. En cambio, la propiedad más firme de definición negativa de d^2z en todas partes es *suficiente*, pero *no necesaria*, para la concavidad estricta de f, porque la concavidad estricta de f es compatible con un valor cero de d^2z en un punto estacionario.

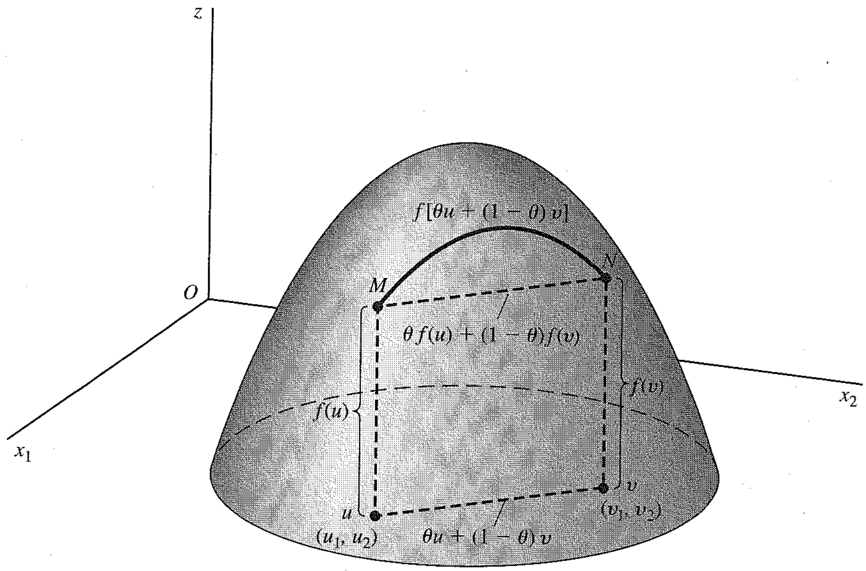
El mensaje más importante que conlleva la figura 11.5 radica en los símbolos ⇒ extendidos que pasan por los dos diamantes. El de la izquierda expresa que, dada una función objetivo *cóncava*, cualquier punto estacionario se identifica de inmediato como un máximo absoluto. Prosiguiendo, se ve que el de la derecha indica que si la función objetivo es *estrictamente* cóncava, el punto estacionario debe ser en realidad un máximo absoluto único. En cualquier caso, una vez que se satisface la condición de primer orden, la concavidad o la concavidad estricta reemplaza de manera eficaz la condición de segundo orden como una condición suficiente para un máximo, no para un máximo absoluto. La fuerza de esta nueva condición suficiente se aclara al recordar que d^2z puede ser cero en una cúspide, lo cual provoca que falle la condición suficiente de segundo orden. Sin embargo, la concavidad o concavidad estricta se encarga incluso de picos problemáticos, porque garantiza que se cumpla una condición suficiente de orden superior incluso si no se satisface la de segundo orden. Por esta razón, los economistas suelen asumir la concavidad desde el principio cuando van a formular un modelo de maximización con una función objetivo *general* (y, de manera similar, con frecuencia se supone convexidad para un modelo de minimización). Porque todo lo que se requiere hacer entonces es aplicar la condición de primer orden. Sin embargo, advierte que si se emplea una función objetivo *específica*, ya no podemos suponer simplemente la propiedad de concavidad o convexidad. En cambio, ésta se debe comprobar.

Comprobación de concavidad o convexidad

La concavidad y convexidad, estricta o no, se puede definir (y comprobar) de varias maneras. Primero se introducirá una definición geométrica de concavidad y convexidad para una función de dos variables $z = f(x_1, x_2)$, similar a la versión de una variable analizada en la sección 9.3:

> La función $z = f(x_1, x_2)$ es *cóncava* (*convexa*) si y sólo si, para cualquier par de puntos distintos M y N en su gráfica, que es una superficie, el segmento de recta MN *toca* o está *abajo* (*arriba*) de la superficie. La función es *estrictamente cóncava* (*estrictamente convexa*) si y sólo si el segmento de recta MN se ubica por completo *abajo* (*arriba*) de la superficie, excepto en M y N.

FIGURA 11.6



En la figura 11.6 se ilustra el caso de una función estrictamente cóncava, donde M y N, dos puntos arbitrarios que pertenecen a la superficie, se unen mediante un segmento de recta discontinuo así como un arco representado por una línea continua, el cual está formado por los puntos pertenecientes a la superficie que yacen directamente arriba del segmento de recta. Puesto que la concavidad estricta requiere que el segmento de recta MN esté directamente abajo del arco MN (excepto en M y N) para *cualquier* par de puntos M y N, la superficie, por lo común, debe tener forma de domo. De manera análoga, la superficie de una función estrictamente convexa normalmente debe tener forma de tazón. En cuanto a las funciones (no estrictamente) cóncavas o convexas, puesto que se permite que el segmento de recta MN sea parte de la superficie misma, cierta porción de la superficie, o incluso toda, podría ser plana en vez de curva.

Para facilitar la generalización al caso no graficable de n dimensiones, la definición geométrica requiere ser trasladada a una versión algebraica equivalente. Volviendo a la figura 11.6, sean $u = (u_1, u_2)$ y $v = (v_1, v_2)$ dos pares ordenados distintos (dos vectores) en el dominio de $z = f(x_1, x_2)$, entonces los valores de z (altura de la superficie) que les corresponden serán $f(u) = f(u_1, u_2)$ y $f(v) = f(v_1, v_2)$, respectivamente. Hemos supuesto que las variables pueden tomar valores reales, así que si u y v están en el dominio, entonces todos los puntos del segmento de recta uv están también en el dominio. Ahora cada punto de dicho segmento es de la naturaleza de un "promedio ponderado" de u y v, por lo tanto este segmento de recta podemos denotarlo mediante $\theta u + (1 - \theta)v$, donde θ (la letra griega theta), a diferencia de u y v, es un escalar (variable) con el intervalo de valores $0 \leq \theta \leq 1$.[10] De la misma manera, el segmento de recta MN, que representa el conjunto de los promedios ponderados de $f(u)$ y $f(v)$, podemos expresarlo mediante $\theta f(u) + (1 - \theta) f(v)$, con θ variando de nuevo de 0 a 1. ¿Y qué pasa con el arco MN que está a lo largo de la superficie? Puesto que el arco muestra los

[10] La expresión de promedio ponderado $\theta u + (1 - \theta)v$, para algún valor específico de θ entre 0 y 1, se conoce técnicamente como una *combinación convexa* de los dos vectores u y v. A reserva de una explicación posterior más detallada sobre esto en un punto de esta sección, se puede notar aquí que cuando $\theta = 0$, la expresión dada se reduce al vector v y de modo similar que cuando $\theta = 1$, la expresión se reduce al vector u. Un valor intermedio de θ, por otro lado, da un promedio de los dos vectores u y v.

valores de la función f evaluada en los distintos puntos del segmento de recta uv, podemos escribir simplemente como $f[\theta u + (1-\theta)v]$. Usando estas expresiones, podemos enunciar la siguiente definición algebraica:

Una función f es $\left\{ \begin{matrix} \text{cóncava} \\ \text{convexa} \end{matrix} \right\}$ si y sólo si para cualquier par de puntos distintos u y v en el dominio de f, y para $0 < \theta < 1$,

$$\underbrace{\theta f(u) + (1-\theta)f(v)}_{\text{altura del segmento de recta}} \quad \left\{ \begin{matrix} \leq \\ \geq \end{matrix} \right\} \quad \underbrace{f[\theta u + (1-\theta)v]}_{\text{altura del arco}} \qquad \textbf{(11.20)}$$

Tenga en cuenta que a fin de excluir los dos puntos finales M y N de la comparación de altura, hemos restringido a θ sólo al intervalo abierto (0, 1).

Esta definición se adapta con facilidad a la concavidad y convexidad *estrictas* al cambiar las desigualdades débiles $\leq$ y $\geq$ a las desigualdades estrictas $<$ y $>$, respectivamente. La ventaja de la definición algebraica es que se puede aplicar a una función de cualquier número de variables, porque los vectores u y v de la definición se pueden interpretar muy bien como vectores n dimensionales en lugar de vectores bidimensionales.

Los tres teoremas siguientes sobre concavidad y convexidad se pueden deducir fácilmente de (11.20). Se expresarán en términos de funciones $f(x)$ y $g(x)$, pero x se puede interpretar como un vector de variables, es decir, los teoremas son válidos para funciones de cualquier número de variables.

Teorema I (función lineal) Si $f(x)$ es una función cóncava, entonces es tanto una función cóncava como una función convexa, pero no en sentido estricto.

Teorema II (negativo de una función) Si $f(x)$ es una función cóncava, entonces $-f(x)$ es una función convexa, y viceversa. De manera similar, si $f(x)$ es una función estrictamente cóncava, entonces $-f(x)$ es una función estrictamente convexa, y viceversa.

Teorema III (suma de funciones) Si $f(x)$ y $g(x)$ son funciones cóncavas (convexas), entonces $f(x) + g(x)$ es también una función cóncava (convexa). Si $f(x)$ y $g(x)$ son cóncavas (convexas) y, además, una o ambas son estrictamente cóncavas (estrictamente convexas), entonces $f(x) + g(x)$ es estrictamente cóncava (estrictamente convexa).

El teorema I se deduce del hecho de que una función lineal se traza como una recta, plano o hiperplano, de modo que el "segmento de recta MN" coincide siempre con el "arco MN". En consecuencia, la parte de igualdad de las dos desigualdades débiles de (11.20) se satisface de forma simultánea, lo cual hace que la función califique como cóncava y convexa. Sin embargo, puesto que no se cumple la parte de desigualdad estricta de la definición, la función lineal no es ni estrictamente cóncava ni estrictamente convexa.

El fundamento del teorema II reside en el hecho de que las definiciones de concavidad y convexidad difieren sólo en el sentido de desigualdad. Suponga que $f(x)$ es cóncava; entonces,

$$\theta f(u) + (1-\theta)f(v) \leq f[\theta u + (1-\theta)v]$$

Al multiplicar todo por -1 e invertir como es debido el sentido de la desigualdad, obtenemos

$$\theta[-f(u)] + (1-\theta)[-f(v)] \geq -f[\theta u + (1-\theta)v]$$

Esto, sin embargo, es precisamente la condición para que $-f(x)$ sea convexa. Por lo tanto, el teorema se prueba para el caso de $f(x)$ cóncava. La interpretación de este resultado es muy sencilla: la imagen especular de una colina respecto al plano base o hiperplano es un valle. El caso opuesto se comprueba de manera similar.

Para constatar el fundamento del teorema III, suponga que tanto $f(x)$ como $g(x)$ son cóncavas. Entonces se cumplen las dos desigualdades siguientes:

$$\theta f(u)+(1-\theta)f(v)\le f[\theta u+(1-\theta)v] \tag{11.21}$$

$$\theta g(u)+(1-\theta)g(v)\le g[\theta u+(1-\theta)v] \tag{11.22}$$

Al sumarlas obtenemos una nueva desigualdad

$$\begin{aligned}\theta[f(u)+g(u)]&+(1-\theta)[f(v)+g(v)]\\ &\le f[\theta u+(1-\theta)v]+g[\theta u+(1-\theta)v]\end{aligned} \tag{11.23}$$

Pero ésta es precisamente la condición para que $[f(x)+g(x)]$ sea cóncava. Por lo tanto, se demuestra el teorema para el caso cóncavo. La demostración para el caso convexo es similar.

Al pasar a la segunda parte del teorema III, dejemos que $f(x)$ sea *estrictamente* cóncava. Entonces (11.21) se transforma en una desigualdad *estricta*:

$$\theta f(u)+(1-\theta)f(v)< f[\theta u+(1-\theta)v] \tag{11.21'}$$

Si esto se suma a (11.22), notamos que la suma de las expresiones del lado izquierdo en estas dos desigualdades es *estrictamente* menor que la suma de las expresiones del lado derecho, sin importar si se cumple el signo < o el signo = en (11.22). Esto significa que (11.23) se convierte ahora en una desigualdad *estricta*, también, con lo cual se logra que $[f(x)+g(x)]$ sea *estrictamente* cóncava. Además, la misma conclusión surge *a fortiori* si $g(x)$ se hace estrictamente cóncava junto con $f(x)$, es decir, si (11.22) se convierte en una desigualdad estricta junto con (11.21). Esto prueba la segunda parte del teorema para el caso cóncavo. La demostración para el caso convexo es similar.

Este teorema, que también es válido para una suma de más de dos funciones cóncavas (convexas), podría demostrar que es útil algunas veces, porque hace posible la fragmentación de la tarea de comprobar la concavidad y convexidad de una función que consta de términos aditivos. Si se advierte que los términos aditivos son cada uno cóncavos (convexos), eso sería suficiente para que la función de suma sea cóncava (convexa).

Ejemplo 1

Compruebe la concavidad o convexidad de $z=x_1^2+x_2^2$. Para aplicar (11.20), sean $u=(u_1,u_2)$ y $v=(v_1,v_2)$ dos puntos cualesquiera distintos en el dominio. Entonces, tenemos

$$f(u)=f(u_1,u_2)=u_1^2+u_2^2$$
$$f(v)=f(v_1,v_2)=v_1^2+v_2^2$$

$$\text{y}\quad f[\theta u+(1-\theta)v]=f\Big[\underbrace{\theta u_1+(1-\theta)v_1}_{\text{valor de } x_1},\ \underbrace{\theta u_2+(1-\theta)v_2}_{\text{valor de } x_2}\Big]$$
$$=[\theta u_1+(1-\theta)v_1]^2+[\theta u_2+(1-\theta)v_2]^2$$

Al sustituir estas ecuaciones en (11.20), restar la expresión del lado derecho del izquierdo y reunir términos, encontramos que su diferencia es

$$\begin{aligned}\theta(1-\theta)\left(u_1^2+u_2^2\right)+\theta(1-\theta)\left(v_1^2+v_2^2\right)&-2\theta(1-\theta)(u_1v_1+u_2v_2)\\ &=\theta(1-\theta)[(u_1-v_1)^2+(u_2-v_2)^2]\end{aligned}$$

Puesto que θ es una fracción positiva, $\theta(1-\theta)$ debe ser positivo. Además, como (u_1, u_2) y (v_1, v_2) son puntos distintos, de modo que $u_1 \neq v_1$ o $u_2 \neq v_2$ (o ambas), la expresión entre corchetes también debe ser positiva. Por lo tanto, se cumple la desigualdad estricta $>$ en (11.20), y $z = x_1^2 + x_2^2$ es estrictamente convexa.

Por otro lado, los términos x_1^2 y x_2^2 se pueden comprobar por separado. Puesto que cada uno de ellos es por sí solo estrictamente convexo, su suma es también estrictamente convexa.

Debido a que esta función es estrictamente convexa, posee un mínimo absoluto único. Es fácil comprobar que dicho mínimo es $z^* = 0$, obtenido en $x_1^* = x_2^* = 0$, y que de hecho es el único mínimo absoluto porque cualquier par ordenado $(x_1, x_2) \neq (0, 0)$ produce un valor z mayor que cero.

Ejemplo 2 Compruebe la concavidad o convexidad de $z = -x_1^2 - x_2^2$. Esta función es el negativo de la función del ejemplo 1. Así, el teorema II es estrictamente convexo.

Ejemplo 3 Compruebe la concavidad o convexidad de $z = (x+y)^2$. Aunque las variables se denotan por x y y en lugar de x_1 y x_2, aún se puede hacer que $u = (u_1, u_2)$ y $v = (v_1, v_2)$ denoten dos puntos distintos en el dominio, donde el subíndice i se refiere a la i-ésima variable. Entonces, se tiene

$$\begin{aligned} f(u) &= f(u_1, u_2) = (u_1 + u_2)^2 \\ f(v) &= f(v_1, v_2) = (v_1 + v_2)^2 \end{aligned}$$

y

$$\begin{aligned} f[\theta u + (1-\theta)v] &= [\theta u_1 + (1-\theta)v_1 + \theta u_2 + (1-\theta)v_2]^2 \\ &= [\theta(u_1 + u_2) + (1-\theta)(v_1 + v_2)]^2 \end{aligned}$$

Al sustituirlas en (11.20), restar la expresión derecha de la izquierda y simplificar, encontramos que su diferencia es

$$\begin{aligned} \theta(1-\theta)(u_1+u_2)^2 - 2\theta(1-\theta)(u_1+u_2)(v_1+v_2) + \theta(1-\theta)(v_1+v_2)^2 \\ = \theta(1-\theta)[(u_1+u_2) - (v_1+v_2)]^2 \end{aligned}$$

Como en el ejemplo 1, $\theta(1-\theta)$ es positivo. El cuadrado de la expresión que está entre corchetes es no negativo (esta vez no se puede eliminar el cero). Por lo tanto, se cumple la desigualdad $\geq$ en (11.20), y la función $(x+y)^2$ es convexa, aunque no en sentido estricto.

En consecuencia, esta función tiene un mínimo absoluto que podría no ser único. Es fácil comprobar que el mínimo absoluto es $z^* = 0$, el cual se obtiene siempre que $x^* + y^* = 0$. Que sea un mínimo absoluto resulta claro del hecho de que si $x + y \neq 0$, z será mayor que $z^* = 0$. Que no sea único se deduce del hecho de que un número infinito de pares (x^*, y^*) satisfacen la condición $x^* + y^* = 0$.

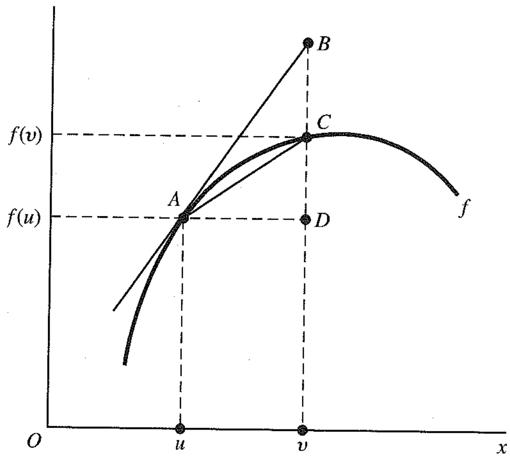
Funciones diferenciables

Como se estableció en (11.20), la definición de concavidad y convexidad no emplea derivadas, por lo tanto no requiere diferenciabilidad. Sin embargo, si la función es diferenciable, la concavidad y la convexidad también se pueden definir en términos de sus primeras derivadas. En el caso de una variable, la definición es:

Una función diferenciable $f(x)$ es $\left\{\begin{matrix}\text{cóncava}\\ \text{convexa}\end{matrix}\right\}$ si y sólo si para algún punto dado u y algún otro punto v en el dominio,

$$f(v) \quad \left\{\begin{matrix}\leq\\ \geq\end{matrix}\right\} \quad f(u) + f'(u)(v-u) \tag{11.24}$$

FIGURA 11.7



La concavidad y la convexidad serán *estrictas* si las desigualdades débiles de (11.24) se reemplazan por las desigualdades *estrictas* < y >, respectivamente. Interpretada desde el punto de vista geométrico, esta definición ilustra una curva cóncava (convexa) como una que queda abajo (arriba) o tiene puntos de coincidencia con todas su rectas tangentes. Para calificar como una curva estrictamente cóncava, ésta debe quedar estrictamente abajo (arriba) de las rectas tangentes, excepto en los puntos de tangencia.

En la figura 11.7, sea el punto A cualquier punto dado sobre la curva, con altura $f(u)$ y con recta tangente AB. Dejemos que x se incremente desde el valor u. Entonces una curva estrictamente cóncava (como está trazada), a fin de formar una colina, debe curvarse en forma progresiva lejos de la recta tangente AB, de modo que el punto C, con altura $f(v)$, se ubique debajo del punto B. En este caso, la pendiente del segmento de recta AC es menor que la de la tangente AB. Por otro lado, si la curva *no es* estrictamente cóncava, podría contener un segmento de recta, de tal manera que, por ejemplo, el arco AC podría volverse un segmento de recta y coincidir con el segmento de recta AB, como una porción lineal de la curva. En el último caso, la pendiente de AC es igual a la de AB. Juntas, estas dos situaciones implican que

$$\left(\text{Pendiente del segmento } AC = \frac{DC}{AD} =\right) \frac{f(v)-f(u)}{v-u} \leq (\text{pendiente de } AB =)\ f'(u)$$

Cuando se multiplica por la cantidad positiva $(v-u)$, esta desigualdad produce el resultado de (11.24) para la función cóncava. Si se consideran valores de x menores que u, obtenemos el mismo resultado.

Cuando hay dos o más variables independientes, la definición requiere una modificación:

Una función diferenciable $f(x) = f(x_1, \ldots, x_n)$ es $\begin{Bmatrix} \text{cóncava} \\ \text{convexa} \end{Bmatrix}$ si y sólo si para algún punto dado $u = (u_1, \ldots, u_n)$ y algún otro punto $v = (v_1, \ldots, v_n)$ en el dominio,

$$f(v) \quad \begin{Bmatrix} \leq \\ \geq \end{Bmatrix} \quad f(u) + \sum_{j=1}^{n} f_j(u)(v_j - u_j) \qquad \textbf{(11.24')}$$

donde $f_j(u) \equiv \partial f/\partial x_j$ se evalúa en $u = (u_1, \ldots, u_n)$.

Esta definición requiere que la gráfica de una función cóncava (convexa) $f(x)$ se ubique en o debajo (arriba) de todos sus planos o hiperplanos tangentes. Para la concavidad y convexidad *estrictas*, las desigualdades débiles de (11.24′) se deben cambiar a desigualdades *estrictas*, lo cual requeriría que la gráfica de una función estrictamente cóncava (o convexa) estuviera precisamente abajo (arriba) de todos sus planos o hiperplanos tangentes, excepto en los puntos de tangencia.

Por último, considere una función $z = f(x_1, \ldots, x_n)$ que es dos veces diferenciable en forma continua. Para esta función existen derivadas parciales de segundo orden, por lo que d^2z está definida. La concavidad y la convexidad se pueden comprobar entonces mediante el signo de d^2z:

Una función diferenciable en forma continua dos veces $z = f(x_1, \ldots, x_n)$ es $\left\{\begin{matrix}\text{cóncava}\\ \text{convexa}\end{matrix}\right\}$ si y sólo si d^2z es semidefinita $\left\{\begin{matrix}\text{negativa}\\ \text{positiva}\end{matrix}\right\}$ en todas partes. Dicha función es estrictamente $\left\{\begin{matrix}\text{cóncava}\\ \text{convexa}\end{matrix}\right\}$ si (pero *no* sólo si) d^2z es definida $\left\{\begin{matrix}\text{negativa}\\ \text{positiva}\end{matrix}\right\}$ en todas partes. **(11.25)**

Recuerde que los aspectos de concavidad y concavidad estricta de (11.25) ya se incorporaron en la figura 11.5.

Ejemplo 4

Compruebe la concavidad o convexidad de $z = -x^4$ mediante condiciones de derivadas. Primero aplique (11.24). Las expresiones izquierda y derecha en esa desigualdad son en este caso $-v^4$ y $-u^4 - 4u^3(v-u)$, respectivamente. Al restar el último del primero, advierta que su diferencia es

$$\begin{aligned} -v^4 + u^4 + 4u^3(v-u) &= (v-u)\left(-\frac{v^4-u^4}{v-u} + 4u^3\right) && \text{[factorizando]} \\ &= (v-u)[-(v^3+v^2u+vu^2+u^3)+4u^3] && \text{[por (7.2)]} \end{aligned}$$

sería bueno si la expresión que está entre corchetes resultara divisible entre $(v-u)$, porque entonces se podría factorizar $(v-u)$ y obtener un término cuadrado $(v-u)^2$ para facilitar la evaluación de signo. De acuerdo con el resultado, de hecho éste es el caso. Así, la ecuación en diferencias anterior se puede reescribir como

$$-(v-u)^2[v^2+2vu+3u^2] = -(v-u)^2[(v+u)^2+2u^2]$$

Dado que $v \neq u$, el signo de esta expresión debe ser negativo. Si en (11.24) se cumple la desigualdad estricta $<$, la función $z = -x^4$ es estrictamente cóncava. Esto significa que tiene un máximo absoluto único. Como se comprueba fácilmente, ese máximo es $z^* = 0$, que se obtiene en $x^* = 0$.

Debido a que esta función es diferenciable en forma continua dos veces, se podría aplicar también (11.25). Puesto que sólo hay una variable, (11.25) produce

$$d^2z = f''(x)\,dx^2 = -12x^2\,dx^2 \qquad \text{[mediante (11.2)]}$$

Se sabe que dx^2 es positiva (sólo se consideran cambios no cero en x), pero $-12x^2$ puede ser negativo o cero. Así, lo mejor que podemos hacer es concluir que d^2z es *semi*definida negativa en todas partes, y que $z = -x^4$ es cóncava (no en sentido estricto). Esta conclusión de (11.25) es obviamente más débil que la obtenida antes a partir de (11.24), a saber, $z = -x^4$. Lo que en este caso limita a la conclusión más débil es el mismo culpable que hace que en ocasiones falle la prueba de la segunda derivada: el hecho de que d^2z pueda tomar un valor cero en un punto estacionario de una función que se sabe que es estrictamente cóncava, o estrictamente convexa. Ésta es la razón de que la definición negativa (positiva) de d^2z se presente en (11.25) sólo como una condición suficiente, pero no necesaria, para concavidad estricta (convexidad estricta).

Ejemplo 5 Compruebe la concavidad o convexidad de $z = x_1^2 + x_2^2$ mediante condiciones de derivada. Esta vez tenemos que usar (11.24′) en lugar de (11.24). Con $u = (u_1, u_2)$ y $v = (v_1, v_2)$ como dos puntos cualesquiera en el dominio, los dos lados de (11.24′) son

$$\text{Lado izquierdo} = v_1^2 + v_2^2$$
$$\text{Lado derecho} = u_1^2 + u_2^2 + 2u_1(v_1 - u_1) + 2u_2(v_2 - u_2)$$

Al restar la última expresión de la primera, y simplificar, podemos expresar su diferencia como

$$v_1^2 - 2v_1u_1 + u_1^2 + v_2^2 - 2v_2u_2 + u_2^2 = (v_1 - u_1)^2 + (v_2 - u_2)^2$$

Dado que $(v_1, v_2) \neq (u_1, u_2)$, esta diferencia siempre es positiva. Así, se cumple la desigualdad estricta $>$ de (11.24′), y $z = x_1^2 + x_2^2$ es estrictamente convexa. Toma en cuenta que el presente resultado sólo reafirma lo que encontramos antes en el ejemplo 1.

En cuanto al uso de (11.25), puesto que $f_1 = 2x_1$ y $f_2 = 2x_2$, tenemos

$$f_{11} = 2 > 0 \qquad \text{y} \qquad \begin{vmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{vmatrix} = \begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix} = 4 > 0$$

sin importar dónde se evalúen las derivadas parciales de segundo orden. Por lo tanto, d^2z es definida positiva en todas partes, lo cual satisface la condición suficiente para la convexidad estricta. En el caso presente, (11.24′) y (11.25) llevan a la misma conclusión.

Funciones convexas contra conjuntos convexos

Una vez aclarado el significado del adjetivo *convexo* aplicado a una función, debemos explicar su significado cuando se emplea para describir un *conjunto*. Aunque los conjuntos convexos y las funciones convexas tienen relación, son conceptos distintos, y es importante no confundirlos.

Para lograr una comprensión intuitiva más fácil, comenzaremos con la caracterización geométrica de un conjunto convexo. Sea S un conjunto de puntos en un espacio de dos o tres dimensiones. Para dos puntos cualesquiera del conjunto S, si el segmento de recta que los une queda por completo en S, se dice entonces que S es un *conjunto convexo*. Debe ser evidente que una recta cumple esta definición y constituye un conjunto convexo. Por convención, un conjunto que consta de un solo punto se considera convexo y también se considera convexo al conjunto vacío (sin ningún punto). En la figura 11.8 se ilustran más ejemplos. El disco, es decir, el círculo "sólido", un círculo más todos los puntos que hay dentro de él, es un conjunto convexo, porque un segmento de recta que une dos puntos cualesquiera del disco queda por completo contenido en él, como se ejemplifica por ab (que une dos puntos límite) y cd (que une dos puntos internos). Sin embargo, tome en cuenta que un círculo (hueco) *no* es un conjunto convexo. De manera similar, un triángulo, o un pentágono, no es en sí mismo un conjun-

FIGURA 11.8

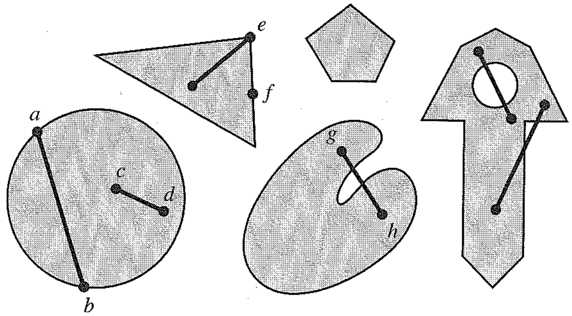
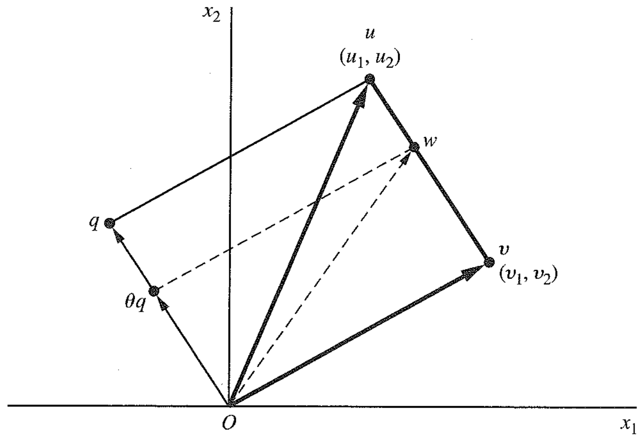


FIGURA 11.9



to convexo, pero sí el área que contiene. Las dos figuras sólidas restantes de la figura 11.8 no son conjuntos convexos. La figura en forma de paleta tiene un hundimiento así, un segmento de recta como *gh* no queda por completo en el conjunto. La figura de forma de llave no es un conjunto convexo por dos causas; la característica peculiar de su contorno, y la presencia de un hueco. En términos generales, para calificar como un conjunto convexo, el conjunto de puntos no debe contener huecos, y su límite no debe estar mordido en ningún lado.

La definición geométrica de convexidad se aplica también con facilidad a conjuntos de puntos que están en un espacio tridimensional. Por ejemplo, un cubo sólido es un conjunto convexo, mientras que un cilindro hueco no lo es. Cuando se tiene un espacio de cuatro dimensiones o más, la interpretación geométrica se vuelve menos obvia. Entonces, es necesario ocuparse de la definición algebraica de conjuntos convexos.

Para este fin, es útil introducir el concepto de *combinación convexa* de vectores (puntos), que es un tipo especial de combinación lineal. Una combinación lineal de dos vectores u y v se puede escribir como

$$k_1 u + k_2 v$$

donde k_1 y k_2 son dos escalares. Cuando estos dos escalares están en el intervalo cerrado [0, 1] y su suma es la unidad, se dice que la combinación lineal es una combinación convexa, y se puede expresar como

$$\theta u + (1-\theta)v \qquad (0 \leq \theta \leq 1) \tag{11.26}$$

Como una ilustración, la combinación $\frac{1}{3}\begin{bmatrix} 2 \\ 0 \end{bmatrix} + \frac{2}{3}\begin{bmatrix} 4 \\ 9 \end{bmatrix}$ es una combinación lineal. En vista de que estos dos multiplicadores escalares son fracciones positivas cuya suma es igual a la unidad, tal combinación convexa se puede interpretar como un *promedio ponderado* de los dos vectores.[11]

La única característica de la combinación de (11.26) es que, para cada valor aceptable de θ, el vector suma resultante queda en el segmento de recta que une los puntos u y v. Lo anterior se puede demostrar por medio de la figura 11.9, donde se han graficado dos vectores $u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ y $v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ como dos puntos con coordenadas (u_1, u_2) y (v_1, v_2), respectivamente.

[11] Esta interpretación se usó antes en la explicación de las funciones cóncavas y convexas.

Si trazamos otro vector q tal que $Oquv$ forme un paralelogramo, entonces tenemos (en virtud de la explicación en la figura 4.3)

$$u = q + v \qquad \text{o} \qquad q = u - v$$

Se deduce que una combinación convexa de vectores u y v (llámesele w) se puede expresar en términos del vector q, porque

$$w = \theta u + (1-\theta)v = \theta u + v - \theta v = \theta(u - v) + v = \theta q + v$$

Por consiguiente, para graficar el vector w, se suma simplemente θq y v mediante el conocido método del paralelogramo. Si el escalar θ es una fracción positiva, el vector θq será solamente una contracción del vector q; así, θq debe estar en el segmento de recta Oq. Por lo tanto, al sumar θq y v se debe encontrar que el vector w queda sobre el segmento de recta uv, porque el nuevo paralelogramo más pequeño no es sino el paralelogramo original con el lado qu desplazado hacia abajo. La ubicación exacta del vector w, por supuesto, variará de acuerdo con el escalar θ; al variar θ de cero a la unidad, la ubicación de w se desplaza de v a u. Así que el conjunto de los puntos que están sobre el segmento de recta uv, incluidos u y v, corresponde al conjunto de combinaciones convexas de los vectores u y v.

En vista de lo anterior, un conjunto convexo se podría redefinir como: un conjunto S es convexo si y sólo si para dos puntos cualesquiera $u \in S$ y $v \in S$, y para todo escalar $\theta \in [0, 1]$, se cumple que $w = \theta u + (1-\theta)v \in S$. Como esta definición es algebraica, es aplicable sin importar la dimensión del espacio en el cual se localizan los vectores u y v. Al comparar esta definición de un conjunto convexo con la de una función convexa de (11.20), vemos que, aunque se use en ambas el mismo adjetivo *convexo*, su significado cambió en forma radical de un contexto a otro. Al describir una *función*, el término *convexa* especifica cómo se dobla por sí misma una curva o superficie: debe formar un valle. Pero al describir un *conjunto*, el término especifica cómo se agrupan los puntos en el conjunto: no deben permitir que aparezcan huecos, y la frontera del conjunto no debe estar mordida. Así, las funciones convexas y los conjuntos convexos son claramente entidades matemáticas distintas.

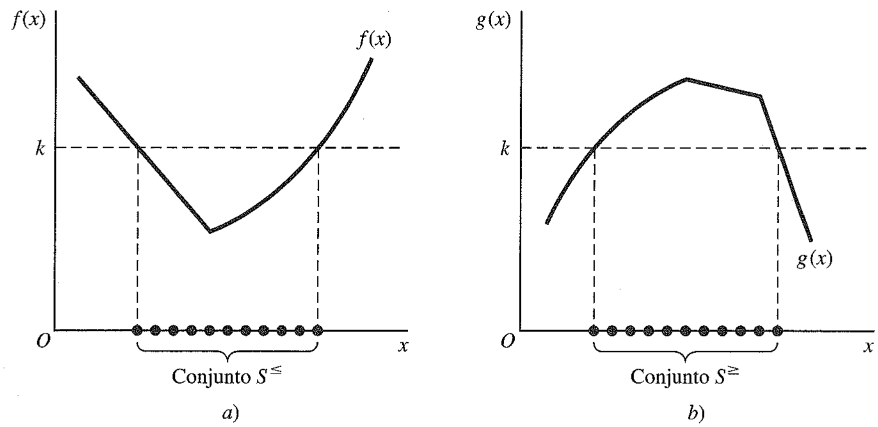
Sin embargo, las funciones convexas y los conjuntos convexos tienen relación. Entre otras cosas, al definir una función convexa, se necesita un conjunto convexo para el dominio. Esto se debe a que la definición (11.20) requiere que, para dos puntos cualesquiera u y v en el dominio, todas las combinaciones convexas de u y v, en particular $\theta u + (1-\theta)v$, $0 \leq \theta \leq 1$, también deben estar en el dominio, lo cual es otra forma de decir que el dominio debe ser un conjunto convexo. Para cumplir este requerimiento, adoptamos antes la suposición bastante firme de que el dominio consta de todo el espacio n (donde n es el número de variables de elección), que es de hecho un conjunto convexo. Sin embargo, ahora que ya contamos con el concepto de conjuntos convexos, podemos debilitar en forma sustancial esa suposición. Todo lo que necesitamos es suponer que el dominio es un subconjunto convexo de R^n, en vez de R^n mismo.

Hay otra forma en la que las funciones convexas se relacionan con los conjuntos convexos. Si $f(x)$ es una función convexa, entonces para cualquier constante k, da lugar a un conjunto convexo

$$S^{\leq} \equiv \{x \mid f(x) \leq k\} \qquad [f(x) \text{ convexa}] \qquad \textbf{(11.27)}$$

Eso se ilustra en la figura 11.10a para el caso de una variable. El conjunto $S^{\leq}$ consta de todos los valores de x relacionados con el segmento de la curva $f(x)$ que queda en o debajo de la línea horizontal discontinua. Por consiguiente, es el segmento de recta que está sobre el eje horizontal marcado por los puntos sólidos, el cual es un conjunto convexo. Tome en cuenta

FIGURA 11.10



que si se cambia el valor k, el conjunto $S^{\leq}$ se convertirá en un segmento de recta diferente sobre el eje horizontal, pero aún será un conjunto convexo.

Al ir un poco más allá, se puede observar que incluso una función *cóncava* se relaciona con conjuntos convexos en formas similares. Primero, la definición de una función cóncava de (11.20), como en el caso de la función convexa, se basa en un dominio que es un conjunto convexo. Además, incluso una función cóncava —por ejemplo $g(x)$— genera un conjunto convexo relacionado, dada alguna constante k. Ese conjunto convexo es

$$S^{\geq} \equiv \{x \mid g(x) \geq k\} \qquad [g(x) \text{ cóncava}] \tag{11.28}$$

en la que el signo $\geq$ aparece en lugar de $\leq$. Desde el punto de vista geométrico, como se ilustra en la figura 11.10*b* para el caso de una variable, el conjunto $S^{\geq}$ contiene todos los valores de x que corresponden al segmento de la curva $g(x)$ que queda en o arriba de la recta horizontal discontinua. Así, de nuevo se trata de un segmento de recta sobre el eje horizontal, un conjunto convexo.

Aunque en la figura 11.10 se ilustra en particular el caso de una variable, las definiciones de $S^{\leq}$ y $S^{\geq}$ en (11.27) y (11.28) no se limitan a funciones de una sola variable. Son igualmente válidas si se interpreta x como un vector, es decir, sea $x = (x_1, \ldots, x_n)$. En ese caso, sin embargo, (11.27) y (11.28) definirán conjuntos convexos en el espacio de n dimensiones. Es importante recordar que si bien una función convexa implica (11.27), y una función cóncava implica (11.28), lo contrario no es cierto porque (11.27) se puede satisfacer también mediante una función no convexa y (11.28) mediante una función no cóncava. Esto se analiza con más detalle en la sección 12.4.

EJERCICIO 11.5

1. Utilice (11.20) para comprobar si las siguientes funciones son cóncavas, convexas, estrictamente cóncavas, estrictamente convexas o ninguna:
   (*a*) $z = x^2$ (*b*) $z = x_1^2 + 2x_2^2$ (*c*) $z = 2x^2 - xy + y^2$
2. Use (11.24) u (11.24′) para comprobar si las siguientes funciones son cóncavas, convexas, estrictamente cóncavas, estrictamente convexas o ninguna:
   (*a*) $z = -x^2$ (*b*) $z = (x_1 + x_2)^2$ (*c*) $z = -xy$
3. De acuerdo con su respuesta al problema 2*c*, ¿se podría hacer uso del teorema III de esta sección para dividir la tarea de comprobar la función $z = 2x^2 - xy + y^2$ en el problema 1*c*? Explique su respuesta.

4. ¿Constituyen las formas siguientes conjuntos convexos en el espacio tridimensional?
   (*a*) Una dona (*b*) Un pino de boliche (*c*) Una canica perfecta
5. La ecuación $x^2 + y^2 = 4$ representa un círculo con centro en (0, 0) y con radio de 2.
   (*a*) Interprete geométricamente el conjunto $\{(x, y) \mid x^2 + y^2 \leq 4\}$.
   (*b*) ¿Es convexo el conjunto?
6. Grafique cada uno de los siguientes conjuntos e indique si es convexo:
   (*a*) $\{(x, y) \mid y = e^x\}$ (*c*) $\{(x, y) \mid y \leq 13 - x^2\}$
   (*b*) $\{(x, y) \mid y \geq e^x\}$ (*d*) $\{(x, y) \mid xy \geq 1; x > 0, y > 0\}$
7. Dado $u = \begin{bmatrix} 10 \\ 6 \end{bmatrix}$ y $v = \begin{bmatrix} 4 \\ 8 \end{bmatrix}$, ¿cuáles de las siguientes son combinaciones *convexas* de u y v?
   (*a*) $\begin{bmatrix} 7 \\ 7 \end{bmatrix}$ (*b*) $\begin{bmatrix} 5.2 \\ 7.6 \end{bmatrix}$ (*c*) $\begin{bmatrix} 6.2 \\ 8.2 \end{bmatrix}$
8. Dados dos vectores u y v en el espacio bidimensional, determine y bosqueje:
   (*a*) El conjunto de las combinaciones lineales de u y v.
   (*b*) El conjunto de las combinaciones lineales no negativas de u y v.
   (*c*) El conjunto de las combinaciones convexas de u y v.
9. (*a*) Reescriba (11.27) y (11.28) en particular para los casos donde las funciones f y g tienen n variables independientes.
   (*b*) Sea $n = 2$, y sea que la función f tiene forma de cono de helado (sostenido verticalmente) mientras que la función g tiene forma de pirámide. Describa los conjuntos $S^{\leq}$ y $S^{\geq}$.

11.6 Aplicaciones económicas

Al comienzo de este capítulo, expusimos el caso de una compañía multiproducto como ejemplo del problema general de optimización con más de una variable de elección. Ahora ya contamos con las herramientas para manejar ese problema y otros de naturaleza similar.

Problema de una empresa multiproducto

Ejemplo 1

Analizaremos primero una empresa de dos productos bajo circunstancias de competencia pura. Puesto que con competencia pura los precios de ambos artículos se deben tomar como exógenos, éstos se identificarán por medio de P_{10} y P_{20}, respectivamente. En consecuencia, la función de ingreso de la empresa será,

$$R_1 = P_{10}Q_1 + P_{20}Q_2$$

donde Q_i representa el nivel de producción del i-ésimo producto por unidad de tiempo. Se supone que la función de costo de la empresa es

$$C = 2Q_1^2 + Q_1Q_2 + 2Q_2^2$$

Tome en cuenta que $\partial C/\partial Q_1 = 4Q_1 + Q_2$ (el costo marginal del primer producto) es una función no sólo de Q_1 sino también de Q_2. De manera similar, el costo marginal del segundo producto también depende, en parte, del nivel de producción del primer producto. Así, de acuerdo con la función de costo supuesta, se ve que los dos artículos están relacionados técnicamente en la producción.

La función de ganancia de esta empresa hipotética se puede escribir fácilmente como

$$\pi = R - C = P_{10}Q_1 + P_{20}Q_2 - 2Q_1^2 - Q_1Q_2 - 2Q_2^2$$

una función de dos variables de elección (Q_1 y Q_2) y dos parámetros de precio. La tarea es hallar los niveles de (Q_1 y Q_2) que, juntos, maximizarán π. Para este fin, hallamos primero las derivadas parciales de primer orden de la función de ganancia:

$$\begin{aligned} \pi_1 \left(\equiv \frac{\partial \pi}{\partial Q_1} \right) &= P_{10} - 4Q_1 - Q_2 \\ \pi_2 \left(\equiv \frac{\partial \pi}{\partial Q_2} \right) &= P_{20} - Q_1 - 4Q_2 \end{aligned} \tag{11.29}$$

Si igualamos ambas expresiones a cero, a fin de satisfacer la condición necesaria para un máximo, obtenemos dos ecuaciones simultáneas

$$\begin{aligned} 4Q_1 + Q_2 &= P_{10} \\ Q_1 + 4Q_2 &= P_{20} \end{aligned}$$

que producen la solución única

$$Q_1^* = \frac{4P_{10} - P_{20}}{15} \qquad \text{y} \qquad Q_2^* = \frac{4P_{20} - P_{10}}{15}$$

De esta forma, $P_{10} = 12$ y $P_{20} = 18$, por ejemplo, tenemos $Q_1^* = 2$ y $Q_2^* = 4$, lo que implica una ganancia óptima $\pi^* = 48$ por unidad de tiempo.

Para tener la seguridad de que esto representa una ganancia máxima, comprobamos la condición de segundo orden. Las segundas derivadas parciales, las cuales se obtienen mediante diferenciación parcial de (11.29), nos dan el siguiente hessiano:

$$|H| = \begin{vmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{vmatrix} = \begin{vmatrix} -4 & -1 \\ -1 & -4 \end{vmatrix}$$

Puesto que $|H_1| = -4 < 0$ y $|H_2| = 15 > 0$, la matriz hessiana (o d^2z) es definida negativa, y la solución maximiza la ganancia. De hecho, puesto que los signos de los menores principales directores no dependen de dónde se evalúen, d^2z en este caso es definida negativa en *todas partes*. Por consiguiente, según (11.25), la función objetivo debe ser estrictamente cóncava, y la ganancia máxima encontrada es en realidad un máximo absoluto único.

Ejemplo 2

Traslademos el problema del ejemplo 1 al entorno de un mercado monopolista. Con base en esta nueva suposición de estructura de mercado, la función de ingreso se debe modificar para reflejar el hecho de que los precios de los dos productos ahora variarán en función de sus niveles de producción (se supone que toda la producción se vende, no hay acumulación de inventario). La manera exacta en que variarán los precios en función de los niveles de producción se encontrará en las funciones de demanda para los dos productos de la empresa.

Suponga que las demandas que enfrenta la empresa monopolista son:

$$\begin{aligned} Q_1 &= 40 - 2P_1 + P_2 \\ Q_2 &= 15 + \ \ P_1 - P_2 \end{aligned} \tag{11.30}$$

Estas ecuaciones revelan que los dos artículos se relacionan en *consumo*; en particular, son bienes sustitutos, porque un incremento en el precio de uno aumentará la demanda del otro. De acuerdo con (11.30), ahí se expresan las cantidades demandadas de Q_1 y Q_2 como funciones de precios, pero para los fines presentes será más conveniente tener los precios P_1 y P_2 expresados en términos de los volúmenes de ventas Q_1 y Q_2, es decir, tener funciones de ingreso promedio para los dos productos. Puesto que (11.30) podemos reescribirla como

$$\begin{aligned} -2P_1 + P_2 &= Q_1 - 40 \\ P_1 - P_2 &= Q_2 - 15 \end{aligned}$$

podemos (considerando Q_1 y Q_2 como parámetros) aplicar la regla de Cramer para determinar P_1 y P_2 de la siguiente forma:

$$\begin{aligned} P_1 &= 55 - Q_1 - \ \ Q_2 \\ P_2 &= 70 - Q_1 - 2Q_2 \end{aligned} \qquad \textbf{(11.30')}$$

Éstas constituyen las funciones de ingreso promedio deseadas, puesto que $P_1 \equiv \text{AR}_1$ y $P_2 \equiv \text{AR}_2$.*

En consecuencia, la función de ingreso total de la empresa se puede escribir como

$$\begin{aligned} R &= P_1Q_1 + P_2Q_2 \\ &= (55 - Q_1 - Q_2)Q_1 + (70 - Q_1 - 2Q_2)Q_2 \qquad \text{[por (11.30')]} \\ &= 55Q_1 + 70Q_2 - 2Q_1Q_2 - Q_1^2 - 2Q_2^2 \end{aligned}$$

Si de nuevo se supone que la función de costo total es

$$C = Q_1^2 + Q_1Q_2 + Q_2^2$$

entonces la función de ganancia será

$$\pi = R - C = 55Q_1 + 70Q_2 - 3Q_1Q_2 - 2Q_1^2 - 3Q_2^2 \qquad \textbf{(11.31)}$$

la cual es una función objetivo con dos variables de elección. Una vez que determinamos los niveles de producción de maximización de ganancia Q_1^* y Q_2^*, es bastante fácil hallar los precios óptimos P_1^* y P_2^* a partir de (11.30').

La función objetivo produce las siguientes derivadas parciales primera y segunda:

$$\pi_1 = 55 - 3Q_2 - 4Q_1 \qquad \pi_2 = 70 - 3Q_1 - 6Q_2$$
$$\pi_{11} = -4 \qquad \pi_{12} = \pi_{21} = -3 \qquad \pi_{22} = -6$$

A fin de satisfacer la condición de primer orden para un máximo de π, se debe tener $\pi_1 = \pi_2 = 0$; es decir,

$$\begin{aligned} 4Q_1 + 3Q_2 &= 55 \\ 3Q_1 + 6Q_2 &= 70 \end{aligned}$$

De esta manera, los niveles de producción óptimos (por unidad de tiempo) son

$$(Q_1^*, Q_2^*) = \left(8, 7\tfrac{2}{3}\right)$$

Al sustituir este resultado en (11.30') y (11.31), respectivamente, encontramos que

$$P_1^* = 39\tfrac{1}{3} \qquad P_2^* = 46\tfrac{2}{3} \qquad \text{y} \qquad \pi^* = 488\tfrac{1}{3} \qquad \text{(por unidad de tiempo)}$$

En vista de que el hessiano es $\begin{vmatrix} -4 & -3 \\ -3 & -6 \end{vmatrix}$, tenemos

$$|H_1| = -4 < 0 \qquad \text{y} \qquad |H_2| = 15 > 0$$

de modo que el valor de π^* representa la ganancia máxima. Aquí, los signos de los menores principales directores son de nuevo independientes de dónde se evalúen. Por lo tanto, la matriz hessiana es definida negativa en todas partes, lo cual significa que la función objetivo es estrictamente cóncava y que tiene un máximo absoluto único.

Discriminación de precio

Incluso en una empresa de un solo producto puede surgir un problema de optimización en el que intervienen dos o más variables de elección. Un caso sería, por ejemplo, cuando una empresa monopolista vende un solo producto en dos o más mercados separados (por ejemplo, doméstico

* Nota de la revisión técnica de la versión en español: AR son las iniciales de "Average-Revenue" en inglés que significa ingreso promedio.

y extranjero) y, por lo tanto, debe decidir sobre las cantidades (Q_1, Q_2, etc.) que se proveerán a los mercados respectivos a fin de maximizar la ganancia. Los distintos mercados, en general, tendrán condiciones de demanda diferentes, y si las elasticidades de demanda difieren en los diversos mercados, la maximización de ganancia estará vinculada con la práctica de discriminación de precios. A continuación procedemos a deducir matemáticamente esta conclusión ya conocida.

Ejemplo 3

Para un cambio de ritmo, esta vez usaremos tres variables de elección, es decir, tres mercados separados; asimismo, trabajaremos con funciones generales en vez de numéricas. En consecuencia, supondremos que la empresa monopolista tiene las siguientes funciones de costo total y de ingreso total:

$$R = R_1(Q_1) + R_2(Q_2) + R_3(Q_3)$$
$$C = C(Q) \qquad \text{donde} \qquad Q = Q_1 + Q_2 + Q_3$$

Tome en cuenta que el símbolo R_i representa aquí la función de ingreso del i-ésimo mercado, en vez de una derivada en el sentido de f_i. Cada función de ingreso implica una estructura de demanda particular, la cual suele ser diferente de las que prevalecen en los otros dos mercados. Por otra parte, del lado del costo, sólo se postula la función de costo, puesto que una sola empresa está produciendo para los tres mercados. En vista del hecho de que $Q = Q_1 + Q_2 + Q_3$, el costo total C es también básicamente una función de Q_1, Q_2 y Q_3, que constituyen las variables de elección del modelo. También se puede reescribir $C(Q)$ como $C(Q_1 + Q_2 + Q_3)$. Sin embargo, se debe notar que aun cuando la última versión contiene tres variables independientes, se debe considerar que la función tiene un solo argumento, porque la suma de Q_i es en realidad una sola entidad. Por el contrario, si la función aparece en la forma $C(Q_1, Q_2, Q_3)$, entonces se pueden contar ahí tantos argumentos como variables independientes.

Ahora, la función de ganancia es

$$\pi = R_1(Q_1) + R_2(Q_2) + R_3(Q_3) - C(Q)$$

con primeras derivadas parciales $\pi_i \equiv \partial\pi/\partial Q_i$ (para $i = 1, 2, 3$) como sigue:[12]

$$\begin{aligned}
\pi_1 &= R_1'(Q_1) - C'(Q)\frac{\partial Q}{\partial Q_1} = R_1'(Q_1) - C'(Q) \quad \left[\text{puesto que } \frac{\partial Q}{\partial Q_1} = 1\right] \\
\pi_2 &= R_2'(Q_2) - C'(Q)\frac{\partial Q}{\partial Q_2} = R_2'(Q_2) - C'(Q) \quad \left[\text{puesto que } \frac{\partial Q}{\partial Q_2} = 1\right] \\
\pi_3 &= R_3'(Q_3) - C'(Q)\frac{\partial Q}{\partial Q_3} = R_3'(Q_3) - C'(Q) \quad \left[\text{puesto que } \frac{\partial Q}{\partial Q_3} = 1\right]
\end{aligned} \qquad \textbf{(11.32)}$$

Al igualar a cero estas ecuaciones en forma simultánea, obtenemos

$$C'(Q) = R_1'(Q_1) = R_2'(Q_2) = R_3'(Q_3)$$

es decir,

$$\text{MC} = \text{MR}_1 = \text{MR}_2 = \text{MR}_3$$

Así que los niveles de Q_1, Q_2 y Q_3 se deben elegir de tal manera que el ingreso marginal de cada mercado se iguale al costo marginal de la producción total Q.

[12] Tome en cuenta que, para hallar $\partial C/\partial Q_i$, se usa la regla de la cadena:

$$\frac{\partial C}{\partial Q_i} = \frac{dC}{dQ}\frac{\partial Q}{\partial Q_i}$$

Para ver las implicaciones de esta condición en relación con la discriminación de precio, investigamos primero cómo el MR de cualquier mercado se relaciona en particular con el precio en ese mercado. Puesto que el ingreso en cada mercado es $R_i = P_i Q_i$, deducimos que el ingreso marginal debe ser

$$\begin{aligned} \text{MR}_i \equiv \frac{dR_i}{dQ_i} &= P_i \frac{dQ_i}{dQ_i} + Q_i \frac{dP_i}{dQ_i} \\ &= P_i\left(1 + \frac{dP_i}{dQ_i}\frac{Q_i}{P_i}\right) = P_i\left(1 + \frac{1}{\varepsilon_{di}}\right) \qquad \text{[por (8.4)]} \end{aligned}$$

donde ε_{di}, la elasticidad puntual de demanda en el i-ésimo mercado, por lo común es negativa. En consecuencia, la relación entre MR_i y P_i podemos expresarla de otro modo mediante la ecuación

$$\text{MR}_i = P_i\left(1 - \frac{1}{|\varepsilon_{di}|}\right) \tag{11.33}$$

Recuerde que $|\varepsilon_{di}|$ es, en general, una función de P_i, de modo que cuando se elige Q_i^*, y por lo tanto se especifica P_i^*, $|\varepsilon_{di}|$ asumirá un valor específico, el cual puede ser mayor que, menor que o igual a uno. Pero si $|\varepsilon_{di}| < 1$ (con la demanda inelástica en un punto), entonces su recíproco será mayor que la unidad y la expresión entre paréntesis de (11.33) será negativa, lo cual indica un valor negativo para MR_i. De manera similar, si $|\varepsilon_{di}| = 1$ (elasticidad unitaria), entonces MR_i tomará un valor cero. En vista de que el MC de una empresa es positivo, la condición de primer orden $\text{MC} = \text{MR}_i$ requiere que la empresa opere a un nivel positivo de MR_i. Por consiguiente, los niveles de ventas elegidos de la empresa, Q_i, deben ser tales que la elasticidad puntual correspondiente de demanda en cada mercado sea mayor que uno.

La condición de primer orden $\text{MR}_1 = \text{MR}_2 = \text{MR}_3$ se puede traducir ahora, vía (11.33), en lo siguiente:

$$P_1\left(1 - \frac{1}{|\varepsilon_{d1}|}\right) = P_2\left(1 - \frac{1}{|\varepsilon_{d2}|}\right) = P_3\left(1 - \frac{1}{|\varepsilon_{d3}|}\right)$$

De esto se puede inferir fácilmente que mientras *más pequeño* sea el valor de $|\varepsilon_d|$ (en el nivel elegido de producción) en un mercado particular, *mayor* debe ser el precio cargado en ese mercado, por lo tanto, la discriminación de precio, si se va a maximizar la ganancia.

Para asegurar la maximización, examinaremos la condición de segundo orden. De (11.32) se desprende que las segundas derivadas parciales son

$$\begin{aligned} \pi_{11} &= R_1''(Q_1) - C''(Q)\frac{\partial Q}{\partial Q_1} = R_1''(Q_1) - C''(Q) \\ \pi_{22} &= R_2''(Q_2) - C''(Q)\frac{\partial Q}{\partial Q_2} = R_2''(Q_2) - C''(Q) \\ \pi_{33} &= R_3''(Q_3) - C''(Q)\frac{\partial Q}{\partial Q_3} = R_3''(Q_3) - C''(Q) \end{aligned}$$

y
$$\pi_{12} = \pi_{21} = \pi_{13} = \pi_{31} = \pi_{23} = \pi_{32} = -C''(Q) \qquad \left[\text{puesto que } \frac{\partial Q}{\partial Q_i} = 1\right]$$

de modo que tenemos (después de abreviar la notación de segunda derivada)

$$|H| = \begin{vmatrix} R_1'' - C'' & -C'' & -C'' \\ -C'' & R_2'' - C'' & -C'' \\ -C'' & -C'' & R_3'' - C'' \end{vmatrix}$$

La condición suficiente de segundo orden se satisfará como es debido, siempre que tengamos:

1. $|H_1| = R_1'' - C'' < 0$; es decir, la pendiente de MR_1 es menor que la pendiente de MC de toda la producción [cf. la situación del punto L en la figura 9.6*c*]. (Puesto que cualquiera de los tres mercados se puede tomar como el "primero", esto en efecto también implica $R_2'' - C'' < 0$ y $R_3'' - C'' < 0$.)
2. $|H_2| = (R_1'' - C'')(R_2'' - C'') - (C'')^2 > 0$; o bien, $R_1''R_2'' - (R_1'' + R_2'')C'' > 0$.
3. $|H_3| = R_1''R_2''R_3'' - (R_1''R_2'' + R_1''R_3'' + R_2''R_3'')C'' < 0$.

Las dos últimas partes de esta condición no son tan fáciles de interpretar desde el punto de vista económico como la primera. Tome en cuenta que hemos supuesto que las funciones generales $R_i(Q_i)$ son cóncavas y la función general $C(Q)$ es convexa, de modo que $-C(Q)$ es cóncava, entonces la función de ganancia, la suma de las funciones cóncavas, podría haber sido tomada como cóncava, obviando así la necesidad de comprobar la condición de segundo orden.

Ejemplo 4

A fin de concretar más el ejemplo anterior, vamos a dar una versión numérica. Imagine que la empresa monopolista tiene las funciones específicas de ingreso promedio

$$\begin{aligned} P_1 &= 63 - 4Q_1 &\quad \text{de modo que} \quad R_1 &= P_1Q_1 = 63Q_1 - 4Q_1^2 \\ P_2 &= 105 - 5Q_2 & R_2 &= P_2Q_2 = 105Q_2 - 5Q_2^2 \\ P_3 &= 75 - 6Q_3 & R_3 &= P_3Q_3 = 75Q_3 - 6Q_3^2 \end{aligned}$$

y que la función de costo total es

$$C = 20 + 15Q$$

Entonces las funciones de costo marginal serán

$$R_1' = 63 - 8Q_1 \qquad R_2' = 105 - 10Q_2 \qquad R_3' = 75 - 12Q_3 \qquad C' = 15$$

Cuando se iguala cada ingreso marginal R_i' con el costo marginal C' de la producción total, se encuentra que las cantidades de equilibrio son

$$Q_1^* = 6 \qquad Q_2^* = 9 \qquad \text{y} \qquad Q_3^* = 5$$

Por lo tanto, $$Q^* = \sum_{i=1}^{3} Q_1^* = 20$$

Al sustituir estas soluciones en las ecuaciones de ingreso y costo obtenemos $\pi^* = 679$ como la ganancia total de la operación de negocios de triple mercado.

Debido a que éste es un modelo específico, tenemos que comprobar la condición de segundo orden (o la concavidad de la función objetivo). Puesto que las segundas derivadas son

$$R_1'' = -8 \qquad R_2'' = -10 \qquad R_3'' = -12 \qquad C'' = 0$$

las tres partes de las condiciones suficientes de segundo orden dadas en el ejemplo 3 se satisfacen como es debido.

Partiendo de las funciones de ingreso promedio es fácil ver que la empresa debe cargar los precios discriminatorios $P_1^* = 39$, $P_2^* = 60$, y $P_3^* = 45$ en los tres mercados. Como se comprueba fácilmente, la elasticidad puntual de demanda es la más baja en el segundo mercado, en el cual se carga el precio más alto.

Decisiones de una empresa relacionadas con los insumos

En lugar de los niveles de producción Q_i, las variables de elección de una empresa podrían aparecer también como niveles de insumos.

Ejemplo 5 Consideremos una empresa competitiva con la siguiente función de ganancia

$$\pi = R - C = PQ - wL - rK \tag{11.34}$$

donde $P =$ precio
$Q =$ producción
$L =$ mano de obra
$K =$ capital
$w, r =$ precios de insumos para L y K, respectivamente

Puesto que la empresa opera en un mercado competitivo, las variables exógenas son P, w y r (escritas aquí sin el subíndice cero). Hay tres variables endógenas, K, L y Q; sin embargo, la producción Q resulta ser una función de K y L vía la función de producción

$$Q = Q(K, L)$$

Supongamos que se trata de una función de Cobb-Douglas (analizada con más detalle en la sección 12.6) de la forma

$$Q = L^{\alpha} K^{\beta}$$

donde α y β son parámetros positivos. Si además suponemos que el rendimiento a la medida es decreciente, entonces $\alpha + \beta < 1$. Por simplicidad, se considerará el caso simétrico donde $\alpha = \beta < \frac{1}{2}$

$$Q = L^{\alpha} K^{\alpha} \tag{11.35}$$

Al sustituir (11.35) en (11.34), obtenemos

$$\pi(K, L) = PL^{\alpha}K^{\alpha} - wL - rK$$

La condición de primer orden para la maximización de ganancia es

$$\begin{aligned} \frac{\partial \pi}{\partial L} &= P\alpha L^{\alpha-1}K^{\alpha} - w = 0 \\ \frac{\partial \pi}{\partial K} &= P\alpha L^{\alpha}K^{\alpha-1} - r = 0 \end{aligned} \tag{11.36}$$

Este sistema de ecuaciones define a L y K óptimas para la maximización de ganancia. Pero primero comprobaremos la condición de segundo orden para demostrar que se tiene un máximo.

El hessiano para este problema es

$$|H| = \begin{vmatrix} \pi_{LL} & \pi_{LK} \\ \pi_{KL} & \pi_{KK} \end{vmatrix} = \begin{vmatrix} P\alpha(\alpha-1)L^{\alpha-2}K^{\alpha} & P\alpha^2 L^{\alpha-1}K^{\alpha-1} \\ P\alpha^2 L^{\alpha-1}K^{\alpha-1} & P\alpha(\alpha-1)L^{\alpha}K^{\alpha-2} \end{vmatrix}$$

La condición suficiente para un máximo es que $|H_1| < 0$ y $|H| > 0$:

$$\begin{aligned} |H_1| &= P\alpha(\alpha-1)L^{\alpha-2}K^{\alpha} < 0 \\ |H| &= P^2\alpha^2(\alpha-1)^2 L^{2\alpha-2}K^{2\alpha-2} - P^2\alpha^4 L^{2\alpha-2}K^{2\alpha-2} \\ &= P^2\alpha^2 L^{2\alpha-2}K^{2\alpha-2}(1-2\alpha) > 0 \end{aligned}$$

Por lo tanto, para $\alpha < \frac{1}{2}$, se satisface la condición suficiente de segundo orden.

Ahora podemos volver a la condición de primer orden para determinar los valores óptimos de K y L. Reescribiendo la primera ecuación de (11.36) para aislar K, obtenemos

$$P\alpha L^{\alpha-1}K^{\alpha} = w$$

$$K = \left(\frac{w}{P\alpha} L^{1-\alpha}\right)^{\frac{1}{\alpha}}$$

Al sustituir esto en la segunda ecuación de (11.36), tenemos

$$P\alpha L^{\alpha}K^{\alpha-1} - r = P\alpha L^{\alpha}\left[\left(\frac{w}{P\alpha}L^{1-\alpha}\right)^{\frac{1}{\alpha}}\right]^{\alpha-1} - r = 0$$

o bien,

$$P^{\frac{1}{\alpha}}\alpha^{\frac{1}{\alpha}}w^{(\alpha-1)/\alpha}L^{(2\alpha-1)/\alpha} = r$$

Si reordenamos la ecuación anterior para despejar L, obtenemos

$$L^* = (P\alpha w^{\alpha-1}r^{-\alpha})^{1/(1-2\alpha)}$$

Al aprovechar la simetría de este modelo, podemos escribir rápidamente la K óptima como

$$K^* = (P\alpha r^{\alpha-1}w^{-\alpha})^{1/(1-2\alpha)}$$

L^* y K^* son las ecuaciones de demanda de insumo de la empresa.

Si sustituimos L^* y K^* en la función de producción, descubrimos que

$$\begin{aligned} Q^* &= (L^*)^{\alpha}(K^*)^{\alpha} \\ &= (P\alpha w^{\alpha-1}r^{-\alpha})^{\alpha/(1-2\alpha)}(P\alpha r^{\alpha-1}w^{-\alpha})^{\alpha/(1-2\alpha)} \\ &= \left(\frac{\alpha^2 P^2}{wr}\right)^{\alpha/(1-2\alpha)} \end{aligned} \qquad \textbf{(11.37)}$$

Esto da una expresión para la producción óptima como una función de las variables exógenas P, w y r.

Ejemplo 6

Supongamos las siguientes circunstancias: (1) En la producción de un solo producto Q de una empresa hipotética se utilizan dos insumos, a y b. (2) Los precios de ambos insumos, P_a y P_b, no dependen del control de la empresa, al igual que el precio de producción P; aquí los identificaremos mediante P_{a0}, P_{b0} y P_0, respectivamente. (3) El proceso de producción tarda t_0 años (con t_0 que representa alguna constante positiva) en completarse; así, el ingreso de ventas debe ser descontado debidamente antes de que se pueda comparar de manera apropiada con el costo de producción incurrido en el tiempo presente. La tasa de descuento, sobre una base continua, se supone que está dada en r_0.

Con base en la suposición 1, podemos escribir una función de producción general $Q = Q(a, b)$, con productos físicos marginales Q_a y Q_b. La suposición 2 nos permite expresar el costo total como

$$C = P_{a0}a + P_{b0}b$$

y el ingreso total como

$$R = P_0 Q(a, b)$$

Sin embargo, para escribir la función de ganancia debemos descontar primero el ingreso multiplicándolo por la constante $e^{-r_0 t_0}$, la cual, para evitar subíndices complicados con subíndices, la escribiremos como e^{-rt}. Así, la función de ganancia es

$$\pi = P_0 Q(a, b)e^{-rt} - P_{a0}a - P_{b0}b$$

en la que a y b son las únicas variables de elección.

Para maximizar la ganancia, es necesario que las primeras derivadas parciales

$$\begin{aligned} \pi_a\left(\equiv \frac{\partial \pi}{\partial a}\right) &= P_0 Q_a e^{-rt} - P_{a0} \\ \pi_b\left(\equiv \frac{\partial \pi}{\partial b}\right) &= P_0 Q_b e^{-rt} - P_{b0} \end{aligned} \qquad \textbf{(11.38)}$$

sean cero. Esto significa que

$$P_0 Q_a e^{-rt} = P_{a0} \qquad \text{y} \qquad P_0 Q_b e^{-rt} = P_{b0} \tag{11.39}$$

Puesto que $P_0 Q_a$ (el precio del producto multiplicado por el producto marginal del insumo a) representa el *valor de producto marginal del insumo a* (VMP_a), la primera ecuación solamente dice que el valor presente de VMP_a se debe igualar con el precio dado del insumo a. La segunda ecuación es el mismo prerrequisito aplicado al insumo b.

Considere que, para cumplir (11.39), ambos productos marginales Q_a y Q_b deben ser positivos, porque P_0, P_{a0}, P_{b0} y e^{-rt} tienen valores positivos. Esto tiene una interpretación importante en términos de una *isocuanta*, definida como el conjunto de combinaciones de insumos que producen el mismo nivel de producción. Cuando se grafican en el plano ab, las isocuantas suelen parecerse a las dibujadas en la figura 11.11. En vista de que cada una pertenece a un nivel de producción fijo, a lo largo de cualquier isocuanta se debe tener

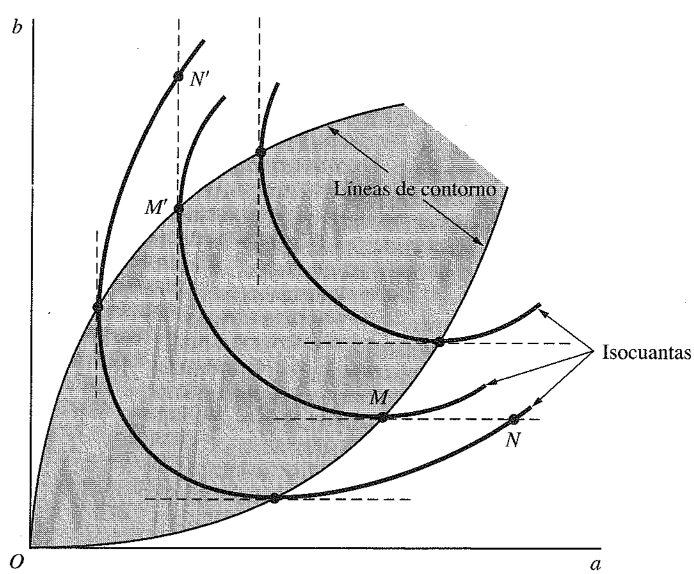
$$dQ = Q_a\,da + Q_b\,db = 0$$

lo que significa que la pendiente de una isocuanta podemos expresarla como

$$\frac{db}{da} = -\frac{Q_a}{Q_b} \qquad \left(= -\frac{\text{MPP}_a}{\text{MPP}_b}\right) \tag{11.40}$$

Por lo tanto, tener tanto a Q_a como Q_b positivas es confinar la elección de insumo de la empresa a los segmentos con pendiente negativa de las isocuantas solamente. En la figura 11.11, la región pertinente de operación, en consecuencia, se restringe al área sombreada definida mediante las denominadas líneas de contorno. Fuera del área sombreada, donde las isocuantas se caracterizan por pendientes positivas, el producto marginal de un insumo debe ser negativo. El movimiento de la combinación de insumo en M a una en N, por ejemplo, indica que con el insumo b mantenido constante el *incremento* en el insumo a conduce a una isocuanta *menor* (una producción más pequeña); así, Q_a debe ser negativa. De manera similar, un movimiento de M' a N' ilustra la negatividad de Q_b. Tome en cuenta que cuando se confina la atención al área sombreada, cada isocuanta se puede tomar como una función de la forma $b = \phi(a)$, porque para cada valor admisible de a, la isocuanta determina un valor único de b.

FIGURA 11.11



La condición de segundo orden gira en torno a las segundas derivadas parciales de p, que se obtienen de (11.38). Teniendo en mente que Q_a y Q_b, siendo derivadas, son por sí mismas funciones de las variables a y b, podemos hallar π_{aa}, $\pi_{ab} = \pi_{ba}$ y π_{bb}, y arreglándolas en un hessiano:

$$|H| = \begin{vmatrix} \pi_{aa} & \pi_{ab} \\ \pi_{ab} & \pi_{bb} \end{vmatrix} = \begin{vmatrix} P_0 Q_{aa} e^{-rt} & P_0 Q_{ab} e^{-rt} \\ P_0 Q_{ab} e^{-rt} & P_0 Q_{bb} e^{-rt} \end{vmatrix} \tag{11.41}$$

Para que un valor estacionario de π sea un máximo, es suficiente que

$|H_1| < 0$ [es decir, $\pi_{aa} < 0$, lo cual ocurre si y sólo si $Q_{aa} < 0$]

$|H_2| = |H| > 0$ $\left[\text{es decir, } \pi_{aa}\pi_{bb} > \pi_{ab}^2, \text{ lo cual ocurre si y sólo si } Q_{aa}Q_{bb}\ Q_{ab}^2\right]$

Así, se puede probar la condición de segundo orden, ya sea con las derivadas π_{ij} o las derivadas Q_{ij}, las que sean más convenientes.

El símbolo Q_{aa} denota la tasa de cambio de Q_a ($\equiv \text{MPP}_a$) cuando el insumo a cambia mientras se mantiene fijo el insumo b; de manera similar, Q_{bb} denota la tasa de cambio de Q_b ($\equiv \text{MPP}_b$) cuando sólo cambia el insumo b. Por lo tanto, la condición suficiente de segundo orden estipula, en parte, que el MPP de ambos insumos es *decreciente* a los niveles de insumo elegidos, a^* y b^*. Sin embargo, observe que disminuir MPP_a y MPP_b *no* garantiza que se satisfaga la condición de segundo orden, porque la última condición también tiene que ver con la magnitud de $Q_{ab} = Q_{ba}$, que mide la tasa de cambio de MPP de un insumo cuando varía la cantidad del otro insumo.

Al llevar a cabo un examen más minucioso resulta que, así como la condición de primer orden especifica que la isocuanta tendrá pendiente negativa en la combinación de insumos elegida (como se muestra en el área sombreada de la figura 11.11), la condición suficiente de segundo orden sirve para especificar que la misma isocuanta es estrictamente convexa en la combinación de insumos elegida. La curvatura de la isocuanta se relaciona con el signo de la segunda derivada d^2b/da^2. Para obtener esta última, (11.40) se debe derivar por completo respecto al parámetro a, sin olvidar que tanto Q_a como Q_b son funciones derivadas de a y b y sin embargo, en una isocuanta, b es por sí misma una función de a; es decir,

$$Q_a = Q_a(a, b) \qquad Q_b = Q_b(a, b) \qquad \text{y} \qquad b = \phi(a)$$

Entonces la diferenciación total procede como sigue:

$$\frac{d^2b}{da^2} = \frac{d}{da}\left(-\frac{Q_a}{Q_b}\right) = -\frac{1}{Q_b^2}\left[Q_b\frac{dQ_a}{da} - Q_a\frac{dQ_b}{da}\right] \tag{11.42}$$

Puesto que b es una función de a en la isocuanta, la fórmula de derivada total (8.9) produce

$$\begin{aligned} \frac{dQ_a}{da} &= \frac{\partial Q_a}{\partial b}\frac{db}{da} + \frac{\partial Q_a}{\partial a} = Q_{ba}\frac{db}{da} + Q_{aa} \\ \frac{dQ_b}{da} &= \frac{\partial Q_b}{\partial b}\frac{db}{da} + \frac{\partial Q_b}{\partial a} = Q_{bb}\frac{db}{da} + Q_{ab} \end{aligned} \tag{11.43}$$

Después de sustituir (11.40) en (11.43) y sustituir esta última en (11.42), podemos reescribir la segunda derivada como

$$\begin{aligned} \frac{d^2b}{da^2} &= -\frac{1}{Q_b^2}\left[Q_{aa}Q_b - Q_{ba}Q_a - Q_{ab}Q_a + Q_{bb}Q_a^2\left(\frac{1}{Q_b}\right)\right] \\ &= -\frac{1}{Q_b^3}[Q_{aa}(Q_b)^2 - 2Q_{ab}(Q_a)(Q_b) + Q_{bb}(Q_a)^2] \end{aligned} \tag{11.44}$$

En (11.44) observamos que la expresión entre corchetes (último renglón) es una forma cuadrática en las dos variables Q_a y Q_b. Si se satisface la condición suficiente de segundo orden, de modo que

$$Q_{aa} < 0 \qquad \text{y} \qquad \begin{vmatrix} Q_{aa} & -Q_{ab} \\ -Q_{ab} & Q_{bb} \end{vmatrix} > 0$$

entonces, en virtud de (11.11′), dicha forma cuadrática debe ser negativa definida. A su vez, esto hará que d^2b/da^2 sea positiva, porque se ha restringido a Q_b a ser positiva por la condición de primer orden. Así, la satisfacción de la condición suficiente de segundo orden significa que la isocuanta pertinente (con pendiente negativa) es estrictamente convexa en la combinación de insumos elegida, como se afirmó.

El concepto de convexidad estricta, aplicado a una isocuanta $b = \phi(a)$, que se traza en el plano *ab* bidimensional, se debe distinguir cuidadosamente del mismo concepto que se aplicó a la función de producción $Q(a, b)$, que se traza en un espacio tridimensional *abQ*. Tome en cuenta, especialmente, que si aplicamos el concepto de concavidad o concavidad estricta a la función de producción en el presente contexto, entonces, para producir la forma de isocuanta deseada, la estipulación apropiada es que $Q(a, b)$ sea estrictamente *cóncava* en el espacio de tres dimensiones (con forma de domo), lo cual tiene un marcado contraste con la estipulación de que la isocuanta pertinente sea estrictamente *convexa* en el espacio bidimensional (con forma de U o parte de una U).

Ejemplo 7

A continuación, supongamos que el interés se compone *trimestralmente*, a una tasa de interés de i_0 por trimestre. Asimismo, supongamos que el proceso de producción toma un trimestre del año. Por lo tanto, la función de ganancia se convierte en

$$\pi = P_0 Q(a, b)(1 + i_0)^{-1} - P_{a0}a - P_{b0}b$$

La condición de primer orden es

$$P_0 Q_a(1 + i_0)^{-1} - P_{a0} = 0$$
$$P_0 Q_b(1 + i_0)^{-1} - P_{b0} = 0$$

con una interpretación analítica igual a la del ejemplo 6, excepto por la manera distinta de descontar.

Se ve sin dificultad que la misma condición suficiente obtenida en el ejemplo 6 también se debe aplicar aquí.

EJERCICIO 11.6

1. Si la empresa competitiva del ejemplo 1 tuviera una función de costo $C = 2Q_1^2 + 2Q_2^2$, entonces:
   (*a*) ¿Aún estaría técnicamente relacionada la producción de los dos bienes?
   (*b*) ¿Cuáles serían los nuevos niveles óptimos de Q_1 y Q_2?
   (*c*) ¿Cuál sería el valor de π_{12}? ¿Qué implicaría esto económicamente?
2. Una empresa de dos productos enfrenta las siguientes funciones de demanda y costo:
   $Q_1 = 40 - 2P_1 - P_2 \qquad Q_2 = 35 - P_1 - P_2 \qquad C = Q_1^2 + 2Q_2^2 + 10$
   (*a*) Encuentre los niveles de producción que satisfacen la condición de primer orden para ganancia máxima (usa fracciones).
   (*b*) Compruebe la condición suficiente de segundo orden. ¿Se puede concluir que este problema posee un máximo absoluto único?
   (*c*) ¿Cuál es la ganancia máxima?

3. Con base en el precio y la cantidad de equilibrio del ejemplo 4, calcule la elasticidad puntual de demanda $|\varepsilon_{di}|$ (para $i = 1, 2$). ¿Cuál mercado tiene las elasticidades de demanda máxima y mínima?
4. Si la función de costo del ejemplo 4 se cambia a $C = 20 + 15Q + Q^2$
   (*a*) Encuentre la nueva función de costo marginal.
   (*b*) Encuentre las nuevas cantidades de equilibrio (use fracciones).
   (*c*) Encuentre los nuevos precios de equilibrio.
   (*d*) Compruebe que se satisface la condición suficiente de segundo orden.
5. En el ejemplo 7, ¿cómo reescribiría la función de ganancia si se cumplen las condiciones siguientes?:
   (*a*) El interés se compone semianualmente a una tasa de interés de i_0 por año y el proceso de producción toma un año.
   (*b*) El interés se compone trimestralmente a una tasa de interés de i_0 por año y el proceso de producción toma nueve meses.
6. Dada $Q = Q(a, b)$, ¿cómo expresaría en forma algebraica la isocuanta para el nivel de producción de, por ejemplo, 260?

11.7 Aspectos estáticos comparativos de la optimización

La optimización, que es una variedad especial del análisis de equilibrio estático, está sujeta también a investigaciones del tipo estático comparativo. La idea es, una vez más, encontrar cómo un cambio en algún parámetro afecta la posición de equilibrio del modelo, que en el presente contexto se refiere a los valores óptimos de las variables de elección (y al valor óptimo de la función objetivo). Puesto que no se utiliza ninguna nueva técnica más allá de las explicadas en la parte 3, podemos proceder de modo directo con algunas ilustraciones con base en los ejemplos presentados en la sección 11.6.

Soluciones de forma reducida

El ejemplo 1 de la sección 11.6 contiene dos parámetros (o variables exógenas), P_{10} y P_{20}; no es sorprendente, pues, que los niveles de producción óptima de esta empresa de dos productos se expresen de modo estricto en términos de estos parámetros:

$$Q_1^* = \frac{4P_{10} - P_{20}}{15} \qquad \text{y} \qquad Q_2^* = \frac{4P_{20} - P_{10}}{15}$$

Éstas son soluciones de forma reducida, y la diferenciación parcial sola es suficiente para indicar las propiedades estáticas comparativas del modelo, a saber,

$$\frac{\partial Q_1^*}{\partial P_{10}} = \frac{4}{15} \qquad \frac{\partial Q_1^*}{\partial P_{20}} = -\frac{1}{15} \qquad \frac{\partial Q_2^*}{\partial P_{10}} = -\frac{1}{15} \qquad \frac{\partial Q_2^*}{\partial P_{20}} = \frac{4}{15}$$

Para la ganancia máxima, cada producto de la empresa se debe producir en una cantidad mayor si se eleva su precio de mercado o si cae el precio de mercado del otro producto.

Estas conclusiones se deducen sólo de las suposiciones particulares del modelo en cuestión. En particular, podríamos señalar que los efectos de un cambio en P_{10} sobre Q_2^* y de P_{20}

sobre Q_1^*, son consecuencia de la relación técnica supuesta en el lado de producción de estos dos artículos, y que en ausencia de tal relación se tendrá

$$\frac{\partial Q_1^*}{\partial P_{20}} = \frac{\partial Q_2^*}{\partial P_{10}} = 0$$

Pasando al ejemplo 2, vemos que los niveles de producción óptima se expresan ahí en forma numérica como $Q_1^* = 8$ y $Q_2^* = 7\frac{2}{3}$, no aparece ningún parámetro. De hecho, todas las constantes en las ecuaciones del modelo son numéricas en vez de paramétricas, de modo que al momento de alcanzar la etapa de solución esas constantes han perdido sus identidades respectivas por el proceso de manejo aritmético. Esto subraya la carencia fundamental de generalidad en el uso de constantes numéricas y la falta resultante de contenido estático comparativo en la solución de equilibrio.

Por otro lado, el hecho de *no* usar constantes numéricas no garantiza que el problema será asequible al análisis estático comparativo. El problema de discriminación de precios (ejemplo 3), por ejemplo, se planteó sobre todo para el estudio de la condición de equilibrio (ganancia-maximización), y no se introdujo absolutamente ningún parámetro. En consecuencia, aunque se expresó en términos de funciones generales, será necesaria una reformulación si se está interesado en un estudio estático comparativo.

Modelos de función general

El problema de decisión acerca del insumo del ejemplo 6 ilustra el caso donde una formulación de función general abarca varios parámetros, de hecho no menos de cinco (P_0, P_{a0}, P_{b0}, r y t), donde, como antes, se han omitido los subíndices de las variables exógenas r_0 y t_0. ¿Cómo se obtienen las propiedades estáticas comparativas de este modelo?

La respuesta radica de nuevo en la aplicación del teorema de la función implícita. Pero, a diferencia de los casos de modelos de equilibrio intermedio del mercado o de determinación de ingreso nacional, donde trabajamos con las condiciones de equilibrio del modelo, el presente contexto de equilibrio final establece que se trabaja con las condiciones de optimización de primer orden. Para el ejemplo 6, estas condiciones se expresaron en (11.39). Si se reúnen todos los términos en (11.39) a la izquierda de los signos de igualdad y se hace explícito que Q_a y Q_b son funciones de las variables (de elección) endógenas a y b, se pueden reescribir las condiciones de primer orden en el formato de (8.24) de la siguiente manera:

$$\begin{aligned} F^1(a, b; P_0, P_{a0}, P_{b0}, r, t) &= P_0 Q_a(a, b)e^{-rt} - P_{a0} = 0 \\ F^2(a, b; P_0, P_{a0}, P_{b0}, r, t) &= P_0 Q_b(a, b)e^{-rt} - P_{b0} = 0 \end{aligned} \qquad \textbf{(11.45)}$$

Supongamos que las funciones F^1 y F^2 poseen derivadas continuas. Así, se podría aplicar el teorema de la función implícita, siempre que el jacobiano de este sistema respecto a las variables endógenas a y b no se anulen en el equilibrio inicial. Este jacobiano no es otra cosa que el determinante hessiano de la función π del ejemplo 6:

$$|J| = \begin{vmatrix} \dfrac{\partial F^1}{\partial a} & \dfrac{\partial F^1}{\partial b} \\ \dfrac{\partial F^2}{\partial a} & \dfrac{\partial F^2}{\partial b} \end{vmatrix} = \begin{vmatrix} P_0 Q_{aa} e^{-rt} & P_0 Q_{ab} e^{-rt} \\ P_0 Q_{ab} e^{-rt} & P_0 Q_{bb} e^{-rt} \end{vmatrix} = |H| \quad [\text{por } (11.46)] \qquad \textbf{(11.46)}$$

Por consiguiente, si suponemos que se satisface la condición suficiente de segundo orden para la maximización de ganancia, entonces $|H|$ debe ser positiva, lo mismo que $|J|$, en el equilibrio inicial u óptimo. En ese caso, el teorema de la función implícita permite escribir el par de funciones implícitas

$$\begin{aligned} a^* &= a^*(P_0, P_{a0}, P_{b0}, r, t) \\ b^* &= b^*(P_0, P_{a0}, P_{b0}, r, t) \end{aligned} \tag{11.47}$$

así como el par de identidades

$$\begin{aligned} P_0 Q_a(a^*, b^*)e^{-rt} - P_{a0} &\equiv 0 \\ P_0 Q_b(a^*, b^*)e^{-rt} - P_{b0} &\equiv 0 \end{aligned} \tag{11.48}$$

Para estudiar la estática comparativa del modelo, primero toma la diferencial total de cada identidad de (11.48). Por lo pronto, se permitirá que varíen todas las variables exógenas, de modo que en el resultado de la diferenciación total aparecerán da^*, db^*, así como dP_0, dP_{a0}, dP_{b0}, dr y dt. Si colocamos del lado izquierdo de los signos de igualdad sólo los términos donde aparecen da^* y db^*, el resultado será

$$\begin{aligned} P_0 Q_{aa} e^{-rt} da^* + P_0 Q_{ab} e^{-rt} db^* &= -\, Q_a e^{-rt} dP_0 + dP_{a0} \\ &\quad + P_0 Q_a t e^{-rt} dr + P_0 Q_a r e^{-rt} dt \\ P_0 Q_{ab} e^{-rt} da^* + P_0 Q_{bb} e^{-rt} db^* &= -\, Q_b e^{-rt} dP_0 + dP_{b0} \\ &\quad + P_0 Q_b t e^{-rt} dr + P_0 Q_b r e^{-rt} dt \end{aligned} \tag{11.49}$$

donde se observa que las derivadas primera y segunda de Q se evaluarán en el equilibrio, es decir, en a^* y b^*; se observa también que los coeficientes de da^* y db^* que están a la izquierda son precisamente los elementos del jacobiano de (11.46).

Para deducir las derivadas estáticas comparativas, de las cuales hay un total de 10 (¿por qué?), permitiremos ahora que varíe cada vez una sola variable exógena. Imagine que permitimos que solamente varíe P_0. Entonces, $dP_0 \neq 0$, pero $dP_{a0} = dP_{b0} = dr = dt = 0$, de modo que sólo el primer término permanecerá del lado derecho de cada ecuación en (11.49). Al dividir entre dP_0 e interpretar la relación da^*/dP_0 como la derivada estática comparativa $(\partial a^*/\partial P_0)$, y de manera similar para la relación db^*/dP_0, podemos escribir la ecuación matricial

$$\begin{bmatrix} P_0 Q_{aa} e^{-rt} & P_0 Q_{ab} e^{-rt} \\ P_0 Q_{ab} e^{-rt} & P_0 Q_{bb} e^{-rt} \end{bmatrix} \begin{bmatrix} (\partial a^*/\partial P_0) \\ (\partial b^*/\partial P_0) \end{bmatrix} = \begin{bmatrix} -Q_a e^{-rt} \\ -Q_b e^{-rt} \end{bmatrix}$$

Se descubre que la solución, por la regla de Cramer, es

$$\begin{aligned} \left(\frac{\partial a^*}{\partial P_0}\right) &= \frac{(Q_b Q_{ab} - Q_a Q_{bb}) P_0 e^{-2rt}}{|J|} \\ \left(\frac{\partial b^*}{\partial P_0}\right) &= \frac{(Q_a Q_{ab} - Q_b Q_{aa}) P_0 e^{-2rt}}{|J|} \end{aligned} \tag{11.50}$$

También existe otro método para obtener estos resultados: simplemente se podrían diferenciar las dos identidades de (11.48) *por completo* respecto a P_0 (mientras se mantienen fijas las otras cuatro variables exógenas), sin olvidar que P_0 puede afectar a a^* y b^* vía (11.47).

Procedamos ahora a analizar los signos de las derivadas estáticas comparativas de (11.50). Con la suposición de que se satisface la condición suficiente de segundo orden, el jacobiano del denominador debe ser positivo. La condición de segundo orden implica también que Q_{aa} y Q_{bb} sean negativas, así como la condición de primer orden implica que Q_a y Q_b sean positivas. Además, la expresión P_0e^{-2rt} es de hecho positiva. Así, si $Q_{ab} > 0$ (si aumenta un insumo crecerá el MPP del otro insumo), podemos concluir que tanto $(\partial a^*/\partial P_0)$ como $(\partial b^*/\partial P_0)$ serán positivas, lo cual significa que un incremento en el precio del producto dará como resultado un mayor uso de ambos insumos en el equilibrio. Por otro lado, si $Q_{ab} < 0$, el signo de cada derivada de (11.50) dependerá de la resistencia relativa de la fuerza negativa y la fuerza positiva en la expresión que está entre paréntesis de la derecha.

A continuación, deje que sólo varíe la variable exógena r. Entonces, todos los términos de la derecha de (11.49) se anulan excepto los que tienen dr. Al dividir entre $dr \neq 0$, se obtiene la siguiente ecuación matricial

$$\begin{bmatrix} P_0Q_{aa}e^{-rt} & P_0Q_{ab}e^{-rt} \\ P_0Q_{ab}e^{-rt} & P_0Q_{bb}e^{-rt} \end{bmatrix} \begin{bmatrix} (\partial a^*/\partial r) \\ (\partial b^*/\partial r) \end{bmatrix} = \begin{bmatrix} P_0Q_ate^{-rt} \\ P_0Q_bte^{-rt} \end{bmatrix}$$

con la solución

$$\begin{aligned} \left(\frac{\partial a^*}{\partial r}\right) &= \frac{t(Q_aQ_{bb} - Q_bQ_{ab})(P_0e^{-rt})^2}{|J|} \\ \left(\frac{\partial b^*}{\partial r}\right) &= \frac{t(Q_bQ_{aa} - Q_aQ_{ab})(P_0e^{-rt})^2}{|J|} \end{aligned} \qquad \textbf{(11.51)}$$

Estas dos derivadas estáticas comparativas serán negativas si Q_{ab} es positiva, pero serán indeterminadas en cuanto a signo si Q_{ab} es negativa.

Los efectos de cambios en los demás parámetros se pueden definir mediante un procedimiento similar. En realidad, en vista de la simetría entre r y t en (11.48) es obvio que tanto $(\partial a^*/\partial t)$ como $(\partial b^*/\partial t)$ deben ser similares en apariencia a (11.51).

Los efectos de cambios en P_{a0} y P_{b0} se dejan para que los analice por su cuenta. Como verá, la restricción de signo de la condición suficiente de segundo orden será útil de nuevo en la evaluación de derivadas estáticas comparativas, porque indica los signos de Q_{aa} y Q_{bb}, así como el jacobiano $|J|$ en el equilibrio inicial (óptimo). Por lo tanto, aparte de distinguir entre máximo y mínimo, la condición de segundo orden también tiene un papel vital en el estudio de cambios en las posiciones de equilibrio.

EJERCICIO 11.7

Para los problemas 1 al 3, suponga que $Q_{ab} > 0$.

1. Con base en el modelo descrito en (11.45) a (11.48), encuentre las derivadas estáticas comparativas $(\partial a^*/\partial P_{a0})$ y $(\partial b^*/\partial P_{a0})$. Interprete el significado económico del resultado. Luego, analice los efectos sobre a^* y b^* de un cambio en P_{b0}.
2. Para el problema del ejemplo 7 de la sección 11.6:
   (*a*) ¿Cuántos parámetros hay? Enumérelos.
   (*b*) Siguiendo el procedimiento descrito en (11.45) a (11.50), y suponiendo que se satisface la condición suficiente de segundo orden, encuentre las derivadas estáticas com-

parativas $(\partial \alpha^*/\partial P_0)$ y $(\partial b^*/\partial P_0)$. Evalúe sus signos e interprete sus significados económicos.

(c) Encuentre $(\partial \alpha^*/\partial i_0)$ como $(\partial b^*/\partial i_0)$, evalúe sus signos, e interprete sus significados económicos.

3. Demuestre que los resultados de (11.50) se pueden obtener alternativamente mediante diferenciación *total* de las dos identidades de (11.48) respecto a P_0, mientras se mantienen fijas las otras variables exógenas. No olvide que P_0 puede afectar a α^* y b^* en virtud de (11.47).
4. Un determinante jacobiano, como se define en (7.27), está constituido por derivadas parciales de *primer* orden. Por otro lado, un determinate hessiano, como se define en las secciones 11.3 y 11.4, tiene como sus elementos a las derivadas parciales de *segundo* orden. ¿Cómo, entonces, puede ser que $|J| = |H|$, como en (11.46)?

Capítulo 12

Optimización con restricciones de igualdad

En el capítulo 11 presentamos un método general para encontrar los extremos relativos de una función objetivo de dos o más variables seleccionadas. Un aspecto importante de esa problemática es que todas las variables seleccionadas son *independientes* entre sí, en el sentido de que la decisión tomada en relación con una variable no afecta la selección de las variables restantes. Por ejemplo, una empresa que maneje dos productos puede seleccionar cualquier valor de Q_1 y Q_2 que desee, sin que las dos elecciones se limiten entre sí.

Si se requiriera que esta empresa observara una restricción (tal como una cuota de producción) en la forma de $Q_1 + Q_2 = 950$, se perdería la independencia entre las variables seleccionadas. En ese caso, los niveles de producto Q_1^* y Q_2^* de la empresa que maximizan la ganancia no solamente son simultáneos sino también dependientes, porque cuanto mayor sea Q_1^*, menor debe ser Q_2^*, para que la cuota combinada sea 950. El nuevo óptimo que satisface la cuota de producción constituye un *óptimo restringido*, el cual, en general, puede esperarse que difiera del *óptimo libre* que estudiamos en el capítulo 11.

Una restricción, tal como la cuota de producción que se acaba de mencionar, establece una relación entre las dos variables de elección, pero debe diferenciarse de otros tipos de relación que puedan enlazar a las variables. Como en el ejemplo 2 de la sección 11.6, donde los dos productos de la firma están relacionados por el consumo (sustitutos), así como por la producción (como se refleja en la función costo), pero ese hecho no caracteriza al problema como de optimización con restricciones, ya que las dos variables de la producción, *como variables de elección*, siguen siendo *independientes*. Solamente la dependencia de las variables en su papel de variables de elección da lugar a un óptimo restringido.

En este capítulo consideramos solamente las restricciones de igualdad, tales como $Q_1 + Q_2 = 950$. Nuestro objetivo principal son los extremos restringidos *relativos*, aunque también presentamos los *absolutos* en la sección 12.4.

12.1 Efectos de una restricción

El propósito principal de la imposición de una restricción es reconocer ciertos factores limitantes presentes en el problema de optimización que se estudia.

Ya hemos visto la limitación sobre la selección de productos resultante de una cuota de producción. Como ejemplo adicional, consideremos a un consumidor con la función simple de utilidad (índice)

$$U = x_1x_2 + 2x_1 \tag{12.1}$$

Como las utilidades marginales —las derivadas parciales $U_1 \equiv \partial U/\partial x_1$ y $U_2 \equiv \partial U/\partial x_2$— son positivas aquí para todos los niveles positivos de x_1 y x_2, al hacer que U se maximice sin ninguna restricción el consumidor debe comprar una cantidad *infinita* de ambos bienes, una solución que, obviamente, tiene muy poca importancia práctica. Para hacer que el problema de optimización tenga un significado económico, también debemos considerar el poder de compra del consumidor; es decir, debemos incorporar al problema una *restricción presupuestaria*. Si el consumidor desea gastar una suma dada, por ejemplo, \$60, para los dos bienes y si los precios vigentes son $P_{10} = 4$ y $P_{20} = 2$, entonces la restricción presupuestaria puede expresarse mediante la ecuación lineal

$$4x_1 + 2x_2 = 60 \tag{12.2}$$

Esta restricción, al igual que la cuota de producción mencionada, hace mutuamente dependientes a x_1^* y x_2^*.

Ahora, el problema es maximizar (12.1), sujeto a la restricción establecida en (12.2). Matemáticamente, lo que hace la restricción (también denominada *limitación*, *relación lateral* o *condición subsidiaria*) es reducir el dominio, y con ello el rango de la función objetivo. Normalmente el dominio de (12.1) sería el conjunto $\{(x_1, x_2) \mid x_1 \geq 0, x_2 \geq 0\}$. Gráficamente, el dominio está representado por el cuadrante no negativo del plano x_1x_2 de la figura 12.1*a*. Sin embargo, después de añadir la restricción presupuestaria (12.2), podemos admitir sólo aquellos valores de las variables que satisfagan esta última ecuación, de modo que el dominio se reduce inmediatamente al conjunto de puntos situados en la línea del presupuesto. Esto también afecta automáticamente al rango de la función objetivo; ahora sólo es relevante el subconjunto de la superficie de utilidad situado directamente arriba de la línea de la restricción presupuestaria. Este subconjunto (una sección transversal de la superficie) se vería como la curva de la figura 12.1*b*, donde U corresponde al eje vertical, con la línea del presupuesto del diagrama *a* situada en el eje horizontal. Entonces nuestro interés radica sólo en localizar el máximo en la curva del diagrama *b*.

En general, para una función $z = f(x, y)$, la diferencia entre un extremo restringido y un extremo libre se ilustra en la gráfica tridimensional de la figura 12.2. El extremo libre de esta

FIGURA 12.1

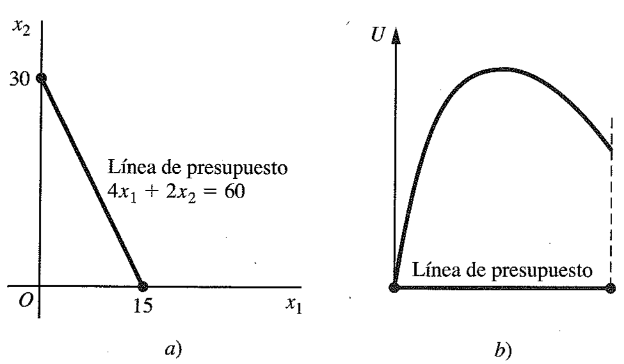
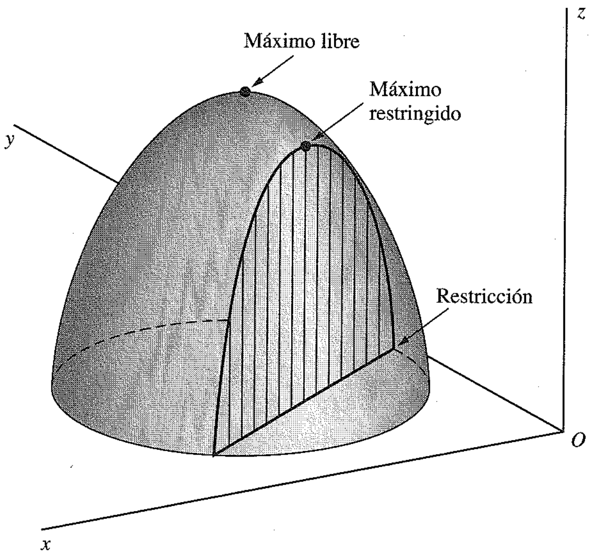


FIGURA 12.2



gráfica específica es el punto cúspide del domo completo, pero el extremo restringido está en la cúspide de la curva en forma de U invertida situada en la parte superior de la línea de restricciones (es decir, situada directamente arriba). En general, podemos esperar que un máximo restringido tenga un valor inferior al máximo libre, aunque, excepcionalmente puede ocurrir que los dos máximos tengan el mismo valor. Pero el máximo restringido nunca puede sobrepasar al máximo libre.

Es interesante observar que si hubiéramos añadido otra restricción que interceptara la primera restricción en un solo punto en el plano xy, las dos restricciones juntas habrían restringido el dominio a ese punto único. Entonces, la localización del extremo sería una cuestión trivial. En un problema relevante, el número y la naturaleza de las restricciones serían tales que restringirían pero no eliminarían la posibilidad de selección. Generalmente, el número de restricciones sería menor que el número de variables seleccionadas.

12.2 Cómo encontrar los valores estacionarios

Aun sin ninguna técnica nueva de solución, podemos encontrar fácilmente el máximo restringido en el ejemplo sencillo definido por (12.1) y (12.2). Como la restricción (12.2) implica

$$x_2 = \frac{60 - 4x_1}{2} = 30 - 2x_1 \qquad \textbf{(12.2')}$$

podemos combinar la restricción con la función objetivo al sustituir (12.2′) en (12.1). El resultado es una función objetivo con sólo una variable:

$$U = x_1(30 - 2x_1) + 2x_1 = 32x_1 - 2x_1^2$$

que podemos manejarla con el método ya aprendido. Al hacer $dU/dx_1 = 32 - 4x_1$ igual a cero, obtenemos la solución $x_1^* = 8$, que de acuerdo con (12.2′) conduce inmediatamente a $x_2^* = 30 - 2(8) = 14$. De (12.1) podemos encontrar el valor estacionario $U^* = 128$; y puesto que la segunda derivada es $d^2U/dx_1^2 = -4 < 0$, ese valor estacionario constituye un máximo de U^* (restringido).[1]

Sin embargo, cuando la restricción en sí misma es una función complicada, o cuando hay varias restricciones que deben considerarse, la técnica de sustitución y eliminación de variables podría convertirse en una tarea fatigante. Aún más importante, si la restricción presenta una forma tal que no podemos resolverla para expresar una variable (x_2) como una función explícita de la otra (x_1), el método de eliminación sería inútil –aun sabiendo que x_2 es una función implícita de x_1, es decir, incluso si se satisficieran las condiciones del teorema de la función implícita. En estos casos, podemos recurrir a un método, conocido como el *método de los multiplicadores de Lagrange (indeterminados)*, el cual, como veremos, tiene ventajas analíticas notorias.

El método de los multiplicadores de Lagrange

La esencia del método de los multiplicadores de Lagrange es convertir un problema de extremo restringido a una forma tal que todavía sea aplicable la condición de primer orden del problema de extremo libre.

Dado el problema de maximizar $U = x_1x_2 + 2x_1$, sujeto a la restricción $4x_1 + 2x_2 = 60$ [de (12.1) y (12.2)], podemos escribir lo que se denomina la *función lagrangiana*, que es una versión modificada de la función objetivo que incorpora a la restricción, como:

$$Z = x_1x_2 + 2x_1 + \lambda(60 - 4x_1 - 2x_2) \qquad \textbf{(12.3)}$$

El símbolo λ (la letra minúscula griega lambda), que representa un número todavía indeterminado, se llama *multiplicador de Lagrange (indeterminado)*. Si de alguna manera podemos asegurarnos de que $4x_1 + 2x_2 = 60$, de modo que se satisfaga la restricción, entonces el último término de (12.3) se anula, independientemente del valor que tenga λ. En ese caso, Z es idéntico a U. Aún más, si no se toma en cuenta la restricción, sólo tenemos que buscar el máximo *libre* de Z, en lugar del máximo *restringido* de U respecto a las dos variables x_1 y x_2. La pregunta, pues, es: ¿cómo podemos hacer que se anule la expresión que está entre paréntesis en (12.3)?

La táctica que logra esto consiste simplemente en tratar a λ como una variable de elección adicional en (12.3), es decir, considerar $Z = Z(\lambda, x_1, x_2)$. Porque entonces la condición de primer orden para el extremo libre consistirá en el conjunto de ecuaciones simultáneas

$$\begin{aligned} Z_\lambda(\equiv \partial Z/\partial\lambda) &= 60 - 4x_1 - 2x_2 = 0 \\ Z_1(\equiv \partial Z/\partial x_1) &= x_2 + 2 - 4\lambda = 0 \\ Z_2(\equiv \partial Z/\partial x_2) &= x_1 - 2\lambda = 0 \end{aligned} \qquad \textbf{(12.4)}$$

y la primera ecuación garantiza automáticamente la satisfacción de la restricción. Entonces, al incorporar la restricción en la función lagrangiana Z y al tratar al multiplicador de Lagrange como una variable extra, podemos obtener el extremo restringido U^* (dos variables de elección) simplemente discriminando los valores estacionarios de Z, tomada ésta como una función *libre* de tres variables de elección.

Al resolver (12.4) para los valores críticos de las variables, encontramos $x_1^* = 8$, $x_2^* = 14$ (y $\lambda^* = 4$). Como se esperaba, los valores de x_1^* y x_2^* coinciden con las respuestas obtenidas

[1] Recuerde que para el problema del campo de flores del ejercicio 9.4-2 se aplicó la misma técnica de sustitución para encontrar el área máxima, usando una restricción (la cantidad disponible de red de alambre) para eliminar una de las dos variables (la longitud o el ancho del campo de flores).

mediante el método de sustitución. Aún más, partiendo de (12.3) es evidente que $Z^* = 128$; esto es idéntico al valor de U^* encontrado anteriormente, como debe ser.

En general, dada una función objetivo

$$z = f(x, y) \tag{12.5}$$

sujeta a la restricción

$$g(x, y) = c \tag{12.6}$$

donde c es una constante,[2] podemos escribir la función lagrangiana como

$$Z = f(x, y) + \lambda[c - g(x, y)] \tag{12.7}$$

Para los valores estacionarios de Z, considerada como una función de las tres variables λ, x y y, la condición necesaria es

$$\begin{aligned} Z_\lambda &= c - g(x, y) = 0 \\ Z_x &= f_x - \lambda g_x = 0 \\ Z_y &= f_y - \lambda g_y = 0 \end{aligned} \tag{12.8}$$

Como la primera ecuación de (12.8) es simplemente un replanteamiento de (12.6), los valores estacionarios de la función lagrangiana Z satisfacen automáticamente la restricción de la función original z. Y como la expresión $\lambda[c - g(x, y)]$ ahora es con toda certeza igual a cero, los valores estacionarios de Z de (12.7) deben ser idénticos a los de (12.5), sujeta a (12.6).

Ilustremos el método con dos ejemplos adicionales.

Ejemplo 1

Encuentre el extremo de

$$z = xy \quad \text{sujeto a} \quad x + y = 6$$

El primer paso es escribir la función lagrangiana

$$Z = xy + \lambda(6 - x - y)$$

Para un valor estacionario de Z, es necesario que

$$\left.\begin{aligned} Z_\lambda &= 6 - x - y = 0 \\ Z_x &= y - \lambda = 0 \\ Z_y &= x - \lambda = 0 \end{aligned}\right\} \quad \text{o} \quad \left\{\begin{array}{rcrcrl} & & x & + & y & = 6 \\ -\lambda & & & + & y & = 0 \\ -\lambda & + & x & & & = 0 \end{array}\right.$$

Entonces, mediante la regla de Cramer o por algún otro método, encontramos

$$\lambda^* = 3 \qquad x^* = 3 \qquad y^* = 3$$

El valor estacionario es $Z^* = z^* = 9$, que debe comprobarse respecto a una condición de segundo orden antes de poder afirmar si es un máximo o un mínimo (o ninguno de los dos). Eso lo consideraremos en la sección 12.3.

[2] También se puede incluir la constante c bajo la función de restricción de modo que (12.6) aparezca como $G(x, y) = 0$, donde $G(x, y) = g(x, y) - c$. En ese caso, (12.7) debe cambiarse a $Z = f(x, y) + \lambda[0 - G(x, y)] = f(x, y) - \lambda G(x, y)$. Se escoge la versión de (12.6) porque posteriormente facilita el estudio del efecto estático comparativo de un cambio en la constante de restricción [ver (12.16)].

Ejemplo 2 Encuentre el extremo de

$$z = x_1^2 + x_2^2 \qquad \text{sujeto a} \qquad x_1 + 4x_2 = 2$$

La función lagrangiana es

$$Z = x_1^2 + x_2^2 + \lambda(2 - x_1 - 4x_2)$$

la condición necesaria para un valor estacionario es

$$\left.\begin{aligned} Z_\lambda &= 2 - x_1 - 4x_2 = 0 \\ Z_1 &= 2x_1 - \lambda = 0 \\ Z_2 &= 2x_2 - 4\lambda = 0 \end{aligned}\right\} \qquad \text{o} \qquad \left\{\begin{aligned} x_1 + 4x_2 &= 2 \\ -\lambda + 2x_1 \qquad &= 0 \\ -4\lambda \qquad + 2x_2 &= 0 \end{aligned}\right.$$

El valor estacionario de Z, definido por la solución

$$\lambda^* = \tfrac{4}{17} \qquad x_1^* = \tfrac{2}{17} \qquad x_2^* = \tfrac{8}{17}$$

es, por lo tanto, $Z^* = z^* = \frac{4}{17}$. Nuevamente deberemos consultar una condición de segundo orden antes de poder afirmar si z^* es un máximo o un mínimo.

El enfoque de la diferencial total

En el estudio del extremo libre de $z = f(x, y)$, aprendimos que la condición necesaria de primer orden puede establecerse en términos de la diferencial total dz como sigue:

$$dz = f_x\, dx + f_y\, dy = 0 \tag{12.9}$$

Esta proposición permanece válida después de añadir una restricción $g(x, y) = c$. Sin embargo, con la intervención de la restricción, ya no podemos considerar dx y dy como cambios "arbitrarios" como antes. Si $g(x, y) = c$, entonces dg debe ser igual a dc, el cual es cero ya que c es constante. Entonces,

$$(dg =) g_x\, dx + g_y\, dy = 0 \tag{12.10}$$

y esta relación hace que dx y dy sean dependientes entre sí. Por lo tanto, la condición necesaria de primer orden se convierte en $dz = 0$ [(12.9)], sujeto a $g = c$, y, por lo tanto, sujeto también a $dg = 0$ [(12.10)]. Tomando en cuenta (12.9) y (12.10), debe ser evidente que con objeto de satisfacer la condición necesaria, debemos tener

$$\frac{f_x}{g_x} = \frac{f_y}{g_y} \tag{12.11}$$

Este resultado puede verificarse al despejar dy de (12.10) y sustituir su resultado en (12.9). La condición (12.11), junto con la restricción $g(x, y) = c$, proporcionan dos ecuaciones a partir de las cuales se encuentran los valores críticos de x y y.[3]

¿Aporta el enfoque de la diferencial total la misma condición de primer orden que el método de los multiplicadores de Lagrange? Comparemos (12.8) con el resultado recién obtenido. La primera ecuación de (12.8) simplemente repite la restricción; el nuevo resultado requiere

[3] Observe que todavía debe considerarse la restricción $g = c$ junto con (12.11), aun cuando hemos utilizado la ecuación $dg = 0$ —es decir, (12.10)— al derivar (12.11). Aun cuando $g = c$ implica necesariamente $dg = 0$, el inverso no es verdad: $dg = 0$ implica meramente $g =$ a una constante (no necesariamente c). A menos que la restricción se considere implícitamente, por lo tanto, alguna información inadvertidamente se dejará fuera del problema.

también que se satisfaga. Las últimas dos ecuaciones de (12.8) pueden reescribirse, respectivamente, como

$$\frac{f_x}{g_x} = \lambda \qquad \text{y} \qquad \frac{f_y}{g_y} = \lambda \tag{12.11'}$$

y éstas proporcionan precisamente la misma información que (12.11). Observe, sin embargo, que mientras que el enfoque de la diferencial total suministra solamente los valores de x^* y y^*, el método de los multiplicadores de Lagrange también da el valor de λ^* como un subproducto directo. Como resultado, λ^* proporciona una medición de la sensibilidad de Z^* (y z^*) a un cambio en la restricción, como se demostrará. Por lo tanto, el método de los multiplicadores de Lagrange ofrece la ventaja de contener cierta información de estática comparativa incorporada en la solución.

Una interpretación de los multiplicadores de Lagrange

Para mostrar que λ^* realmente mide la sensibilidad de Z^* a los cambios en la restricción, realicemos un análisis comparativo estático de la condición de primer orden (12.8). Como λ, x y y son endógenas, la única variable exógena disponible es el parámetro de restricción c. Un cambio en c provocaría un desplazamiento de la curva de restricciones en el plano xy y con ello se alteraría la solución óptima. En especial, el efecto de un *incremento* en c (un presupuesto mayor, o una cuota de producción mayor) indicaría la forma en que una *relajación* de la restricción afecta la solución óptima.

Para hacer el análisis comparativo estático, empleamos nuevamente el teorema de la función implícita. Haciendo que las tres ecuaciones de (12.8) adopten la forma de $F^j(\lambda, x, y; c) = 0$ (con $j = 1, 2, 3$), y suponiendo que tengan derivadas parciales continuas, debemos verificar primero que el siguiente jacobiano de variables endógenas (donde $f_{xy} = f_{yx}$ y $g_{xy} = g_{yx}$)

$$|J| = \begin{vmatrix} \dfrac{\partial F^1}{\partial \lambda} & \dfrac{\partial F^1}{\partial x} & \dfrac{\partial F^1}{\partial y} \\[2ex] \dfrac{\partial F^2}{\partial \lambda} & \dfrac{\partial F^2}{\partial x} & \dfrac{\partial F^2}{\partial y} \\[2ex] \dfrac{\partial F^3}{\partial \lambda} & \dfrac{\partial F^3}{\partial x} & \dfrac{\partial F^3}{\partial y} \end{vmatrix} = \begin{vmatrix} 0 & -g_x & -g_y \\ -g_x & f_{xx} - \lambda g_{xx} & f_{xy} - \lambda g_{xy} \\ -g_y & f_{xy} - \lambda g_{xy} & f_{yy} - \lambda g_{yy} \end{vmatrix} \tag{12.12}$$

no se anule en el estado óptimo. En este momento, ciertamente no hay sospecha de que éste sería el caso; sin embargo, nuestra experiencia previa con la estática comparativa de los problemas de optimización [ver la controversia de (11.46)] sugeriría que este jacobiano está estrechamente relacionado con la condición suficiente de segundo orden, y si se satisface la condición suficiente, entonces el jacobiano será diferente de cero para el equilibrio (óptimo). Dejando por el momento la demostración completa de este hecho para la sección 12.3, prosigamos con la hipótesis de que $|J| \neq 0$. Si así es, entonces podemos expresar a λ^*, x^* y y^* como funciones implícitas del parámetro c:

$$\lambda^* = \lambda^*(c) \qquad x^* = x^*(c) \qquad \text{y} \qquad y^* = y^*(c) \tag{12.13}$$

las cuales, todas, tienen derivadas continuas. También tenemos las identidades de equilibrio

$$\begin{aligned} c - g(x^*, y^*) &\equiv 0 \\ f_x(x^*, y^*) - \lambda^* g_x(x^*, y^*) &\equiv 0 \\ f_y(x^*, y^*) - \lambda^* g_y(x^*, y^*) &\equiv 0 \end{aligned} \tag{12.14}$$

Ahora como el valor óptimo de Z depende de λ^*, x^* y y^*, es decir,

$$Z^* = f(x^*, y^*) + \lambda^*[c - g(x^*, y^*)] \qquad \textbf{(12.15)}$$

en vista de (12.13), podemos considerar que Z^* es sólo una función de c. Al obtener la diferencial total de Z^* respecto a c, encontramos

$$\begin{aligned}\frac{dZ^*}{dc} &= f_x\frac{dx^*}{dc} + f_y\frac{dy^*}{dc} + [c - g(x^*, y^*)]\frac{d\lambda^*}{dc} + \lambda^*\left(1 - g_x\frac{dx^*}{dc} - g_y\frac{dy^*}{dc}\right)\\ &= (f_x - \lambda^* g_x)\frac{dx^*}{dc} + (f_y - \lambda^* g_y)\frac{dy^*}{dc} + [c - g(x^*, y^*)]\frac{d\lambda^*}{dc} + \lambda^*\end{aligned}$$

donde f_x, f_y, g_x, y g_y deben evaluarse en el óptimo. Sin embargo, mediante (12.14) se cancelan los primeros tres términos de la derecha. Entonces, queda el resultado sencillo

$$\frac{dZ^*}{dc} = \lambda^* \qquad \textbf{(12.16)}$$

que valida nuestra afirmación de que el valor solución de los multiplicadores de Lagrange constituye una medida del efecto de un cambio en la restricción vía el parámetro c en el valor óptimo de la función objetivo.

Sin embargo, aquí debemos tener precaución. Para esta interpretación de λ^*, debe expresarse a Z específicamente como en (12.7). En especial, escribe el último término como $\lambda[c - g(x, y)]$, *no como* $\lambda[g(x, y) - c]$.

Casos de *n* variables y de restricciones múltiples

La generalización del método de los multiplicadores de Lagrange a n variables puede desarrollarse fácilmente si escribimos las variables de elección con notación de subíndice. Entonces, la función objetivo adopta la forma

$$z = f(x_1, x_2, \ldots, x_n)$$

sujeto a la restricción

$$g(x_1, x_2, \ldots, x_n) = c$$

De lo que se concluye que la función de Lagrange será

$$Z = f(x_1, x_2, \ldots, x_n) + \lambda[c - g(x_1, x_2, \ldots, x_n)]$$

para la cual la condición de primer orden consiste en las siguientes $(n + 1)$ ecuaciones simultáneas

$$\begin{aligned}Z_\lambda &= c - g(x_1, x_2, \ldots, x_n) = 0\\ Z_1 &= f_1 - \lambda g_1 = 0\\ Z_2 &= f_2 - \lambda g_2 = 0\\ &\cdots\cdots\cdots\cdots\cdots\cdots\\ Z_n &= f_n - \lambda g_n = 0\end{aligned}$$

Nuevamente, la primera de estas ecuaciones nos asegura que la restricción se cumple, aún cuando debamos centrar nuestra atención en la función lagrangiana *libre*.

Cuando hay más de una restricción, es igualmente aplicable el método de los multiplicadores de Lagrange, pero siempre que introduzcamos tantos multiplicadores como restricciones haya en la función lagrangiana. Sea una función de n variables que esté sujeta simultáneamente a las dos restricciones

$$g(x_1, x_2, \dots, x_n) = c \qquad \text{y} \qquad h(x_1, x_2, \dots, x_n) = d$$

Entonces, al adoptar λ y μ (la letra griega mi) como los dos multiplicadores indeterminados, podemos construir una función lagrangiana como sigue:

$$Z = f(x_1, x_2, \dots, x_n) + \lambda[c - g(x_1, x_2, \dots, x_n)] + \mu[d - h(x_1, x_2, \dots, x_n)]$$

Esta función tiene el mismo valor que la función objetivo original f si se satisfacen ambas restricciones; es decir, si se anulan los dos últimos términos en la función lagrangiana. Considerando a λ y μ como variables de elección, contamos ahora $(n+2)$ variables; entonces, en este caso la condición de primer orden consiste en las siguientes $(n+2)$ ecuaciones simultáneas:

$$\begin{aligned} Z_\lambda &= c - g(x_1, x_2, \dots, x_n) = 0 \\ Z_\mu &= d - h(x_1, x_2, \dots, x_n) = 0 \\ Z_i &= f_i - \lambda g_i - \mu h_i = 0 \qquad (i = 1, 2, \dots, n) \end{aligned}$$

Normalmente, estas ecuaciones nos permiten resolver para los valores de x_i, así como de λ y μ. Como antes, las dos primeras ecuaciones de la condición necesaria representan la reiteración de las dos restricciones.

EJERCICIO 12.2

1. Use el método de los multiplicadores de Lagrange para encontrar los valores estacionarios de z:
   (*a*) $z = xy$, sujeto a $x + 2y = 2$.
   (*b*) $z = x(y+4)$, sujeto a $x + y = 8$.
   (*c*) $z = x - 3y - xy$, sujeto a $x + y = 6$.
   (*d*) $z = 7 - y + x^2$, sujeto a $x + y = 0$.
2. En el problema anterior, encuentre si una ligera relajación de la restricción aumenta o disminuye el valor óptimo de z. ¿A qué tasa?
3. Escriba la función lagrangiana y la condición de primer orden para los valores estacionarios (sin resolver las ecuaciones) para cada una de las siguientes funciones objetivo z:
   (*a*) $z = x + 2y + 3w + xy - yw$, sujeto a $x + y + 2w = 10$.
   (*b*) $z = x^2 + 2xy + yw^2$, sujeto a $2x + y + w^2 = 24$ y $x + w = 8$.
4. Si en lugar de $g(x, y) = c$, la restricción se escribe en la forma de $G(x, y) = 0$, ¿cómo deben modificarse la función lagrangiana y, como consecuencia, la condición de primer orden?
5. Al estudiar el enfoque de la diferencial total, señalamos que, dada la restricción $g(x, y) = c$, podemos deducir que $dg = 0$; por la misma razón, podemos deducir además que $d^2g = d(dg) = d(0) = 0$. Sin embargo, en nuestro estudio anterior del extremo no restringido de una función $z = f(x, y)$, teníamos una situación en la cual $dz = 0$ está acompañada por un d^2z positivo definido o negativo definido, en lugar de $d^2z = 0$. ¿Cómo explicaría esta disparidad de tratamiento en los dos casos?
6. Si la función de Lagrange se escribe como $Z = f(x, y) + \lambda[g(x, y) - c]$ en vez de como en (12.7), ¿puede interpretarse todavía al multiplicador de Lagrange como en (12.16)? Complete la nueva interpretación, si hay alguna.

12.3 Condiciones de segundo orden

La introducción de un multiplicador de Lagrange como una variable adicional permite aplicar al problema del extremo restringido la misma condición de primer orden que se usa en el problema del extremo libre. Podríamos dar un paso adicional y tomar prestadas también las condiciones suficientes y necesarias de segundo orden, pero no debemos hacerlo. Porque aun cuando Z^* sea en realidad un tipo estándar de extremo respecto a las variables de elección, *no* lo es respecto al multiplicador de Lagrange. Específicamente, partiendo de (12.15) podemos ver que, a diferencia de x^* y y^*, si λ^* se reemplaza por cualquier otro valor de λ, no se produce ningún efecto sobre Z^*, ya que $[c - g(x^*, y^*)]$ es idénticamente cero. Entonces, el papel que juega λ en la solución óptima difiere básicamente del de x y y.[4] Aun cuando no hay problema en tratar a λ como solamente otra variable de elección en el estudio de la condición de primer orden, debemos tener cuidado de no aplicar a ciegas las condiciones de segundo orden desarrolladas para el problema del extremo libre al caso restringido presente. En lugar de esto, debemos obtener un conjunto de condiciones nuevas. Como veremos, las condiciones nuevas pueden establecerse una vez más en términos de la diferencial total de segundo orden d^2z. Sin embargo, la presencia de la restricción implica ciertas modificaciones importantes en el criterio.

Diferencial total de segundo orden

Hemos mencionado que, como la restricción $g(x, y) = c$ significa que $dg = g_x\, dx + g_y\, dy = 0$, como en (12.10), dx y dy ya no son arbitrarias. Por supuesto que todavía podemos tomar (por ejemplo) a dx como un cambio arbitrario, pero entonces dy debe considerarse como dependiente de dx, que debe escogerse siempre de modo que satisfaga a (12.10); es decir, que satisfaga a $dy = -(g_x/g_y)\, dx$. Visto desde otro ángulo, una vez que se especifica el valor de dx, dy depende de g_x y g_y, pero ya que estas últimas derivadas dependen, a su vez, de las variables x y y, dy también depende de x y y. Es obvio, pues, que la fórmula anterior de d^2z de (11.6), que se basa en la condición arbitraria de dx y dy, ya no es aplicable.

A fin de encontrar una nueva expresión apropiada para d^2z, debemos tratar a dy como una variable que depende de x y y durante la diferenciación (si dx va a considerarse como una constante). Entonces,

$$\begin{aligned}
d^2z = d(dz) &= \frac{\partial(dz)}{\partial x}\, dx + \frac{\partial(dz)}{\partial y}\, dy \\
&= \frac{\partial}{\partial x}(f_x\, dx + f_y\, dy)dx + \frac{\partial}{\partial y}(f_x\, dx + f_y\, dy)\, dy \\
&= \left[f_{xx}\, dx + \left(f_{xy}\, dy + f_y\, \frac{\partial dy}{\partial x}\right)\right] dx + \left[f_{yx}\, dx + \left(f_{yy}\, dy + f_y\, \frac{\partial dy}{\partial y}\right)\right] dy \\
&= f_{xx}\, dx^2 + f_{xy}\, dy\, dx + f_y \frac{\partial(dy)}{\partial x}\, dx + f_{yx}\, dx\, dy + f_{yy}\, dy^2 + f_y \frac{\partial(dy)}{\partial y}\, dy
\end{aligned}$$

[4] En un marco más general de optimización restringida conocida como "programación no lineal", que se va a estudiar en el capítulo 13, se mostrará que, con restricciones de desigualdad, si Z^* es un máximo (mínimo), respecto a x y y, entonces de hecho será un mínimo (máximo) respecto a λ. En otras palabras, el punto (λ^*, x^*, y^*) es un punto silla. El presente caso —donde Z^* es un extremo genuino respecto a x y y, pero es invariante respecto a λ— puede considerarse como un caso degenerado del punto silla. La naturaleza de punto silla de la solución (λ^*, x^*, y^*) también conduce al importante concepto de "dualidad". Pero es preferible tratar este tema posteriormente.

Puesto que el tercero y sexto términos pueden reducirse a

$$f_y\left[\frac{\partial(dy)}{\partial x}\,dx+\frac{\partial(dy)}{\partial y}\,dy\right]=f_y\,d(dy)=f_y\,d^2y$$

la expresión deseada para d^2z es

$$d^2z = f_{xx}\,dx^2 + 2f_{xy}\,dx\,dy + f_{yy}\,dy^2 + f_y\,d^2y \qquad \textbf{(12.17)}$$

que difiere de (11.6) sólo por el último término, $f_y\,d^2y$.

Observe que este último término está en *primer* grado [d^2y *no* es lo mismo que $(dy)^2$], así que su presencia en (12.17) descalifica a d^2z como forma cuadrática. Sin embargo, d^2z puede transformarse en una forma cuadrática en virtud de la restricción $g(x, y) = c$. Dado que la restricción implica que $dg = 0$ y también que $d^2g = d(dg) = 0$, entonces, por el procedimiento usado al obtener (12.17) tendremos

$$(d^2g =)g_{xx}\,dx^2 + 2g_{xy}\,dx\,dy + g_{yy}\,dy^2 + g_y\,d^2y = 0$$

Al despejar d^2y de esta última ecuación y al sustituir el resultado en (12.17), podemos eliminar la expresión de primer grado d^2y y escribir d^2z como la siguiente forma cuadrática:

$$d^2z = \left(f_{xx} - \frac{f_y}{g_y}g_{xx}\right)dx^2 + 2\left(f_{xy} - \frac{f_y}{g_y}g_{xy}\right)dx\,dy + \left(f_{yy} - \frac{f_y}{g_y}g_{yy}\right)dy^2$$

A causa de (12.11′), el primer coeficiente que está entre paréntesis se reduce a $(f_{xx} - \lambda g_{xx})$, y en forma similar en los otros términos. Sin embargo, al diferenciar parcialmente las derivadas en (12.8), encontrarás que las siguientes segundas derivadas

$$\begin{aligned} Z_{xx} &= f_{xx} - \lambda g_{xx} \\ Z_{xy} &= f_{xy} - \lambda g_{xy} = Z_{yx} \\ Z_{yy} &= f_{yy} - \lambda g_{yy} \end{aligned} \qquad \textbf{(12.18)}$$

son precisamente iguales a estos coeficientes que están entre paréntesis. Entonces, al usar la ecuación lagrangiana, podemos expresar finalmente a d^2z en una forma más conveniente, como sigue:

$$\begin{aligned} d^2z = \quad & Z_{xx}\,dx^2 \quad + Z_{xy}\,dx\,dy \\ & + Z_{yx}\,dy\,dx + Z_{yy}\,dy^2 \end{aligned} \qquad \textbf{(12.17′)}$$

Los coeficientes de (12.17′) son simplemente las segundas derivadas parciales de Z respecto a las variables de elección x y y; por lo tanto, al juntarse, pueden dar origen a un determinante hessiano.

Condiciones de segundo orden

Para un extremo restringido de $z = f(x, y)$, sujeto a $g(x, y) = c$, las condiciones necesarias y suficientes de segundo orden todavía giran alrededor del signo algebraico de la diferencial total de segundo orden d^2z, evaluada en un punto estacionario. Sin embargo, hay un cambio importante. En el presente contexto nos interesa la definitividad o semidefinitivad del signo de d^2z, *no* para todos los valores posibles de dx y dy (ambos no pueden ser cero), sino *solamente* para los valores de dx y dy (ambos no pueden ser cero) que satisfagan la restricción lineal (12.10), $g_x dx + g_y dy = 0$. Entonces, las condiciones *necesarias* de segundo orden son

Para el máximo de z: d^2z negativo semidefinido, sujeto a $dg = 0$

Para el mínimo de z: d^2z positivo semidefinido, sujeto a $dg = 0$

y las condiciones *suficientes* de segundo orden son:

Para el máximo de z: d^2z negativo definido, sujeto a $dg = 0$

Para el mínimo de z: d^2z positivo definido, sujeto a $dg = 0$

En los párrafos siguientes nos concentraremos en las condiciones suficientes de segundo orden.

Siempre que los pares (dx, dy) que satisfacen la restricción $g_x\, dx + g_y\, dy = 0$ constituyan sólo un subconjunto del conjunto de todos los dx y dy posibles, la definitividad del signo de la restricción es menos estricta –es decir, más fácil de satisfacer– que la definitividad del signo sin restricción que estudiamos en el capítulo 11. En otras palabras, la condición suficiente de segundo orden para un problema de extremo restringido es una condición más débil que la de un problema de extremo libre. Estas son buenas noticias porque, a diferencia de las condiciones necesarias que deben ser estrictas para que sirvan como dispositivos efectivos de discriminación, las condiciones suficientes deben ser débiles para que realmente sean útiles.[5]

El hessiano orlado

Como en el caso del extremo libre, la condición suficiente de segundo orden se puede expresar en forma de determinante. Sin embargo, en lugar del determinante hessiano $|H|$, en el caso del extremo restringido encontraremos lo que se conoce como *hessiano orlado*.

Como preparación para el desarrollo de esta idea, analicemos primero las condiciones para la definitividad del signo de una forma cuadrática de dos variables, sujeta a una restricción lineal, por ejemplo,

$$q = au^2 + 2\,huv + bv^2 \qquad \text{sujeto a} \qquad \alpha u + \beta v = 0$$

Como la restricción implica $v = -(\alpha/\beta)u$, podemos reescribir q como una función sólo de una variable:

$$q = au^2 - 2h\frac{\alpha}{\beta}u^2 + b\frac{\alpha^2}{\beta^2}u^2 = (a\beta^2 - 2h\alpha\beta + b\alpha^2)\frac{u^2}{\beta^2}$$

Es obvio que q es positivo (negativo) definido si y sólo si la expresión que está entre paréntesis es positiva (negativa). Ahora resulta que el siguiente determinante simétrico

$$\begin{vmatrix} 0 & \alpha & \beta \\ \alpha & a & h \\ \beta & h & b \end{vmatrix} = 2h\alpha\beta - a\beta^2 - b\alpha^2$$

es exactamente el *negativo* de la expresión que está entre paréntesis. En consecuencia, podemos afirmar que

$$q \text{ si } \begin{Bmatrix} \text{positivo definido} \\ \text{negativo definido} \end{Bmatrix} \text{ sujeto a } \quad \alpha u + \beta v = 0 \text{ si y sólo si } \begin{vmatrix} 0 & \alpha & \beta \\ \alpha & a & h \\ \beta & h & b \end{vmatrix} \begin{cases} < 0 \\ > 0 \end{cases}$$

[5] "Un depósito bancario de un millón de dólares" es claramente una condición suficiente para "poder costearse una cena con filete". Pero la aplicabilidad muy limitada de esa condición la hace prácticamente inútil. Una condición suficiente más significativa puede ser algo como "cincuenta dólares", que es un requerimiento financiero mucho menos demandante.

Vale la pena observar que el determinante usado para este criterio no es otra cosa más que el discriminante de la forma cuadrática original $\begin{vmatrix} a & h \\ h & b \end{vmatrix}$, con una orla colocada en la parte superior y una orla similar en la izquierda. Aún más, la orla está compuesta simplemente por los dos coeficientes α y β que provienen de la restricción, más un cero en la diagonal principal. Este discriminante orlado es simétrico.

Ejemplo 1

Determine si $q = 4u^2 + 4uv + 3v^2$, sujeto a $u - 2v = 0$, es positivo o negativo definido. Primero formamos el discriminante orlado $\begin{vmatrix} 0 & 1 & -2 \\ 1 & 4 & 2 \\ -2 & 2 & 3 \end{vmatrix}$, que se hace simétrico dividiendo el coeficiente de uv en dos partes iguales para su inserción en el determinante. Siempre que el determinante tenga un valor negativo (-27), q debe ser positivo definido.

Cuando se aplica a la forma cuadrática d^2z en (12.17′), las variables u y v se transforman en dx y dy, respectivamente, y el determinante (simple) consiste en el hessiano $\begin{vmatrix} Z_{xx} & Z_{xy} \\ Z_{yx} & Z_{yy} \end{vmatrix}$. Aún más, al ser $g_x\,dx + g_y\,dy = 0$ la restricción para la forma cuadrática, , tenemos $\alpha = g_x$ y $\beta = g_y$. Entonces, para los valores de dx y dy que satisfagan esta restricción, disponemos ahora del siguiente criterio del determinante para la definitividad del signo de d^2z:

$$d^2z \text{ si } \begin{Bmatrix} \text{positivo definido} \\ \text{negativo definido} \end{Bmatrix} \text{ sujeto a } \quad dg = 0 \text{ si y sólo si } \begin{vmatrix} 0 & g_x & g_y \\ g_x & Z_{xx} & Z_{xy} \\ g_y & Z_{yx} & Z_{yy} \end{vmatrix} \begin{Bmatrix} < 0 \\ > 0 \end{Bmatrix}$$

El determinante de la derecha, con frecuencia denominado hessiano orlado, se denota por $|\overline{H}|$, donde la barra superior simboliza la orla. Con base en esto, concluimos que, dado un valor estacionario de $z = f(x, y)$ o de $Z = f(x, y) + \lambda[c - g(x, y)]$, un valor positivo de $|\overline{H}|$ es suficiente para establecerlo como un máximo relativo de z; en forma similar, un valor negativo de $|\overline{H}|$ es suficiente para establecerlo como un mínimo —evaluándose a todas las derivadas que intervienen en $|\overline{H}|$ para los valores críticos de x y y.

Ahora que hemos llegado a la condición suficiente de segundo orden, es muy sencillo verificar que, como se explicó anteriormente, la satisfacción de esta condición garantiza que el jacobiano de variables endógenas (12.12) no se anula en el estado óptimo. Al sustituir (12.18) en (12.12), y al multiplicar por -1 tanto la primera columna como la primera fila del jacobiano (que deja inalterado el valor del determinante), vemos que

$$|J| = \begin{vmatrix} 0 & g_x & g_y \\ g_x & Z_{xx} & Z_{xy} \\ g_y & Z_{yx} & Z_{yy} \end{vmatrix} = |\overline{H}| \qquad \textbf{(12.19)}$$

es decir, el jacobiano de variables endógenas es idéntico al hessiano orlado —un resultado similar a (11.42) en el cual se mostró que, en el contexto del extremo libre, el jacobiano de variables endógenas es idéntico al hessiano simple. En cumplimiento de la condición suficiente, si tenemos $|\overline{H}| \neq 0$ para el óptimo, entonces $|J|$ también debe ser diferente de cero. En consecuencia, al aplicar el teorema de la función implícita en este contexto, se podría sustituir la condición $|\overline{H}| \neq 0$ en lugar de la condición habitual $|J| \neq 0$. Seguiremos esta práctica cuando analicemos la estática comparativa de los problemas de optimización restringida en la sección 12.5.

Ejemplo 2 Regresemos al ejemplo 1 de la sección 12.2 y analicemos si el valor estacionario encontrado ahí da un máximo o un mínimo. Como $Z_x = y - \lambda$ y $Z_y = x - \lambda$, las derivadas parciales de segundo orden son $Z_{xx} = 0$, $Z_{xy} = Z_{yx} = 1$, y $Z_{yy} = 0$. Los elementos de la orla que necesitamos son $g_x = 1$ y $g_y = 1$. Entonces, encontramos que

$$|\overline{H}| = \begin{vmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{vmatrix} = 2 > 0$$

que establece como máximo el valor $z^* = 9$.

Ejemplo 3 Continuando con el ejemplo 2 de la sección 12.2, vemos que $Z_1 = 2x_1 - \lambda$ y $Z_2 = 2x_2 - 4\lambda$. De aquí se obtiene que $Z_{11} = 2$, $Z_{12} = Z_{21} = 0$, y $Z_{22} = 2$. De la restricción $x_1 + 4x_2 = 2$, obtenemos $g_1 = 1$ y $g_2 = 4$. Se concluye que el hessiano orlado es

$$|\overline{H}| = \begin{vmatrix} 0 & 1 & 4 \\ 1 & 2 & 0 \\ 4 & 0 & 2 \end{vmatrix} = -34 < 0$$

y el valor $z^* = \frac{4}{17}$ es un mínimo.

Ejemplo 4 Consideremos un modelo sencillo de dos periodos en el cual la utilidad del consumidor es una función del consumo para ambos periodos. Sea la función de utilidad del consumidor

$$U(x_1, x_2) = x_1 x_2$$

donde x_1 es el consumo en el periodo 1 y x_2 es el consumo en el periodo 2. El consumidor también dispone de un presupuesto B al inicio del periodo 1.

Sea r la tasa de interés de mercado con la cual el consumidor puede elegir pedir prestado o prestar a través de los dos periodos. La restricción del presupuesto intertemporal del consumidor es que x_1 y el valor presente de x_2 suman B. Entonces,

$$x_1 + \frac{x_2}{1+r} = B$$

El lagrangiano para este problema de maximización de utilidad es

$$Z = x_1 x_2 + \lambda\left(B - x_1 - \frac{x_2}{1+r}\right)$$

con las condiciones de primer orden

$$\frac{\partial Z}{\partial \lambda} = B - x_1 - \frac{x_2}{1+r} = 0$$

$$\frac{\partial Z}{\partial x_1} = x_2 - \lambda = 0$$

$$\frac{\partial Z}{\partial x_2} = x_1 - \frac{\lambda}{1+r} = 0$$

La combinación de las dos últimas ecuaciones de primer orden para eliminar λ nos da

$$\frac{x_2}{x_1} = \frac{\lambda}{\lambda/(1+r)} = 1 + r$$

La sustitución de esta ecuación en la restricción del presupuesto proporciona entonces la solución

$$x_1^* = \frac{B}{2} \qquad \text{y} \qquad x_2^* = \frac{B(1+r)}{2}$$

Debemos verificar enseguida la condición suficiente de segundo orden para encontrar un máximo. El hessiano orlado para este problema es

$$|\overline{H}| = \begin{vmatrix} 0 & -1 & -\dfrac{1}{1+r} \\ -1 & 0 & 1 \\ -\dfrac{1}{1+r} & 1 & 0 \end{vmatrix} = \frac{2}{1+r} > 0$$

Entonces la condición suficiente de segundo orden se satisface para un máximo U.

El caso de *n* variables

Cuando la función objetivo adopta la forma

$$z = f(x_1, x_2, \ldots, x_n) \qquad \text{sujeto a} \qquad g(x_1, x_2, \ldots, x_n) = c$$

la condición de segundo orden todavía depende del signo de d^2z. Como esta última es una forma cuadrática restringida para las variables $dx_1, dx_2, \ldots, dx_n$, sujeta a la relación

$$(dg =)g_1\,dx_1 + g_2\,dx_2 + \cdots + g_n\,dx_n = 0$$

las condiciones para la definitividad positiva o negativa de d^2z nuevamente incluyen a un hessiano orlado. Pero en esta ocasión estas condiciones deben expresarse en términos de los menores principales orlados directores del hessiano.

Dado un hessiano orlado

$$|\overline{H}| = \begin{vmatrix} 0 & g_1 & g_2 & \cdots & g_n \\ g_1 & Z_{11} & Z_{12} & \cdots & Z_{1n} \\ g_2 & Z_{21} & Z_{22} & \cdots & Z_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ g_n & Z_{n1} & Z_{n2} & \cdots & Z_{nn} \end{vmatrix}$$

sus menores principales orlados directores se definen como

$$|\overline{H}_2| \equiv \begin{vmatrix} 0 & g_1 & g_2 \\ g_1 & Z_{11} & Z_{12} \\ g_2 & Z_{21} & Z_{22} \end{vmatrix} \qquad |\overline{H}_3| \equiv \begin{vmatrix} 0 & g_1 & g_2 & g_3 \\ g_1 & Z_{11} & Z_{12} & Z_{13} \\ g_2 & Z_{21} & Z_{22} & Z_{23} \\ g_3 & Z_{31} & Z_{32} & Z_{33} \end{vmatrix} \qquad \text{(etc.)}$$

siendo el último $|\overline{H}_n| = |\overline{H}|$. Para los símbolos de nueva introducción, la barra horizontal arriba de H nuevamente significa orlado, y el subíndice indica el orden del menor principal director orlado. Por ejemplo, $|\overline{H}_2|$ incluye el segundo menor principal director del hessiano (simple), orlado con 0, g_1, y g_2; y en forma similar para los demás. Entonces, las condiciones para la definitividad positiva y negativa de d^2z son

$$d^2z \text{ si} \left\{ \begin{matrix} \text{positivo definido} \\ \text{negativo definido} \end{matrix} \right\} \text{sujeto a} \quad dg = 0 \text{ si y sólo si} \left\{ \begin{matrix} |\overline{H}_2|, |\overline{H}_3|, \ldots, |\overline{H}_n| < 0 \\ |\overline{H}_2| > 0;\ |\overline{H}_3| < 0;\ |\overline{H}_4| > 0; \text{ etc.} \end{matrix} \right.$$

En el primer caso, todos los menores principales directores orlados, comenzando con $|\overline{H}_2|$, deben ser negativos; en el último caso, deben alternar el signo. Como antes, un d^2z definido

TABLA 12.1 **Prueba del determinante para el extremo restringido relativo: $z = f(x_1, x_2, \ldots, x_n)$, sujeto a $g(x_1, x_2, \ldots, x_n) = c$; con $Z = f(x_1, x_2, \ldots, x_n) + \lambda[c - g(x_1, x_2, \ldots, x_n)]$**

Condición	Máximo	Mínimo
Condición necesaria de primer orden	$Z_\lambda = Z_1 = Z_2 = \cdots = Z_n = 0$	$Z_\lambda = Z_1 = Z_2 = \cdots = Z_n = 0$
Condición suficiente de segundo orden[1]	$\lvert\overline{H}_2\rvert > 0$; $\lvert\overline{H}_3\rvert < 0$; $\lvert\overline{H}_4\rvert > 0; \ldots; (-1)^n\lvert\overline{H}_n\rvert > 0$	$\lvert\overline{H}_2\rvert, \lvert\overline{H}_3\rvert, \ldots, \lvert\overline{H}_n\rvert < 0$

[1] Aplicable solamente después de haber satisfecho la condición necesaria de primer orden.

positivo es suficiente para establecer un valor estacionario de z como su mínimo, mientras que un d^2z definido negativo es suficiente para establecerlo como máximo.

Si agrupamos los conceptos de este estudio, en la tabla 12.1 podemos resumir las condiciones para un extremo relativo restringido. Sin embargo, debemos reconocer que el criterio establecido en la tabla no es completo. Debido a que la condición suficiente de segundo orden no es necesaria, el hecho de que no se satisfagan los criterios establecidos no excluye la posibilidad de que el valor estacionario sea de todas maneras un máximo o un mínimo, cualquiera que sea el caso. Sin embargo, en muchas aplicaciones económicas, esta condición suficiente de segundo orden (relativamente menos estricta) se satisface, o se supone que se satisface, de tal manera que la información de la tabla es adecuada. Sugerimos que compare los resultados contenidos en la tabla 12.1, con los de la tabla 11.2 que se refiere al caso de extremos libres.

El caso de las restricciones múltiples

Cuando aparece más de una restricción en el problema, la condición de segundo orden incluye un hessiano con más de una orla. Suponga que hay n variables de elección y m restricciones $m < n$ de forma $g^j(x_1, \ldots, x_n) = c_j$. Entonces, la función lagrangiana es

$$Z = f(x_1, \ldots, x_n) + \sum_{j=1}^{m} \lambda_j [c_j - g^j(x_1, \ldots, x_n)]$$

y el hessiano orlado resulta ser

$$|\overline{H}| \equiv \left|\begin{array}{cccc:cccc} 0 & 0 & \cdots & 0 & g_1^1 & g_2^1 & \cdots & g_n^1 \\ 0 & 0 & \cdots & 0 & g_1^2 & g_2^2 & \cdots & g_n^2 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & g_1^m & g_2^m & \cdots & g_n^m \\ \hdashline g_1^1 & g_1^2 & \cdots & g_1^m & Z_{11} & Z_{12} & \cdots & Z_{1n} \\ g_2^1 & g_2^2 & \cdots & g_2^m & Z_{21} & Z_{22} & \cdots & Z_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ g_n^1 & g_n^2 & \cdots & g_n^m & Z_{n1} & Z_{n2} & \cdots & Z_{nn} \end{array}\right|$$

donde $g_i^j \equiv \partial g^j/\partial x_i$ son las derivadas parciales de las funciones de restricción, y los símbolos con subíndice doble denotan, como antes, las derivadas parciales de segundo orden de la función lagrangiana. Observe que, para claridad visual, se ha dividido al hessiano orlado en cuatro *áreas*. El área superior izquierda consta solamente de ceros, y el área inferior derecha es el hessiano simple. Las otras dos áreas, que contienen las derivadas g_i^j, guardan entre sí una relación de imagen de espejo con referencia a la diagonal principal, dando como resultado un arreglo simétrico de elementos en el hessiano orlado completo.

A partir de $|\overline{H}|$ pueden formarse diversos menores principales directores orlados. El que contiene a Z_{22} como el último elemento en la diagonal principal se denomina $|\overline{H}_2|$, como antes. Al incluir una fila y una columna adicionales, de modo que entra en escena Z_{33}, tendremos $|\overline{H}_3|$, etc. Con esta simbología, podemos establecer la condición suficiente de segundo orden en términos de los signos de los siguientes $(n - m)$ menores principales directores orlados:

$$|\overline{H}_{m+1}|, |\overline{H}_{m+2}|, \ldots, |\overline{H}_n| (= |\overline{H}|)$$

Para un máximo de z, una condición suficiente es que en estos menores principales directores orlados se alterne el signo, siendo el signo de $|\overline{H}_{m+1}|$ igual a el de $(-1)^{m+1}$. Para un mínimo de z, una condición suficiente es que estos menores principales tomen todos el mismo signo, el de $(-1)^m$.

Observe que marca una diferencia importante el que tengamos un número impar o par de restricciones, ya que (-1) elevado a una potencia nos proporciona el signo opuesto al caso de una potencia par. Observe también que cuando $m = 1$, la condición recién establecida se reduce a la presentada en la tabla 12.1.

EJERCICIO 12.3

1. Use el hessiano orlado para determinar si el valor estacionario de z obtenido en cada parte del ejercicio 12.2-1 es un máximo o un mínimo.
2. Al establecer las condiciones suficientes de segundo orden para el máximo y el mínimo restringidos, especificamos los signos algebraicos de $|\overline{H}_2|$, $|\overline{H}_3|$, $|\overline{H}_4|$, etc., pero no de $|\overline{H}_1|$. Escriba una expresión apropiada para $|\overline{H}_1|$, y verifique que invariablemente adopta el signo negativo.
3. Recordando la propiedad II de los determinantes (sección 5.3), muestra que:

   (*a*) Al intercambiar en forma apropiada dos filas o dos columnas de $|\overline{H}_2|$ y al alterar apropiadamente el signo del determinante después de cada intercambio, puede trasformarse en

   $$\begin{vmatrix} Z_{11} & Z_{12} & g_1 \\ Z_{21} & Z_{22} & g_2 \\ g_1 & g_2 & 0 \end{vmatrix}$$

   (*b*) Mediante un procedimiento similar, $|\overline{H}_3|$ se transforma en

   $$\begin{vmatrix} Z_{11} & Z_{12} & Z_{13} & g_1 \\ Z_{21} & Z_{22} & Z_{23} & g_2 \\ Z_{31} & Z_{32} & Z_{33} & g_3 \\ g_1 & g_2 & g_3 & 0 \end{vmatrix}$$

   ¿Qué forma alternativa de "orlar" los menores principales del hessiano sugieren estos resultados?
4. Escriba el hessiano orlado para un problema de optimización restringida con cuatro variables de elección y dos restricciones. Entonces, establezca específicamente la condición suficiente de segundo orden para un máximo y para un mínimo de z, respectivamente.

12.4 Cuasiconcavidad y cuasiconvexidad

En la sección 11.5 se mostró que, para un problema de extremo libre, un conocimiento de la concavidad o la convexidad de la función objetivo evita la necesidad de verificar la condición de segundo orden. En el contexto de la optimización restringida, nuevamente es posible prescindir de la condición de segundo orden si la superficie o la hipersuperficie tienen el tipo apropiado de configuración. Pero esta vez la configuración deseada es la cuasiconcavidad (en lugar de concavidad) para un máximo, y la cuasiconvexidad (en lugar de convexidad) para un mínimo. Como demostraremos, la cuasiconcavidad (cuasiconvexidad) es una condición más débil que la concavidad (convexidad). Esto es de esperarse, ya que la condición suficiente de segundo orden que se va a omitir es también más débil para el problema de optimización restringida (d^2z definido en el signo solamente para aquellos dx_i que satisfagan $dg = 0$) que para el problema libre (d^2z definido en el signo para *todo* dx_i).

Caracterización geométrica

La cuasiconcavidad y la cuasiconvexidad, al igual que la concavidad y la convexidad, puede ser estricta o no estricta. Presentaremos primero la caracterización geométrica de estos conceptos:

> Sean u y v dos puntos cualquiera diferentes en el dominio (un conjunto convexo) de una función f, y sea el segmento de línea uv en el dominio que origina el arco MN en la gráfica de la función, tal que el punto N esté más alto que o tenga la misma altura que el punto M. Entonces, se dice que la función f es *cuasicóncava* (*cuasiconvexa*) si todos los puntos en el arco MN diferentes de M y N están más altos que o tienen la misma altura que el punto M (están más bajos que o tienen la misma altura que el punto N). Se dice que la función f es *estrictamente cuasicóncava* (*estrictamente cuasiconvexa*) si todos los puntos del arco MN diferentes de M y N están estrictamente a mayor altura que el punto M (estrictamente a una menor altura que el punto N).

A partir de esto debe ser claro que cualquier función estrictamente cuasicóncava (estrictamente cuasiconvexa) es cuasicóncava (cuasiconvexa), pero el inverso no es verdad.

Para comprenderlo mejor, examinemos las ilustraciones de la figura 12.3, trazadas para el caso de una variable. En la figura 12.3*a* el segmento de línea uv del dominio origina el arco MN en la curva tal que N está a mayor altura que M. Como todos los puntos entre M y N en el arco están estrictamente a mayor altura que M, este arco específico satisface la condición de la cuasiconcavidad estricta. Sin embargo, para que la curva califique como estrictamente cuasicóncava, *todos* los pares posibles (u, v) deben tener arcos que satisfagan la misma condición. Éste es realmente el caso para la función de la figura 12.3*a*. Observe que esta función también cumple la condición de la cuasiconcavidad (no estricta). Pero no cumple con la condición de cuasiconvexidad, porque algunos puntos del arco MN están a mayor altura que N, lo cual está prohibido para una función cuasiconvexa. La función de la figura 12.3*b* tiene la configuración opuesta. Todos los puntos del arco $M'N'$ están a menor altura que N', que es el más alto de los dos extremos, y lo mismo es cierto de todos los arcos que pueden trazarse. Entonces, la fun-

FIGURA 12.3

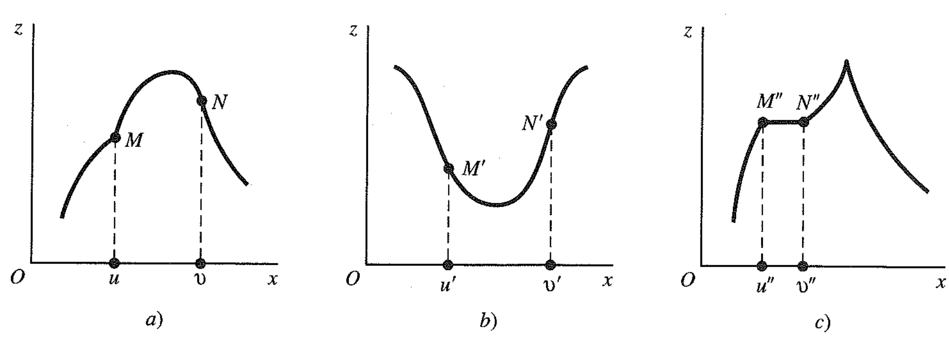
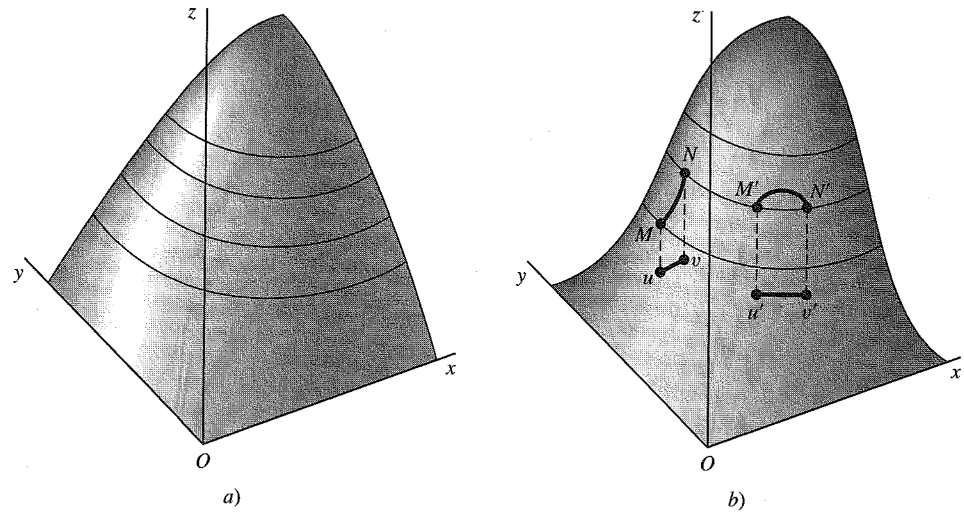


FIGURA 12.4



ción de la figura 12.3*b* es estrictamente cuasiconvexa. Como se puede verificar, esta también satisface la condición de cuasiconvexidad (no estricta), pero no cumple la condición de cuasiconcavidad. Lo que distingue a la figura 12.3*c* es la presencia de un segmento de línea horizontal $M''N''$, donde todos los puntos tienen la misma altura. Como resultado, ese segmento de línea —y, por lo tanto, la curva completa— puede cumplir solamente la condición de cuasiconcavidad, pero no la cuasiconcavidad estricta.

Desde un punto de vista general, una función cuasicóncava que tampoco es cóncava tiene una gráfica parecida a una campana, o a una parte de ella; y una función cuasiconvexa tiene una gráfica parecida a una campana invertida, o a una parte de ella. En una campana se admite (aunque no se requiere) tener segmentos cóncavos y convexos. Esta naturaleza más permisiva de la caracterización hace de la cuasiconcavidad (cuasiconvexidad) una condición más débil que la concavidad (convexidad). En la figura 12.4, contrastamos la concavidad estricta con la cuasiconcavidad estricta para el caso de dos variables. Tal como están dibujadas, ambas superficies describen funciones crecientes, ya que contienen solamente la parte ascendente de un domo y de una campana, respectivamente. La superficie de la figura 12.4*a* es estrictamente cóncava, pero la de la figura 12.4*b* no lo es, ya que contiene partes convexas cerca de la base de la campana. Aun así, es estrictamente cuasicóncava; todos los arcos de la superficie, ejemplificados por MN y por $M'N'$, satisfacen la condición de que todos los puntos en cada arco entre los dos puntos terminales están a mayor altura que el punto terminal más bajo. Regresando a la figura 12.4*a*, observamos que la superficie interior también es estrictamente cuasicóncava. Aunque no hemos dibujado ningún arco ilustrativo MN y $M'N'$ en la figura 12.4*a*, no es difícil verificar que en realidad todos los arcos posibles satisfacen la condición de cuasiconcavidad estricta. En general, una función estrictamente cóncava debe ser estrictamente cuasicóncava, aunque el inverso no es verdad. Demostraremos esto formalmente en el párrafo que sigue.

Definición algebraica

La caracterización geométrica anterior puede traducirse a una definición algebraica para una generalización más sencilla a casos de dimensiones superiores:

Una función f es $\left\{\begin{matrix}\text{cuasicóncava}\\ \text{cuasiconvexa}\end{matrix}\right\}$ si y sólo si para cualquier par de puntos diferentes u y v en dominio (conjunto convexo) de f, y para $0 < \theta < 1$,

$$f(v) \geq f(u) \Rightarrow f[\theta u + (1-\theta)v] \left\{\begin{matrix}\geq f(u)\\ \leq f(v)\end{matrix}\right\} \qquad \textbf{(12.20)}$$

Para adaptar esta definición a la cuasiconcavidad y cuasiconvexidad *estrictas*, las dos desigualdades débiles que están a la derecha deben transformarse en desigualdades estrictas $\left\{\begin{matrix} > f(u) \\ < f(v) \end{matrix}\right\}$. Sugerimos comparar (12.20) con (11.20).

A partir de esta definición, siguen inmediatamente los tres teoremas. Éstos se establecen en términos de una función $f(x)$, donde x puede interpretarse como un vector de variables, $x = (x_1, \ldots, x_n)$

Teorema I (negativo de una función) Si $f(x)$ es cuasicóncava (estrictamente cuasicóncava), entonces $-f(x)$ es cuasiconvexa (estrictamente cuasiconvexa).

Teorema II (concavidad contra cuasiconcavidad) Cualquier función cóncava (convexa) es cuasicóncava (cuasiconvexa), pero el inverso no es verdad. En forma similar, cualquier función estrictamente cóncava (estrictamente convexa) es estrictamente cuasicóncava (estrictamente cuasiconvexa), pero el inverso no es verdad.

Teorema III (función lineal) Si $f(x)$ es una función lineal, entonces es cuasicóncava y cuasiconvexa al mismo tiempo.

El teorema I proviene del hecho de que multiplicar una desigualdad por -1 invierte el sentido de la desigualdad. Sea $f(x)$ cuasicóncava, con $f(v) \geq f(u)$. Entonces, mediante (12.20), $f[\theta u + (1-\theta)v] \geq f(u)$. Sin embargo, respecto a la función $-f(x)$, tenemos (después de multiplicar las dos desigualdades por -1) $-f(u) \geq -f(v)$ y $-f[\theta u + (1-\theta)v] \leq -f(u)$. Interpretando $-f(u)$ como la altura del punto N, y $-f(v)$ como la altura de M, vemos que la función $-f(x)$ satisface la condición de cuasiconvexidad de (12.20). Esto prueba uno de los cuatro casos citados en el teorema I; las pruebas de los otros tres son similares.

Para el teorema II, probaremos solamente que la concavidad implica la cuasiconcavidad. Sea $f(x)$ cóncava; entonces, mediante (11.20),

$$f[\theta u + (1-\theta)v] \geq \theta f(u) + (1-\theta) f(v)$$

Supongamos ahora que $f(v) \geq f(u)$; entonces, cualquier promedio ponderado de $f(v)$ y $f(u)$ no puede ser menor que $f(u)$, es decir,

$$\theta f(u) + (1-\theta) f(v) \geq f(u)$$

De la combinación de estos dos resultados encontramos que, por la transitividad,

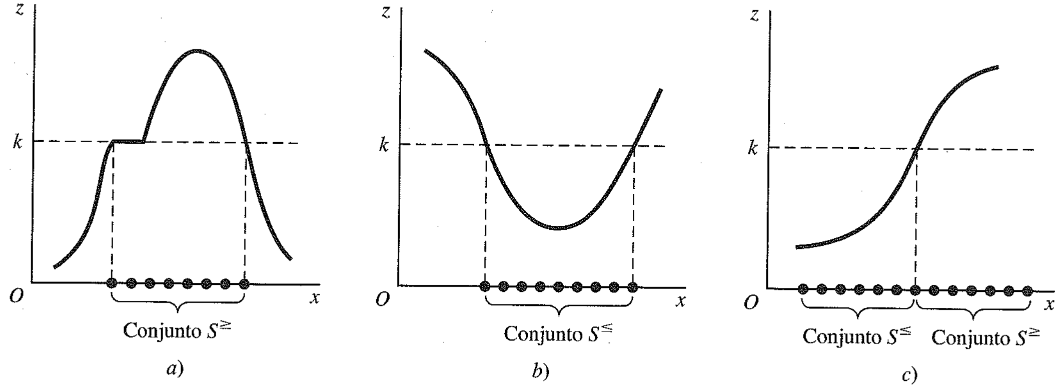
$$f[\theta u + (1-\theta)v] \geq f(u) \qquad \text{para } f(v) \geq f(u)$$

que satisface la definición de cuasiconcavidad en (12.20). Observe, sin embargo, que la condición de cuasiconcavidad no puede garantizar la concavidad.

Una vez que se establece el teorema II, sigue inmediatamente el teorema III. Ya sabemos que una función lineal es tanto cóncava como convexa, aunque no estrictamente. En vista del teorema II, una función lineal debe ser también cuasicóncava y cuasiconvexa, aunque no estrictamente.

En el caso de funciones cóncavas y convexas, hay un teorema útil para que la suma de funciones cóncavas (convexas) sea también cóncava (convexa). Desafortunadamente, este teorema no puede generalizarse a las funciones cuasicóncavas y cuasiconvexas. Por ejemplo, la suma de dos funciones cuasicóncavas *no es necesariamente* cuasicóncava (ver el ejercicio 12.4-3).

FIGURA 12.5



Algunas veces puede ser mas fácil verificar la cuasiconcavidad y la cuasiconvexidad mediante la siguiente definición alterna:

> Una función $f(x)$, donde x es un vector de variables, es $\left\{ \begin{matrix} \text{cuasicóncava} \\ \text{cuasiconvexa} \end{matrix} \right\}$ si y sólo si, para cualquier constante k, el conjunto
>
> $$\left\{ \begin{matrix} S^{\geq} \equiv \{x \mid f(x) \geq k\} \\ S^{\leq} \equiv \{x \mid f(x) \leq k\} \end{matrix} \right\} \text{ es un conjunto convexo} \qquad \textbf{(12.21)}$$

Los conjuntos $S^{\geq}$ y $S^{\leq}$, que son subconjuntos del dominio, se introdujeron anteriormente (figura 11.10) para mostrar que una función convexa (o aun una función cóncava) puede originar un conjunto convexo. Empleamos aquí estos dos conjuntos como prueba de la cuasiconcavidad y la cuasiconvexidad. Las tres funciones de la figura 12.5 contienen tanto segmentos cóncavos como convexos y, por lo tanto, no son ni convexos ni cóncavos. Pero la función de la figura 12.5*a* es cuasicóncava, ya que para cualquier valor de k (solamente se ha ilustrado uno de ellos), el conjunto $S^{\geq}$ es convexo. Por otro lado, la función de la figura 12.5*b* es cuasiconvexa ya que el conjunto $S^{\leq}$ es convexo. La función de la figura 12.5*c* —una función monótona— difiere de las otras dos en que tanto $S^{\geq}$ como $S^{\leq}$ son conjuntos convexos. Por tanto, esa función es cuasicóncava y cuasiconvexa.

Observe que mientras (12.21) puede usarse para verificar la cuasiconcavidad y la cuasiconvexidad, no puede distinguir entre las variedades estricta y no estricta de estas propiedades. Observe también que las propiedades que surgen a partir de la definición (12.21) no son en sí mismas suficientes para la concavidad y la convexidad, respectivamente. En especial, dada una función cóncava que debe forzosamente ser cuasicóncava, concluimos que $S^{\geq}$ es un conjunto convexo; pero dado que $S^{\geq}$ es un conjunto convexo, concluimos solamente que la función f es *cuasi*cóncava (pero no necesariamente cóncava).

Ejemplo 1

Verifica $z = x^2$ $(x \geq 0)$ respecto a la cuasiconcavidad y la cuasiconvexidad. Se verifica fácilmente en forma geométrica que esta función es convexa, de hecho, estrictamente convexa. Por tanto, es cuasiconvexa. Resulta interesante que también es cuasicóncava. Por su gráfica —la mitad derecha de una curva con forma de U, que inicia desde el punto de origen y crece a una tasa creciente— tiene la capacidad, en forma similar a la figura 12.5*c*, de generar un $S^{\geq}$ convexo, así como un $S^{\leq}$ convexo.

Si en lugar de esto deseamos aplicar (12.20), primero consideramos que u y v son dos valores no negativos diferentes de x. Entonces,

$$f(u) = u^2 \quad f(v) = v^2 \quad \text{y} \quad f[\theta u + (1-\theta)v] = [\theta u + (1-\theta)v]^2$$

Suponga que $f(v) \geq f(u)$, es decir, $v^2 \geq u^2$; entonces $v \geq u$, o más específicamente, $v > u$ (ya que u y v son diferentes). Como el promedio ponderado $[\theta u + (1-\theta)v]$ debe estar entre u y v, podemos escribir la desigualdad continua

$$v^2 > [\theta u + (1-\theta)v]^2 > u^2 \qquad \text{para } 0 < \theta < 1$$

o

$$f(v) > f[\theta u + (1-\theta)v] > f(u) \qquad \text{para } 0 < \theta < 1$$

Mediante (12.20), este resultado convierte la función f tanto en cuasicóncava como cuasiconvexa —verdaderamente en forma estricta.

Ejemplo 2 Muestre que $z = f(x, y) = xy$ (con $x, y \geq 0$) es cuasicóncava. Usaremos el criterio de (12.21) y estableceremos que el conjunto $S^{\geq} = \{(x, y) \mid xy \geq k\}$ es un conjunto convexo para cualquier k. Para este propósito, hacemos $xy = k$ para obtener una curva de isovalor para cada valor de k. Al igual que x y y, k debe ser no negativo. Para el caso $k > 0$, la curva de isovalor es una hipérbola rectangular en el primer cuadrante del plano xy. El conjunto $S^{\geq}$, que está formado por todos los puntos que están sobre o arriba de una hipérbola rectangular, es un conjunto convexo. Para el otro caso, con $k = 0$, la curva de isovalor definida por $xy = 0$ tiene forma de L, con la L que coincide con los segmentos no negativos de los ejes x y y. El conjunto $S^{\geq}$, que esta vez está formado por el cuadrante completo no negativo, nuevamente es un conjunto convexo. Entonces, mediante (12.21), la función $z = xy$ (witℓcon $' \geq 0$) es cuasicóncava.

Debe tener cuidado de no confundir la forma de las curvas de isovalor $xy = k$ (que se define en el plano xy) con la forma de la superficie $z = xy$ (que se define en el espacio xyz). La característica de la superficie z (cuasicóncava en el espacio 3) es lo que queremos indagar. La forma de las curvas de isovalor (convexa en el espacio bidimensional para k positivo) es de interés aquí sólo como un medio para delinear los conjuntos $S^{\geq}$ con objeto de aplicar el criterio de (12.21).

Ejemplo 3 Muestre que $z = f(x, y) = (x-a)^2 + (y-b)^2$ es cuasiconvexa. Apliquemos nuevamente (12.21). Haciendo $(x-a)^2 + (y-b)^2 = k$, vemos que k debe ser no negativo. Para cada k, la curva de isovalor es un círculo en el plano xy con su centro en (a, b) y con radio $\sqrt{k}$. Como $S^{\leq} = \{(x, y) \mid (x-a)^2 + (y-b)^2 \leq k\}$ es el conjunto de todos los puntos en o dentro del círculo, constituye un conjunto convexo. Esto es verdad aun cuando $k = 0$ —cuando el círculo degenera hasta un punto individual, (a, b)—, ya que por convención un punto individual se considera como un conjunto convexo. Entonces, la función dada es cuasiconvexa.

Funciones diferenciables

Las definiciones (12.20) y (12.21) no requieren la diferenciabilidad de la función f. Sin embargo, si f es diferenciable, la cuasiconcavidad y la cuasiconvexidad pueden definirse alternativamente en términos de sus primeras derivadas:

Una función diferenciable de una variable, $f(x)$, es $\left\{\begin{matrix}\text{cuasicóncava}\\\text{cuasiconvexa}\end{matrix}\right\}$ si y sólo si, para cualquier par de puntos diferentes u y v en el dominio

$$f(v) \geq f(u) \Rightarrow \left\{\begin{matrix} f'(u)(v-u) \\ f'(v)(v-u) \end{matrix}\right\} \geq 0 \tag{12.22}$$

La cuasiconcavidad y la cuasiconvexidad serán *estrictas* si la desigualdad débil de la derecha se transforma en la desigualdad estricta > 0. Cuando hay dos o más variables independientes, la definición debe modificarse como sigue:

Una función diferenciable $f(x_1, \ldots, x_n)$ es $\left\{\begin{matrix}\text{cuasicóncava}\\\text{cuasiconvexa}\end{matrix}\right\}$ si y sólo si, para cualquiera dos puntos diferentes $u = (u_1, \ldots, u_n)$ y $v = (v_1, \ldots, v_n)$ en el dominio,

$$f(v) \geq f(u) \Rightarrow \begin{Bmatrix} \sum_{j=1}^{n} f_j(u)(v_j - u_j) \\ \sum_{j=1}^{n} f_j(v)(v_j - u_j) \end{Bmatrix} \geq 0 \qquad \textbf{(12.22')}$$

donde $f_j \equiv \partial f/\partial x_j$, que debe evaluarse para u o v según sea el caso.

Nuevamente, para la cuasiconcavidad y la cuasiconvexidad *estrictas*, la desigualdad débil de la derecha debe transformarse en la desigualdad estricta > 0.

Finalmente, si una función $z = f(x_1, \ldots, x_n)$ es dos veces continuamente diferenciable, la cuasiconcavidad y la cuasiconvexidad pueden verificarse mediante la primera y la segunda derivadas parciales de la función, arregladas en el determinante orlado

$$|B| = \begin{vmatrix} 0 & f_1 & f_2 & \cdots & f_n \\ f_1 & f_{11} & f_{12} & \cdots & f_{1n} \\ f_2 & f_{21} & f_{22} & \cdots & f_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ f_n & f_{n1} & f_{n2} & \cdots & f_{nn} \end{vmatrix} \qquad \textbf{(12.23)}$$

Este determinante orlado se parece al hessiano orlado $|\overline{H}|$ introducido en la sección 12.3. Pero a diferencia de este último, la orla de $|B|$ está compuesta por las primeras derivadas de la función f en lugar de una función restrictiva g externa. Debido a que $|B|$ depende exclusivamente de las derivadas de la función f misma podemos usar $|B|$, junto con sus menores principales directores

$$|B_1| = \begin{vmatrix} 0 & f_1 \\ f_1 & f_{11} \end{vmatrix} \quad |B_2| = \begin{vmatrix} 0 & f_1 & f_2 \\ f_1 & f_{11} & f_{12} \\ f_2 & f_{21} & f_{22} \end{vmatrix} \quad \cdots \quad |B_n| = |B| \qquad \textbf{(12.24)}$$

para caracterizar la configuración de esa función.

Estableceremos aquí dos condiciones; una es necesaria, y la otra suficiente. Ambas se relacionan con la cuasiconcavidad en un dominio que está formado solamente por el *n-cuadrante no negativo* (la analogía n-dimensional del cuadrante no negativo); es decir, con $x_1, \ldots, x_n \geq 0$.[6]

Para que $z = f(x_1, \ldots, x_n)$ sea cuasicóncava en el n-cuadrante no negativo, es necesario que

$$|B_1| \leq 0, \quad |B_2| \geq 0, \quad \ldots, \quad |B_n| \begin{Bmatrix} \leq \\ \geq \end{Bmatrix} 0 \text{ si } n \text{ es} \begin{cases} \text{impar} \\ \text{par} \end{cases} \qquad \textbf{(12.25)}$$

siempre que las derivadas parciales se evalúen en el n-cuadrante no negativo.

[6] Mientras que la concavidad (convexidad) de una función en un dominio convexo siempre puede generalizarse a la concavidad (convexidad) sobre el espacio completo, no es posible hacerlo con la cuasiconcavidad y la cuasiconvexidad. Por ejemplo, nuestras conclusiones de los ejemplos 1 y 2 no son válidas si se permite que las variables adopten valores negativos. Las dos condiciones dadas aquí se basan en Kenneth J. Arroz y Alain C. Enthoven, "Quasi-Concave Programming", *Econometrica*, octubre de 1961, p. 797 (teorema 5); y Akira Takayama, *Analytical Methods in Economics*, University of Michigan Press, 1993, p. 65 (teorema 1.12).

Una condición suficiente para que f sea estrictamente cuasicóncava en el n-cuadrante no negativo es que

$$|B_1| < 0, \quad |B_2| > 0, \quad \ldots, \quad |B_n| \begin{Bmatrix} < \\ > \end{Bmatrix} 0 \text{ si } n \text{ es} \begin{cases} \text{impar} \\ \text{par} \end{cases} \tag{12.26}$$

siempre que las derivadas parciales se evalúen en el n-cuadrante no negativo.

Observe que la condición $|B_1| \leq 0$ de (12.25) se satisface automáticamente porque $|B_1| = -f_1^2$; se lista aquí solamente para tomar en cuenta la simetría. Lo mismo pasa con la condición $|B_1| < 0$ de (12.26).

Ejemplo 4 La función $z = f(x_1, x_2) = x_1 x_2$ $(x_1, x_2 \geq 0)$ es cuasicóncava (ver el ejemplo 2). Ahora verificaremos esto mediante (12.22'). Sean $u = (u_1, u_2)$ y $v = (v_1, v_2)$ dos puntos cualquiera en el dominio. Entonces, $f(u) = u_1 u_2$ y $f(v) = v_1 v_2$. Suponga que

$$f(v) \geq f(u) \quad \text{or} \quad v_1 v_2 \geq u_1 u_2 \quad (v_1, v_2, u_1, u_2 \geq 0) \tag{12.27}$$

Como las derivadas parciales de f son $f_1 = x_2$ y $f_2 = x_1$, (12.22') apunta a la condición que

$$f_1(u)(v_1 - u_1) + f_2(u)(v_2 - u_2) = u_2(v_1 - u_1) + u_1(v_2 - u_2) \geq 0$$

o al arreglar,

$$u_2(v_1 - u_1) \geq u_1(u_2 - v_2) \tag{12.28}$$

Necesitamos considerar cuatro posibilidades respecto a los valores de u_1 y u_2. Primero, si $u_1 = u_2 = 0$, entonces (12.28) se satisface trivialmente. Segundo, si $u_1 = 0$ pero $u_2 > 0$, entonces (12.28) se reduce a la condición $u_2 v_1 \geq 0$, que se satisface nuevamente, ya que u_2 y v_1 son no negativos. Tercero, si $u_1 > 0$ y $u_2 = 0$, entonces (12.28) se reduce a la condición $0 \geq -u_1 v_2$, que todavía se satisface. Cuarto y último, suponga que u_1 y u_2 son positivos, de modo que v_1 y v_2 también son positivos. Al sustraer $v_2 u_1$ de ambos lados de (12.27), obtenemos

$$v_2(v_1 - u_1) \geq u_1(u_2 - v_2) \tag{12.29}$$

Ahora se presentan por sí mismas tres posibilidades subyacentes:

1. Si $u_2 = v_2$, entonces $v_1 \geq u_1$. De hecho, deberíamos tener $v_1 > u_1$, ya que (u_1, u_2) y (v_1, v_2) son puntos diferentes. El hecho de que $u_2 = v_2$ y $v_1 > u_1$ implica que se satisface la condición (12.28).
2. Si $u_2 > v_2$, entonces también debemos tener $v_1 > u_1$ mediante (12.29). Al multiplicar ambos lados de (12.29) por u_2/v_2, obtenemos

$$u_2(v_1 - u_1) \geq \frac{u_2}{v_2} u_1(u_2 - v_2) > u_1(u_2 - v_2) \quad \left[\text{ya que } \frac{u_2}{v_2} > 1\right] \tag{12.30}$$

   Entonces, (12.28) se satisface nuevamente.
3. La posibilidad final subyacente es que $u_2 < v_2$, lo que implica que u_2/v_2 es una fracción positiva. En este caso, todavía es válido el primer renglón de (12.30). También es válido el segundo renglón, pero ahora por una razón diferente: una fracción (u_2/v_2) de un número negativo $(u_2 - v_2)$ es mayor que el mismo número último.

Siempre que (12.28) se satisfaga en toda situación posible que pueda surgir, la función $z = x_1 x_2$ $(x_1, x_2 \geq 0)$ es cuasicóncava. Por lo tanto, la condición necesaria (12.25) debe ser válida. Debido a que las derivadas parciales de f son

$$f_1 = x_2 \quad f_2 = x_1 \quad f_{11} = f_{22} = 0 \quad f_{12} = f_{21} = 1$$

los menores principales directores relevantes resultan ser

$$|B_1| = \begin{vmatrix} 0 & x_2 \\ x_2 & 0 \end{vmatrix} = -x_2^2 \leq 0 \quad |B_2| = \begin{vmatrix} 0 & x_2 & x_1 \\ x_2 & 0 & 1 \\ x_1 & 1 & 0 \end{vmatrix} = 2x_1 x_2 \geq 0$$

Entonces (12.25) se satisface realmente. Observe, sin embargo, que la condición suficiente (12.26) se satisface solamente para el n-cuadrante positivo.

Ejemplo 5 Muestre que $z = f(x, y) = x^a y^b$ $(x, y > 0; 0 < a, b < 1)$ es estrictamente cuasicóncava. Las derivadas parciales de esta función son

$$f_x = ax^{a-1} y^b \quad f_y = bx^a y^{b-1}$$
$$f_{xx} = a(a-1)x^{a-2} y^b \quad f_{xy} = f_{yx} = abx^{a-1} y^{b-1} \quad f_{yy} = b(b-1)x^a y^{b-2}$$

entonces los menores principales directores de $|B|$ tienen los siguientes signos:

$$|B_1| = \begin{vmatrix} 0 & f_x \\ f_x & f_{xx} \end{vmatrix} = -(ax^{a-1} y^b)^2 < 0$$

$$|B_2| = \begin{vmatrix} 0 & f_x & f_y \\ f_x & f_{xx} & f_{xy} \\ f_y & f_{yx} & f_{yy} \end{vmatrix} = [2a^2 b^2 - a(a-1)b^2 - a^2 b(b-1)]x^{3a-2} y^{3b-2} > 0$$

Esto satisface la condición suficiente para la cuasiconcavidad estricta de (12.26).

Una mirada adicional al hessiano orlado

El determinante orlado $|B|$, como se define en (12.23), difiere del hessiano orlado

$$|\overline{H}| = \begin{vmatrix} 0 & g_1 & g_2 & \cdots & g_n \\ g_1 & Z_{11} & Z_{12} & \cdots & Z_{1n} \\ g_2 & Z_{21} & Z_{22} & \cdots & Z_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ g_n & Z_{n1} & Z_{n2} & \cdots & Z_{nn} \end{vmatrix}$$

de dos maneras: (1) los elementos situados en la orla de $|B|$ son las derivadas parciales de primer orden de la función f en vez de g; y (2) los elementos restantes de $|B|$ son las derivadas parciales de segundo orden de f en vez de la función lagrangiana Z. Sin embargo, en el caso especial de una ecuación restrictiva lineal, $g(x_1, \ldots, x_n) = a_1 x_1 + \cdots + a_n x_n = c$ —un caso que se encuentra frecuentemente en la economía (ver la sección 12.5)— Z_{ij} se reduce a f_{ij}. Como entonces la función lagrangiana es

$$Z = f(x_1, \ldots, x_n) + \lambda(c - a_1 x_1 - \cdots - a_n x_n)$$

de modo que

$$Z_j = f_j - \lambda a_j \qquad \text{y} \qquad Z_{ij} = f_{ij}$$

Regresando a las orlas, observamos que la función restrictiva lineal suministra la primera derivada $g_j = a_j$. Aún más, cuando se satisface la condición de primer orden, tenemos $Z_j = f_j - \lambda a_j = 0$, de modo que $f_j = \lambda a_j$ o $f_j = \lambda g_j$. Entonces, la orla de $|B|$ es simplemente la de $|\overline{H}|$ multiplicada por un escalar positivo λ. Al factorizar λ sucesivamente desde las orlas horizontal y vertical de $|\overline{H}|$ (ver la sección 5.3, ejemplo 5), tenemos

$$|B| = \lambda^2 |\overline{H}|$$

En consecuencia, en el caso de la restricción lineal, los dos determinantes orlados siempre poseen el mismo signo en el punto estacionario de Z. Por la misma razón, los menores principales directores $|B_i|$ y $|\overline{H}_i|$ $(i = 1, \ldots, n)$ también deben compartir el mismo signo en ese punto. Así pues, se concluye que si el determinante orlado $|B|$ satisface la condición suficiente para la cuasiconcavidad estricta de (12.26), el hessiano orlado $|\overline{H}|$ debe entonces satisfacer la condición suficiente de segundo orden para la maximización restringida en la tabla 12.1.

Extremos absolutos contra extremos relativos

En la figura 12.6 se presenta una visión esquemática de la relación entre la cuasiconcavidad y las condiciones de segundo orden. (Una modificación adecuada adaptará la figura para el caso de la cuasiconvexidad.) Construida con la misma intención —y para leerse de la misma manera— que la figura 11.5, esta figura relaciona la cuasiconcavidad con los máximos restringidos *absolutos* así como *relativos* de una función diferenciable dos veces $z = f(x_1, \ldots, x_n)$. Los tres óvalos de la parte superior resumen las condiciones de primero y segundo órdenes para un máximo restringido relativo. Y los rectángulos de la columna media, como los de la figura 11.5, enlazan entre sí los conceptos de máximo relativo, máximo absoluto y máximo absoluto único.

Pero la información realmente interesante es la que está en los dos diamantes y en los símbolos alargados $\Rightarrow$ que pasan por ellos. El que está a la izquierda nos dice que, una vez que se satisface la condición de primer orden, y si también se satisfacen las dos condiciones listadas en el diamante, tenemos una condición suficiente para un máximo restringido absoluto. La primera condición es que la función f sea explícitamente cuasicóncava, un nuevo término que debemos apresurarnos a definir.

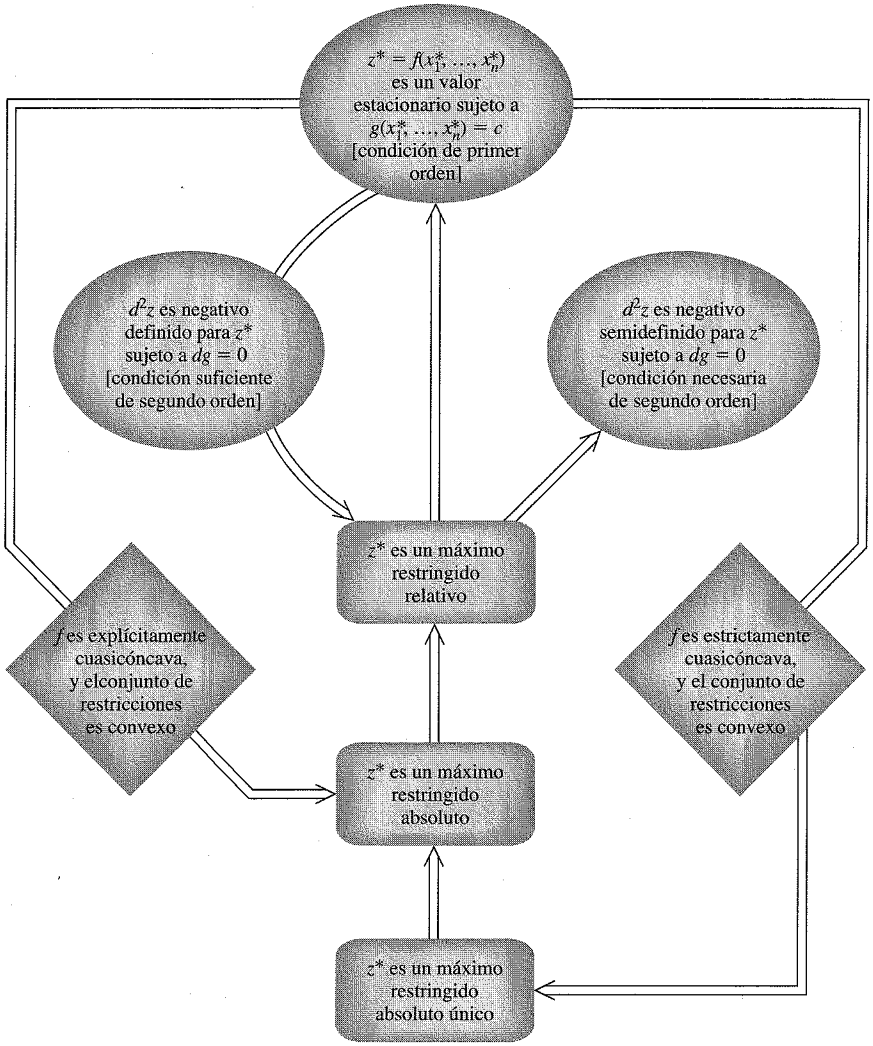
Una función cuasicóncava f es *explícitamente cuasicóncava* si tiene la propiedad adicional de que

$$f(v) > f(u) \Rightarrow f[\theta u + (1 - \theta)v] > f(u)$$

Esta propiedad definida significa que siempre que un punto de la superficie, $f(v)$, esté a una altura mayor que otro, $f(u)$, entonces todos los puntos intermedios —los puntos en la superficie que está situada directamente arriba del segmento de línea uv en el dominio— también deben tener mayor altura que $f(u)$. Lo que este enunciado hace es descartar cualquiera segmento de plano *horizontal* de la superficie con excepción de una plataforma en la parte superior de la superficie.[7] Observe que la condición de cuasiconcavidad *explícita* no es tan fuerte como la condición de cuasiconcavidad *estricta*, ya que esta última requiere que $f[\theta u + (1 - \theta)v] > f(u)$ aun para $f(v) = f(u)$, lo que también implica que se descartan

[7] Sea que la superficie contiene un segmento de plano horizontal P tal que $f(u) \in P$ y $f(v) \notin P$. Entonces, aquellos puntos intermedios que se localizan sobre P tendrán la misma altura que $f(u)$, incumpliendo con ello la primera disposición.

FIGURA 12.6



los segmentos de planos *no* horizontales.[8] La otra condición del diamante del lado izquierdo es que el conjunto $\{(x_1, \ldots, x_n) \mid g(x_1, \ldots, x_n) = c\}$ sea convexo. Cuando se cumplen ambas condiciones, estaremos tratando con la porción de una superficie con forma de campana libre de segmentos horizontales (o hipersuperficie) situada directamente arriba de un conjunto convexo en el dominio. Un máximo local encontrado en un subconjunto de este tipo en la superficie debe ser un máximo restringido absoluto.

El diamante de la derecha de la figura 12.6 incluye la condición más fuerte de cuasiconcavidad *estricta*. Una función estrictamente cuasicóncava debe ser explícitamente cuasicóncava, aunque el inverso no sea verdad. Por tanto, cuando la cuasiconcavidad estricta reemplace a la

[8] Sea que la superficie contiene un segmento de plano oblicuo P' tal que $f(u) = f(v)$ estén ambos localizados sobre P'. Entonces, todos los puntos intermedios también estarán sobre P' y tendrán la misma altura que $f(u)$, incumpliendo con ello el requerimiento citado de cuasiconcavidad estricta.

cuasiconcavidad explícita, todavía se asegura un máximo restringido absoluto. Pero esta vez ese máximo restringido absoluto también debe ser único, ya que la ausencia de cualquier segmento plano en cualquier parte de la superficie excluye definitivamente la posibilidad de máximos restringidos múltiples.

EJERCICIO 12.4

1. Dibuje una curva estrictamente cuasicóncava $z = f(x)$ que sea
   (*a*) también cuasiconvexa (*d*) no cóncava
   (*b*) no cuasiconvexa (*e*) ni cóncava ni convexa
   (*c*) no convexa (*f*) tanto cóncava como convexa
2. ¿Las siguientes funciones son cuasicóncavas? ¿Lo son estrictamente? Revise primero gráficamente y luego algebraicamente mediante (12.20). Suponga que $x \geq 0$.
   (*a*) $f(x) = a$ (*b*) $f(x) = a + bx$ $(b > 0)$ (*c*) $f(x) = a + cx^2$ $(c < 0)$
3. (*a*) Sea $z = f(x)$ cuya gráfica tiene la forma de una curva con pendiente negativa como la mitad derecha de una campana en el primer cuadrante, que pasa por los puntos (0, 5), (2, 4), (3, 2) y (5, 1). Sea $z = g(x)$ cuya gráfica es una recta con pendiente positiva a 45°. ¿Son $f(x)$ y $g(x)$ cuasicóncavas?
   (*b*) Ahora grafique la suma $f(x) + g(x)$. ¿Es la función suma cuasicóncava?
4. Mediante el examen de sus gráficas y con el uso de (12.21), verifique si las siguientes funciones son cuasicóncavas, cuasiconvexas, ambas o ninguna de ellas:
   (*a*) $f(x) = x^3 - 2x$ (*b*) $f(x_1, x_2) = 6x_1 - 9x_2$ (*c*) $f(x_1, x_2) = x_2 - \ln x_1$
5. (*a*) Verifique que una función cúbica $z = ax^3 + bx^2 + cx + d$ no es en general ni cuasicóncava ni cuasiconvexa.
   (*b*) ¿Se puede imponer restricciones a los parámetros de modo que la función se transforme en cuasicóncava y cuasiconvexa a la vez para $x \geq 0$?
6. Use (12.22) para verificar $z = x^2 (x \geq 0)$ en cuanto a cuasiconcavidad y cuasiconvexidad.
7. Muestre que $z = xy$ $(x, y \geq 0)$ no es cuasiconvexa.
8. Use determinantes orlados para verificar las siguientes funciones en cuanto a cuasiconcavidad y ciasiconvexidad:
   (*a*) $z = -x^2 - y^2$ $(x, y > 0)$ (*b*) $z = -(x+1)^2 - (y+2)^2$ $(x, y > 0)$

12.5 Maximización de utilidad y demanda del consumidor

En la sección 12.1 hablamos de la maximización de una función de utilidad como un ejemplo de optimización restringida. Examinemos este problema con mayor detalle. Por simplicidad, le permitiremos a nuestro consumidor hipotético la elección entre dos bienes, los cuales tienen funciones de utilidad marginal continuas y positivas. Los precios de ambos bienes los determina el mercado; por lo tanto, son exógenos, aunque en esta sección omitiremos el subíndice cero en los símbolos de precios. Si el poder de compra del consumidor en una cantidad dada B (de presupuesto), el problema que se enfrenta será el de la maximización de una función suave de utilidad (índice)

$$U = U(x, y) \qquad (U_x, U_y > 0)$$

sujeto a

$$xP_x + yP_y = B$$

Condición de primer orden

La función lagrangiana de este modelo de optimización es

$$Z = U(x, y) + \lambda(B - xP_x - yP_y)$$

Como condición de primer orden, tenemos el siguiente conjunto de ecuaciones simultáneas:

$$\begin{aligned} Z_\lambda &= B - xP_x - yP_y = 0 \\ Z_x &= U_x - \lambda P_x = 0 \\ Z_y &= U_y - \lambda P_y = 0 \end{aligned} \tag{12.31}$$

Como las dos últimas ecuaciones equivalen a

$$\frac{U_x}{P_x} = \frac{U_y}{P_y} = \lambda \tag{12.31'}$$

la condición de primer orden, en efecto, requiere la satisfacción de (12.31′), sujeta a la restricción presupuestaria: la primera ecuación de (12.31). Lo que establece (12.31′) es simplemente la proposición familiar de la teoría clásica del consumidor de que, con objeto de maximizar la utilidad, los consumidores deben asignar sus presupuestos para igualar la razón de utilidad marginal entre el precio para cada artículo. Específicamente, en el equilibrio o en el óptimo, estas razones deben tener el valor común de λ^*. Como aprendimos anteriormente, λ^* mide el efecto estático comparativo de la constante de restricción sobre el valor óptimo de la función objetivo. Por lo tanto, en el contexto presente tenemos $\lambda^* = (\partial U^*/\partial B)$; es decir, el valor óptimo de los multiplicadores de Lagrange puede interpretarse como la *utilidad marginal de dinero* (dinero presupuestario) cuando se maximiza la utilidad del consumidor.

Si restablecemos la condición de (12.31′) de la forma

$$\frac{U_x}{U_y} = \frac{P_x}{P_y} \tag{12.31''}$$

podemos darle una interpretación alterna a la condición de primer orden, en términos de curvas de indiferencia.

Una *curva de indiferencia* se define como el lugar geométrico de las combinaciones de x y y que proporciona un nivel constante de U. Esto significa que en una curva de indiferencia debemos encontrar

$$dU = U_x\,dx + U_y\,dy = 0$$

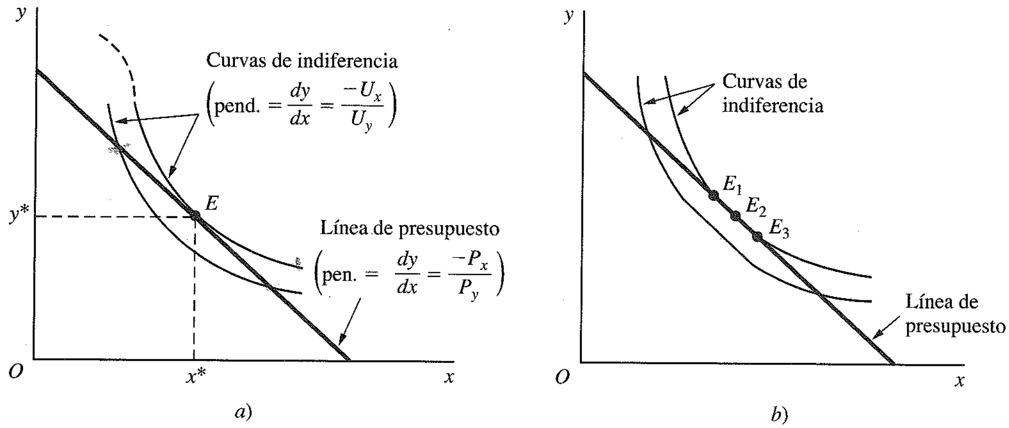
con la implicación de que $dy/dx = -U_x/U_y$. De acuerdo con esto, si graficamos una curva de indiferencia en el plano xy, como en la figura 12.7, su pendiente, dy/dx, debe ser igual al negativo de la razón de utilidad marginal U_x/U_y. (Como suponemos que U_x, $U_y > 0$, la pendiente de la curva de indiferencia debe ser negativa.) Observe que U_x/U_y, el negativo de la pendiente de la curva de indiferencia, se llama la *tasa marginal de sustitución* entre los dos bienes.

¿Cuál es el significado de P_x/P_y? Como veremos ahora, esta tasa representa el negativo de la pendiente de la gráfica de la restricción presupuestaria. La restricción presupuestaria, $xP_x + yP_y = B$, puede escribirse en forma alterna como

$$y = \frac{B}{P_y} - \frac{P_x}{P_y}x$$

de modo que, cuando se grafica en el plano xy como en la figura 12.7, surge como una línea recta con pendiente $-P_x/P_y$ (e intercepción vertical B/P_y).

FIGURA 12.7



Ante esto, la nueva versión de la condición de primer orden —(12.31″) más la restricción presupuestaria— revela que para maximizar la utilidad un consumidor debe asignar el presupuesto de modo que la pendiente de la línea presupuestaria (sobre la cual debe permanecer el consumidor) sea igual a la pendiente de alguna curva de indiferencia. Esta condición se cumple en el punto E de la figura 12.7*a*, donde la línea presupuestaria es tangente a una curva de indiferencia.

Condición de segundo orden

Si el hessiano orlado de este problema es positivo, es decir, si

$$|\overline{H}| = \begin{vmatrix} 0 & P_x & P_y \\ P_x & U_{xx} & U_{xy} \\ P_y & U_{yx} & U_{yy} \end{vmatrix} = 2P_xP_yU_{xy} - P_y^2U_{xx} - P_x^2U_{yy} > 0 \qquad \textbf{(12.32)}$$

(con todas las derivadas evaluadas para los valores críticos x^* y y^*), entonces el valor estacionario de U con seguridad será un máximo. La presencia de las derivadas U_{xx}, U_{yy}, y U_{xy} de (12.32) sugiere claramente que el cumplimiento de esta condición implicaría ciertas restricciones sobre la función de utilidad y, por lo tanto, sobre la forma de las curvas de indiferencia. ¿Cuáles son estas restricciones?

Considerando primero la forma de las curvas de indiferencia, podemos mostrar que un $|\overline{H}|$ positivo implica la convexidad estricta de la curva de indiferencia (con pendiente hacia abajo) en el punto de tangencia E. Así como la pendiente hacia abajo de una curva de indiferencia está garantizada por un dy/dx $(= -U_x/U_y)$ negativo, su convexidad estricta estaría asegurada por un d^2y/dx^2 positivo. Para obtener la expresión para d^2y/dx^2, podemos diferenciar $-U_x/U_y$ respecto a x; pero al hacerlo, debemos recordar que U_x y U_y (siendo derivadas) no sólo son funciones de x y de y sino también que, a lo largo de una curva de indiferencia dada, y es en sí misma una función de x. De acuerdo con esto, tanto U_x como U_y pueden considerarse como funciones de x solamente; por lo tanto, podemos obtener una derivada total

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\left(-\frac{U_x}{U_y}\right) = -\frac{1}{U_y^2}\left(U_y\frac{dU_x}{dx} - U_x\frac{dU_y}{dx}\right) \qquad \textbf{(12.33)}$$

Como x puede afectar a U_x y a U_y no sólo directamente sino también indirectamente, vía el intermediario de y, tenemos

$$\frac{dU_x}{dx} = U_{xx} + U_{yx}\frac{dy}{dx} \qquad \frac{dU_y}{dx} = U_{xy} + U_{yy}\frac{dy}{dx} \qquad \textbf{(12.34)}$$

donde dy/dx se refiere a la pendiente de la curva de indiferencia. Ahora, en el punto de tangencia E —el único punto relevante para el estudio de la condición de segundo orden— esta pendiente es idéntica a la de la restricción presupuestaria; es decir, $dy/dx = -P_x/P_y$. Entonces, podemos escribir (12.34) como

$$\frac{dU_x}{dx} = U_{xx} - U_{yx}\frac{P_x}{P_y} \qquad \frac{dU_y}{dx} = U_{xy} - U_{yy}\frac{P_x}{P_y} \tag{12.34'}$$

Sustituyendo (12.34′) en (12.33) y utilizando la información de que

$$U_x = \frac{U_y P_x}{P_y} \quad [\text{ de } (12.31'')]$$

y luego factorizando U_y/P_y^2, finalmente podemos transformar (12.33) en

$$\frac{d^2y}{dx^2} = \frac{2P_xP_yU_{xy} - P_y^2U_{xx} - P_x^2U_{yy}}{U_yP_y^2} = \frac{|\overline{H}|}{U_yP_y^2} \tag{12.33'}$$

Es evidente que cuando se satisface la condición suficiente de segundo orden (12.32), la segunda derivada de (12.33′) es positiva, y la curva de indiferencia relevante es estrictamente convexa en el punto de tangencia. En este contexto, también es verdad que la convexidad estricta de la curva de indiferencia en la tangencia implica la satisfacción de la condición suficiente (12.32). Esto se debe a que puesto que las curvas de indiferencia tienen pendiente negativa, sin puntos estacionarios en ningún lado, se descarta la posibilidad de un valor cero de d^2y/dx^2 en una curva estrictamente convexa. Esta convexidad estricta puede conducir solamente a un d^2y/dx^2 positivo y, por tanto, a un $|\overline{H}|$ positivo, mediante (12.33′).

Sin embargo, recuerde que las derivadas en $|\overline{H}|$ deben evaluarse solamente para los valores críticos x^* y y^*. Entonces, la convexidad estricta de la curva de indiferencia, como una condición suficiente, pertenece solamente al punto de tangencia, y no es inconcebible que la curva contenga un segmento cóncavo alejado del punto E, como se ilustra en el segmento interrumpido de curva en la figura 12.7*a*. Por otro lado, si sabemos que la función utilidad es una función suave y creciente estrictamente cuasicóncava, entonces todas las curvas de indiferencia serán estrictamente convexas en todos lados. Una función utilidad de este tipo tiene una superficie como la de la figura 12.4*b*. Cuando una superficie de este tipo la corta un plano paralelo al plano *xy*, para cada uno de estos cortes obtenemos una sección transversal que, cuando se proyecta sobre el plano *xy*, se transforma en una curva de indiferencia estrictamente convexa con pendiente hacia abajo. En ese caso, no importa dónde pueda presentarse el punto de tangencia, siempre se cumplirá la condición suficiente de segundo orden. Además, puede haber solamente un punto de tangencia, uno que proporcione el nivel único máximo absoluto de utilidad alcanzable sobre el presupuesto lineal dado. Este resultado, por supuesto, concuerda perfectamente con lo que establece el diamante de la derecha de la figura 12.6.

Hemos recordado repetidamente que la condición suficiente de segundo orden no es necesaria. Ilustremos aquí la maximización de la utilidad mientras que (12.32) no se verifica. Suponga que, como se ilustra en la figura 12.7*b*, la curva de indiferencia relevante contiene un segmento lineal que coincide con una parte de la línea presupuestaria. Entonces, es claro que tenemos máximos múltiples, ya que la condición de primer orden $U_x/U_y = P_x/P_y$ se cumple ahora para todos los puntos sobre el segmento lineal de la curva de indiferencia, incluyendo E_1, E_2 y E_3. De hecho, éstos son máximos restringidos absolutos. Pero como d^2y/dx^2 es cero sobre un segmento de línea, tenemos $|\overline{H}| = 0$ mediante (12.33′). De esta forma, en este caso se logra la maximización, aun cuando se incumpla la condición suficiente de segundo orden (12.32).

El hecho de que un segmento lineal aparezca sobre la curva de indiferencia sugiere la presencia de un segmento de plano oblicuo sobre la superficie de utilidad. Esto ocurre cuando la función de utilidad es explícitamente cuasicóncava en vez de estrictamente cuasicóncava. Como muestra la figura 12.7*b*, los puntos E_1, E_2 y E_3, los cuales están ubicados sobre la misma curva de indiferencia, proporcionan la misma utilidad máxima absoluta bajo la restricción presupuestaria lineal dada. Haciendo referencia nuevamente a la figura 12.6, observamos que este resultado es perfectamente consistente con el mensaje transmitido en el diamante a la izquierda.

Análisis estático comparativo

En nuestro modelo del consumidor, los precios P_x y P_y son exógenos, así como la cantidad del presupuesto B. Si damos por hecho el cumplimiento de la condición suficiente de segundo orden, podemos analizar las propiedades estático-comparativas del modelo sobre la base de la condición de primer orden (12.31), vista como un conjunto de ecuaciones $F^j = 0\,(j = 1, 2, 3)$, donde cada función F^j tiene derivadas parciales continuas. Como señalamos en (12.19), el jacobiano de variables endógenas de este conjunto de ecuaciones debe tener el mismo valor que el hessiano orlado; es decir, $|J| = |\overline{H}|$. Entonces, cuando se cumpla la condición de segundo orden (12.32), $|J|$ debe ser positivo y no se anula en el óptimo inicial. En consecuencia, es aplicable el teorema de función implícita, y podemos expresar los valores óptimos de las variables endógenas como funciones implícitas de las variables exógenas:

$$\begin{aligned} \lambda^* &= \lambda^*(P_x, P_y, B) \\ x^* &= x^*(P_x, P_y, B) \\ y^* &= y^*(P_x, P_y, B) \end{aligned} \qquad \textbf{(12.35)}$$

Sabemos que estas variables poseen derivadas continuas que dan información estático-comparativa. En particular, las derivadas de las dos últimas funciones x^* y y^*, que describen el comportamiento de la demanda del consumidor, nos muestran cómo va a reaccionar el consumidor a los cambios de los precios y del presupuesto. Sin embargo, para encontrar estas derivadas, primero debemos convertir (12.31) en un conjunto de identidades de equilibrio, como sigue:

$$\begin{aligned} B - x^* P_x - y^* P_y &\equiv 0 \\ U_x(x^*, y^*) - \lambda^* P_x &\equiv 0 \\ U_y(x^*, y^*) - \lambda^* P_y &\equiv 0 \end{aligned} \qquad \textbf{(12.36)}$$

Al calcular la diferencial total de cada identidad en turno (permitiendo que cambie cada una de las variables), y observando que $U_{xy} = U_{yx}$, llegamos al sistema lineal

$$\begin{aligned} - P_x\, dx^* - P_y\, dy^* &= x^* dP_x + y^* dP_y - dB \\ -P_x\, d\lambda^* + U_{xx}\, dx^* + U_{xy}\, dy^* &= \lambda^* dP_x \\ -P_y\, d\lambda^* + U_{yx}\, dx^* + U_{yy}\, dy^* &= \qquad\qquad \lambda^* dP_y \end{aligned} \qquad \textbf{(12.37)}$$

Para estudiar el efecto de un cambio en el monto del presupuesto (también denominado *ingreso* del consumidor), sean $dP_x = dP_y = 0$, pero conservando $dB \neq 0$. Entonces, después de dividir (12.37) entre dB, e interpretando cada tasa de diferenciales como una derivada parcial, podemos escribir la ecuación matricial[9]

$$\begin{bmatrix} 0 & -P_x & -P_y \\ -P_x & U_{xx} & U_{xy} \\ -P_y & U_{yx} & U_{yy} \end{bmatrix} \begin{bmatrix} (\partial\lambda^*/\partial B) \\ (\partial x^*/\partial B) \\ (\partial y^*/\partial B) \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} \qquad \textbf{(12.38)}$$

[9] La ecuación matricial (12.38) también puede obtenerse al obtener la diferencial total (12.36) respecto a *B*, al tener presentes las soluciones implícitas de (12.35).

Como se puede verificar, el arreglo de elementos en la matriz de coeficientes es exactamente el mismo que aparecería en el jacobiano $|J|$, el cual tiene el mismo valor que el hessiano orlado $|\overline{H}|$ aunque éste tiene P_x y P_y (en vez de $-P_x$ y $-P_y$) en el primer renglón y en la primera columna. Por la regla de Cramer, podemos despejar las tres derivadas estático-comparativas, pero concentraremos nuestra atención en las dos siguientes:

$$\left(\frac{\partial x^*}{\partial B}\right) = \frac{1}{|J|}\begin{vmatrix} 0 & -1 & -P_y \\ -P_x & 0 & U_{xy} \\ -P_y & 0 & U_{yy} \end{vmatrix} = \frac{1}{|J|}\begin{vmatrix} -P_x & U_{xy} \\ -P_y & U_{yy} \end{vmatrix} \tag{12.39}$$

$$\left(\frac{\partial y^*}{\partial B}\right) = \frac{1}{|J|}\begin{vmatrix} 0 & -P_x & -1 \\ -P_x & U_{xx} & 0 \\ -P_y & U_{yx} & 0 \end{vmatrix} = \frac{-1}{|J|}\begin{vmatrix} -P_x & U_{xx} \\ -P_y & U_{yx} \end{vmatrix} \tag{12.40}$$

Por la condición de segundo orden, $|J| = |\overline{H}|$ es positivo, como lo son P_x y P_y. Desafortunadamente, en ausencia de información adicional acerca de las magnitudes relativas de P_x, P_y, y los U_{ij}, todavía no podemos conocer los signos de estas dos derivadas estático-comparativas. Esto significa que, a medida que aumenta el presupuesto del consumidor (o ingreso), sus compras óptimas x^* y y^* *pueden* aumentar *o* decrecer. Por ejemplo, en el caso de que x^* disminuya a medida que B aumenta, el producto x se denomina un *bien inferior* en contraste con un *bien normal.*

A continuación analizamos el efecto de un cambio en P_x. Haciendo $dP_y = dB = 0$ esta vez, pero conservando $dP_x \neq 0$, y luego dividiendo (12.37) entre dP_x, obtenemos otra ecuación matricial:

$$\begin{bmatrix} 0 & -P_x & -P_y \\ -P_x & U_{xx} & U_{xy} \\ -P_y & U_{yx} & U_{yy} \end{bmatrix}\begin{bmatrix} (\partial\lambda^*/\partial P_x) \\ (\partial x^*/\partial P_x) \\ (\partial y^*/\partial P_x) \end{bmatrix} = \begin{bmatrix} x^* \\ \lambda^* \\ 0 \end{bmatrix} \tag{12.41}$$

A partir de esto, surgen las siguientes derivadas estático-comparativas:

$$\begin{aligned} \left(\frac{\partial x^*}{\partial P_x}\right) &= \frac{1}{|J|}\begin{vmatrix} 0 & x^* & -P_y \\ -P_x & \lambda^* & U_{xy} \\ -P_y & 0 & U_{yy} \end{vmatrix} \\ &= \frac{-x^*}{|J|}\begin{vmatrix} -P_x & U_{xy} \\ -P_y & U_{yy} \end{vmatrix} + \frac{\lambda^*}{|J|}\begin{vmatrix} 0 & -P_y \\ -P_y & U_{yy} \end{vmatrix} \\ &\equiv T_1 + T_2 \qquad [T_i \text{ representa al término } i\text{-ésimo}] \end{aligned} \tag{12.42}$$

$$\begin{aligned} \left(\frac{\partial y^*}{\partial P_x}\right) &= \frac{1}{|J|}\begin{vmatrix} 0 & -P_x & x^* \\ -P_x & U_{xx} & \lambda^* \\ -P_y & U_{yx} & 0 \end{vmatrix} \\ &= \frac{x^*}{|J|}\begin{vmatrix} -P_x & U_{xx} \\ -P_y & U_{yx} \end{vmatrix} - \frac{\lambda^*}{|J|}\begin{vmatrix} 0 & -P_x \\ -P_y & U_{yx} \end{vmatrix} \\ &\equiv T_3 + T_4 \end{aligned} \tag{12.43}$$

¿Cómo interpretamos estos dos resultados? El primero, $(\partial x^*/\partial P_x)$, nos muestra cómo afecta un cambio de P_x a la compra óptima de x; por tanto, nos pone la base para el estudio de nuestra función de demanda del consumidor para x. Para este efecto, hay dos términos componentes. El primer término, T_1, puede reescribirse, con el uso de (12.39), como $-(\partial x^*/\partial B)x^*$.

Como resultado, T_1 parece ser una medida del efecto de un cambio de B (presupuesto, o ingreso) sobre la compra óptima x^*, con x^* mismo que sirve como factor de ponderación. Sin embargo, como esta derivada obviamente tiene que ver con un cambio en los precios, T_1 debe interpretarse como el *efecto de ingreso* de un *cambio en el precio*. A medida que aumenta P_x, la declinación del ingreso real del consumidor produce un efecto sobre x^* similar al de un decremento real de B; de ahí el uso del término $-(\partial x^*/\partial B)$. Lógicamente, cuanto más prominente sea el lugar del artículo x en el presupuesto total, mayor será el efecto de este ingreso, y de ahí la aparición del factor de ponderación x^* en T_1. Esta interpretación puede demostrarse más formalmente expresando la pérdida del ingreso efectivo del consumidor por la diferencial $dB = -x^* dP_x$. Entonces, tenemos

$$x^* = -\frac{dB}{dP_x} \tag{12.44}$$

y

$$T_1 = -\left(\frac{\partial x^*}{\partial B}\right)x^* = \left(\frac{\partial x^*}{\partial B}\right)\frac{dB}{dP_x}$$

que muestra que T_1 es la medida del efecto de dP_x sobre x^* vía B, el efecto de ingreso.

Si ahora compensamos al consumidor por la pérdida del ingreso efectivo mediante un pago en efectivo numéricamente igual a dB, debido a la neutralización del efecto de ingreso, el componente restante en la derivada estático-comparativa $(\partial x^*/\partial P_x)$, es decir, T_2, va a medir el cambio de x^* debido totalmente a la sustitución inducida del precio de un artículo por otro; es decir, el *efecto de sustitución* del cambio de P_x. Para ver esto más claramente, regresemos a (12.37), y consideremos cómo va a modificar la situación la compensación del ingreso. Al estudiar sólo el efecto de dP_x (con $dP_y = dB = 0$), la primera ecuación de (12.37) puede escribirse como $-P_x\, dx^* - P_y\, dy^* = x^*\, dP_x$. Ya que la indicación de la pérdida del ingreso efectivo para el consumidor radica en la expresión $x^* dP_x$ (la cual, incidentalmente, aparece solamente en la primera ecuación), la compensación para el consumidor equivale a igualar este término a cero. Si así es, el vector de constantes de (12.41) debe cambiar de $\begin{bmatrix} x^* \\ \lambda^* \\ 0 \end{bmatrix}$ a $\begin{bmatrix} 0 \\ \lambda^* \\ 0 \end{bmatrix}$, y la versión de compensación del ingreso de la derivada $(\partial x^*/\partial P_x)$ será

$$\left(\frac{\partial x^*}{\partial P_x}\right)_{\text{compensado}} = \frac{1}{|J|}\begin{vmatrix} 0 & 0 & -P_y \\ -P_x & \lambda^* & U_{xy} \\ -P_y & 0 & U_{yy} \end{vmatrix} = \frac{\lambda^*}{|J|}\begin{vmatrix} 0 & -P_y \\ -P_y & U_{yy} \end{vmatrix} = T_2$$

Por tanto, podemos expresar (12.42) como la forma

$$\left(\frac{\partial x^*}{\partial P_x}\right) = T_1 + T_2 = \underbrace{-\left(\frac{\partial x^*}{\partial B}\right)x^*}_{\text{efecto de ingreso}} + \underbrace{\left(\frac{\partial x^*}{\partial P_x}\right)_{\text{compensado}}}_{\text{efecto de sustitución}} \tag{12.42$'$}$$

Este resultado, que descompone a la derivada estático-comparativa $(\partial x^*/\partial P_x)$ en dos componentes, un efecto de ingreso y un efecto de sustitución, es la versión para dos bienes de la así llamada ecuación de Slutsky.

¿Qué podemos decir acerca del signo de $(\partial x^*/\partial P_x)$? El efecto de sustitución T_2 es claramente negativo, porque $|J| > 0$ y $\lambda^* > 0$ [ver (12.31$'$)]. Por otro lado, el efecto de ingreso T_1 está indeterminado en el signo de acuerdo con (12.39). Si fuera negativo, reforzaría a T_2; en ese caso, un incremento de P_x2 debe disminuir la compra de x, y la curva de demanda del consumidor que maximiza la utilidad tendría pendiente negativa. Si fuera positiva, pero de magnitud relativamente pequeña, disminuiría el efecto de sustitución, aun cuando el resultado total todavía sería una curva de demanda con pendiente hacia abajo. Pero en caso de que T_1 sea positivo y domine a T_2 (como cuando x^* es un elemento importante en el presupuesto del consumidor, suministrando así un factor de ponderación demoledor), entonces un aumento de P_x va a conducir en realidad a una compra *mayor* de x, una situación especial de demanda característica de lo que se llama los *bienes de Giffen*. Normalmente, por supuesto, esperaríamos que $(\partial x^*/\partial P_x)$ fuera negativo.

Finalmente, examinemos la derivada estático-comparativa de (12.43), $(\partial y^*/\partial P_x) = T_3 + T_4$, lo cual tiene que ver con el *efecto cruzado* de un cambio en el precio de x para la compra óptima de y. El término T_3 se parece mucho al término T_1 y nuevamente tiene la interpretación de un efecto de ingreso.[10] Observa que el factor de ponderación es nuevamente x^* (en vez de y^*); esto se debe a que estamos estudiando el efecto de un cambio de P_x sobre el ingreso efectivo, que depende para su magnitud de la importancia relativa de x^* (no y^*) en el presupuesto del consumidor. Naturalmente, el término restante, T_4, es otra vez una medida del efecto de sustitución.

El signo de T_3, de acuerdo con (12.40), depende de factores tales como U_{xx}, U_{yx}, etc., y es indeterminado sin restricciones adicionales en el modelo. Sin embargo, el efecto de sustitución T_4 seguramente será positivo en nuestro modelo, ya que λ^*, P_x, P_y y $|J|$ son todos positivos. Esto significa que, a menos que sea desplazado por un efecto de ingreso negativo, un incremento del precio de x siempre aumentará la compra de y en nuestro modelo de dos artículos. En otras palabras, en el contexto del presente modelo, donde el consumidor puede escoger sólo entre dos bienes, estos bienes deben estar relacionados entre sí como sustitutos.

Aun cuando el análisis anterior se relaciona con los efectos de un cambio de P_x, nuestros resultados son rápidamente adaptables al caso de un cambio de P_y. Nuestro modelo resulta ser tal que las posiciones ocupadas por las variables x y y son perfectamente simétricas. Así, para inferir los efectos de un cambio de P_y, todo lo que se necesita es intercambiar los papeles de x y y en los resultados ya obtenidos.

Cambios proporcionales de los precios y del ingreso

También es conveniente preguntarnos como serán afectados x^* y y^* cuando los tres parámetros P_x, P_y y B se modifiquen en la misma proporción. Esta interrogante todavía está situada en el reino de la estática comparativa, pero a diferencia del análisis anterior, la presente averiguación implica el cambio simultáneo de todos los parámetros.

Cuando aumentan ambos precios, junto con el ingreso, por el mismo múltiplo j, todos los términos de la restricción presupuestaria van a aumentar j veces para transformarse en

$$jB - jxP_x - jyP_y = 0$$

Sin embargo, como el factor común j puede cancelarse, esta nueva restricción de hecho es idéntica a la anterior. Aún más, la función utilidad es independiente de estos parámetros. En consecuencia, los niveles de equilibrio anteriores de x y y continúan prevaleciendo; es decir, la posición del equilibrio del consumidor en muestro modelo es invariante para cambios proporcionales *iguales* de todos los precios y en el ingreso. Así, en este modelo, se ve que el consumidor está libre de cualquier "ilusión monetaria".

[10] Si necesita una dosis más grande de certeza de que T_3 representa el efecto de ingreso, puede usar (12.40) y (12.44) para escribir

$$T_3 = -\left(\frac{\partial y^*}{\partial B}\right) x^* = \left(\frac{\partial y^*}{\partial B}\right) \frac{dB}{dP_x}$$

Entonces, T_3 es el efecto de un cambio de P_x sobre y^* vía el factor de ingreso B.

En forma simbólica, esta situación puede describirse por las ecuaciones

$$x^*(P_x, P_y, B) = x^*(jP_x, jP_y, jB)$$
$$y^*(P_x, P_y, B) = y^*(jP_x, jP_y, jB)$$

Las funciones x^* y y^*, con la propiedad de *invariancia* recién citada, no son funciones ordinarias, sino ejemplos de una clase especial de función conocida como *funciones homogéneas*, que tienen aplicaciones económicas interesantes, las examinaremos en la sección 12.6.

EJERCICIO 12.5

1. Dado $U = (x+2)(y+1)$ y $P_x = 4$, $P_y = 6$, y $B = 130$:
   (*a*) Escriba la función lagrangiana.
   (*b*) Encuentre los niveles óptimos de compra x^* y y^*.
   (*c*) ¿Se satisface la condición suficiente de segundo orden para un máximo?
   (*d*) ¿La respuesta de (*b*) da alguna información estático-comparativa?
2. Suponga que $U = (x+2)(y+1)$, pero esta vez no asigne valores numéricos específicos a los parámetros de precio e ingreso.
   (*a*) Escriba la función lagrangiana.
   (*b*) Encuentre x^*, y^* y λ^* en los términos de los parámetros P_x, P_y y B.
   (*c*) Verfifique la condición suficiente del orden segundo para el máximo.
   (*d*) Tomando $P_x = 4$, $P_y = 6$ y $B = 130$ revise la validación de tu respuesta al problema 1.
3. ¿Puede su solución (x^* y y^*) del problema 2 suministrar alguna información estático-comparativa? Encuentre todas las derivadas estático-comparativas que pueda, evalúe sus signos, e interprete su significado económico.
4. De la función de utilidad $U = (x+2)(y+1)$ y de la restricción $xP_x + yP_y = B$ del problema 2, ya hemos encontrado los U_{ij} y $|\overline{H}|$, así como x^* y λ^*. Aún más, recordemos que $|J| = |\overline{H}|$.
   (*a*) Sustitúyalos en (12.39) y (12.40) para encontrar $(\partial x^*/\partial B)$ y $(\partial y^*/\partial B)$.
   (*b*) Sustitúyalos en (12.42) y (12.43) para encontrar $(\partial x^*/\partial P_x)$ y $(\partial y^*/\partial P_x)$.
   ¿Están de acuerdo estos resultados con los obtenidos en el problema 3?
5. Comente la validez de esta afirmación: "Si la derivada $(\partial x^*/\partial P_x)$ es negativa, entonces x no puede representar de ninguna manera a un bien inferior."
6. Al estudiar el efecto de dP_x, la primera ecuación de (12.37) se reduce a $-P_x\,dx^* - P_y\,dy^* = x^*\,dP_x$, y cuando compensamos la pérdida del ingreso efectivo del consumidor al eliminar el término x^*dP_x, la ecuación se transforma en $-P_x\,dx^* - P_y\,dy^* = 0$. Muestre que este resultado puede obtenerse en forma alternada a partir de un procedimiento de compensación mediante el cual tratamos de conservar al nivel de utilidad U^* óptimo del consumidor (en vez del ingreso efectivo) sin cambio, de modo que el término T_2 pueda interpretarse de manera alterna como $(\partial x^*/\partial P_x)_{U^*=\text{constante}}$. [*Sugerencia:* usa (12.31″).]
7. (*a*) ¿La hipótesis de la disminución de la utilidad marginal de los bienes x y y implica curvas de indiferencia estrictamente convexas?
   (*b*) ¿La hipótesis de convexidad estricta en las curvas de indiferencia implica la disminución de la utilidad marginal de los bienes x y y?

12.6 Funciones homogéneas

Se dice que una función es homogénea de grado r, si la multiplicación de cada una de las variables independientes por una constante j altera el valor de la función en la proporción j^r; es decir, si

$$f(jx_1, \ldots, jx_n) = j^r f(x_1, \ldots, x_n)$$

En general, j puede adoptar cualquier valor. Sin embargo, con objeto de que esta ecuación tenga sentido $(jx_1, \ldots, jx_n)$ no debe estar situado fuera del dominio de la función f. Por esta razón, en las aplicaciones económicas la constante j generalmente se considera positiva, ya que la mayoría de las variables económicas no admiten valores negativos.

Ejemplo 1

Dada la función $f(x, y, w) = x/y + 2w/3x$, si multiplicamos cada variable por j, obtenemos

$$f(jx, jy, jw) = \frac{(jx)}{(jy)} + \frac{2(jw)}{3(jx)} = \frac{x}{y} + \frac{2w}{3x} = f(x, y, w) = j^0 f(x, y, w)$$

En este ejemplo específico, el valor de la función *no* será afectado de ninguna manera por los cambios proporcionales iguales en todas las variables independientes; también se puede decir que el valor de la función se modifica por un múltiplo de j^0 $(= 1)$. Esto hace que la función f sea homogénea de grado cero.

Observe que las funciones x^* y y^* citadas al final de la sección 12.5 son ambas homogéneas de grado cero.

Ejemplo 2

Cuando multiplicamos cada variable de la función

$$g(x, y, w) = \frac{x^2}{y} + \frac{2w^2}{x}$$

por j, obtenemos

$$g(jx, jy, jw) = \frac{(jx)^2}{(jy)} + \frac{2(jw)^2}{(jx)} = j\left(\frac{x^2}{y} + \frac{2w^2}{x}\right) = jg(x, y, w)$$

La función g es homogénea de grado uno (o de primer grado); la multiplicación de cada variable por j también altera el valor de la función exactamente j veces.

Ejemplo 3

Ahora, considere la función $h(x, y, w) = 2x^2 + 3yw - w^2$. Esta vez una multiplicación similar nos da

$$h(jx, jy, jw) = 2(jx)^2 + 3(jy)(jw) - (jw)^2 = j^2 h(x, y, w)$$

Así, la función h es homogénea de grado dos; en este caso, duplicar todas las variables, cuadruplica el valor de la función.

Homogeneidad lineal

En el estudio de las funciones de producción se hace amplio uso de las funciones homogéneas de primer grado. Con frecuencia, a éstas se les denomina funciones *linealmente homogéneas*, con el adverbio *linealmente* que modifica al adjetivo *homogéneas*. Sin embargo, algunos autores parecen preferir la terminología un poco confusa de funciones homogéneas *lineales*, o aun funciones homogéneas *y lineales*, lo que tiende a trasmitir en forma errónea la impresión de que las funciones mismas son lineales. Basándonos en la función g del ejemplo 2, sabemos que una función que es homogénea en primer grado *no es necesariamente* lineal en sí misma. Por tanto, se debe evitar el uso de los términos "funciones homogéneas lineales" y "funciones homogéneas y lineales" a menos que, por supuesto, las funciones en cuestión sean realmente lineales. Sin embargo, observe que no es incorrecto hablar de "homogeneidad lineal", lo que significa homogeneidad de grado uno, porque la modificación de un sustantivo (homogeneidad) requiere del uso de un adjetivo (lineal).

Como el campo primario de aplicación de las funciones linealmente homogéneas está en la teoría de la producción, adoptemos como el marco de nuestro estudio una función de producción de la forma, por ejemplo,

$$Q = f(K, L) \tag{12.45}$$

Ya sea que se aplique al nivel *micro* o *macro*, la hipótesis matemática de homogeneidad lineal apuntaría a la hipótesis económica de retornos constantes a escala, ya que la homogeneidad lineal significa que la elevación de todos los insumos (variables independientes) j veces siempre va a elevar el producto (valor de la función) exactamente j veces.

¿Cuáles son las propiedades únicas que caracterizan a esta función de producción linealmente homogénea?

Propiedad I Dada la función de producción linealmente homogénea $Q = f(K, L)$, el producto físico promedio del trabajo (APP_L) y del capital (APP_K) pueden expresarse sólo como funciones de la relación capital-trabajo, $k \equiv K/L$.

Para probar esto, multiplicamos cada variable independiente de (12.45) por un factor $j = 1/L$. En virtud de la homogeneidad lineal, esto va a modificar al producto de Q a $jQ = Q/L$. El lado derecho de (12.45) se transforma correspondientemente en

$$f\left(\frac{K}{L}, \frac{L}{L}\right) = f\left(\frac{K}{L}, 1\right) = f(k, 1)$$

Como las variables K y L en la función original deben reemplazarse (siempre que aparezcan) por k y 1, respectivamente, el lado derecho se convierte sólo en una función de la relación capital-trabajo k, por ejemplo, $\phi(k)$, que es una función con un solo argumento, k, aun cuando en realidad intervienen dos variables independientes K y L en ese argumento. Al igualar ambos lados, tenemos

$$\text{APP}_L \equiv \frac{Q}{L} = \phi(k) \tag{12.46}$$

Entonces, se encuentra que la expresión para APP_K es

$$\text{APP}_K \equiv \frac{Q}{K} = \frac{Q}{L}\frac{L}{K} = \frac{\phi(k)}{k} \tag{12.47}$$

Como ambos productos promedio dependen solamente de k, la homogeneidad lineal implica que, siempre que la relación K/L se mantenga constante (cualesquiera que sean los niveles absolutos de K y L), los productos promedio también serán constantes. Por lo tanto, mientras que la función de producción es homogénea de grado uno, tanto APP_L como APP_K son homogéneas de grado cero en las variables K y L, ya que los cambios proporcionalmente iguales en K y L (manteniendo k constante) no alteran las magnitudes de los productos promedio.

Propiedad II Dada una función de producción linealmente homogénea $Q = f(K, L)$, los productos físicos marginales MPP_L y MPP_K pueden expresarse como funciones de k como única variable.

Para encontrar los productos marginales, escribimos primero el producto total como

$$Q = L\phi(k) \qquad [\text{por } (12.46)] \tag{12.45'}$$

y luego diferenciamos Q respecto a K y L. Para este propósito, nos sirven los dos siguientes resultados preliminares:

$$\frac{\partial k}{\partial K} = \frac{\partial}{\partial K}\left(\frac{K}{L}\right) = \frac{1}{L} \qquad \frac{\partial k}{\partial L} = \frac{\partial}{\partial L}\left(\frac{K}{L}\right) = \frac{-K}{L^2} \tag{12.48}$$

Los resultados de la diferenciación son

$$\begin{aligned}
\text{MPP}_K \equiv \frac{\partial Q}{\partial K} &= \frac{\partial}{\partial K}[L\phi(k)] \\
&= L\frac{\partial \phi(k)}{\partial K} = L\frac{d\phi(k)}{dk}\frac{\partial k}{\partial K} && [\text{regla de la cadena}] \\
&= L\phi'(k)\left(\frac{1}{L}\right) = \phi'(k) && [\text{por } (12.48)] \quad \textbf{(12.49)}
\end{aligned}$$

$$\begin{aligned}
\text{MPP}_L \equiv \frac{\partial Q}{\partial L} &= \frac{\partial}{\partial L}[L\phi(k)] \\
&= \phi(k) + L\frac{\partial \phi(k)}{\partial L} && [\text{regla del producto}] \\
&= \phi(k) + L\phi'(k)\frac{\partial k}{\partial L} && [\text{regla de la cadena}] \\
&= \phi(k) + L\phi'(k)\frac{-K}{L^2} && [\text{por } (12.48)] \\
&= \phi(k) - k\phi'(k) && \textbf{(12.50)}
\end{aligned}$$

lo que muestra realmente que MPP_K y MPP_L son funciones de k como única variable .

Al igual que los productos promedio, los productos marginales permanecen iguales siempre que la relación capital-trabajo se mantenga constante; ellos son homogéneos de grado cero en las variables K y L.

Propiedad III (teorema de Euler) Si $Q = f(K, L)$ es linealmente homogénea, entonces

$$K\frac{\partial Q}{\partial K} + L\frac{\partial Q}{\partial L} \equiv Q$$

PRUEBA

$$\begin{aligned}
K\frac{\partial Q}{\partial K} + L\frac{\partial Q}{\partial L} &= K\phi'(k) + L[\phi(k) - k\phi'(k)] && [\text{por } (12.49), (12.50)] \\
&= K\phi'(k) + L\phi(k) - K\phi'(k) && [k \equiv K/L] \\
&= L\phi(k) = Q && [\text{por } (12\text{-}45')]
\end{aligned}$$

Observe que este resultado es válido para cualquier valor de K y L; por esto puede escribirse la propiedad como una identidad. Lo que afirma esta propiedad es que el valor de una función linealmente homogénea siempre puede expresarse como una suma de términos, cada uno de los cuales es el producto de una de las variables independientes y de la derivada parcial de primer orden respecto a esa variable, independientemente de los niveles de los dos insumos que realmente se empleen. Sin embargo, tenga cuidado en distinguir entre la identidad $K\frac{\partial Q}{\partial K} + L\frac{\partial Q}{\partial L} \equiv Q$ [teorema de Euler, que se aplica solamente al caso de retornos constantes a escala de $Q = f(K, L)$] y la ecuación $dQ = \frac{\partial Q}{\partial K}dK + \frac{\partial Q}{\partial L}dL$ [diferencial total de Q, para *cualquier* función $Q = f(K, L)$].

Económicamente, esta propiedad implica que en condiciones de retornos constantes a escala, si a cada factor de insumo se le paga la cantidad de su producto marginal, el producto total será distribuido exactamente por las participaciones de todos los factores de insumo, o en forma equivalente, la ganancia económica pura será cero. Como esta situación describe el equilibrio a largo plazo en competencia pura, se pensaba que solamente las funciones de producción linealmente homogéneas tendrían sentido en economía. Por supuesto que éste no es el caso. La ganancia económica cero en el equilibrio a largo plazo es causada por las fuerzas de competencia a través de la entrada y la salida de las firmas, independientemente de la naturaleza específica de las funciones de producción que realmente prevalecen. Entonces, no es obligatorio tener una función de producción que asegure el agotamiento del producto para cualquiera y para todos los pares (K, L). Aún más, cuando existe la competencia imperfecta en los mercados de factores, la remuneración de los factores puede no ser igual a los productos marginales y, en consecuencia, el teorema de Euler resulta ser irrelevante para el panorama de la distribución. Sin embargo, las funciones de producción linealmente homogéneas con frecuencia son convenientes para trabajar con ellas debido a las diversas propiedades matemáticas que poseen.

Función de producción de Cobb-Douglas

Una función de producción específica muy usada en análisis económico (citada en la sección 11.6, ejemplo 5) es la *función de producción de Cobb-Douglas*:

$$Q = AK^{\alpha}L^{1-\alpha} \tag{12.51}$$

donde A es una constante positiva y α es una fracción positiva. Lo que se va a considerar aquí en primer lugar es una versión generalizada de esta función,

$$Q = AK^{\alpha}L^{\beta} \tag{12.52}$$

donde β es otra fracción positiva que puede ser igual o no a $1 - \alpha$. Algunas de las principales características de esta función son: (1) es homogénea de grado $(\alpha + \beta)$; (2) en el caso especial de $\alpha + \beta = 1$, es linealmente homogénea; (3) sus isocuantas tienen pendiente negativa en todos lados y son estrictamente convexas para valores positivos de K y L; y (4) es estrictamente cuasicóncava para K y L positivos.

Su homogeneidad se ve fácilmente a partir del hecho de que, al cambiar K y L a jK y jL, respectivamente, el producto se transformará en

$$A(jK)^{\alpha}(jL)^{\beta} = j^{\alpha+\beta}(AK^{\alpha}L^{\beta}) = j^{\alpha+\beta}Q$$

es decir, la función es homogénea de grado $(\alpha + \beta)$. En el caso $\alpha + \beta = 1$, habrá retornos constantes a escala, ya que la función será linealmente homogénea. (¡Observe, sin embargo, que esta función *no* es lineal! Entonces, sería confuso referirse a ella como una función "homogénea *lineal*" o "*lineal* y homogénea".) El hecho de que sus isocuantas tienen pendiente negativa puede verificarse a partir de los signos de las derivadas dK/dL y d^2K/dL^2 (o los signos de dL/dK y d^2L/dK^2). Para cualquier producto positivo Q_0, (12.52) puede escribirse como

$$AK^{\alpha}L^{\beta} = Q_0 \qquad (A, K, L, Q_0 > 0)$$

Al tomar el logaritmo natural de ambos lados y al trasponer, encontramos que

$$\ln A + \alpha \ln K + \beta \ln L - \ln Q_0 = 0$$

lo que implícitamente define a K como una función de L.[11] Por la regla de la función implícita y la regla de los logaritmos, tenemos

$$\frac{dK}{dL} = -\frac{\partial F/\partial L}{\partial F/\partial K} = -\frac{(\beta/L)}{(\alpha/K)} = -\frac{\beta K}{\alpha L} < 0$$

Entonces, se sigue que

$$\frac{d^2K}{dL^2} = \frac{d}{dL}\left(-\frac{\beta K}{\alpha L}\right) = -\frac{\beta}{\alpha}\frac{d}{dL}\left(\frac{K}{L}\right) = -\frac{\beta}{\alpha}\frac{1}{L^2}\left(L\frac{dK}{dL} - K\right) > 0$$

Los signos de estas derivadas establecen que la isocuanta (cualquier isocuanta) tiene pendiente hacia abajo en todos lados y es estrictamente convexa en el plano LK para valores positivos de K y L. Esto, por supuesto, puede esperarse solamente de una función que sea estrictamente cuasicóncava para K y L positivos. Para la característica de cuasiconcavidad estricta de esta función, consulta el ejemplo 5 de la sección 12.4, donde estudiamos una función similar.

Examinemos ahora el caso $\alpha + \beta = 1$ (propiamente la función de Cobb-Douglas), para verificar las tres propiedades de homogeneidad lineal citadas anteriormente. Primero que todo, digamos que el producto total en este caso especial se expresa como

$$Q = AK^{\alpha}L^{1-\alpha} = A\left(\frac{K}{L}\right)^{\alpha} L = LAk^{\alpha} \qquad \textbf{(12.51')}$$

donde la expresión Ak^{α} es una versión específica de la expresión general $\phi(k)$ que usamos anteriormente. Por lo tanto, los productos promedio son

$$\begin{aligned} \text{APP}_L &= \frac{Q}{L} = Ak^{\alpha} \\ \text{APP}_K &= \frac{Q}{K} = \frac{Q}{L}\frac{L}{K} = \frac{Ak^{\alpha}}{k} = Ak^{\alpha-1} \end{aligned} \qquad \textbf{(12.53)}$$

los cuales son ahora funciones solamente de k.

Segundo, la diferenciación de $Q = AK^{\alpha}L^{1-\alpha}$ proporciona los productos marginales:

$$\begin{aligned} \frac{\partial Q}{\partial K} &= A\alpha K^{\alpha-1}L^{-(\alpha-1)} = A\alpha\left(\frac{K}{L}\right)^{\alpha-1} = A\alpha k^{\alpha-1} \\ \frac{\partial Q}{\partial L} &= AK^{\alpha}(1-\alpha)L^{-\alpha} = A(1-\alpha)\left(\frac{K}{L}\right)^{\alpha} = A(1-\alpha)k^{\alpha} \end{aligned} \qquad \textbf{(12.54)}$$

y éstas también son funciones solamente de k.

[11] Se satisfacen las condiciones del teorema de la función implícita porque F (la expresión del lado izquierdo) tiene derivadas parciales continuas, y porque $\partial F/\partial K = \alpha/K \neq 0$ para valores positivos de K.

Por último, podemos verificar el teorema de Euler usando (12.54) como sigue:

$$\begin{aligned} K\frac{\partial Q}{\partial K}+L\frac{\partial Q}{\partial L} &= KA\alpha k^{\alpha-1}+LA(1-\alpha)k^{\alpha} \\ &= LAk^{\alpha}\left(\frac{K\alpha}{Lk}+1-\alpha\right) \\ &= LAk^{\alpha}(\alpha+1-\alpha)=LAk^{\alpha}=Q \qquad \text{[por (12.51')]} \end{aligned}$$

Pueden asignarse significados económicos interesantes a los exponentes α y $(1-\alpha)$ en la función de producción Cobb-Douglas linealmente homogénea. Si se supone que cada insumo se paga por la cantidad de su producto marginal, la participación relativa del producto total acumulado al capital será

$$\frac{K(\partial Q/\partial K)}{Q}=\frac{KA\alpha k^{\alpha-1}}{LAk^{\alpha}}=\alpha$$

En forma similar, la participación relativa del trabajo será

$$\frac{L(\partial Q/\partial L)}{Q}=\frac{LA(1-\alpha)k^{\alpha}}{LAk^{\alpha}}=1-\alpha$$

De esta manera, el exponente de cada variable de insumo indica la participación relativa de ese insumo en el producto total. Viéndolo de otra manera, también podemos interpretar el exponente de cada variable de insumo como la elasticidad parcial del producto respecto a ese insumo. Esto se debe a que la expresión de participación de capital recién dada es equivalente a la expresión $\dfrac{\partial Q/\partial K}{Q/K}\equiv\varepsilon_{QK}$ y, en forma similar, la expresión de participación de trabajo recién dada es precisamente aquella de ε_{QL}.

¿Qué podemos esperar del significado de la constante A? Para valores dados de K y L, la magnitud de A afectará proporcionalmente al nivel de Q. Entonces, A puede considerarse como un *parámetro de eficiencia*, es decir, como un indicador del estado de la tecnología.

Extensiones de los resultados

Hemos estudiado la homogeneidad lineal en el contexto específico de las funciones de producción, pero las propiedades citadas son igualmente válidas en otros contextos, siempre que las variables K, L y Q se reinterpreten apropiadamente.

Aún más, podemos extender nuestros resultados al caso de más de dos variables. Con una función linealmente homogénea

$$y=f(x_1,x_2,\ldots,x_n)$$

nuevamente podemos dividir cada variable entre x_1 (es decir, multiplicarla por $1/x_1$) y obtener el resultado

$$y=x_1\phi\left(\frac{x_2}{x_1},\frac{x_3}{x_1},\ldots,\frac{x_n}{x_1}\right) \qquad \text{[homogeneidad de grado 1]}$$

que es comparable con (12.45′). Aún más, el teorema de Euler se extiende fácilmente a la forma

$$\sum_{i=1}^{n}x_i f_i\equiv y \qquad \text{[teorema de Euler]}$$

donde las derivadas parciales de la función original f (a saber, f_i) son nuevamente homogéneas de grado cero en las variables x_i, como en el caso de dos variables.

Las extensiones anteriores de hecho también pueden generalizarse con relativa facilidad a una función homogénea de grado r. En primer lugar, por definición de homogeneidad, podemos escribir en este caso

$$y = x_1^r \phi\left(\frac{x_2}{x_1}, \frac{x_3}{x_1}, \ldots, \frac{x_n}{x_1}\right) \qquad \text{[homogeneidad de grado } r\text{]}$$

La versión modificada del teorema de Euler aparece ahora en la forma

$$\sum_{i=1}^{n} x_i f_i \equiv ry \qquad \text{[teorema de Euler]}$$

donde se ha añadido una constante multiplicativa r a la variable dependiente y a la derecha. Y finalmente, las derivadas parciales de la función original f, las f_i, serán todas homogéneas de grado $(r - 1)$ en las variables x_i. Entonces, se puede ver que el caso homogéneo lineal es meramente un caso especial en el cual $r = 1$.

EJERCICIO 12.6

1. Determine si las siguientes funciones son homogéneas. Si lo son, ¿de qué grado?
   (*a*) $f(x, y) = \sqrt{xy}$
   (*b*) $f(x, y) = (x^2 - y^2)^{1/2}$
   (*c*) $f(x, y) = x^3 - xy + y^3$
   (*d*) $f(x, y) = 2x + y + 3\sqrt{xy}$
   (*e*) $f(x, y, w) = \dfrac{xy^2}{w} + 2xw$
   (*f*) $f(x, y, w) = x^4 - 5yw^3$
2. Muestre que la función (12.45) puede expresarse de manera alterna como $Q = K\psi\left(\frac{L}{K}\right)$ en lugar de $Q = L\phi\left(\frac{K}{L}\right)$.
3. Deduzca del teorema de Euler que, con retornos constantes a escala:
   (*a*) Cuando $\text{MPP}_K = 0$, APP_L es igual a MPP_L.
   (*b*) Cuando $\text{MPP}_L = 0$, APP_K es igual a MPP_K.
4. Tomando como base de (12.46) a (12.50), verifique si lo que sigue es verdad bajo condiciones de retornos constantes de escala:
   (*a*) Una curva APP_L puede graficarse contra $k\ (= K/L)$ como la variable independiente (sobre el eje horizontal).
   (*b*) MPP_K se mide con la pendiente de curva APP_L.
   (*c*) APP_K se mide por la pendiente del radio del vector de la curva APP_L.
   (*d*) $\text{MPP}_L = \text{APP}_L - k(\text{MPP}_K) = \text{APP}_L - k$ (pendiente de APP_L).
5. Use (12.53) y (12.54) para verificar que las relaciones descritas en el problema 4*b*, *c* y *d* están de acuerdo con la función de producción de Cobb-Douglas.
6. Dada la función de producción $Q = AK^{\alpha}L^{\beta}$, muestre que:
   (*a*) $\alpha + \beta > 1$ implica retornos crecientes a escala.
   (*b*) $\alpha + \beta < 1$ implica retornos decrecientes a escala.
   (*c*) α y β son, respectivamente, las elasticidades parciales del producto respecto al capital y a los productos del trabajo.

7. Sea el producto una función de tres insumos: $Q = AK^aL^bN^c$.
   (*a*) ¿Es homogénea esta función?
   (*b*) ¿Bajo qué condición habría retornos constantes a escala? ¿Retornos crecientes a escala?
   (*c*) Encuentre la participación del producto para el insumo N, si se paga por la cantidad de su producto marginal.
8. Sea la función de producción $Q = g(K, L)$ homogénea de grado 2.
   (*a*) Escriba una ecuación para expresar la propiedad de homogeneidad de segundo grado de esta función.
   (*b*) Encuentre una expresión para Q en términos de $\phi(k)$, de acuerdo con (12.45′).
   (*c*) Encuentre la función MPP_K. ¿Es todavía MPP_K una función solamente de k, como en el caso de homogeneidad lineal?
   (*d*) ¿Es homogénea la función MPP_K en K y L? Si lo es, ¿de qué grado?

12.7 Combinación de insumos de costo mínimo

Como otro ejemplo de optimización restringida, estudiemos el problema de encontrar la combinación de insumos de costo mínimo para la producción de un nivel especificado de producto Q_0 que represente, por ejemplo, una orden especial de un cliente. Aquí vamos a trabajar con una función de producción general; sin embargo, posteriormente haremos referencia a las funciones de producción homogéneas.

Condición de primer orden

Supongamos una función de producción suave con dos variables como insumos, $Q = Q(a, b)$, donde $Q_a, Q_b > 0$, y suponiendo que los dos precios de los insumos son exógenos (aunque nuevamente se omite el subíndice cero), podemos formular el problema como uno de minimización del costo

$$C = aP_a + bP_b$$

sujeto a la restricción de producto

$$Q(a, b) = Q_0$$

Entonces, la función lagrangiana es

$$Z = aP_a + bP_b + \mu[Q_0 - Q(a, b)]$$

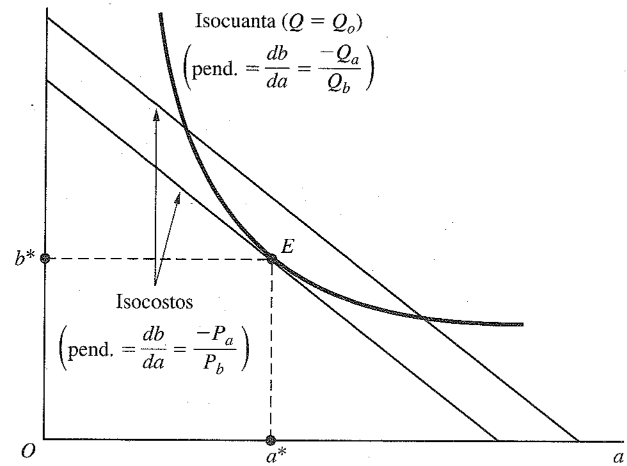
Para satisfacer la condición de primer orden para un mínimo C, los niveles de los insumos (las variables de elección) deben satisfacer las siguientes ecuaciones simultáneas:

$$\begin{aligned} Z_\mu &= Q_0 - Q(a, b) = 0 \\ Z_a &= P_a - \mu Q_a = 0 \\ Z_b &= P_b - \mu Q_b = 0 \end{aligned}$$

La primera ecuación de este conjunto es simplemente la restricción reformulada, y las dos últimas implican la condición

$$\frac{P_a}{Q_a} = \frac{P_b}{Q_b} = \mu \qquad \textbf{(12.55)}$$

FIGURA 12.8



En el punto de la combinación óptima de insumos, la relación producto marginal precio-insumo debe ser la misma para cada insumo. Como esta relación mide la cantidad de gasto por unidad de producto marginal del insumo en cuestión, al multiplicador de Lagrange puede dársele la interpretación del costo marginal de producción en el estado óptimo. Por supuesto que esta interpretación es enteramente consistente con nuestro descubrimiento anterior de (12.16) de que el valor óptimo de los multiplicadores de Lagrange mide el efecto estático comparativo de la constante de restricción sobre el valor óptimo de la función objetivo; es decir, $\mu^* = (\S C^* / \S Q_0)$, donde el símbolo § indica que ésta es una derivada parcial total.

La ecuación (12.55) puede escribirse de manera alterna en la forma

$$\frac{P_a}{P_b} = \frac{Q_a}{Q_b} \tag{12.55$'$}$$

que debería compararse con (12.31″). Presentada en esta forma, la condición de primer orden puede explicarse en términos de isocuantas e isocostos. Como aprendimos en (11.36), la relación Q_a/Q_b es el negativo de la pendiente de una isocuanta; es decir, es una medida de la *tasa marginal de la sustitución técnica de a por b* (MRTS_{ab}). En este modelo, el nivel de producto se especifica en Q_0; entonces, solamente se incluye una isocuanta, como se muestra en la figura 12.8, con una pendiente negativa.

Por otro lado, la relación P_a/P_b representa el negativo de la pendiente de los *isocostos* (una noción comparable con la línea de presupuesto en la teoría del consumidor). Un isocosto, definido como el lugar geométrico de las combinaciones de insumos que causan el mismo costo total, se expresa mediante la ecuación

$$C_0 = aP_a + bP_b \qquad \text{o} \qquad b = \frac{C_0}{P_b} - \frac{P_a}{P_b}a$$

donde C_0 representa una cifra de costo (paramétrica). Cuando se grafica en el plano *ab*, como en la figura 12.8; por lo tanto, ofrece una familia de líneas rectas con pendiente (negativa) $-P_a/P_b$ (e intercepción vertical C_0/P_b). Por lo tanto, la igualdad de las dos relaciones apunta a la igualdad de las pendientes de la isocuanta y de un isocosto seleccionado. Como estamos obligados a permanecer en la isocuanta dada, esta condición nos conduce al punto de tangencia E y a la combinación de insumos (a^*, b^*).

Condición de segundo orden

Para asegurar un costo *mínimo*, es suficiente (después de cumplir la condición de primer orden) tener un hessiano orlado negativo, es decir, tener

$$|\overline{H}| = \begin{vmatrix} 0 & Q_a & Q_b \\ Q_a & -\mu Q_{aa} & -\mu Q_{ab} \\ Q_b & -\mu Q_{ba} & -\mu Q_{bb} \end{vmatrix} = \mu\left(Q_{aa}Q_b^2 - 2Q_{ab}Q_aQ_b + Q_{bb}Q_a^2\right) < 0$$

Como el valor óptimo de μ (costo marginal) es positivo, esto se reduce a la condición de que la expresión que está entre paréntesis sea negativa cuando se evalúa para E.

De (11.44) recordamos que la curvatura de una isocuanta se representa por la segunda derivada

$$\frac{d^2b}{da^2} = \frac{-1}{Q_b^3}\left(Q_{aa}Q_b^2 - 2Q_{ab}Q_aQ_b + Q_{bb}Q_a^2\right)$$

donde aparece la misma expresión entre paréntesis. Siempre que Q_b sea positivo, la satisfacción de la condición suficiente de segundo orden implicaría que d^2b/da^2 es positivo –es decir, la isocuanta es estrictamente convexa– en el punto de tangencia. En este contexto, la convexidad estricta de la isocuanta también implicaría la satisfacción de la condición suficiente de segundo orden. Como la isocuanta tiene pendiente negativa, la convexidad estricta puede significar solamente un d^2b/da^2 positivo (d^2b/da^2 cero es posible solamente para un punto estacionario en la isocuanta), lo que a su vez aseguraría que $|\overline{H}| < 0$. Sin embargo, nuevamente debemos tener presente que la condición suficiente $|\overline{H}| < 0$ (y, por lo tanto, la convexidad estricta de la isocuanta) en la tangencia es, por sí misma, no necesaria para la minimización de C. Específicamente, C puede minimizarse aún cuando la isocuanta sea convexa (no estrictamente), en una situación de mínimos múltiples análoga a la figura 12.7*b*, con $d^2b/da^2 = 0$ y $|\overline{H}| = 0$ para cada mínimo.

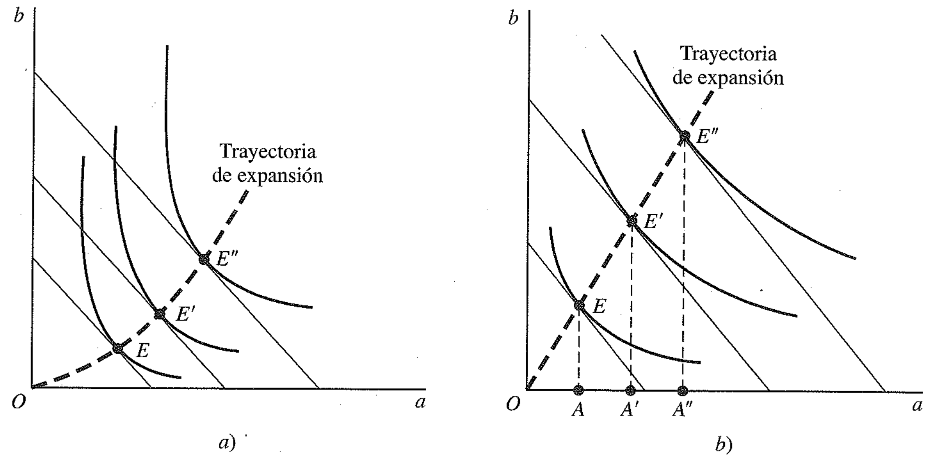
Al estudiar el modelo de maximización de utilidad (sección 12.5), señalamos que una función de utilidad suave y creciente estrictamente cuasicóncava $U = U(x, y)$ origina curvas de indiferencia estrictamente convexas con pendiente hacia abajo en todos lados en el plano xy. Como la noción de las isocuantas es casi idéntica a la de las curvas de indiferencia,[12] podemos razonar por analogía que una función suave y creciente de producción estrictamente cuasicóncava $Q = Q(a, b)$ puede generar isocuantas con pendiente hacia abajo estrictamente convexas en todos lados en el plano ab. Si suponemos una función de producción de este tipo, entonces obviamente siempre se satisfará la condición suficiente de segundo orden. Aún más, debe ser claro que el C^* resultante será un mínimo restringido absoluto único.

La trayectoria de expansión

Veamos ahora uno de los aspectos estático-comparativo de este modelo. Suponiendo una relación *fija* de los dos precios de los insumos, postulemos incrementos sucesivos de Q_0 (ascensión a isocuantas cada vez más altas) y rastreemos el efecto de la combinación b^*/a^* de costo mínimo. Por supuesto que cada desplazamiento de la isocuanta conducirá a un nuevo punto de tangencia, con un isocosto más alto. El lugar geométrico de estos puntos de tangencia, conocido como la *trayectoria de expansión* de la compañía, sirve para describir las combinaciones de costo mínimo requeridas para producir diversos niveles de Q_0. En la figura 12.9 se muestran dos formas posibles de la trayectoria de expansión.

[12] Ambas tienen la naturaleza de curvas de "isovalor". Difieren solamente en el campo de aplicación; las curvas de indiferencia se usan en los modelos de consumo, y las isocuantas, en los modelos de producción.

FIGURA 12.9



Si suponemos la convexidad estricta de las isocuantas (por tanto, la satisfacción de la condición de segundo orden), la trayectoria de expansión se podrá derivar directamente de la condición de primer orden (12.55′). Ilustremos esto para la versión generalizada de la función de producción de Cobb-Douglas.

La condición (12.55′) requiere la igualdad de la relación insumo-precio y de la relación de producto marginal. Para la función $Q = Aa^{\alpha}b^{\beta}$, esto significa que cada punto de la trayectoria de expansión debe satisfacer

$$\frac{P_a}{P_b} = \frac{Q_a}{Q_b} = \frac{A\alpha a^{\alpha-1}b^{\beta}}{Aa^{\alpha}\beta b^{\beta-1}} = \frac{\alpha b}{\beta a} \tag{12.56}$$

lo que implica que la relación de insumo óptima debería ser

$$\frac{b^*}{a^*} = \frac{\beta P_a}{\alpha P_b} = \text{son constantes} \tag{12.57}$$

ya que α, β, y los precios de los insumos son constantes. Como resultado, todos los puntos en la trayectoria de expansión deben mostrar la misma relación de insumos *fija*; es decir, la trayectoria de expansión debe ser una línea recta que emana del punto de origen. Esto se ilustra en la figura 12.9*b*, donde las relaciones de insumos para los diferentes puntos de tangencia (AE/OA, $A'E'/OA'$, y $A''E''/OA''$) son iguales.

La linealidad de la trayectoria de expansión es característica de la función generalizada de Cobb-Douglas independientemente de que $\alpha + \beta = 1$, ya que la derivación del resultado de (12.57) no depende de la hipótesis $\alpha + \beta = 1$. De hecho, cualquier función de producción homogénea (no necesariamente la de Cobb-Douglas) originará una trayectoria de expansión lineal para cada conjunto de precios de insumos, por la siguiente razón: si es homogénea de grado r (por ejemplo), ambas funciones de producto marginal Q_a y Q_b deben ser homogéneas de grado $(r-1)$ para los insumos a y b; entonces, un incremento de j veces en ambos insumos producirá un cambio de j^{r-1} veces en los valores de Q_a y Q_b, que dejará intacta la relación Q_a/Q_b. Por lo tanto, si la condición de primer orden $P_a/P_b = Q_a/Q_b$ se cumple para precios de insumos dados por una combinación específica de los insumos (a_0, b_0), también debe satisfacerse por una combinación (ja_0, jb_0) —precisamente como lo ilustra la trayectoria de expansión lineal de la figura 12.9*b*.

Aunque *cualquier* función de producción homogénea puede originar una trayectoria de expansión lineal, el grado específico de homogeneidad no tiene mucha importancia para la interpretación de la trayectoria de expansión. En la figura 12.9*b* hemos dibujado la distancia OE igual a la EE', de modo que el punto E' implica la duplicación de la escala del punto E. Ahora, si la función de producción es homogénea de grado *uno*, el producto en E' debe ser el doble ($2^1 = 2$) del de E. Pero si el grado de homogeneidad es *dos*, el producto en E' será cuatro veces ($2^2 = 4$) el de E. Entonces, el espaciamiento de las isocuantas para $Q = 1$, $Q = 2, \ldots,$ será ampliamente diferente para diferentes grados de homogeneidad.

Funciones homotéticas

Hemos explicado que, dado un conjunto de precios de insumos, la homogeneidad (de cualquier grado) de la función de producción produce una trayectoria de expansión lineal. Pero las trayectorias de expansión lineal no son privativas de las funciones de producción homogénea, ya que una clase más general de funciones, conocidas como *funciones homotéticas*, también las puede producir.

La homotecia puede surgir de una función compuesta en la forma

$$H = h[Q(a, b)] \qquad [h'(Q) \neq 0] \tag{12.58}$$

donde $Q(a, b)$ es homogénea de grado r. Aunque se deriva de una función homogénea, la función $H = H(a, b)$ es en general *no* homogénea en las variables a y b. Sin embargo, las trayectorias de expansión de $H(a, b)$, como las de $Q(a, b)$, son lineales. La clave de este resultado es que, para cualquier punto dado en el plano ab, la isocuanta H comparte la misma pendiente que la isocuanta Q:

$$\begin{aligned} \text{Pendiente de la isocuanta } H &= -\frac{H_a}{H_b} = -\frac{h'(Q)Q_a}{h'(Q)Q_b} \\ &= -\frac{Q_a}{Q_b} = \text{Pendiente de la isocuanta } Q \end{aligned} \tag{12.59}$$

Ahora, la linealidad de las trayectorias de expansión de $Q(a, b)$ implica, y está implicada por, la condición

$$-\frac{Q_a}{Q_b} = \text{constante para cualquier } \frac{b}{a} \text{ dado}$$

Sin embargo, en vista de (12.59), tenemos inmediatamente

$$-\frac{H_a}{H_b} = \text{constante para cualquier } \frac{b}{a} \text{ dado} \tag{12.60}$$

también. Y esto establece que $H(a, b)$ de igual manera produce trayectorias de expansión lineal.

El concepto de homotecia es más general que el de homogeneidad. De hecho, toda función homogénea forma parte automáticamente de la familia homotética, pero una función homotética puede no pertenecer a la familia homogénea. El hecho de que una función homogénea siempre es homotética puede verse en (12.58), donde si hacemos que la función $H = h(Q)$ adopte la forma específica $H = Q$, con $h'(Q) = dH/dQ = 1$, entonces la función Q, siendo idéntica a la misma función H, obviamente es homotética. El que una función homotética no sea homogénea lo ilustraremos en el ejemplo 2.

Al definir la función homotética H, especificamos en (12.58) que $h'(Q) \neq 0$. Esto nos permite evitar la división por cero en (12.59). Aun cuando la especificación $h'(Q) \neq 0$ es el único requerimiento desde el punto de vista matemático, las consideraciones económicas sugerirían la restricción más drástica $h'(Q) > 0$. Ya que si $H(a, b)$, al igual que $Q(a, b)$, debe servir como una función de producción; es decir, si H va a denotar al producto, entonces deberá hacerse que H_a y H_b, respectivamente, vayan en la misma dirección que Q_a y Q_b en la función $Q(a, b)$. Entonces, $H(a, b)$ necesita restringirse a ser una transformación crecientemente monótona de $Q(a, b)$.

Las funciones de producción homotéticas (incluyendo el caso especial de las homogéneas) poseen la interesante propiedad de que la elasticidad (parcial) del nivel óptimo de los insumos respecto al nivel de producto es uniforme para todos los insumos. Para ver esto, recordemos que la linealidad de las trayectorias de expansión de las funciones homotéticas significa que la relación de insumos óptima b^*/a^* no se afecta por un cambio del nivel exógeno H_0 del producto. Entonces $\partial(b^*/a^*)/\partial H_0 = 0$ o

$$\frac{1}{a^{*2}}\left(a^*\frac{\partial b^*}{\partial H_0} - b^*\frac{\partial a^*}{\partial H_0}\right) = 0 \qquad \text{[regla del cociente]}$$

Al multiplicar por $a^{*2}H_0$ y reordenar, obtenemos

$$\frac{\partial a^*}{\partial H_0}\frac{H_0}{a^*} = \frac{\partial b^*}{\partial H_0}\frac{H_0}{b^*} \qquad \text{o} \qquad \varepsilon_{a^*H_0} = \varepsilon_{b^*H_0}$$

que es lo que se afirmó anteriormente.

Ejemplo 1

Sea $H = Q^2$, donde $Q = Aa^\alpha b^\beta$. Ya que $Q(a, b)$ es homogénea y $h'(Q) = 2Q$ es positiva para un producto positivo, $H(a, b)$ es homotética para $Q > 0$. Verificaremos que cumple con (12.60). Por sustitución, primero tenemos

$$H = Q^2 = (Aa^\alpha b^\beta)^2 = A^2a^{2\alpha}b^{2\beta}$$

Entonces, la pendiente de las isocuantas de H se expresa como

$$-\frac{H_a}{H_b} = -\frac{A^2 2\alpha a^{2\alpha-1}b^{2\beta}}{A^2a^{2\alpha}2\beta b^{2\beta-1}} = -\frac{\alpha b}{\beta a} \qquad \textbf{(12.61)}$$

Este resultado cumple con (12.60) e implica trayectorias lineales de expansión. Una comparación de (12.61) con (12.56) también muestra que la función H cumple con (12.59).

En este ejemplo, $Q(a, b)$ es homogénea de grado $(\alpha + \beta)$. Resulta que $H(a, b)$ también es homogénea pero de grado $2(\alpha + \beta)$. Sin embargo, como regla, una función homotética no necesariamente es homogénea.

Ejemplo 2

Sea $H = e^Q$, donde $Q = Aa^\alpha b^\beta$. Ya que $Q(a, b)$ es homogénea y $h'(Q) = e^Q$ es positiva, $H(a, b)$ es homotética. De esta función

$$H(a, b) = \exp(Aa^\alpha b^\beta)$$

se encuentra fácilmente que

$$-\frac{H_a}{H_b} = -\frac{A\alpha a^{\alpha-1}b^\beta \exp(Aa^\alpha b^\beta)}{Aa^\alpha\beta b^{\beta-1}\exp(Aa^\alpha b^\beta)} = -\frac{\alpha b}{\beta a}$$

Este resultado, por supuesto, es idéntico a (12.61) del ejemplo 1. Sin embargo, esta vez la función homotética no es homogénea, ya que

$$\begin{aligned} H(ja, jb) &= \exp[A(ja)^\alpha(jb)^\beta] = \exp(Aa^\alpha b^\beta j^{\alpha+\beta}) \\ &= [\exp(Aa^\alpha b^\beta)]^{j^{\alpha+\beta}} = [H(a, b)]^{j^{\alpha+\beta}} \neq j^r H(a, b) \end{aligned}$$

Elasticidad de la sustitución

Otro aspecto de la estática comparativa tiene que ver con el efecto de un cambio de la relación P_a/P_b sobre la combinación de insumos de costo mínimo b^*/a^* para producir el mismo producto dado Q_0 (es decir, siempre que permanezcamos en la misma isocuanta).

Cuando la relación P_a/P_b insumo-precio (exógena) aumenta, normalmente podemos esperar que la relación óptima de b^*/a^* también aumente, ya que el insumo b (ahora relativamente más barato) tenderá a ser sustituido por el insumo a. La *dirección* de la sustitución es clara, pero ¿qué pasa con su *alcance*? El alcance de la sustitución de insumos puede medirse por la siguiente expresión elasticidad-punto, llamada la *elasticidad de sustitución* y denotada por σ (la letra griega minúscula sigma —"ese"—, que representa a "sustitución"):

$$\sigma \equiv \frac{\text{cambio relativo de } (b^*/a^*)}{\text{cambio relativo de } (P_a/P_b)} = \frac{\dfrac{d(b^*/a^*)}{b^*/a^*}}{\dfrac{d(P_a/P_b)}{P_a/P_b}} = \frac{\dfrac{d(b^*/a^*)}{d(P_a/P_b)}}{\dfrac{b^*/a^*}{P_a/P_b}} \qquad \textbf{(12.62)}$$

El valor de σ puede ser cualquiera entre 0 y ∞; cuanto mayor sea σ, mayor será la sustituibilidad entre los dos insumos. El caso límite de $\sigma = 0$ es donde los dos insumos deben usarse en una proporción fija como complementos entre sí. El otro caso límite, con σ infinito, es donde los dos insumos son sustitutos perfectos entre sí. Observa que, si (b^*/a^*) se considera como una función de (P_a/P_b), entonces la elasticidad σ será nuevamente la relación de una función *marginal* entre una función *promedio*.[13]

A modo de ejemplo, calculemos la elasticidad de sustitución para la función de producción generalizada de Cobb-Douglas. Anteriormente aprendimos que, para este caso, la combinación de insumos de costo mínimo se especifica por

$$\left(\frac{b^*}{a^*}\right) = \frac{\beta}{\alpha}\left(\frac{P_a}{P_b}\right) \qquad \text{[de (12.57)]}$$

Esta ecuación tiene la forma $y = ax$, para la cual dy/dx (el marginal) y y/x (el promedio) son ambos iguales a la constante a, es decir,

$$\frac{d(b^*/a^*)}{d(P_a/P_b)} = \frac{\beta}{\alpha} \qquad \text{y} \qquad \frac{b^*/a^*}{P_a/P_b} = \frac{\beta}{\alpha}$$

Al sustituir estos valores en (12.62), encontramos inmediatamente que $\sigma = 1$; es decir, la función de producción generalizada de Cobb-Douglas se caracteriza por una elasticidad de sustitución *constante unitaria*. Observe que la derivación de este resultado no depende de ninguna manera de la hipótesis de que $\alpha + \beta = 1$. Entonces, la elasticidad de sustitución de la función de producción $Q = Aa^\alpha b^\beta$ será unitaria aun si $\alpha + \beta \neq 1$.

[13] Hay una forma alterna de expresar a σ. Puesto que en el punto de tangencia siempre tenemos

$$\frac{P_a}{P_b} = \frac{Q_a}{Q_b} = \text{MRTS}_{ab}$$

la elasticidad de sustitución se define en forma equivalente como

$$\sigma = \frac{\text{cambio relativo en } (b^*/a^*)}{\text{cambio relativo en MRTS}_{ab}} = \frac{\dfrac{d(b^*/a^*)}{b^*/a^*}}{\dfrac{d(Q_a/Q_b)}{Q_a/Q_b}} = \frac{\dfrac{d(b^*/a^*)}{d(Q_a/Q_b)}}{\dfrac{b^*/a^*}{Q_a/Q_b}} \qquad \textbf{(12.62')}$$

La función de producción de CES

Más recientemente, se ha generalizado el uso de otra forma de función de producción que, aun cuando se caracteriza por una elasticidad de sustitución constante (CES), puede proporcionar una σ con un valor (constante) diferente de 1.[14] La ecuación de esta función, conocida como la *función de producción CES*, es

$$Q = A[\delta K^{-\rho} + (1-\delta)L^{-\rho}]^{-1/\rho} \qquad (A > 0; 0 < \delta < 1; -1 < \rho \neq 0) \qquad \textbf{(12.63)}$$

donde K y L representan dos factores de producción, y A, δ y ρ (la letra griega minúscula "ro") son tres parámetros. El parámetro A (el *parámetro de eficiencia*) juega el mismo papel que el coeficiente A en la función de Cobb-Douglas; sirve como indicador del estado de la tecnología. El parámetro δ (el *parámetro de distribución*), al igual que la α en la función de Cobb-Douglas, tiene que ver con las participaciones del factor relativo en el producto. Y el parámetro ρ (el *parámetro de sustitución*) –que no tiene equivalente en la función de Cobb-Douglas– es lo que determina el valor de la elasticidad de sustitución (constante), como se mostrará posteriormente en esta sección.

Sin embargo, observemos primero que esta función es homogénea de grado uno. Si reemplazamos a K y L por jK y jL, respectivamente, el producto va a cambiar de Q a

$$\begin{aligned} A[\delta(jK)^{-\rho} + (1-\delta)(jL)^{-\rho}]^{-1/\rho} &= A\{j^{-\rho}[\delta K^{-\rho} + (1-\delta)L^{-\rho}]\}^{-1/\rho} \\ &= (j^{-\rho})^{-1/\rho} Q = jQ \end{aligned}$$

En consecuencia, la función CES, al igual que todas las funciones de producción linealmente homogéneas, exhibe retornos constantes de escala, califica para la aplicación del teorema de Euler, y posee productos promedios y productos marginales que son homogéneos de grado cero en las variables K y L.

Podemos también observar que las isocuantas generadas por la función de producción CES siempre tienen pendiente negativa y son estrictamente convexas para los valores positivos de K y L. Para mostrar esto, encontremos primero las expresiones para los productos marginales Q_L y Q_K. Usando la notación $[\ldots]$ como una abreviatura para $[\delta K^{-\rho} + (1-\delta)L^{-\rho}]$, tenemos

$$\begin{aligned} Q_L \equiv \frac{\partial Q}{\partial L} &= A\left(-\frac{1}{\rho}\right)[\cdots]^{-(1/\rho)-1}(1-\delta)(-\rho)L^{-\rho-1} \\ &= (1-\delta)A[\cdots]^{-(1+\rho)/\rho}L^{-(1+\rho)} \\ &= (1-\delta)\frac{A^{1+\rho}}{A^{\rho}}[\cdots]^{-(1+\rho)/\rho}L^{-(1+\rho)} \\ &= \frac{(1-\delta)}{A^{\rho}}\left(\frac{Q}{L}\right)^{1+\rho} > 0 \qquad \text{[por (12.63)]} \end{aligned} \qquad \textbf{(12.64)}$$

y en forma similar,

$$Q_K \equiv \frac{\partial Q}{\partial K} = \frac{\delta}{A^{\rho}}\left(\frac{Q}{K}\right)^{1+\rho} > 0 \qquad \textbf{(12.65)}$$

[14] K. J. Arrow, H. B. Chenery, B. S. Minhas y R. M. Solow, "Capital-Labor Substitution and Economic Efficiency", *Review of Economics and Statistics*, agosto de 1961, pp. 225-250.

que se definen para los valores positivos de K y L. Entonces, la pendiente de las isocuantas (con K graficado en sentido vertical y L en sentido horizontal) es

$$\frac{dK}{dL} = -\frac{Q_L}{Q_K} = -\frac{(1-\delta)}{\delta}\left(\frac{K}{L}\right)^{1+\rho} < 0 \qquad \text{[ver (11.36)]} \qquad \textbf{(12.66)}$$

Entonces, puede verificarse fácilmente que $d^2K/dL^2 > 0$ (lo cual te lo dejamos como ejercicio), lo que implica que las isocuantas son estrictamente convexas para K y L positivos.

También puede mostrarse que la función de producción CES es estrictamente cuasicóncava para K y L positivos. La diferenciación adicional de (12.64) y (12.65) muestra que las segundas derivadas de la función tienen los siguientes signos:

$$Q_{LL} = \frac{\partial}{\partial L}Q_L = \frac{(1-\delta)(1+\rho)}{A^\rho}\left(\frac{Q}{L}\right)^\rho \frac{Q_L L - Q}{L^2} < 0$$

$$[Q_L L - Q < 0, \text{ por el teorema de Euler}]$$

$$Q_{KK} = \frac{\partial}{\partial K}Q_K = \frac{\delta(1+\rho)}{A^\rho}\left(\frac{Q}{K}\right)^\rho \frac{Q_K K - Q}{K^2} < 0$$

$$[Q_K K - Q < 0, \text{ por el teorema de Euler}]$$

$$Q_{KL} = Q_{LK} = \frac{(1-\delta)(1+\rho)}{A^\rho}\left(\frac{Q}{L}\right)^\rho \frac{Q_K}{L} > 0$$

Estos signos de las derivadas, válidos para K y L positivos, nos permiten verificar la condición suficiente para la cuasiconcavidad estricta (12.26). Como se puede verificar,

$$|B_1| = -Q_K^2 < 0$$

y

$$|B_2| = 2Q_K Q_L Q_{KL} - Q_K^2 Q_{LL} - Q_L^2 Q_{KK} > 0$$

Entonces, la función CES es estrictamente cuasicóncava para K y L positivos.

Por último, usaremos los productos marginales de (12.64) y (12.65) para encontrar la elasticidad de sustitución de la función CES. Para satisfacer la condición de combinación de costo mínimo $Q_L/Q_K = P_L/P_K$, donde P_L y P_K denotan los precios del servicio laboral (tarifa del salario) y el servicio del capital (cargo de la renta por los bienes de capital), respectivamente, debemos tener

$$\frac{1-\delta}{\delta}\left(\frac{K}{L}\right)^{1+\rho} = \frac{P_L}{P_K} \qquad \text{[vea (12.66)]}$$

Entonces, la relación óptima de insumos es (introduciendo el símbolo c para abreviar)

$$\left(\frac{K^*}{L^*}\right) = \left(\frac{\delta}{1-\delta}\right)^{1/(1+\rho)}\left(\frac{P_L}{P_K}\right)^{1/(1+\rho)} \equiv c\left(\frac{P_L}{P_K}\right)^{1/(1+\rho)} \qquad \textbf{(12.67)}$$

Tomando a (K^*/L^*) como una función de (P_L/P_K), encontramos que las funciones asociadas marginal y promedio son

$$\text{función marginal} = \frac{d(K^*/L^*)}{d(P_L/P_K)} = \frac{c}{1+\rho}\left(\frac{P_L}{P_K}\right)^{1/(1+\rho)-1}$$

$$\text{función promedio} = \frac{K^*/L^*}{P_L/P_K} = c\left(\frac{P_L}{P_K}\right)^{1/(1+\rho)-1}$$

Por lo tanto, la elasticidad de sustitución es[15]

$$\sigma = \frac{\text{función marginal}}{\text{función promedio}} = \frac{1}{1+\rho} \qquad \textbf{(12.68)}$$

Lo que muestra esto es que σ es una constante cuya magnitud depende del valor del parámetro ρ como sigue:

$$\left.\begin{aligned} -1 < \rho < 0 \\ \rho = 0 \\ 0 < \rho < \infty \end{aligned}\right\} \quad \Rightarrow \quad \left\{\begin{aligned} \sigma > 1 \\ \sigma = 1 \\ \sigma < 1 \end{aligned}\right.$$

La función de Cobb-Douglas como un caso especial de la función CES

En este último resultado, el caso medio de $\rho = 0$ conduce a una elasticidad de sustitución unitaria que, como sabemos, es característica de la función de Cobb-Douglas. Esto sugiere que la función de Cobb-Douglas (linealmente homogénea) es un caso especial de la función CES (linealmente homogénea). La dificultad es que la función CES, como está dada en (12.63), es indefinida cuando $\rho = 0$, ya que la división por cero no es posible. Sin embargo, podemos demostrar que, cuando $\rho \to 0$, la función CES se aproxima a la función Cobb-Douglas.

Para esta demostración, emplearemos una técnica conocida como la *regla de L'Hôpital*. Esta regla tiene que ver con la evaluación del límite de una función $f(x) = \dfrac{m(x)}{n(x)}$ cuando $x \to a$ (donde a puede ser finito o infinito), cuando el numerador $m(x)$ y el denominador $n(x)$ (1) ambos tienden a cero cuando $x \to a$, lo que conduce a una expresión de la forma 0/0, o (2) ambos tienden a $\pm\infty$ cuando $x \to a$, conduciendo así a una expresión de la forma de ∞/∞ (o $\infty/-\infty$, o $-\infty/\infty$, o $-\infty/-\infty$). Aun cuando el límite de $f(x)$ no puede evaluarse tal como está la expresión en estas dos circunstancias, no obstante su valor puede encontrarse usando la fórmula

$$\lim_{x\to a} \frac{m(x)}{n(x)} = \lim_{x\to a} \frac{m'(x)}{n'(x)} \qquad \text{[regla de L'Hôpital]} \qquad \textbf{(12.69)}$$

Ejemplo 3

Encuentre el límite de $(1 - x^2)/(1 - x)$ cuando $x \to 1$. Aquí, tanto $m(x)$ como $n(x)$ se aproximan a cero cuando x se aproxima a la unidad, ejemplificando la circunstancia (1). Como $m'(x) = -2x$ y $n'(x) = -1$, podemos escribir

$$\lim_{x\to 1} \frac{1-x^2}{1-x} = \lim_{x\to 1} \frac{-2x}{-1} = \lim_{x\to 1} 2x = 2$$

Esta respuesta es idéntica a la obtenida por otro método en el ejemplo 2 de la sección 6.4.

[15] Por supuesto, pudimos haber obtenido el mismo resultado tomando primero logaritmos a ambos lados de (12.67):

$$\ln\left(\frac{K^*}{L^*}\right) = \ln c + \frac{1}{1+\rho} \ln\left(\frac{P_L}{P_K}\right)$$

y luego aplicando la fórmula de elasticidad de (10.28) para obtener

$$\sigma = \frac{d\,(\ln K^*/L^*)}{d\,(\ln P_L/P_K)} = \frac{1}{1+\rho}$$

Ejemplo 4 Encuentre el límite de $(2x+5)/(x+1)$ cuando $x \to \infty$. Cuando x tiende a infinito, tanto $m(x)$ como $n(x)$ tienden a infinito en el caso presente; entonces, tenemos aquí un ejemplo de la circunstancia (2). Ya que $m'(x) = 2$ y $n'(x) = 1$, podemos escribir

$$\lim_{x\to\infty} \frac{2x+5}{x+1} = \lim_{x\to\infty} \frac{2}{1} = 2$$

Nuevamente, esta respuesta es idéntica a la obtenida por otro método en el ejemplo 3 de la sección 6.4.

Puede resultar que la expresión del lado derecho de (12.69) nuevamente se sitúe en el formato $0/0$ o ∞/∞, al igual que la expresión del lado izquierdo. En este caso, podemos volver a aplicar la regla de L'Hôpital; es decir, podemos buscar el límite de $m''(x)/n''(x)$ cuando $x \to a$, y tomar ese límite como nuestra respuesta. Puede resultar que aun cuando la función dada $f(x)$, cuyo límite queremos evaluar, no está originalmente en la forma de $m(x)/n(x)$ que se sitúa en el formato de $0/0$ o ∞/∞ al calcular el límite, una transformación adecuada hará que $f(x)$ sea asequible a la aplicación de la regla de (12.69). Esta última posibilidad puede ilustrarse con el problema de encontrar el límite de la función CES (12.63) —ahora considerada como una función $Q(\rho)$— cuando $\rho \to 0$.

Tal como está dada, $Q(\rho)$ no está en la forma de $m(\rho)/n(\rho)$. Sin embargo, al dividir ambos lados de (12.63) entre A, y al tomar el logaritmo normal, ciertamente obtenemos una expresión de esa forma, es decir,

$$\ln\frac{Q}{A} = \frac{-\ln[\delta K^{-\rho} + (1-\delta)L^{-\rho}]}{\rho} \equiv \frac{m(\rho)}{n(\rho)} \qquad \textbf{(12.70)}$$

Aún más, cuando $\rho \to 0$, encontramos que $m(\rho) \to -\ln(\delta + 1 - \delta) = -\ln 1 = 0$, y $n(\rho) \to 0$, también. Entonces, podemos usar la regla de L'Hôpital para encontrar el límite de $\ln(Q/A)$. Una vez hecho esto, también podemos encontrar el límite de Q, ya que $Q/A = e^{\ln(Q/A)}$, de modo que $Q = Ae^{\ln(Q/A)}$, se sigue que

$$\text{lím}\, Q = \text{lím}\, Ae^{\ln(Q/A)} = Ae^{\text{lím} \ln(Q/A)} \qquad \textbf{(12.71)}$$

De (12.70), encontramos primero $m'(\rho)$ y $n'(\rho)$, como lo requiere la regla de L'Hôpital. La última es simplemente $n'(\rho) = 1$. La primera es

$$m'(\rho) = \frac{-1}{[\delta K^{-\rho} + (1-\delta)L^{-\rho}]} \frac{d}{d\rho}[\delta K^{-\rho} + (1-\delta)L^{-\rho}] \qquad \text{[regla de la cadena]}$$

$$= \frac{-[-\delta K^{-\rho}\ln K - (1-\delta)L^{-\rho}\ln L]}{[\delta K^{-\rho} + (1-\delta)L^{-\rho}]} \qquad \text{[por (10.21')]}$$

Por lo tanto, por la regla de L'Hôpital tenemos

$$\lim_{\rho\to 0} \ln\frac{Q}{A} = \lim_{\rho\to 0} \frac{m'(\rho)}{n'(\rho)} = \frac{\delta\ln K + (1-\delta)\ln L}{1} = \ln(K^{\delta}L^{1-\delta})$$

En vista de este resultado, cuando e se eleva a la potencia de $\lim_{\rho\to 0} \ln(Q/A)$, el resultado es simplemente $K^{\delta}L^{1-\delta}$. Entonces, por (12.71), finalmente llegamos al resultado

$$\lim_{\rho\to 0} Q = AK^{\delta}L^{1-\delta}$$

mostrando que, cuando $\rho \to 0$, la función CES tiende realmente a la función de Cobb-Douglas.

EJERCICIO 12.7

1. Suponga que las isocuantas de la figura 12.9*b* se obtienen de una función de producción homogénea específica $Q = Q(a, b)$. Observando que $OE = EE' = E'E''$, ¿cuáles deben ser las razones entre los niveles de producto representados por las tres isocuantas si la función Q es homogénea

   (*a*) de grado uno? (*b*) de grado dos?

2. Para el caso generalizado de Cobb-Douglas, si graficamos la relación b^*/a^* contra la relación P_a/P_b, ¿qué tipo de curva va a resultar? ¿Depende este resultado de la hipótesis de que $\alpha + \beta = 1$? Interprete gráficamente la elasticidad de sustitución de esta curva.
3. ¿Está caracterizada la función de producción CES por la disminución de los retornos de cada insumo para todos los niveles positivos de insumos?
4. Muestre que, para una isocuanta de la función CES, $d^2K/dL^2 > 0$.
5. (*a*) Para la función CES, si a cada factor de producción se le paga de acuerdo con su producto marginal, ¿cuál es la razón de la participación del trabajo del producto entre la participación del capital del producto? ¿Un valor mayor de δ implicaría una mayor participación relativa del capital?

   (*b*) Para la función Cobb-Douglas ¿la razón de la participación del trabajo sobre el capital depende de la relación K/L? ¿Es la misma respuesta para la función CES?

6. (*a*) La función de producción CES descarta $\rho = -1$. Sin embargo, si $\rho = -1$, ¿cuál sería la forma general de las isocuantas para K y L positivos?

   (*b*) ¿Está definido σ para $\rho = -1$? ¿Cuál es el límite de σ cuando $\rho \to -1$?

   (*c*) Interprete económicamente los resultados para las partes (*a*) y (*b*).

7. Muestre que al escribir la función CES como $Q = A[\delta K^{-\rho} + (1-\delta)L^{-\rho}]^{-r/\rho}$, donde $r > 0$ es un nuevo parámetro, podemos introducir retornos crecientes a escala y retornos decrecientes a escala.
8. Evalúe lo siguiente:

   (*a*) $\lim\limits_{x\to 4} \dfrac{x^2 - x - 12}{x - 4}$ (*c*) $\lim\limits_{x\to 0} \dfrac{5^x - e^x}{x}$

   (*b*) $\lim\limits_{x\to 0} \dfrac{e^x - 1}{x}$ (*d*) $\lim\limits_{x\to \infty} \dfrac{\ln x}{x}$

9. Usando la regla de L'Hôpital, muestre que

   (*a*) $\lim\limits_{x\to\infty} \dfrac{x^n}{e^x} = 0$ (*b*) $\lim\limits_{x\to 0^+} x \ln x = 0$ (*c*) $\lim\limits_{x\to 0^+} x^x = 1$

Capítulo 13

Temas adicionales de optimización

Este capítulo trata dos temas principales. El primero es la programación lineal, que amplía las técnicas de la optimización restringida del capítulo 12 al incorporar las *restricciones de desigualdad*. En el capítulo 12, las restricciones debían satisfacerse como igualdades estrictas, es decir, las restricciones siempre eran activas. Ahora consideraremos restricciones que podrían ser inactivas en la solución, es decir, pueden satisfacerse como desigualdades en la solución.

En la segunda parte de este capítulo regresamos a la optimización clásica restringida para estudiar algunos temas no expuestos en los capítulos anteriores. Éstos incluyen la función objetivo indirecta, el teorema de la envolvente y el concepto de la dualidad.

13.1 La programación no lineal y las condiciones de Kuhn-Tucker

En la historia del desarrollo metodológico, los primeros intentos para tratar las restricciones de desigualdad se concentraron solamente en las lineales. Al prevalecer la linealidad en las restricciones, así como en la función objetivo, es natural que se llame *programación lineal* a la metodología resultante. Sin embargo, a pesar de la limitación de la linealidad, podríamos especificar explícitamente por primera vez las variables de elección como no negativas, que es lo apropiado en la mayor parte del análisis económico. Esto representa un avance importante. La programación no lineal, un desarrollo posterior, nos permite manejar las restricciones de las desigualdades no lineales y la función objetivo no lineal, por lo que ocupa un lugar muy importante en la metodología de la optimización.

En el problema clásico de la optimización, sin restricciones explícitas en los signos de las variables de elección, y sin desigualdades en las restricciones, la condición de primer orden para un extremo relativo o local es simplemente que las primeras derivadas parciales de la función lagrangeana (uniforme) respecto a todas las variables de elección y los multiplicadores de Lagrange sean cero. En la programación no lineal existe un tipo similar de condición de primer orden, conocida como las *condiciones de Kuhn-Tucker*.[1] Sin embargo, como se verá, aun cuando siempre es necesaria la condición clásica de primer orden, no puede otorgarse a

[1] H. W. Kuhn y A. W. Tucker, "Nonlinear Programming", en J. Neyman (ed.), *Proceedings of the Second Berkeley Symposium in Mathematical Statistics and Probability*, University of California Press, Berkeley, California, 1951, pp. 481-492.

las condiciones de Kuhn-Tucker el estatus de condiciones necesarias a menos que se satisfaga cierta disposición. Por otro lado, para algunas circunstancias específicas, las condiciones de Kuhn-Tucker son *condiciones suficientes*, o aun también condiciones *necesarias y suficientes*.

Como las condiciones de Kuhn-Tucker son el resultado individual analítico de mayor importancia en la programación no lineal, es esencial lograr una compresión apropiada de esas condiciones, así como también de sus implicaciones. Vamos a desarrollar estas condiciones en dos pasos para favorecer la conveniencia de la exposición.

Paso 1: Efecto de las restricciones de no negatividad

Como primer paso, considere un problema con restricciones de no negatividad sobre las variables de elección, pero sin otras restricciones. Considerando específicamente el caso de una sola variable, tenemos

$$\begin{aligned} &\text{Maximizar} \quad && \pi = f(x_1) \\ &\text{sujeto a} && x_1 \geq 0 \end{aligned} \tag{13.1}$$

donde se supone que la función f es diferenciable. En vista de la restricción $x_1 \geq 0$, pueden surgir tres situaciones posibles. Primero, si se presenta un máximo local para π en el interior de la región de factibilidad sombreada en la figura 13.1, tal como el punto A en la figura 13.1a, entonces tenemos una *solución interior*; en este caso, la condición de primer orden es $d\pi/dx_1 = f'(x_1) = 0$, al igual que en el problema clásico. Segundo, como lo ilustra el punto B en la figura 13.1b, también puede presentarse un máximo local en la columna vertical, para $x_1 = 0$. En este segundo caso, tenemos una *solución de frontera*, sin embargo permanece válida la condición de primer orden $f'(x_1) = 0$. Como una tercera posibilidad, un máximo local puede adoptar en el presente contexto la posición del punto C o del punto D en la figura 13.1c, ya que para calificar como un máximo local en el problema (13.1), el punto candidato solamente tiene que estar a mayor altura que los puntos circundantes *dentro* de la región de factibilidad. En vista de esta última posibilidad, el punto máximo en un problema como el (13.1) puede caracterizarse no solamente por la ecuación $f'(x_1) = 0$, sino también por la desigualdad $f'(x_1) < 0$. Por otro lado, observe que la desigualdad opuesta $f'(x_1) > 0$ puede descartarse sin problema, ya que para un punto en el cual la curva tiene pendiente positiva en la frontera, nunca podremos tener un máximo, como el punto E que se localiza en el eje vertical en la figura 13.1a.

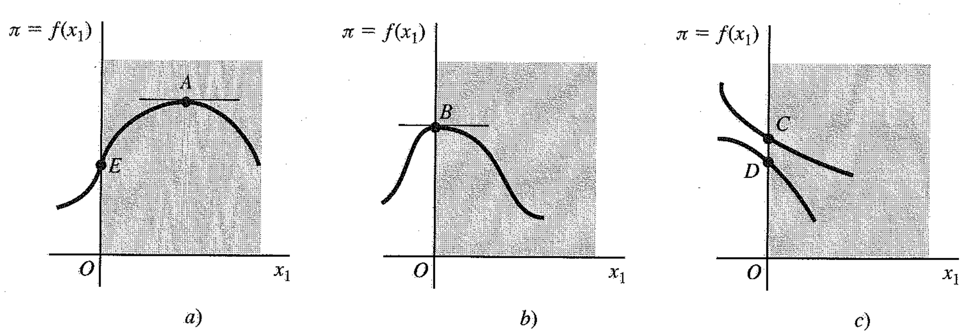
La conclusión es que, con objeto de que un valor de x_1 corresponda a un máximo local de π en el problema (13.1), debe satisfacer una de las tres condiciones siguientes

$$f'(x_1) = 0 \quad \text{y} \quad x_1 > 0 \quad [\text{punto } A] \tag{13.2}$$

$$f'(x_1) = 0 \quad \text{y} \quad x_1 = 0 \quad [\text{punto } B] \tag{13.3}$$

$$f'(x_1) < 0 \quad \text{y} \quad x_1 = 0 \quad [\text{puntos } C \text{ y } D] \tag{13.4}$$

FIGURA 13.1



En realidad, estas tres condiciones pueden reunirse en una sola aseveración

$$f'(x_1) \leq 0 \qquad x_1 \geq 0 \qquad \text{y} \qquad x_1 f'(x_1) = 0 \tag{13.5}$$

La primera desigualdad de (13.5) es un resumen de la información relativa a $f'(x_1)$ enumerada de (13.2) a (13.4). La segunda desigualdad es un resumen similar para x_1; de hecho, simplemente reitera la restricción de no negatividad del problema. Respecto a la tercera parte de (13.5), tenemos una ecuación que expresa una característica importante común a (13.2) hasta (13.4), a saber, que de las dos cantidades x_1 y $f'(x_1)$, *por lo menos una* debe adoptar un valor cero, de modo que el producto de las dos debe ser cero. Esta característica se denomina *holgura complementaria* entre x_1 y $f'(x_1)$. Agrupadas, las tres partes de (13.5) constituyen la condición necesaria de primer orden para un máximo local en un problema en el cual la variable de elección debe ser no negativa. No obstante, al avanzar otro paso adicional, también podemos considerarlas como necesarias para un máximo *global*. Esto se debe a que un máximo global también debe ser un máximo local y, como tal, debe satisfacer la condición necesaria para un máximo local.

Cuando el problema contiene n variables de elección:

$$\begin{array}{ll} \text{Maximizar} & \pi = f(x_1, x_2, \ldots, x_n) \\ \text{sujeto a} & x_j \geq 0 \qquad (j = 1, 2, \ldots, n) \end{array} \tag{13.6}$$

la condición clásica de primer orden $f_1 = f_2 = \cdots = f_n = 0$ debe modificarse en forma similar. Para lograrlo, podemos aplicar el mismo tipo de razonamiento que sustenta a (13.5) a cada variable de elección x_j. Gráficamente, esto nos lleva a ver el eje horizontal en la figura 13.1 como que representa a cada x_j a la vez. Entonces, la modificación requerida de la condición de primer orden se hace evidente por sí misma:

$$f_j \leq 0 \qquad x_j \geq 0 \qquad \text{y} \qquad x_j f_j = 0 \qquad (j = 1, 2, \ldots, n) \tag{13.7}$$

donde f_j es la derivada parcial $\partial\pi/\partial x_j$.

Paso 2: Efecto de las restricciones de desigualdad

Con este antecedente, proseguimos al segundo paso, y tratamos de incluir también las restricciones de desigualdad. Por su sencillez, veamos primero un problema con tres variables de elección ($n = 3$) y dos restricciones ($m = 2$):

$$\begin{array}{ll} \text{Maximizar} & \pi = f(x_1, x_2, x_3) \\ \text{sujeto a} & g^1(x_1, x_2, x_3) \leq r_1 \\ & g^2(x_1, x_2, x_3) \leq r_2 \\ \text{y} & x_1, x_2, x_3 \geq 0 \end{array} \tag{13.8}$$

el cual, con la ayuda de dos variables ficticias s_1 y s_2, puede transformarse en la forma equivalente

$$\begin{array}{ll} \text{Maximizar} & \pi = f(x_1, x_2, x_3) \\ \text{sujeto a} & g^1(x_1, x_2, x_3) + s_1 = r_1 \\ & g^2(x_1, x_2, x_3) + s_2 = r_2 \\ \text{y} & x_1, x_2, x_3, s_1, s_2 \geq 0 \end{array} \tag{13.8'}$$

Si están ausentes las restricciones de no negatividad, podemos utilizar el enfoque clásico y construir la función lagrangeana:

$$\begin{aligned} Z' = f(x_1, x_2, x_3) &+ \lambda_1[r_1 - g^1(x_1, x_2, x_3) - s_1] \\ &+ \lambda_2[r_2 - g^2(x_1, x_2, x_3) - s_2] \end{aligned} \tag{13.9}$$

y escribir la condición de primer orden como

$$\frac{\partial Z'}{\partial x_1} = \frac{\partial Z'}{\partial x_2} = \frac{\partial Z'}{\partial x_3} = \frac{\partial Z'}{\partial s_1} = \frac{\partial Z'}{\partial s_2} = \frac{\partial Z'}{\partial \lambda_1} = \frac{\partial Z'}{\partial \lambda_2} = 0$$

Pero ya que las variables x_j y s_i tienen que ser no negativas, la condición de primer orden para estas variables debe modificarse de acuerdo con (13.7). En consecuencia, obtenemos el siguiente conjunto de condiciones:

$$\begin{aligned} &\frac{\partial Z'}{\partial x_j} \le 0 \qquad x_j \ge 0 \qquad \text{y} \qquad x_j \frac{\partial Z'}{\partial x_j} = 0 \\ &\frac{\partial Z'}{\partial s_i} \le 0 \qquad s_i \ge 0 \qquad \text{y} \qquad s_i \frac{\partial Z'}{\partial s_i} = 0 \\ &\frac{\partial Z'}{\partial \lambda_i} = 0 \qquad\qquad\qquad\qquad \begin{pmatrix} i = 1, 2 \\ j = 1, 2, 3 \end{pmatrix} \end{aligned} \tag{13.10}$$

Observe que las derivadas $\partial Z'/\partial \lambda_i$ todavía tienen que igualarse estrictamente a cero (¿por qué?).

Cada renglón de (13.10) se relaciona con un tipo diferente de variable. Pero podemos consolidar los dos últimos renglones y, en el proceso, eliminar la variable ficticia s_i en la condición de primer orden. Ya que $\partial Z'/\partial s_i = -\lambda_i$, el segundo renglón de (13.10) nos dice que debemos tener $-\lambda_i \le 0$, $s_i \ge 0$ y $-s_i\lambda_i = 0$, o en forma equivalente,

$$s_i \ge 0 \qquad \lambda_i \ge 0 \qquad \text{y} \qquad s_i\lambda_i = 0 \tag{13.11}$$

Pero el tercer renglón —un replanteamiento de las restricciones de (13.8′)— significa que $s_i = r_i - g^i(x_1, x_2, x_3)$. Al sustituir esto último en (13.11), podemos combinar el segundo y tercer renglones de (13.10) como

$$r_i - g^i(x_1, x_2, x_3) \ge 0 \qquad \lambda_i \ge 0 \qquad \text{y} \qquad \lambda_i[r_i - g^i(x_1, x_2, x_3)] = 0$$

Esto nos permite expresar la condición de primer orden (13.10) como una forma equivalente *sin* las variables ficticias. Usando el símbolo g^i_j para denotar a $\partial g^i/\partial x_j$, ahora escribimos

$$\begin{aligned} &\frac{\partial Z'}{\partial x_j} = f_j - \left(\lambda_1 g^1_j + \lambda_2 g^2_j\right) \le 0 \qquad x_j \ge 0 \qquad \text{y} \qquad x_j \frac{\partial Z'}{\partial x_j} = 0 \\ &r_i - g^i(x_1, x_2, x_3) \ge 0 \qquad \lambda_i \ge 0 \qquad \text{y} \qquad \lambda_i[r_i - g^i(x_1, x_2, x_3)] = 0 \end{aligned} \tag{13.12}$$

Éstas son, entonces, las condiciones de Kuhn-Tucker para el problema (13.8), o más exactamente, una versión de las condiciones de Kuhn-Tucker, expresadas en términos de la función lagrangeana Z' en (13.9).

Ahora que conocemos los resultados, podemos obtener más directamente el mismo conjunto de condiciones usando una función lagrangeana diferente. Dado el problema (13.9), ignoremos las restricciones de no negatividad, así como los signos de desigualdad en las restricciones, y escribamos el tipo puramente clásico de función lagrangeana Z:

$$Z = f(x_1, x_2, x_3) + \lambda_1[r_1 - g^1(x_1, x_2, x_3)] + \lambda_2[r_2 - g^2(x_1, x_2, x_3)] \tag{13.13}$$

Proceda de la manera siguiente: (1) haga las derivadas parciales $\partial Z/\partial x_j \leq 0$, pero $\partial Z/\partial \lambda_i \geq 0$; (2) imponga restricciones de no negatividad sobre x_j y λ_i, y (3) haga que prevalezca la holgura complementaria entre cada variable y la derivada parcial de Z respecto a esa variable, es decir, haga que el producto se anule. Con los resultados de estos pasos, se obtienen las ecuaciones

$$\begin{array}{llll} \dfrac{\partial Z}{\partial x_j} = f_j - \left(\lambda_1 g_j^1 + \lambda_2 g_j^2\right) \leq 0 & x_j \geq 0 & \text{y} & x_j \dfrac{\partial Z}{\partial x_j} = 0 \\ \dfrac{\partial Z}{\partial \lambda_i} = r_i - g^i(x_1, x_2, x_3) \geq 0 & \lambda_i \geq 0 & \text{y} & \lambda_i \dfrac{\partial Z}{\partial \lambda_i} = 0 \end{array} \qquad \textbf{(13.14)}$$

que son idénticas a (13.12), las condiciones de Kuhn-Tucker se expresan también en términos de la función lagrangeana Z (en contraste con Z'). Note que al cambiar de Z' a Z, podemos no sólo llegar más directamente a las condiciones de Kuhn-Tucker, sino también identificar la expresión $r_i - g^i(x_1, x_2, x_3)$ —que se dejó sin nombrar en (13.12)— como la derivada parcial $\partial Z/\partial \lambda_i$E. Por lo tanto, en el estudio que sigue, usaremos solamente la versión (13.14) de las condiciones de Kuhn-Tucker, basándonos en la función lagrangeana Z.

Ejemplo 1

Si vaciamos el conocido problema de la maximización de la utilidad en el molde de la programación no lineal, el problema con una restricción de desigualdad es el siguiente:

$$\begin{array}{ll} \text{Maximizar} & U = U(x, y) \\ \text{sujeto a} & P_x x + P_y y \leq B \\ \text{y} & x, y \geq 0 \end{array}$$

Observe que con la restricción de desigualdad ya no se requiere que el consumidor gaste la cantidad total B.

Para añadir una nueva complicación al problema, supongamos que se ha impuesto un racionamiento al artículo x igual a X_0. Entonces, el consumidor enfrentaría una segunda restricción, y el problema se modifica a

$$\begin{array}{ll} \text{Maximizar} & U = U(x, y) \\ \text{sujeto a} & P_x x + P_y y \leq B \\ & x \leq X_0 \\ \text{y} & x, y \geq 0 \end{array}$$

La función lagrangeana es

$$Z = U(x, y) + \lambda_1(B - P_x x - P_y y) + \lambda_2(X_0 - x)$$

y las condiciones de Kuhn-Tucker son

$$\begin{array}{llll} Z_x = U_x - P_x\lambda_1 - \lambda_2 \leq 0 & x \geq 0 & \text{y} & xZ_x = 0 \\ Z_y = U_y - P_y\lambda_1 \leq 0 & y \geq 0 & \text{y} & yZ_y = 0 \\ Z_{\lambda_1} = B - P_x y - P_y y \geq 0 & \lambda_1 \geq 0 & \text{y} & \lambda_1 Z_{\lambda_1} = 0 \\ Z_{\lambda_2} = X_0 - x \geq 0 & \lambda_2 \geq 0 & \text{y} & \lambda_2 Z_{\lambda_2} = 0 \end{array}$$

Es útil examinar las implicaciones de la tercera columna de las condiciones de Kuhn-Tucker. Específicamente, la condición $\lambda_1 Z_{\lambda_1} = 0$, requiere que

$$\lambda_1(B - P_x x - P_y y) = 0$$

por lo tanto, debemos tener ya sea

$$\lambda_1 = 0 \qquad \text{o} \qquad B - P_x x - P_y y = 0$$

Si interpretamos a λ_1 como la utilidad marginal del presupuesto (ingreso), y si la restricción del presupuesto es no activa (que se satisface como una desigualdad en la solución, con dinero sobrante), la utilidad marginal del presupuesto B debe ser cero ($\lambda_1 = 0$).

En forma similar, la condición $\lambda_2 \; Z_{\lambda_2} = 0$ requiere ya sea

$$\lambda_2 = 0 \qquad \text{o} \qquad X_0 - x = 0$$

Puesto que λ_2 puede interpretarse como la utilidad marginal de la relajación de la restricción, vemos que si la restricción por racionamiento es no activa, la utilidad marginal de la relajación de la restricción debe ser cero ($\lambda_2 = 0$).

Esta característica, que se denomina la holgura complementaria, desempeña un papel esencial en la búsqueda de una solución. Ahora lo ilustraremos con un ejemplo numérico:

$$\begin{array}{ll} \text{Maximizar} & U = xy \\ \text{sujeto a} & x + y \leq 100 \\ & x \leq 40 \\ \text{y} & x, y \geq 0 \end{array}$$

La función lagrangeana es

$$Z = xy + \lambda_1(100 - x - y) + \lambda_2(40 - x)$$

y las condiciones de Kuhn-Tucker se transforman en

$$\begin{array}{llll} Z_x = y - \lambda_1 - \lambda_2 \leq 0 & x \geq 0 & \text{y} & xZ_x = 0 \\ Z_y = x - \lambda_1 \leq 0 & y \geq 0 & \text{y} & yZ_y = 0 \\ Z_{\lambda_1} = 100 - x - y \geq 0 & \lambda_1 \geq 0 & \text{y} & \lambda_1 Z_{\lambda_1} = 0 \\ Z_{\lambda_2} = 40 - x \geq 0 & \lambda_2 \geq 0 & \text{y} & \lambda_2 Z_{\lambda_2} = 0 \end{array}$$

Para resolver un problema de programación no lineal, el enfoque típico es el de prueba y error. Por ejemplo, podemos comenzar con un valor cero para una variable de elección. Igualar una variable a cero siempre simplifica las condiciones marginales al originar la cancelación de ciertos términos. Entonces, si pueden encontrarse valores apropiados no negativos de los multiplicadores de Lagrange que satisfagan todas las desigualdades marginales, la solución cero será óptima. Por otro lado, si la solución cero viola alguna de las desigualdades, entonces debemos permitir que una o más de las variables de elección sean positivas. Para cada una de las variables de elección positivas podemos, mediante la holgura complementaria, transformar una condición marginal de desigualdad débil en una igualdad estricta. Si se resuelve apropiadamente, esta igualdad nos conduce, ya sea a una solución o a una contradicción, que nos obligaría a intentar otra cosa. Si existe una solución, estos intentos finalmente nos permitirán descubrirla. También podemos comenzar suponiendo que una de las restricciones es no activa. Entonces, el multiplicador de Lagrange relacionado será cero mediante la holgura complementaria; de esta manera hemos eliminado una variable. Si esta hipótesis nos conduce a una contradicción, entonces debemos tratar la restricción como una igualdad estricta y proseguir sobre esa base.

Para este ejemplo, no tiene sentido intentar $x = 0$ o $y = 0$, porque entonces podríamos tener $U = xy = 0$. Por lo tanto, suponemos que x y y son diferentes de cero, y deducimos $Z_x = Z_y = 0$ a partir de la holgura complementaria. Esto significa

$$y - \lambda_1 - \lambda_2 = x - \lambda_1 (= 0)$$

de modo que $\qquad y - \lambda_2 = x.$

Supongamos ahora que la restricción por racionamiento sea no activa en la solución, lo que implica que $\lambda_2 = 0$. Entonces, tenemos $x = y$ y el presupuesto dado $B = 100$ suministra la solución de prueba $x = y = 50$. Pero esta solución no cumple la restricción por racionamiento $x \leq 40$. Entonces, debemos adoptar la hipótesis alterna de que la restricción por racionamiento es activa con $x^* = 40$. La restricción presupuestaria, así pues, permite que el consumidor tenga $y^* = 60$. Aún más, ya que la holgura complementaria exige que $Z_x = Z_y = 0$, podemos calcular rápidamente que $\lambda_1^* = 40$ y $\lambda_2^* = 20$.

Interpretación de las condiciones de Kuhn-Tucker

Algunas partes de las condiciones de Kuhn-Tucker (13.14) son sólo una reformulación de ciertos aspectos del problema dado. De este modo, las condiciones $x_j \geq 0$ simplemente repiten las restricciones de no negatividad y las condiciones $\partial Z/\partial\lambda_i \geq 0$ sólo reiteran las restricciones. Sin embargo, su inclusión en (13.14) tiene la importante ventaja de revelar más claramente la notable simetría entre los dos tipos de variables, x_j (variable de elección) y λ_i (multiplicadores de Lagrange). A cada variable de cada categoría le corresponde una condición marginal, $\partial Z/\partial x_j \leq 0$ o $\partial Z/\partial\lambda_i \geq 0$, que debe satisfacer la solución óptima. Cada una de estas variables también debe ser no negativa. Finalmente, cada variable se caracteriza por la holgura complementaria en relación con una derivada parcial específica de la función lagrangeana Z. Esto significa que para cada x_j debemos encontrar en la solución óptima que *ya sea* la condición marginal es válida como una igualdad, como en el contexto clásico, *o* la variable de elección en cuestión debe adoptar un valor cero, *o* ambas. En forma análoga, para cada λ_i, debemos encontrar en la solución óptima en la cual la condición marginal sea válida como una igualdad: lo que significa que la restricción i-ésima se satisface exactamente, que anule al multiplicador de Lagrange, o ambas cosas.

Es posible una interpretación aún más explícita cuando examinamos las expresiones expandidas para $\partial Z/\partial x_j$ y $\partial Z/\partial\lambda_i$ (13.14). Supongamos que el problema es el ya conocido de la producción. Entonces, tenemos

$f_j \equiv$ ganancia bruta marginal para el j-ésimo producto

$\lambda_i \equiv$ precio sombra del recurso i-ésimo (el costo de oportunidad del uso de una unidad del recurso i-ésimo)

$g_j^i \equiv$ cantidad del recurso i-ésimo consumido en la producción de la unidad marginal del producto j-ésimo

$\lambda_i g_j^i \equiv$ costo marginal del recurso i-ésimo en el que se incurre al producir una unidad del producto j-ésimo

$\sum_i \lambda_i g_j^i \equiv$ costo marginal del agregado (aquí agregar no significa añadir, sino juntar y formar el t), por aumentar el producto j-ésimo

Entonces, la condición marginal

$$\frac{\partial Z}{\partial x_j} = f_j - \sum_i \lambda_i g_j^i \leq 0$$

requiere que la ganancia bruta marginal del producto j-ésimo no sea mayor que el costo marginal del agregado que se ha añadido . La condición de holgura complementaria significa que, si la solución óptima requiere la producción activa del producto j-ésimo ($x_j^* > 0$), la ganancia bruta marginal debe ser exactamente igual al costo marginal del agregado ($\partial Z/\partial x_j^* = 0$), como sería la situación en el problema clásico de optimización. Por otro lado, si la ganancia bruta marginal se sitúa en forma óptima siendo menor que el costo total del agregado ($\partial Z/\partial x_j^* < 0$), lo que implica una alimentación en *exceso*, entonces ese producto no debe producirse ($x_j^* = 0$).[2] Esta última situación es algo que no puede ocurrir nunca en el contexto clásico, ya que si la pérdida bruta marginal es menor que el costo marginal alimentado, entonces el producto debe reducirse en ese marco de referencia hasta llegar al nivel para el cual la condición marginal se satisface como una igualdad. Lo que hace que la situación de $\partial Z/\partial x_j^* < 0$ aquí califique como óptima, es la especificación explícita de no negatividad en el presente marco de referencia. Como lo más que podemos hacer en el sentido de la reducción del producto es disminuir la producción hasta el nivel $x_j^* = 0$, y si todavía encontramos que $\partial Z/\partial x_j^* < 0$ para producto cero, nos detenemos ahí de cualquier manera.

En cuanto a las condiciones restantes, que están relacionadas con las variables λ_i, su significado es aún más fácil de percibir. Lo primero de todo es que la condición marginal $\partial Z/\partial \lambda_i \geq 0$ simplemente requiere que la compañía permanezca dentro de la limitación de capacidad de cada uno de los recursos de la planta. Así pues, la condición de holgura complementaria estipula que, si el recurso i-ésimo no se usa completamente en la solución óptima ($\partial Z/\partial \lambda_i^* > 0$), el precio de sombra de ese recurso, que nunca puede permitirse que sea negativo, debe igualarse a cero ($\lambda_i^* = 0$). Por otro lado, si un recurso tiene un precio de sombra positivo para la solución óptima ($\lambda_i^* > 0$), entonces es necesariamente un recurso utilizado por completo ($\partial Z/\partial \lambda_i^* = 0$).

Por supuesto que también se puede considerar que el valor λ_i^* del multiplicador de Lagrange es una medida de cómo reacciona el valor óptimo de la función objetivo ante una ligera relajación de la restricción i-ésima. A la luz de esto, la holgura complementaria significaría que, si la restricción i-ésima es no activa óptimamente ($\partial Z/\partial \lambda_i^* > 0$), entonces la relajación de esa restricción específica no va a afectar al valor óptimo de la ganancia bruta ($\lambda_i^* = 0$); del mismo modo que el aflojar un cinturón que no nos oprime la cintura no va a producir una comodidad adicional. Por otro lado, si una ligera relajación de la restricción i-ésima (el aumento de la dotación del recurso i-ésimo) incrementa la ganancia bruta ($\lambda_i^* > 0$), entonces esa restricción del recurso debe ser, de hecho, activa para la solución óptima ($\partial Z/\partial \lambda_i^* = 0$).

El caso de *n* variables, *m* restricciones

La discusión anterior puede generalizarse de manera inmediata cuando hay n variables de elección y m restricciones. La función lagrangeana Z aparece en la forma más general

$$Z = f(x_1, x_2, \ldots, x_n) + \sum_{i=1}^{m} \lambda_i [r_i - g^i(x_1, x_2, \ldots, x_n)] \qquad \textbf{(13.15)}$$

y las condiciones de Kuhn-Tucker son simplemente

$$\begin{aligned} &\frac{\partial Z}{\partial x_j} \leq 0 \qquad x_j \geq 0 \qquad \text{y} \qquad x_j \frac{\partial Z}{\partial x_j} = 0 \qquad [\text{maximización}] \\ &\frac{\partial Z}{\partial \lambda_i} \geq 0 \qquad \lambda_i \geq 0 \qquad \text{y} \qquad \lambda_i \frac{\partial Z}{\partial \lambda_i} = 0 \qquad \begin{pmatrix} i = 1, 2, \ldots, m \\ j = 1, 2, \ldots, n \end{pmatrix} \end{aligned} \qquad \textbf{(13.16)}$$

Con objeto de evitar una apariencia aglomerada, no hemos desarrollado aquí las expresiones expandidas para las derivadas parciales $\partial Z/\partial x_j$ y $\partial Z/\partial \lambda_i$. Pero nos urge desarrollarlas para tener una visión más detallada de las condiciones de Kuhn-Tucker, de forma similar a como se dio en (13.14). Observe que, aparte del cambio de la dimensión del problema, las condiciones de Kuhn-Tucker permanecen completamente iguales. Es natural que la interpretación de estas condiciones también permanezca igual.

[2] Recuerde que, dada la ecuación $ab = 0$, donde a y b son números reales, podemos inferir legítimamente que $a \neq 0$ implica $b = 0$, pero no es verdad que $a = 0$ implica $b \neq 0$, ya que $b = 0$ también es consistente con $a = 0$.

¿Qué pasa si el problema es de *minimización*? Una manera de manejarlo es transformarlo en un problema de maximización y luego aplicar (13.6). Minimizar C equivale a *maximizar* $-C$, si esta conversión siempre fuera factible. Pero por supuesto que también debemos invertir las desigualdades de restricción multiplicando cada una por -1. Sin embargo, en lugar de pasar por el proceso de transformación, podemos —usando de nuevo la función lagrangeana Z tal como se define en (13.15)— aplicar directamente la versión de minimización de las condiciones de Kuhn-Tucker como sigue:

$$\begin{aligned} &\frac{\partial Z}{\partial x_j} \geq 0 \qquad x_j \geq 0 \qquad \text{y} \qquad x_j \frac{\partial Z}{\partial x_j} = 0 \qquad [\text{minimización}] \\ &\frac{\partial Z}{\partial \lambda_i} \leq 0 \qquad \lambda_i \geq 0 \qquad \text{y} \qquad \lambda_i \frac{\partial Z}{\partial \lambda_i} = 0 \qquad \begin{pmatrix} i = 1, 2, \ldots, m \\ j = 1, 2, \ldots, n \end{pmatrix} \end{aligned} \tag{13.17}$$

Esto debe compararlo con (13.16).

Al leer (13.16) y (13.17) en sentido horizontal (por *renglones*), vemos que las condiciones de Kuhn-Tucker para los dos problemas de maximización y minimización consisten en un conjunto de condiciones relacionadas con las variables de elección x_j (primer renglón) y otro conjunto relacionado con los multiplicadores de Lagrange λ_i (segundo renglón). Por otro lado, al leerlos en sentido vertical (por *columnas*) observamos que para cada x_j y λ_i, hay una condición marginal (primera columna), una restricción de no negatividad (segunda columna) y una condición de holgura complementaria (tercera columna). Para cualquier problema, las condiciones marginales pertenecientes a las variables de elección siempre difieren, como grupo, de las condiciones marginales para los multiplicadores de Lagrange en el sentido de desigualdad que toman.

Sujetas a la explicación que haremos en la sección 13.2, las condiciones de Kuhn-Tucker para un máximo (13.16) y condiciones para un mínimo (13.17) son condiciones necesarias para un máximo local y un mínimo local, respectivamente. Pero ya que un máximo global (mínimo) también debe ser un máximo local (mínimo), las condiciones de Kuhn-Tucker también pueden considerarse como condiciones necesarias para un máximo global (mínimo), sujeto a la misma disposición.

Ejemplo 2

Apliquemos las condiciones de Kuhn-Tucker para resolver un problema de minimización:

$$\begin{aligned} \text{Minimizar} \quad & C = (x_1 - 4)^2 + (x_2 - 4)^2 \\ \text{sujeto a} \quad & 2x_1 + 3x_2 \geq 6 \\ & -3x_1 - 2x_2 \geq -12 \\ \text{y} \quad & x_1, x_2 \geq 0 \end{aligned}$$

La función lagrangeana para este problema es

$$Z = (x_1 - 4)^2 + (x_2 - 4)^2 + \lambda_1(6 - 2x_1 - 3x_2) + \lambda_2(-12 + 3x_1 + 2x_2)$$

Puesto que el problema es de minimización, las condiciones apropiadas son (13.17), que incluyen las cuatro condiciones marginales

$$\begin{aligned} \frac{\partial Z}{\partial x_1} &= 2(x_1 - 4) - 2\lambda_1 + 3\lambda_2 \geq 0 \\ \frac{\partial Z}{\partial x_2} &= 2(x_2 - 4) - 3\lambda_1 + 2\lambda_2 \geq 0 \\ \frac{\partial Z}{\partial \lambda_1} &= 6 - 2x_1 - 3x_2 \leq 0 \\ \frac{\partial Z}{\partial \lambda_2} &= -12 + 3x_1 + 2x_2 \leq 0 \end{aligned} \tag{13.18}$$

más las condiciones de no negatividad y holgura complementaria.

Para encontrar una solución, nuevamente usamos el enfoque de prueba y error, y nos damos cuenta, con pocos intentos , que este procedimiento puede conducirnos a un callejón sin salida. Supongamos que intentamos primero $\lambda_1 > 0$ y $\lambda_2 > 0$ y verifiquemos si podemos encontrar los valores correspondientes de x_1 y x_2 que satisfagan ambas restricciones. Con multiplicadores de Lagrange positivos, debemos tener $\partial Z/\partial\lambda_1 = \partial Z/\partial\lambda_2 = 0$. Partiendo de los dos últimos renglones de (13.18), podemos escribir

$$2x_1 + 3x_2 = 6 \qquad \text{y} \qquad 3x_1 + 2x_2 = 12$$

Estas dos ecuaciones suministran la solución de prueba $x_1 = 4\frac{4}{5}$ y $x_2 = -1\frac{1}{5}$, que no cumple la restricción de no negatividad para x_2.

Intentemos ahora $x_1 > 0$ y $x_2 > 0$, lo que implicaría $\partial Z/\partial x_1 = \partial Z/\partial x_2 = 0$ según la holgura complementaria. Entonces, partiendo de los dos primeros renglones de (13.18), podemos escribir

$$2(x_1 - 4) - 2\lambda_1 + 3\lambda_2 = 0 \qquad \text{y} \qquad 2(x_2 - 4) - 3\lambda_1 + 2\lambda_2 = 0 \tag{13.19}$$

Al multiplicar la primera ecuación por 2 y la segunda ecuación por 3, y luego al restar la última de la primera, eliminamos λ_2 y obtenemos el resultado

$$4x_1 - 6x_2 + 5\lambda_1 + 8 = 0$$

Suponiendo además que $\lambda_1 = 0$, podemos obtener la siguiente relación entre x_1 y x_2:

$$x_1 - \frac{3}{2}x_2 = -2 \tag{13.20}$$

Sin embargo, para despejar las dos variables, necesitamos otra relación entre x_1 y x_2. Para este propósito, supongamos que $\lambda_2 \neq 0$, de modo que $\partial Z/\partial\lambda_2 = 0$. Entonces, partiendo de los dos últimos renglones de (13.18) podemos escribir (después de reordenar)

$$3x_1 + 2x_2 = 12 \tag{13.21}$$

Combinados, (13.20) y (13.21) suministran otra solución de prueba

$$x_1 = \frac{28}{13}\left(= 2\frac{2}{13}\right) > 0 \qquad x_2 = \frac{36}{13}\left(= 2\frac{10}{13}\right) > 0$$

Al sustituir estos valores en (13.19) y al despejar los multiplicadores de Lagrange, obtenemos

$$\lambda_1 = 0 \qquad \lambda_2 = \frac{16}{13}\left(= 1\frac{3}{13}\right) > 0$$

Como los valores solución para las cuatro variables son todos no negativos y satisfacen ambas restricciones, son aceptables como solución final.

EJERCICIO 13.1

1. Dibuje un conjunto de diagramas similares a los de la figura 13.1 para el caso de minimización, y deduzca un conjunto de condiciones necesarias para un mínimo local correspondiente a (13.2) mediante (13.4). Condense estas condiciones en una sola aseveración similar a (13.5).
2. (*a*) Muestre que, en (13.16), en lugar de escribir

$$\lambda_i \frac{\partial Z}{\partial \lambda_i} = 0 \qquad (i = 1, \ldots, m)$$

como un conjunto de m condiciones separadas, es suficiente escribir una sola ecuación en forma de

$$\sum_{i=1}^{m} \lambda_i \frac{\partial Z}{\partial \lambda_i} = 0$$

(*b*) ¿Podemos hacer lo mismo para el siguiente conjunto de condiciones?

$$x_j \frac{\partial Z}{\partial x_j} = 0 \qquad (j = 1, \ldots, n)$$

3. Basándose en el razonamiento usado en el problema 2, ¿qué conjunto (o conjuntos) de condiciones en (13.17) puede condensarse en una sola ecuación?
4. Suponga que el problema es

$$\begin{array}{lll} \text{Minimizar} & C = f(x_1, x_2, \ldots, x_n) & \\ \text{sujeto a} & g^i(x_1, x_2, \ldots, x_n) \geq r_i & \\ \text{y} & x_j \geq 0 & \begin{pmatrix} i = 1, 2, \ldots, m \\ j = 1, 2, \ldots, n \end{pmatrix}. \end{array}$$

escriba la función lagrangeana, obtenga las derivadas $\partial Z/\partial x_j$ y $\partial Z/\partial \lambda_i$ y escriba la versión expandida de las condiciones de Kuhn-Tucker (13.17) para un valor mínimo.
5. Transforme el problema de minimización del problema 4 en un problema de maximización, formule la función lagrangeana, obtenga las derivadas respecto a x_j y λ_i, y aplique las condiciones de Kuhn-Tucker (13.16) para un valor máximo. ¿Los resultados son consistentes con los obtenidos en el problema 4?

13.2 Calificación de la restricción

Las condiciones de Kuhn-Tucker son necesarias *solamente si* se satisface una disposición específica. Esa disposición, llamada la *calificación de la restricción*, impone una cierta condición sobre las funciones de restricción de un problema de programación no lineal, para el propósito específico de descartar ciertas irregularidades en la frontera del conjunto factible, que invalidarían las condiciones de Kuhn-Tucker si se presentan en la solución óptima.

Irregularidades en los puntos de frontera

Ilustremos primero la naturaleza de estas irregularidades mediante algunos ejemplos.

Ejemplo 1

$$\begin{array}{ll} \text{Maximizar} & \pi = x_1 \\ \text{sujeto a} & x_2 - (1 - x_1)^3 \leq 0 \\ \text{y} & x_1, x_2 \geq 0 \end{array}$$

Como se muestra en la figura 13.2, la región factible es el conjunto de puntos situados en el primer cuadrante sobre o debajo de la curva $x_2 = (1 - x_1)^3$. Puesto que la función objetivo nos exige maximizar x_1, la solución óptima es el punto (1, 0). Pero la solución no satisface las condiciones de Kuhn-Tucker para un máximo. Para verificar esto, escribimos primero la función lagrangeana

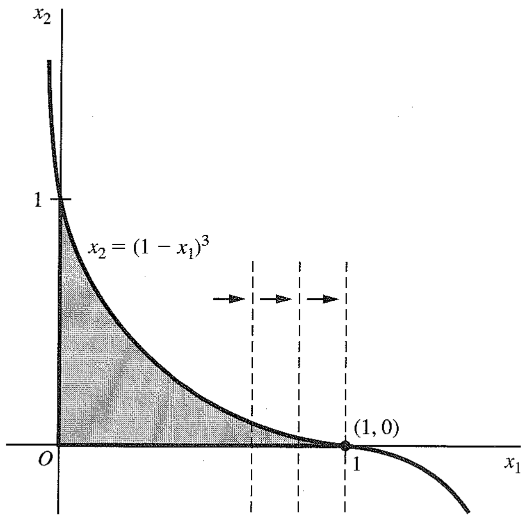
$$Z = x_1 + \lambda_1[-x_2 + (1 - x_1)^3]$$

Como la primera condición marginal, debemos tener entonces

$$\frac{\partial Z}{\partial x_1} = 1 - 3\lambda_1(1 - x_1)^2 \leq 0$$

De hecho, como $x_1^* = 1$ es positivo, la holgura complementaria requiere que esta derivada se anule cuando se evalúa para el punto (1, 0). Sin embargo, el valor real que obtenemos resulta ser $\partial Z/\partial x_1^* = 1$, no cumpliendo así la condición marginal dada.

FIGURA 13.2



La razón de esta anomalía es que la solución óptima (1, 0) se presenta en este ejemplo en un *vértice exterior* con la punta hacia fuera, la cual constituye un tipo de irregularidad que puede invalidar las condiciones de Kuhn-Tucker para una solución óptima de frontera. Un *vértice exterior* es un pico que se forma cuando una curva invierte repentinamente su dirección pero la pendiente de la curva a un lado del punto es la misma que la pendiente de la curva al otro lado del punto. Aquí, la frontera de la región factible sigue en un principio a la curva de restricción, pero cuando se alcanza el punto (1, 0), hace un giro abrupto hacia el oeste y de ahí en adelante sigue el camino del eje horizontal. Como las pendientes tanto del lado curvado como del lado horizontal de la frontera son cero para el punto (1, 0), ese punto es un vértice exterior.

Los vértices son los culpables más frecuentemente citados de la falla de las condiciones de Kuhn-Tucker, pero la verdad es que la presencia de un vértice no es ni necesaria ni suficiente para hacer que esas condiciones fallen para una solución óptima. Los ejemplos 2 y 3 van a confirmarnos esto.

Ejemplo 2

Añadámosle al problema del ejemplo 1 una nueva restricción

$$2x_1 + x_2 \leq 2$$

cuya frontera, $x_2 = 2 - 2x_1$, tiene la gráfica de una línea recta con pendiente -2 que pasa por el punto óptimo de la figura 13.2. Es claro que la región factible permanece igual que antes, y así ocurre con la solución óptima en la del vértice. Pero si escribimos la nueva función lagrangeana

$$Z = x_1 + \lambda_1[-x_2 + (1 - x_1)^3] + \lambda_2[2 - 2x_1 - x_2]$$

y las condiciones marginales

$$\frac{\partial Z}{\partial x_1} = 1 - 3\lambda_1(1 - x_1)^2 - 2\lambda_2 \leq 0$$

$$\frac{\partial Z}{\partial x_2} = -\lambda_1 - \lambda_2 \leq 0$$

$$\frac{\partial Z}{\partial \lambda_1} = -x_2 + (1 - x_1)^3 \geq 0$$

$$\frac{\partial Z}{\partial \lambda_2} = 2 - 2x_1 - x_2 \geq 0$$

resulta que los valores $x_1^* = 1$, $x_2^* = 0$, $\lambda_1^* = 1$ y $\lambda_2^* = \frac{1}{2}$ ciertamente satisfacen estas cuatro desigualdades, así como las condiciones de no negatividad y de holgura complementaria. De hecho, a λ_1^* puede asignársele cualquier valor no negativo (no solamente 1), y todavía pueden satisfacerse todas las condiciones: lo que nos muestra que el valor óptimo de un multiplicador de Lagrange no es necesariamente único. Sin embargo, es más importante el hecho de que este ejemplo muestra que las condiciones de Kuhn-Tucker pueden permanecer válidas a pesar del vértice.

Ejemplo 3

La región factible del problema

$$\begin{array}{ll} \text{Maximizar} & \pi = x_2 - x_1^2 \\ \text{sujeto a} & -\left(10 - x_1^2 - x_2\right)^3 \leq 0 \\ & -x_1 \leq -2 \\ \text{y} & x_1, x_2 \geq 0 \end{array}$$

como se muestra en la figura 13.3, no contiene vértices exteriores en ningún lado. Sin embargo en la solución óptima (2, 6) las condiciones de Kuhn-Tucker no son válidas. Ya que con la función lagrangeana

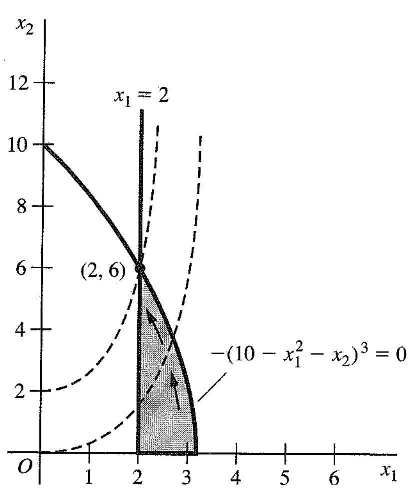
$$Z = x_2 - x_1^2 + \lambda_1\left(10 - x_1^2 - x_2\right)^3 + \lambda_2(-2 + x_1)$$

la segunda condición marginal requeriría que

$$\frac{\partial Z}{\partial x_2} = 1 - 3\lambda_1\left(10 - x_1^2 - x_2\right)^2 \leq 0$$

En realidad, como x_2^* es positivo, esta derivada debe anularse cuando se evalúa para el punto (2, 6). Pero en realidad obtenemos $\partial Z/\partial x_2 = 1$, independientemente del valor asignado a λ_1. De este modo, las condiciones de Kuhn-Tucker pueden fallar aun en ausencia de un vértice exterior: es más, aun cuando la región factible sea un conjunto convexo, como en la figura 13.3. La razón fundamental por la cual los vértices exteriores no son ni necesarios ni suficientes para que fallen las condiciones de Kuhn-Tucker es que las irregularidades anteriores a las cuales se hizo referencia antes se relacionan no con la forma de la región factible por sí misma, sino con las formas de las mismas funciones de restricción.

FIGURA 13.3



Calificación de una restricción

Las irregularidades en las fronteras —con vértices o sin vértices— no se van a presentar si se cumplen ciertos requisitos en las restricciones.

Para explicar esto, sea $x^* \equiv (x_1^*, x_2^*, \ldots, x_n^*)$ un punto frontera de la región factible y un posible candidato para una solución, y sea $dx \equiv (dx_1, dx_2, \ldots, dx_n)$, que representa una dirección particular de movimiento a partir del punto de frontera ya mencionado. La interpretación de dirección del movimiento del vector dx está perfectamente alineada con nuestra interpretación anterior de un vector como un segmento de línea dirigida (una flecha), pero aquí el punto de inicio es el punto x^* en lugar del punto de origen, por lo que entonces el vector dx *no* tiene la naturaleza de un radio vector. Ahora impondremos dos requerimientos sobre el vector dx. Primero, si la j-ésima variable de elección tiene un valor de cero para el punto x^*, entonces nosotros permitiremos solamente un cambio no negativo sobre el eje x_j, es decir,

$$dx_j \geq 0 \qquad \text{si} \qquad x_j^* = 0 \tag{13.22}$$

Segundo, si la restricción i-ésima se satisface exactamente en el punto x^*, entonces solamente permitiremos valores de $dx_1, \ldots, dx_n$ tales que el valor de la función de restricción $g^i(x^*)$ no se va a incrementar (para un problema de maximización) o no va a disminuir (para un problema de minimización), es decir,

$$dg^i(x^*) = g_1^i\, dx_1 + g_2^i\, dx_2 + \cdots + g_n^i\, dx_n \quad \begin{cases} \leq 0 \text{ (máx.)} \\ \geq 0 \text{ (mín.)} \end{cases} \qquad \text{si} \qquad g^i(x^*) = r_i \tag{13.23}$$

donde todas las derivadas parciales de g_j^i deben evaluarse para x^*. Si un vector dx cumple (13.22) y (13.23), lo denominaremos *vector de prueba*. Finalmente, si existe un arco diferenciable que (1) emane del punto x^*, (2) esté contenido enteramente en la región factible y (3) sea tangente a un vector de prueba dado, lo llamaremos un *arco calificador* para ese vector de prueba. Con estos antecedentes, la calificación de restricción puede enunciarse sencillamente como sigue:

> Se satisface la calificación de restricción si, para cualquier punto x^* en la frontera de la región factible, existe un arco calificador para cada vector de prueba dx.

Ejemplo 4

Vamos a mostrar que el punto óptimo (1, 0) del ejemplo 1 de la figura 13.2, que no cumple las condiciones de Kuhn-Tucker, tampoco cumple la calificación de restricción. Para ese punto, $x_2^* = 0$; entonces, el vector de prueba debe satisfacer

$$dx_2 \geq 0 \qquad \text{[por (13.22)]}$$

Aún más, ya que la (única) restricción, $g^1 = x_2 - (1 - x_1)^3 \leq 0$, se satisface exactamente para (1, 0), [mediante (13.23)] debemos dejar que

$$g_1^1\, dx_1 + g_2^1\, dx_2 = 3(1 - x_1^*)^2\, dx_1 + dx_2 = dx_2 \leq 0$$

Estos dos requerimientos juntos implican que debemos dejar que $dx_2 = 0$. En contraste, podemos elegir libremente a dx_1. Así por ejemplo, el vector $(dx_1, dx_2) = (2, 0)$ es un vector de prueba aceptable, como lo es $(dx_1, dx_2) = (-1, 0)$. Este último vector de prueba se graficaría en la figura 13.2 como una flecha que inicia en (1, 0) y apunta en dirección oeste (no está dibujado), y claramente se puede dibujar un arco calificador para él. (La frontera curvada de la región factible puede servir como un arco cualificante.) Por otro lado, el vector de prueba $(dx_1, dx_2) = (2, 0)$ se graficaría como una flecha que inicia en (1, 0) y que apunta en dirección hacia el este (no está dibujado). Como no hay manera de dibujar un arco tangente uniforme para este vector y que esté situado completamente dentro de la región factible, no existe ningún arco calificador para él. Entonces, el punto de solución óptima (1, 0) no cumple la calificación de restricción.

Ejemplo 5 Refiriéndonos al ejemplo 2, observemos que, después de añadir la restricción, $2x_1 + x_2 \leq 2$ el punto (1, 0) de la figura 13.2, va a satisfacer la condición de restricción calificada, revalidando con ello las condiciones de Kuhn-Tucker.

Al igual que en el ejemplo 4, tenemos que requerir $dx_2 \geq 0$ (ya que $x_2^* = 0$) y $dx_2 \leq 0$ (porque la primera restricción se satisface exactamente); entonces, $dx_2 = 0$. Pero la segunda restricción también se satisface exactamente, requiriendo con ello

$$g_1^2\, dx_1 + g_2^2\, dx_2 = 2dx_1 + dx_2 = 2dx_1 \leq 0 \qquad \text{[por (13.23)]}$$

Con dx_1 no positivo y dx_2 cero, los únicos vectores de prueba admisibles —aparte del vector nulo mismo— son aquellos que apuntan en la dirección oeste en la figura 13.2 desde (1, 0). Todos ellos están situados a lo largo del eje horizontal en la región factible, y ciertamente es posible dibujar un arco calificador para cada vector de prueba. Esta vez la calificación de restricción realmente se satisface.

Restricciones lineales

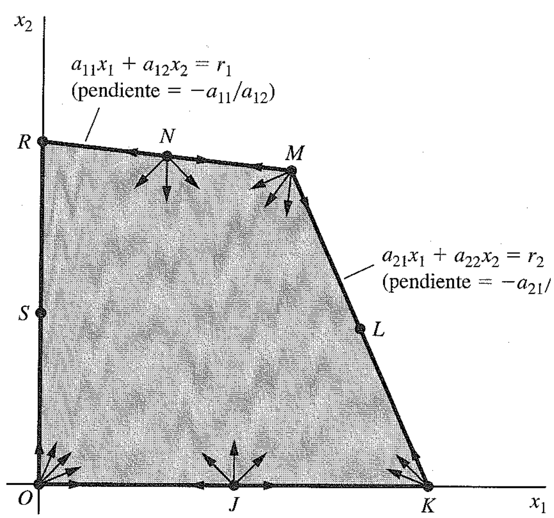
En el ejemplo 3 se demostró que la convexidad del conjunto factible no garantiza la validez de las condiciones de Kuhn-Tucker como condiciones necesarias. Sin embargo, si la región factible es un conjunto convexo formado solamente por restricciones *lineales*, entonces la calificación de restricción invariablemente va a cumplirse, y las condiciones de Kuhn-Tucker siempre serían válidas en una solución óptima. Al ser este el caso, no tenemos que preocuparnos nunca acerca de las irregularidades de frontera cuando se trata el problema de programación no lineal con restricciones lineales.

Ejemplo 6 Ilustremos el resultado de la restricción lineal en el marco de referencia de dos variables y dos restricciones. Para un problema de maximización, las restricciones lineales pueden escribirse como

$$a_{11}x_1 + a_{12}x_2 \leq r_1$$
$$a_{21}x_1 + a_{22}x_2 \leq r_2$$

donde consideraremos que todos los parámetros son positivos. Entonces, como se indica en la figura 13.4, la primera frontera de restricción tendrá una pendiente de $-a_{11}/a_{12} < 0$, y la segunda una pendiente de $-a_{21}/a_{22} < 0$. Los puntos de frontera de la región factible sombreada pueden de los siguientes cinco tipos: (1) el origen, donde se intersecan los dos ejes, (2) los

FIGURA 13.4



puntos situados en un segmento del eje, tales como J y S, (3) los puntos en la intersección de un eje y una línea frontera de restricción, a saber, K y R, (4) los puntos situados en una sola línea frontera de restricción, tales como L y N y (5) el punto de intersección de las dos restricciones, M. Podemos examinar brevemente uno por uno cada tipo de punto frontera en referencia con la satisfacción de la calificación de una restricción.

1. En el origen, no se satisface exactamente ninguna restricción, de modo que podemos ignorar (13.23). Pero ya que $x_1 = x_2 = 0$, debemos escoger vectores de prueba con $dx_1 \geq 0$ y $dx_2 \geq 0$, mediante (13.22). Entonces, todos los vectores de prueba a partir del origen deben apuntar en las direcciones este, norte, o noreste, como se ilustra en la figura 13.4. Todos estos vectores están situados dentro del conjunto factible y puede encontrarse claramente un arco calificador para cada uno.
2. En un punto como J, nuevamente podemos ignorar (13.23). El hecho de que $x_2 = 0$ significa que debemos escoger $dx_2 \geq 0$, pero somos libres de elegir dx_1. Entonces, todos los vectores serían aceptables excepto aquellos que apuntan hacia el sur ($dx_2 < 0$). Nuevamente todos estos vectores se sitúan dentro de la región factible, y hay un arco calificador para cada uno. El análisis del punto S es similar.
3. Para los puntos K y R, deben considerarse tanto (13.22) como (13.23). Específicamente, en K tenemos que escoger $dx_2 \geq 0$ ya que $x_2 = 0$, de modo que debemos descartar todas las flechas que se dirigen hacia el sur. Al satisfacerse exactamente la segunda restricción, los vectores de prueba para el punto K deben satisfacer

$$g_1^2\, dx_1 + g_2^2\, dx_2 = a_{21}\, dx_1 + a_{22}\, dx_2 \leq 0 \tag{13.24}$$

ya que para K también tenemos $a_{21}x_1 + a_{22}x_2 = r_2$ (segunda frontera de restricción); sin embargo, podemos añadir esta igualdad a (13.24) y modificar la restricción sobre el vector de prueba a la forma

$$a_{21}(x_1 + dx_1) + a_{22}(x_2 + dx_2) \leq r_2 \tag{13.24'}$$

Interpretando que $(x_j + dx_j)$ es el nuevo valor de x_j alcanzado en la punta de la flecha de un vector de prueba, podemos inferir que (13.24′) tiene el significado de que todos los vectores de prueba deben tener sus cabezas de flecha ubicados en o debajo de la segunda línea frontera de restricción. En consecuencia, todos estos vectores nuevamente deben situarse dentro de la región factible, y puede encontrarse un arco calificador para cada uno. El análisis del punto R es análogo.
4. Para puntos tales como L y N, ninguna de las variables es cero y (13.22) puede ignorarse. Sin embargo, para el punto N (13.23) determina que

$$g_1^1\, dx_1 + g_2^1\, dx_2 = a_{11}\, dx_1 + a_{12}\, dx_2 \leq 0 \tag{13.25}$$

ya que el punto N satisface $a_{11}\, dx_1 + a_{12}\, dx_2 = r_1$ (primera línea frontera de restricción), podemos añadir esta igualdad a (13.25) y escribir

$$a_{11}(x_1 + dx_1) + a_{12}(x_2 + dx_2) \leq r_1 \tag{13.25'}$$

Esto requeriría que los vectores de prueba tengan cabezas de flecha ubicadas en o debajo de la primera línea frontera de restricción en la figura 13.4. Entonces, obtenemos esencialmente la misma clase de resultado encontrado en los otros casos. El análisis del punto L es análogo.
5. Para el punto M, nuevamente podemos omitir (13.22), pero esta vez (13.23) requiere que todos los vectores de prueba satisfagan tanto (13.24) como (13.25). Como podemos modificar estas últimas condiciones en las formas en (13.24′) y (13.25′), todos los vectores de prueba deben tener ahora sus cabezas de flecha ubicadas en o debajo de la primera y de la segunda línea frontera de restricción. El resultado nuevamente duplica los resultados de los casos anteriores.

En este ejemplo, sucede que para cada tipo de punto de frontera considerado todos los vectores de prueba están situados dentro de la región factible. Aun cuando esta característica de ubicación hace que los arcos calificadores sean fáciles de encontrar, de ninguna manera es un prerrequisito para su existencia. En un problema con una frontera de restricción no lineal, en especial, la frontera de restricción puede servir como un arco calificador para algún vector de prueba que esté situado fuera de la región factible. Un ejemplo de esto puede encontrarse en uno de los siguientes problemas:

EJERCICIO 13.2

1. Verifique si el punto solución $(x_1^*, x_2^*) = (2, 6)$ en el ejemplo 3 satisface la calificación de restricción.
2. Maximizar $\pi = x_1$
   sujeto a $x_1^2 + x_2^2 \le 1$
   y $x_1, x_2 \ge 0$
   Resuelva gráficamente y verifique si el punto de solución óptima satisface (*a*) la calificación de restricción y (*b*) las condiciones de Kuhn-Tucker.
3. Minimizar $C = x_1$
   sujeto a $x_1^2 - x_2 \ge 0$
   y $x_1, x_2 \ge 0$
   Resuelva gráficamente. ¿Se presenta la solución óptima en un vértice? Verifique si la solución óptima cumple (*a*) la cualificación de restricción y (*b*) las condiciones de Kuhn-Tucker para un mínimo.
4. Minimizar $C = x_1$
   sujeto a $-x_2 - (1 - x_1)^3 \ge 0$
   y $x_1, x_2 \ge 0$
   Muestre que (*a*) la solución óptima $(x_1^*, x_2^*) = (1, 0)$ no satisface las condiciones de Kuhn-Tucker, pero (*b*) al introducir un nuevo multiplicador $\lambda_0 \ge 0$, y al modificar la función lagrangeana (13.15) a la forma

$$Z_0 = \lambda_0 f(x_1, x_2, \ldots, x_n) + \sum_{i=1}^{m} \lambda_1 [r_i - g^i(x_1, x_2, \ldots, x_n)]$$

   las condiciones de Kuhn-Tucker se cumplen para (1, 0). (*Nota*: las condiciones de Kuhn-Tucker para los multiplicadores se amplían solamente a $\lambda_1, \ldots, \lambda_m$, pero no a λ_0.)

13.3 Aplicaciones económicas

Racionamiento en tiempo de guerra

Es común que durante una guerra la población civil se someta a alguna forma de racionamiento de los bienes básicos de consumo. Generalmente, el método de racionamiento se controla a través del uso de cupones del gobierno, quien suministra a cada consumidor una dotación de cupones al mes. A su vez, el consumidor deberá canjear un cierto número de cupones en el momento de la compra de un bien racionado. Esto significa efectivamente que el consumidor paga *dos* precios al momento de la compra. Se paga tanto el precio del cupón como el precio monetario del bien racionado. Esto requiere que el consumidor tenga suficientes fondos y suficientes cupones para comprar una unidad del bien racionado.

Considere el caso de un mundo de dos bienes en el cual se racionan los dos bienes x y y, y sea la función de utilidad del consumidor $U = U(x, y)$. El consumidor tiene un presupuesto fijo de dinero B y enfrenta precios exógenos P_x y P_y. Aún más, el consumidor tiene una dotación de cupones, C, que puede usarse para comprar ya sea x o y a un precio de cupón de c_x y c_y. Por lo tanto, el problema de maximización del consumidor es

$$\begin{array}{ll} \text{Maximizar} & U = U(x, y) \\ \text{sujeto a} & P_x x + P_y y \leq B \\ & c_x x + c_y y \leq C \\ \text{y} & x, y \geq 0 \end{array}$$

La función lagrangeana para el problema es

$$Z = U(x, y) + \lambda_1(B - P_x x - P_y y) + \lambda_2(C - c_x x + c_y y)$$

donde λ_1 y λ_2 son los multiplicadores de Lagrange. Como ambas restricciones son lineales, la calificación de restricción se cumple y son necesarias las condiciones de Kuhn-Tucker:

$$\begin{array}{lll} Z_x = U_x - \lambda_1 P_x - \lambda_2 c_x \leq 0 & x \geq 0 & x Z_x = 0 \\ Z_y = U_y - \lambda_1 P_y - \lambda_2 c_y \leq 0 & y \geq 0 & y Z_y = 0 \\ Z_{\lambda_1} = B - P_x x - P_y y \geq 0 & \lambda_1 \geq 0 & \lambda_1 Z_{\lambda_1} = 0 \\ Z_{\lambda_2} = C - c_x x - c_y y \geq 0 & \lambda_2 \geq 0 & \lambda_2 Z_{\lambda_2} = 0 \end{array}$$

Ejemplo 1

Suponga que la función de utilidad es de la forma $U = xy^2$. Además, sea $B = 100$ y $P_x = P_y = 1$, mientras que $C = 120$, $c_x = 2$ y $c_y = 1$.

La función lagrangeana adopta la forma específica de

$$Z = xy^2 + \lambda_1(100 - x - y) + \lambda_2(120 - 2x - y)$$

Ahora, las condiciones de Kuhn-Tucker son

$$\begin{array}{lll} Z_x = y^2 - \lambda_1 - 2\lambda_2 \leq 0 & x \geq 0 & xZ_x = 0 \\ Z_y = 2xy - \lambda_1 - \lambda_2 \leq 0 & y \geq 0 & yZ_y = 0 \\ Z_{\lambda_1} = 100 - x - y \geq 0 & \lambda_1 \geq 0 & \lambda_1 Z_{\lambda_1} = 0 \\ Z_{\lambda_2} = 120 - 2x - y \geq 0 & \lambda_2 \geq 0 & \lambda_2 Z_{\lambda_2} = 0 \end{array}$$

Nuevamente, el procedimiento de solución incluye una cierta cantidad de prueba y error. Primero podemos escoger a una de las restricciones como no activa y despejar x y y. Una vez encontrados, use estos valores para probar si se viola la restricción escogida como no activa. Si se viola, entonces reconstruya el procedimiento escogiendo otra restricción como no activa. Si nuevamente se viola la restricción no activa, entonces podemos suponer que ambas restricciones son activas y la solución se determina solamente por las restricciones.

Paso 1: Suponga que la segunda restricción (racionamiento) es no activa en la solución, de modo que $\lambda_2 = 0$ mediante la holgura complementaria. Pero sean x y y λ_1 positivos de modo que la holgura complementaria nos daría las siguientes tres ecuaciones:

$$\begin{aligned} Z_x &= y^2 - \lambda_1 = 0 \\ Z_y &= 2xy - \lambda_1 = 0 \\ Z_{\lambda_1} &= 100 - x - y = 0 \end{aligned}$$

Al despejar x y y obtenemos una solución de prueba

$$x = 33^1/_3 \qquad y = 66^2/_3$$

Sin embargo, cuando sustituimos estas soluciones en la restricción del cupón encontramos que

$$2(33^1/_3) + 66^2/_3 = 133^1/_3 > 120$$

Esta solución viola la restricción del cupón y debe rechazarse.

Paso 2: Invirtamos ahora la hipótesis para λ_1 y λ_2 de modo que $\lambda_1 = 0$, pero sea $\lambda_2, x, y > 0$. Entonces, a partir de las condiciones marginales, tenemos

$$Z_x = y^2 - 2\lambda_2 = 0$$

$$Z_y = 2xy - \lambda_2 = 0$$

$$Z_{\lambda_1} = 120 - 2x - y = 0$$

La solución de este sistema de ecuaciones suministra otra solución de prueba

$$x = 20 \qquad y = 80$$

lo que implica que $\lambda_2 = 2xy = 3\,200$. Estos valores de solución, junto con $\lambda_1 = 0$, satisfacen ambas restricciones de presupuesto y de racionamiento. Entonces, podemos aceptarlas como la solución final para las condiciones de Kuhn-Tucker.

Sin embargo, esta solución óptima contiene una anormalidad curiosa. Siendo activa la restricción presupuestaria en la solución, normalmente esperaríamos que el multiplicador de Lagrange relacionado fuera positivo, pero en realidad tenemos $\lambda_1 = 0$. Entonces en este ejemplo, mientras que la restricción presupuestaria es *matemáticamente* activa (que se satisface como una igualdad estricta en la solución), es *económicamente* no activa (no necesita una utilidad marginal positiva de dinero).

Fijación de precios a mercados no planeados originalmente

Los precios en el mercado planeado y fuera de él son problemas de planeación en compañías que tienen procesos de producción con capacidad restringida. Generalmente la compañía ha invertido en la capacidad con objeto de centrarse en un mercado planeado o primario; sin embargo, puede haber un mercado secundario en el cual la compañía puede con frecuencia vender su producto. Una vez que se ha comprado el equipo capital para dar servicio al mercado primario de la compañía, está disponible libremente (de acuerdo con la capacidad) para usarse en el mercado secundario. Los ejemplos típicos incluyen escuelas y universidades que se construyen para satisfacer las necesidades diurnas (mercado planeado), pero pueden ofrecer clases nocturnas (no planeado); teatros que ofrecen espectáculos en la tarde (planeado) y matinés (fuera de lo planeado); y empresas de transporte que tienen rutas consagradas pero que pueden escoger entrar a los mercados de "remolques en el camino de regreso". Como el costo de la capacidad es un factor en la decisión de maximización de la ganancia para el mercado planeado y ya está pagado, normalmente no debe ser un factor para el cálculo del precio y la cantidad óptimos para el mercado más pequeño fuera de lo planeado. No obstante, las restricciones pueden ser importantes, especialmente cuando es una práctica común hacer discriminación de precios y cobrar precios más bajos en periodos fuera de lo planeado. Aun cuando el mercado secundario es más reducido que el primario, es posible que para el precio más bajo (maximización de la ganancia) la demanda fuera de lo planeado exceda a la capacidad. En estos casos, deben tomarse decisiones de capacidad que consideren ambos mercados, haciendo del problema una aplicación clásica de la programación no lineal.

Consideremos una compañía que maximiza la ganancia que enfrenta dos curvas de ingreso promedio

$$P_1 = P^1(Q_1) \qquad \text{en horario diurno (periodo planeado)}$$

$$P_2 = P^2(Q_2) \qquad \text{en horario nocturno (periodo no planeado)}$$

Para operar, la compañía debe pagar b por unidad de producto, ya sea de día o de noche. Aún más, la compañía debe comprar capacidad a un costo de c por unidad de capacidad. Sea K la capacidad total medida en unidades de Q. La compañía debe pagar por la capacidad, independientemente de si opera en el periodo fuera del planeado. ¿A quién se le deben cobrar los costos de capacidad: al cliente del periodo planeado, al cliente fuera del plan o a ambos clientes? El problema de maximización de la compañía se transforma en

$$\begin{array}{ll}
\text{Maximizar} & \pi = P_1Q_1 + P_2Q_2 - b(Q_1+Q_2) - cK \\
\text{sujeto a} & Q_1 \le K \\
& Q_2 \le K \\
\text{donde} & P_1 = P^1(Q_1) \\
& P_2 = P^2(Q_2) \\
\text{y} & Q_1, Q_2, K \ge 0
\end{array}$$

En vista de que el ingreso total para Q_i,

$$R_i \equiv P_iQ_i = P^i(Q_i)Q_i$$

es una función solamente de Q_i, podemos simplificar el enunciado del problema a

$$\begin{array}{ll}
\text{Maximizar} & \pi = R_1(Q_1) + R_2(Q_2) - b(Q_1+Q_2) - cK \\
\text{sujeto a} & Q_1 \le K \\
& Q_2 \le K \\
\text{y} & Q_1, Q_2, K \ge 0
\end{array}$$

Observe que ambas restricciones son lineales, por lo que se satisface la calificación de restricción y son necesarias las condiciones de Kuhn-Tucker.

La función lagrangeana es

$$Z = R_1(Q_1) + R_2(Q_2) - b(Q_1+Q_2) - cK + \lambda_1(K - Q_1) + \lambda_2(K - Q_2)$$

y las condiciones de Kuhn-Tucker son

$$\begin{array}{lll}
Z_1 = \mathrm{MR}_1 - b - \lambda_1 \le 0 & Q_1 \ge 0 & Q_1Z_1 = 0 \\
Z_2 = \mathrm{MR}_2 - b - \lambda_2 \le 0 & Q_2 \ge 0 & Q_2Z_2 = 0 \\
Z_K = -c + \lambda_1 + \lambda_2 \le 0 & K \ge 0 & KZ_K = 0 \\
Z_{\lambda_1} = K - Q_1 \ge 0 & \lambda_1 \ge 0 & \lambda_1 Z_{\lambda_1} = 0 \\
Z_{\lambda_2} = K - Q_2 \ge 0 & \lambda_2 \ge 0 & \lambda_2 Z_{\lambda_2} = 0
\end{array}$$

donde MR_i es el ingreso marginal de Q_i $(i = 1, 2)$.

Nuevamente, el procedimiento de solución implica la prueba y el error. Supongamos primero que $Q_1, Q_2, K > 0$. Entonces, por la holgura complementaria, tenemos

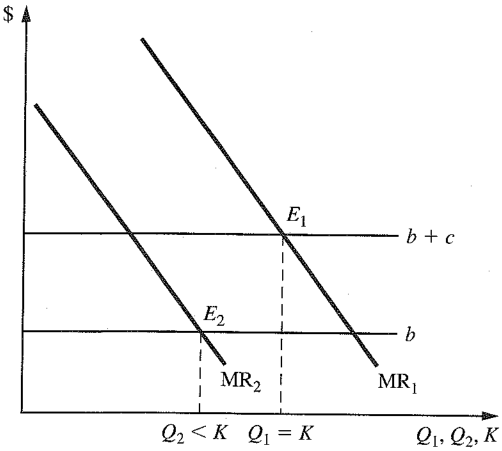
$$\begin{aligned}
&\mathrm{MR}_1 - b - \lambda_1 = 0 \\
&\mathrm{MR}_2 - b - \lambda_2 = 0 \\
&-c + \lambda_1 + \lambda_2 = 0 \quad (\lambda_1 = c - \lambda_2)
\end{aligned} \tag{13.26}$$

que puede agruparse en dos ecuaciones después de eliminar λ_1:

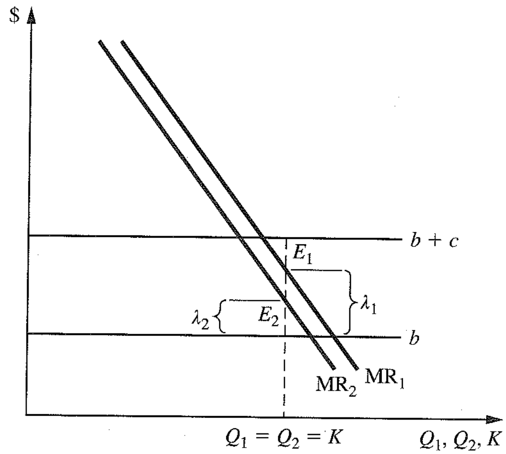
$$\begin{aligned}
\mathrm{MR}_1 &= b + c - \lambda_2 \\
\mathrm{MR}_2 &= b + \lambda_2
\end{aligned} \tag{13.26'}$$

Luego proseguimos en dos pasos.

FIGURA 13.5



a) Restricción no activa fuera de la planeación



b) Restricción activa fuera de la planeación

Paso 1: Como el mercado no planeado es secundario, puede esperarse que la función de ingreso marginal (MR_2) se sitúe debajo de la del mercado primario (MR_1), como se ilustra en la figura 13.5.Y es más probable que la restricción de capacidad sea no activa en el mercado secundario, de modo que lo más probable es que λ_2 sea cero. De modo que intentamos $\lambda_2 = 0$. Entonces, (13.26′) se transforma en

$$\begin{aligned} \mathrm{MR}_1 &= b + c \\ \mathrm{MR}_2 &= b \end{aligned} \tag{13.26''}$$

El hecho de que el mercado primario absorba el costo c de la capacidad total implica que $Q_1 = K$. Sin embargo, todavía necesitamos verificar si se satisface la restricción $Q_2 \leq K$. Si se satisface, hemos encontrado una solución válida. La figura 13.5*a* ilustra el caso donde $Q_1 = K$ y $Q_2 < K$ en la solución. La curva MR_1 intercepta la línea $b + c$ en el punto E_1, y la curva MR_2 intercepta la línea b en el punto E_2.

¿Qué pasa si la solución de prueba anterior implica $Q_2 > K$, lo que ocurriría si la curva MR_2 es muy cercana a MR_1, de modo que intercepte la línea b para un producto mayor que K? Entonces, por supuesto que se viola la segunda restricción y debemos rechazar la hipótesis de $\lambda_2 = 0$ y proseguir al siguiente paso.

Paso 2: Supongamos ahora que ambos multiplicadores de Lagrange son positivos, y entonces $Q_1 = Q_2 = K$. Entonces, al no poder eliminar ninguna de las variables en (13.26), tenemos

$$\begin{aligned} \mathrm{MR}_1 &= b + \lambda_1 \\ \mathrm{MR}_2 &= b + \lambda_2 \\ c &= \lambda_1 + \lambda_2 \end{aligned} \tag{13.26'''}$$

Este caso se ilustra en la figura 13.5*b*, donde los puntos E_1 y E_2 satisfacen las primeras dos ecuaciones en (13.26‴). En la tercera ecuación vemos que el costo c de capacidad es la suma de los dos multiplicadores de Lagrange. Esto significa que λ_1 y λ_2 representan las partes del costo de capacidad sustentadas respectivamente por los dos mercados.

Ejemplo 2

Suponga que la función de ingreso promedio durante las horas planeadas es

$$P_1 = 22 - 10^{-5} Q_1$$

y durante las horas no planeadas es

$$P_2 = 18 - 10^{-5} Q_2$$

La producción de una unidad de producto por medio día requiere una unidad de capacidad que cuesta 8 centavos por día. El costo de una unidad de capacidad es el mismo ya sea que se use solamente en tiempos planeados, o también no planeados. Además de los costos de capacidad, cuesta 6 centavos en costos de operación (mano de obra y combustible) producir una unidad por medio día (tanto en el día como en la tarde).

Si suponemos que la restricción de capacidad es no activa en el mercado secundario ($\lambda_2 = 0$), entonces las condiciones dadas de Kuhn-Tucker se transforman en

$$\begin{aligned} &\lambda_1 = c = 8 \\ &22 - 2 \times 10^{-5} Q_1 \quad = b + c \quad = 14 \\ &\underbrace{18 - 2 \times 10^{-5} Q_2}_{\text{MR}} \quad = \underbrace{b \qquad}_{\text{MC}} \quad = 6 \end{aligned}$$

La solución de este sistema nos da

$$Q_1 = 400\,000$$

$$Q_2 = 600\,000$$

lo que viola la hipótesis de que la segunda restricción es no activa, ya que $Q_2 > Q_1 = K$.

Por lo tanto, supongamos que ambas restricciones son activas. Entonces $Q_1 = Q_2 = Q$ y las condiciones de Kuhn-Tucker se transforman en

$$\lambda_1 + \lambda_2 = 8$$

$$22 - 2 \times 10^{-5} Q = 6 + \lambda_1$$

$$18 - 2 \times 10^{-5} Q = 6 + \lambda_2$$

que aportan la siguiente solución

$$Q_1 = Q_2 = K = 500\,000$$

$$\lambda_1 = 6 \qquad \lambda_2 = 2$$

$$P_1 = 17 \qquad P_2 = 13$$

Como la restricción de capacidad es activa en ambos mercados, el mercado primario paga $\lambda_1 = 6$ del costo de capacidad y el mercado secundario paga $\lambda_2 = 2$.

EJERCICIO 13.3

1. Suponga que en el ejemplo 2 una unidad de capacidad cuesta solamente 3 centavos por día.
   (*a*) ¿Cuáles serían los precios y las cantidades de planeadas y las no planeadas que maximizan la ganancia?
   (*b*) ¿Cuáles serían los valores de los multiplicadores de Lagrange? ¿Qué interpretación le da a esos valores?
2. Un consumidor vive en una isla donde ésta produce dos bienes, x y y, de acuerdo con la frontera de posibilidad de producción $x^2 + y^2 \leq 200$, y ella misma consume todos los bienes. Su función de utilidad es

   $U = xy^3$

   El consumidor también enfrenta una restricción ambiental en su producción total de ambos bienes. La restricción ambiental está dada por $x + y \leq 20$.

(*a*) Escriba las condiciones de primer orden de Kuhn-Tucker.

(*b*) Encuentre el óptimo x y y del consumidor. Identifique cuáles restricciones son activas.

3. Una empresa de electricidad está construyendo una planta de energía en un país extranjero y tiene que planear su capacidad. La demanda de energía del periodo planeado está dada por $P_1 = 400 - Q_1$ y la demanda no planeada está dada por $P_2 = 380 - Q_2$. El costo variable es 20 por unidad (pagadero en ambos mercados) y 10 los costos de capacidad por unidad que se pagan solamente una vez y se usan en ambos periodos.
   (*a*) Escriba la función lagrangeana y las condiciones de Kuhn-Tucker para este problema.
   (*b*) Encuentre la capacidad y la producción óptima para este problema.
   (*c*) ¿Cuánto paga cada mercado por la capacidad (es decir, cuáles son los valores de λ_1 y λ_2)?
   (*d*) Ahora, suponga que el costo de la capacidad es de 30 centavos por unidad (que se paga solamente una vez). Encuentre las cantidades, la capacidad y qué proporción de la capacidad paga cada mercado (es decir, λ_1 y λ_2).

13.4 Los teoremas de suficiencia en la programación no lineal

En las secciones anteriores introdujimos las condiciones de Kuhn-Tucker e ilustramos sus aplicaciones como condiciones *necesarias* en los problemas de optimización con restricciones de desigualdad. Bajo ciertas circunstancias, las condiciones de Kuhn-Tucker también pueden tomarse como condiciones suficientes.

El teorema de suficiencia de Kuhn-Tucker: la programación cóncava

En los problemas clásicos de optimización, las condiciones suficientes para máximos y mínimos se expresan tradicionalmente en términos de los signos de las derivadas o diferenciales de segundo orden. Sin embargo, como hemos mostrado en la sección 11.5, estas condiciones de segundo orden están íntimamente relacionadas con los conceptos de concavidad y convexidad de la función objetivo. Aquí en la programación no lineal, las condiciones suficientes también pueden enunciarse directamente en términos de concavidad y convexidad. De hecho, estos conceptos se aplicarán no solamente a la función objetivo $f(x)$ sino que también a las funciones de restricción $g^i(x)$.

Para el problema de *maximización*, Kuhn y Tucker ofrecen el siguiente enunciado de condiciones suficientes (teorema de la suficiencia):

Dado el problema de programación no lineal

$$\begin{array}{lll} \text{Maximizar} & \pi = f(x) & \\ \text{sujeto a} & g^i(x) \leq r_i & (i = 1, 2, \ldots, m) \\ \text{y} & x \geq 0 & \end{array}$$

si se satisfacen las siguientes condiciones:

a) la función objetivo $f(x)$ es diferenciable y *cóncava* en el cuadrante n-dimensional no negativo

b) cada función de restricción $g^i(x)$ es diferenciable y *convexa* en el cuadrante n-dimensional no negativo

c) el punto x^* satisface las condiciones de máximos de Kuhn-Tucker

entonces x^* da un máximo global de $\pi = f(x)$.

Tome en cuenta que en este teorema, la calificación de restricción no se menciona en ningún lado. Esto se debe a que ya hemos supuesto, en la condición *c*), que las condiciones de Kuhn-

Tucker se satisfacen para x^* y, en consecuencia, el aspecto de la calificación de restricción ya no es un problema.

Tal como está, el teorema anterior indica que las condiciones *a*), *b*) y *c*) son suficientes para establecer que x^* sea una solución óptima. Sin embargo, si se ve desde otro punto de vista, también podemos interpretarlo como que dados *a*) y *b*), entonces las condiciones de máximos de Kuhn-Tucker son condiciones suficientes para un máximo. En la sección anterior aprendimos que las condiciones de Kuhn-Tucker, aun cuando no son necesarias *per se*, se vuelven necesarias cuando se satisface la calificación de restricción. Al combinar esta información con el teorema de suficiencia podemos enunciar que si la calificacion de restricción se satisface y se cumplen las condiciones *a*) y *b*), entonces las condiciones de máximos de Kuhn-Tucker serán *necesarias y suficientes* para un máximo. Éste sería el caso, por ejemplo, cuando todas las restricciones son desigualdades lineales, lo que es suficiente para satisfacer la calificación de la restricción.

El problema de maximización abordado en el teorema de suficiencia anterior frecuentemente se denomina *programación cóncava*. Este nombre se debe a que Kuhn y Tucker adoptan la desigualdad $\geq$ en lugar de la desigualdad $\leq$ en cada restricción, de modo que la condición *b*) requeriría que las funciones $g^i(x)$ fueran *todas cóncavas*, al igual que la función $f(x)$. Pero hemos modificado la formulación con objeto de transmitir la idea de que en un problema de maximización, se impone una restricción para "limitar" (es decir, $\leq$) el intento de ascender a puntos más altos en la función objetivo. Aun cuando de forma diferente, las dos formulaciones son esencialmente equivalentes. Por brevedad, omitimos la prueba.

Como dijimos anteriormente, el teorema de suficiencia trata solamente problemas de maximización, pero la adaptación a problemas de *minimización* no es difícil. Aparte de los cambios apropiados en el teorema para reflejar el inverso del problema mismo, todo lo que tenemos que hacer es intercambiar las dos palabras *cóncavo* y *convexo* en las condiciones *a*) y *b*) y usar las condiciones de Kuhn-Tucker para los valores *mínimos* en la condición *c*) (vea el ejercicio 13.4-1).

El teorema de suficiencia de Arrow-Enthoven: la programación cuasicóncava

Para aplicar el teorema de suficiencia de Kuhn-Tucker deben cumplirse ciertas especificaciones de concavidad-convexidad bastante estrictas. En otro teorema de suficiencia —el de suficiencia de Arrow-Enthoven[3]— estas especificaciones se relajan hasta requerir solamente *cuasiconcavidad* y *cuasiconvexidad* en las funciones objetivo y de restricciones. De esta forma, con menos requerimientos, el alcance de aplicabilidad de las condiciones suficientes se amplía de manera proporcional.

En la formulación original de Arrow-Enthoven, con un problema de maximización y con restricciones en la forma $\geq$, las funciones $f(x)$ y $g^i(x)$ deben ser uniformemente cuasicóncavas para que el teorema sea aplicable. Esto da lugar al nombre *programación cuasicóncava*. Sin embargo, aquí nuevamente usaremos la desigualdad $\leq$ en las restricciones de un problema de maximización y la desigualdad $\geq$ en el problema de minimización.

El teorema es el siguiente:

Dado el programa de programación no lineal

$$\begin{array}{lll} \text{Maximizar} & \pi = f(x) & \\ \text{sujeto a} & g^i(x) \leq r_i & (i = 1, 2, \ldots, m) \\ \text{y} & x \geq 0 & \end{array}$$

[3] Kenneth J. Arrow y Alain C. Enthoven, "Quasi-concave Programming", *Econometrica*, octubre, 1961, pp. 779-800.

si se satisfacen las siguientes condiciones:

a) la función objetivo $f(x)$ es diferenciable y *cuasicóncava* en el cuadrante n-dimensional no negativo

b) cada función de restricción $g^i(x)$ es diferenciable y *cuasiconvexa* en el cuadrante n-dimensional no negativo

c) el punto x^* satisface las condiciones de máximos de Kuhn-Tucker

d) se satisface *cualquiera* de los siguientes:

d-i) $f_j(x^*) < 0$ por cuando menos una variable x_j

d-ii) $f_j(x^*) > 0$ para alguna variable x_j que adopte un valor positivo sin oponerse a las restricciones

d-iii) las n derivadas $f_j(x^*)$ no son todas cero, y la función $f(x)$ es dos veces diferenciable en la vecindad de x^* [es decir, todas las derivadas parciales de segundo orden de $f(x)$ existen para x^*]

d-iv) la función $f(x)$ es cóncava

entonces x^* da un máximo global de $\pi = f(x)$.

Como la prueba de este teorema es muy laboriosa, la omitiremos aquí. Sin embargo, queremos enfocar su atención hacia algunas características importantes de este teorema. Por ejemplo, aun cuando Arrow y Enthoven han tenido éxito en debilitar las especificaciones de concavidad-convexidad hasta la contraparte de cuasiconcavidad-cuasiconvexidad, encuentran necesario adicionar un nuevo requerimiento, *d*). Observe, sin embargo, que solamente se requiere *una* de las cuatro alternativas listadas en *d*) para formar un conjunto completo de condiciones suficientes. En efecto, el teorema anterior contiene *cuatro* diferentes conjuntos de condiciones suficientes para un máximo. En el caso de *d-iv*), con $f(x)$ cóncava, parecería que el teorema de suficiencia de Arrow-Enthoven se vuelve idéntico al teorema de suficiencia de Kuhn-Tucker. Pero esto no es verdad. Como Arrow y Enthoven requieren solamente que las funciones de restricción $g^i(x)$ sean cuasicóncavas, sus condiciones suficientes son todavía más débiles.

Como se estableció, el teorema aglomera las condiciones *a*) hasta *d*) en un conjunto de condiciones suficientes. Pero también es posible interpretarlo de modo que cuando *a*), *b*) y *d*) se satisfacen, entonces las condiciones de máximos de Kuhn-Tucker se hacen condiciones suficientes para un máximo. Aún más, si también se satisface la calificación de restricción, entonces las condiciones de Kuhn-Tucker se harán necesarias y suficientes para un máximo.

Al igual que el de Kuhn-Tucker, el teorema de Arrow-Enthoven puede adaptarse con facilidad al marco de referencia de la *minimización*. Aparte de los cambios obvios que se necesitan para invertir la dirección de la optimización, simplemente tenemos que intercambiar las palabras *cuasicóncavo* y *cuasiconvexo* en las condiciones *a*) y *b*), reemplazar las condiciones de máximos de Kuhn-Tucker por las condiciones mínimas, invertir las desigualdades en *d-i*) y *d-ii*) y cambiar la palabra *cóncavo* por *convexo* en *d-iv*).

Una prueba de calificación de restricción

Mencionamos en la sección 13.2 que si todas las funciones de restricción son lineales, entonces se satisface la calificación de restricción. En el caso de que las funciones $g^i(x)$ sean no lineales, la siguiente prueba ofrecida por Arrow y Enthoven puede resultar útil para determinar si se satisface la calificación de la restricción:

En un problema de maximización, si

a) cada función de restricción $g^i(x)$ es diferenciable y cuasiconvexa

b) existe un punto x^0 en el cuadrante n-dimensional no negativo tal que todas las restricciones se satisfacen como desigualdades estrictas para x^0

c) *una* de las siguientes proposiciones es verdadera:

c-i) cada una de las funciones $g^i(x)$ es convexa

c-ii) las derivadas parciales de cada $g^i(x)$ no son todas cero cuando se evalúan para cada punto x en la región factible

entonces se satisface la calificación de restricción.

Esta prueba puede adaptarse fácilmente al problema de minimización. Para lograrlo, simplemente debemos cambiar la palabra *cuasiconvexo* por *cuasicóncavo* en la condición *a*), y la palabra *convexo* a *cóncavo* en *c-i*).

EJERCICIO 13.4

1. Dado: Minimizar $C = F(x)$
   sujeto a $G^i(x) \geq r_i \quad (i = 1, 2, \ldots, m)$
   y $x \geq 0$
   (*a*) Conviértalo a un problema de maximización.
   (*b*) ¿Cuáles son en este problema los equivalentes de las funciones f y g^i en el teorema de suficiencia de Kuhn-Tucker?
   (*c*) Entonces, ¿qué condiciones de concavidad-convexidad deben imponerse a las funciones F y G^i para hacer aplicables aquí las condiciones suficientes para un máximo?
   (*d*) Basándose en lo anterior, ¿cómo enunciaría las condiciones suficientes de Kuhn-Tucker para un *mínimo*?
2. ¿El teorema de suficiencia de Kuhn-Tucker es aplicable a:
   (*a*) Maximizar $\pi = x_1$
   sujeto a $x_1^2 + x_3^2 \leq 1$
   y $x_1, x_2 \geq 0$
   (*b*) Minimizar $C = (x_1 - 3)^2 + (x_2 - 4)^2$
   sujeto a $x_1 + x_2 \geq 4$
   y $x_1, x_2 \geq 0$
   (*c*) Minimizar $C = 2x_1 + x_2$
   sujeto a $x_1^2 - 4x_1 + x_2 \geq 0$
   y $x_1, x_2 \geq 0$
3. ¿Cuál de las siguientes funciones es matemáticamente aceptable como función objetivo de un problema de *maximización* que califique para la aplicación del teorema de suficiencia de Arrow-Enthoven?
   (*a*) $f(x) = x^3 - 2x$
   (*b*) $f(x_1, x_2) = 6x_1 - 9x_2$
   (*c*) $f(x_1, x_2) = x_2 - \ln x_1$ (*Nota:* vea el ejercicio 12.4-4.)

4. ¿Se satisface la calificación de restricción de Arrow-Enthoven, dado que las restricciones de un problema de maximización son:
(*a*) $x_1^2 + (x_2 - 5)^2 \leq 4$ y $5x_1 + x_2 < 10$
(*b*) $x_1 + x_2 \leq 8$ y $-x_1x_2 \leq -8$ (*Nota:* $-x_1x_2$ no es convexa.)

13.5 Funciones de valor máximo y el teorema de la envolvente[4]

Una función de valor máximo es una función objetivo a cuyas variables de elección se les asignan valores óptimos. Estos valores óptimos de las variables de elección son, a su vez, funciones de las variables y parámetros exógenos del problema. Una vez que los valores óptimos de las variables de elección se han sustituido en la función objetivo original, la función se transforma indirectamente en una función nada más de los parámetros (a través de la influencia de los parámetros sobre los valores óptimos de las variables de elección). De esta forma, la función de valor máximo también se denomina *función objetivo indirecta.*

El teorema de la envolvente para la optimización sin restricciones

¿Cuál es la importancia de la función objetivo indirecta? Tome en cuenta que en cualquier problema de optimización la función objetivo directa se maximiza (o se minimiza) para un conjunto dado de parámetros. La función objetivo indirecta rastrea todos los valores máximos de la función objetivo a medida que los parámetros varían. Entonces, la función objetivo indirecta es una "envolvente" del conjunto de funciones objetivo optimizadas generadas al variar los parámetros del modelo. Para la mayoría de los estudiantes de economía la primera ilustración de esta noción de envolvente surge de la comparación de curvas de costo a corto y a largo plazos. Comúnmente se enseña a los alumnos que la curva de costos promedio a largo plazo es una envolvente de las curvas de costos promedio a corto plazo (¿qué parámetro varía a lo largo de la envolvente en este caso?). Uno de los ejercicios que haremos en esta sección será una derivación formal de este concepto.

Para ejemplificar, considere el siguiente problema de maximización sin restricciones con dos variables de elección x y y y un parámetro ϕ:

$$\text{Maximizar} \qquad U = f(x, y, \phi) \tag{13.27}$$

La condición necesaria de primer orden es

$$f_x(x, y, \phi) = f_y(x, y, \phi) = 0 \tag{13.28}$$

Si se cumplen las condiciones de segundo orden, estas dos ecuaciones definen implícitamente las soluciones

$$x^* = x^*(\phi) \qquad y^* = y^*(\phi) \tag{13.29}$$

Si sustituimos estas soluciones en la función objetivo, obtenemos una nueva función

$$V(\phi) = f(x^*(\phi),\ y^*(\phi), \phi) \tag{13.30}$$

donde esta función es el valor de f cuando los valores de x y y son los que maximizan a $f(x, y, \phi)$. Por lo tanto, $V(\phi)$ es la *función de valor máximo* (o función objetivo indirecta).

[4] Esta sección del capítulo presenta un panorama del teorema de la envolvente. Un tratamiento más profundo de este tema se encuentra en el capítulo 7 de *The Structure of Economics: A Mathematical Analysis* (3a. ed.) de Eugene Silberberg y Wing Suen (McGraw-Hill, 2001) en el cual se basan partes de esta sección.

Si diferenciamos V respecto a ϕ, su único argumento, obtenemos

$$\frac{dV}{d\phi} = f_x \frac{\partial x^*}{\partial \phi} + f_y \frac{\partial y^*}{\partial \phi} + f_\phi \tag{13.31}$$

Sin embargo, a partir de las condiciones de primer orden sabemos que $f_x = f_y = 0$. Por lo tanto, los dos primeros términos desaparecen y el resultado se transforma en

$$\frac{dV}{d\phi} = f_\phi \tag{13.31$'$}$$

Este resultado dice que, para el óptimo, a medida que ϕ varía, permitiendo el ajuste de x^* y y^*, la derivada $dV/d\phi$ da el mismo resultado que si x^* y y^* se trataran como constantes. Observe que ϕ interviene en la función de valor máximo (13.30) en tres ocasiones: una directa y dos indirectas (a través de x^* y y^*). La ecuación (13.31′) muestra que, para el óptimo, solamente importa el efecto directo de ϕ sobre la función objetivo. Ésta es la esencia del teorema de la envolvente. Este teorema dice que solamente es necesario considerar los efectos directos de un cambio en una variable exógena, aun cuando la variable exógena también puede intervenir indirectamente en la función de valor máximo como parte de la solución para las variables de elección endógenas.

La función de ganancia

Apliquemos ahora la noción de la función de valor máximo para derivar la función de ganancia de una empresa competitiva. Tome en cuenta el caso en el cual una compañía usa dos insumos: el capital K y la mano de obra L. La función de ganancia es

$$\pi = Pf(K, L) - wL - rK \tag{13.32}$$

donde P es el precio del producto y w y r son la tasa de salario y la tasa de alquiler, respectivamente.

Las condiciones de primer orden son

$$\begin{aligned} \pi_L &= Pf_L(K, L) - w = 0 \\ \pi_K &= Pf_K(K, L) - r = 0 \end{aligned} \tag{13.33}$$

las cuales definen, respectivamente, a las ecuaciones de insumo-demanda

$$\begin{aligned} L^* &= L^*(w, r, P) \\ K^* &= K^*(w, r, P) \end{aligned} \tag{13.34}$$

La sustitución de las soluciones K^* y L^* en la función objetivo nos da

$$\pi^*(w, r, P) = Pf(K^*, L^*) - wL^* - rK^* \tag{13.35}$$

donde $\pi^*(w, r, P)$ es la *función de ganancia* (una función objetivo indirecta). La función de ganancia da la ganancia máxima como una función de las variables exógenas w, r y P.

Ahora, considere el efecto de un cambio en w de las ganancias de la empresa. Si diferenciamos la función de ganancia original (13.32) respecto a w, manteniendo constantes las demás variables, obtenemos

$$\frac{\partial \pi}{\partial w} = -L \tag{13.36}$$

Sin embargo, este resultado no considera la capacidad de la empresa de maximización de ganancias para hacer una sustitución del capital por la mano de obra y ajustar el nivel de producto de acuerdo con el comportamiento de maximización de la ganancia.

En contraste, como $\pi^*(w, r, P)$ es el valor máximo de las ganancias para cualquier valor de w, r y P, los cambios de π^* debidos a un cambio en w consideran todas las sustituciones del capital por la mano de obra. Para evaluar un cambio en la función de ganancia máxima causado por un cambio en w, diferenciamos $\pi^*(w, r, P)$ respecto a w para obtener

$$\frac{\partial \pi^*}{\partial w} = (Pf_L - w)\frac{\partial L^*}{\partial w} + (Pf_K - r)\frac{\partial K^*}{\partial w} - L^* \qquad \textbf{(13.37)}$$

A partir de las condiciones de primer orden (13.33), los dos términos que están entre paréntesis son iguales a cero. Por lo tanto, la ecuación se transforma en

$$\frac{\partial \pi^*}{\partial w} = -L^*(w, r, P) \qquad \textbf{(13.38)}$$

Este resultado dice que, para la maximización de ganancia, un cambio en las ganancias respecto a un cambio en la tasa de salarios es el mismo independientemente de que los factores se mantengan constantes o se les permita variar a medida que cambia el precio del factor. En ese caso, (13.38) muestra que la derivada de la función de ganancia respecto a w es el negativo de la función $L^*(w, r, P)$ de la demanda de factores. Siguiendo el procedimiento anterior, también podemos mostrar los resultados adicionales de estática comparativa:

$$\frac{\partial \pi^*(w, r, P)}{\partial r} = -K^*(w, r, P) \qquad \textbf{(13.39)}$$

y

$$\frac{\partial \pi^*(w, r, P)}{\partial P} = f(K^*, L^*) \qquad \textbf{(13.40)}$$

Las ecuaciones (13.38), (13.39) y (13.40) se conocen colectivamente como *lema de Hotelling*. Estas derivadas estático-comparativas de la función de ganancia las hemos obtenido permitiendo que K^* y L^* se ajusten a cualquier cambio en los parámetros. Pero es fácil darse cuenta de que van a surgir los mismos resultados si diferenciamos la función de ganancia (13.35) respecto a cada parámetro mientras que K^* y L^* se mantienen constantes. Entonces, el lema de Hotelling es simplemente otra manifestación del teorema de la envolvente que encontramos anteriormente en (13.31′).

La condición de reciprocidad

Considere nuevamente el problema de maximización sin restricciones de dos variables

$$\text{Maximizar} \qquad U = f(x, y, \phi) \qquad [\text{de } (13.27)]$$

donde x y y son las variables de elección y ϕ es un parámetro. Las condiciones de primer orden son $f_x = f_y = 0$, lo que implica que $x^* = x^*(\phi)$ y $y^* = y^*(\phi)$.

Nos interesa la estática comparativa respecto a las direcciones de cambio de $x^*(\phi)$ y $y^*(\phi)$ a medida que cambia ϕ y los efectos sobre la función de valor. La función de valor máximo es

$$V(\phi) = f(x^*(\phi), y^*(\phi), \phi) \qquad \textbf{(13.41)}$$

Por definición, $V(\phi)$ da el valor máximo de f para cualquier ϕ dado.

Ahora, considere una nueva función que describa la diferencia entre el valor real y el valor máximo de U:

$$\Omega(x, y, \phi) = f(x, y, \phi) - V(\phi) \qquad \textbf{(13.42)}$$

Esta nueva función Ω tiene un valor máximo de cero cuando $x = x^*$ y $y = y^*$; para cualquier $x \neq x^*$, $y \neq y^*$ tenemos $f \leq V$. En este marco de referencia $\Omega(x, y, \phi)$ puede considerarse una función de tres variables independientes x, y y ϕ. El máximo de $\Omega(x, y, \phi) = f(x, y, \phi) - V(\phi)$ puede determinarse a través de las condiciones de primer y de segundo orden.

Las condiciones de primer orden son

$$\begin{aligned} \Omega_x(x, y, \phi) &= f_x = 0 \\ \Omega_y(x, y, \phi) &= f_y = 0 \end{aligned} \qquad \textbf{(13.43)}$$

y

$$\Omega_\phi(x, y, \phi) = f_\phi - V_\phi = 0 \qquad \textbf{(13.44)}$$

Podemos ver que las condiciones de primer orden de nuestra nueva función Ω de (13.43) son nada menos que las condiciones máximas originales para $f(x, y, \phi)$ en (13.28), mientras que la condición de (13.44) realmente se reduce al teorema de la envolvente (13.31′). Estas condiciones de primer orden son válidas siempre que $x = x^*(\phi)$ y $y = y^*(\phi)$. Las condiciones suficientes de segundo orden se satisfacen si el hessiano de Ω

$$H = \begin{vmatrix} f_{xx} & f_{xy} & f_{x\phi} \\ f_{yx} & f_{yy} & f_{y\phi} \\ f_{\phi x} & f_{\phi y} & f_{\phi\phi} - V_{\phi\phi} \end{vmatrix}$$

se caracteriza por

$$f_{xx} < 0 \qquad f_{xx}f_{yy} - f_{xy}^2 > 0 \qquad H < 0$$

Al obtener el hessiano anterior, enumeramos las variables en el orden (x, y, ϕ) y en consecuencia el primer elemento en las condiciones de segundo orden $(\Omega_{xx} =)\ f_{xx} < 0$ se relaciona con la variable x. Si hubiéramos adoptado un orden alterno de listado, entonces el primer elemento podría haber sido $\Omega_{yy} = f_{yy} < 0$, o sea,

$$\Omega_{\phi\phi} = f_{\phi\phi} - V_{\phi\phi} < 0 \qquad \textbf{(13.45)}$$

(13.45) puede conducirnos a un resultado que ofrece una forma rápida para alcanzar una conclusión estática comparativa. Primero, sabemos de (13.41) que

$$V_\phi(\phi) = f_\phi(x^*(\phi), y^*(\phi), \phi)$$

Diferenciando ambos lados respecto a ϕ nos da

$$V_{\phi\phi} = f_{\phi x}\frac{\partial x^*}{\partial \phi} + f_{\phi y}\frac{\partial y^*}{\partial \phi} + f_{\phi\phi} \qquad \textbf{(13.46)}$$

Con el uso de (13.45) y el teorema de Young podemos escribir

$$V_{\phi\phi} - f_{\phi\phi} = f_{x\phi}\frac{\partial x^*}{\partial \phi} + f_{y\phi}\frac{\partial y^*}{\partial \phi} > 0 \qquad \textbf{(13.47)}$$

Suponga que ϕ interviene solamente en la condición de primer orden para x, tal que $f_{y\phi} = 0$. Entonces, (13.47) se reduce a

$$f_{x\phi}\frac{\partial x^*}{\partial \phi} > 0 \qquad \textbf{(13.48)}$$

lo que implica que $f_{x\phi}$ y $\partial x^*/\partial\phi$ tendrán el mismo signo. De esta manera, siempre que veamos que el parámetro ϕ aparece solamente en la condición de primer orden relacionada con x, y una vez que hemos determinado el signo de la derivada $f_{x\phi}$ a partir de la función objetivo

$U = f(x, y, \phi)$, podemos deducir inmediatamente el signo de la derivada estática comparativa $\partial x^*/\partial\phi$ sin mayor problema.

Por ejemplo, en el modelo de maximización de ganancia:

$$\pi = Pf(K, L) - wL - rK$$

donde las condiciones de primer orden son

$$\pi_L = Pf_L - w = 0$$
$$\pi_K = Pf_K - r = 0$$

la variable exógena w interviene solamente en la condición de primer orden $Pf_L - w = 0$, con

$$\frac{\partial \pi_L}{\partial w} = -1$$

Por lo tanto, mediante (13.48) podemos concluir que $\partial L^*/\partial w$ también será negativa.

Aún más, si combinamos el teorema de la envolvente con el teorema de Young, podemos derivar una relación conocida como *condición de reciprocidad*: $\partial L^*/\partial r = \partial K^*/\partial w$. A partir de la función de ganancia indirecta $\pi^*(w, r, P)$, el lema de Hotelling nos da

$$\pi_w^* = \frac{\partial \pi^*}{\partial w} = -L^*(w, r, P)$$
$$\pi_r^* = \frac{\partial \pi^*}{\partial r} = -K^*(w, r, P)$$

Diferenciando nuevamente y aplicando el teorema de Young, tenemos

$$\pi_{wr}^* = -\frac{\partial L^*}{\partial r} = -\frac{\partial K^*}{\partial w} = \pi_{rw}^*$$

o sea

$$\frac{\partial L^*}{\partial r} = \frac{\partial K^*}{\partial w} \qquad \textbf{(13.49)}$$

Este resultado se denomina *condición de reciprocidad,* porque muestra la simetría entre el efecto cruzado estático comparativo producido por el precio de un insumo sobre la demanda para el "otro" insumo. Específicamente, en el sentido estático-comparativo, el efecto de r (la tasa de alquiler para el capital K) sobre la demanda óptima para la mano de obra L es el mismo que el efecto de w (la tasa de salarios para la mano de obra L) sobre la demanda óptima para el capital K.

El teorema de la envolvente para la optimización restringida

El teorema de la envolvente también puede derivarse para el caso de la optimización restringida. Nuevamente tenemos una función objetivo (U), dos variables de elección (x y y) y un parámetro (ϕ), pero ahora introducimos la siguiente restricción:

$$g(x, y; \phi) = 0$$

El problema se transforma en:

$$\begin{aligned} \text{Maximizar} \quad & U = f(x, y; \phi) \\ \text{sujeto a} \quad & g(x, y; \phi) = 0 \end{aligned} \qquad \textbf{(13.50)}$$

La función lagrangeana para este problema es

$$Z = f(x, y; \phi) + \lambda[0 - g(x, y; \phi)] \qquad \textbf{(13.51)}$$

con las condiciones de primer orden

$$Z_x = f_x - \lambda g_x = 0$$
$$Z_y = f_y - \lambda g_y = 0$$
$$Z_\lambda = -g(x, y; \phi) = 0$$

La solución de este sistema de ecuaciones nos da

$$x = x^*(\phi) \qquad y = y^*(\phi) \qquad \lambda = \lambda^*(\phi)$$

De la sustitución de las soluciones en la función objetivo, obtenemos

$$U^* = f(x^*(\phi), y^*(\phi), \phi) = V(\phi) \tag{13.52}$$

donde $V(\phi)$ es la función objetivo indirecta, una función de valor máximo. Éste es el valor máximo de y para cualquier ϕ y x_i que satisfaga la restricción.

¿Cómo cambia $V(\phi)$ a medida que cambia ϕ? Primero diferenciamos V respecto a ϕ:

$$\frac{dV}{d\phi} = f_x \frac{\partial x^*}{\partial \phi} + f_y \frac{\partial y^*}{\partial \phi} + f_\phi \tag{13.53}$$

Sin embargo, en este caso (13.53) no se va a simplificar a $dV/d\phi = f_\phi$ ya que en la optimización restringida no es necesario tener $f_x = f_y = 0$ (ver la tabla 12.1). No obstante, si sustituimos las soluciones para x y y en la restricción (produciendo una identidad), obtenemos

$$g(x^*(\phi), y^*(\phi), \phi) \equiv 0$$

diferenciando esto respecto a ϕ nos da

$$g_x \frac{\partial x^*}{\partial \phi} + g_y \frac{\partial y^*}{\partial \phi} + g_\phi \equiv 0 \tag{13.54}$$

Si multiplicamos (13.54) por λ, combinamos el resultado con (13.53), y ordenamos los términos, obtenemos

$$\frac{dV}{d\phi} = (f_x - \lambda g_x)\frac{\partial x^*}{\partial \phi} + (f_y - \lambda g_y)\frac{\partial y^*}{\partial \phi} + f_\phi - \lambda g_\phi = Z_\phi \tag{13.55}$$

donde Z_ϕ es la derivada parcial de la función lagrangeana respecto a ϕ, manteniendo constantes las demás variables. Este resultado tiene la misma intención que (13.31), y en virtud de las condiciones de primer orden se reduce a

$$\frac{dV}{d\phi} = Z_\phi \tag{13.55'}$$

que representa el teorema de la envolvente en el marco de referencia de la optimización restringida. Observe, sin embargo, que en el presente caso la función lagrangeana reemplaza a la función objetivo en la derivación de la función objetivo indirecta.

Aun cuando los resultados de (13.55) se asemejan mucho al caso no restringido, es importante observar que algunos de los resultados estático-comparativos dependen en forma crítica de si los parámetros intervienen solamente en la función objetivo, o solamente en las restricciones, o en ambas. Si un parámetro interviene solamente en la función objetivo, entonces los resultados estático-comparativos son los mismos que para el caso no restringido. Sin embargo, si el parámetro interviene en la restricción, la relación

$$V_{\phi\phi} \geq f_{\phi\phi}$$

ya no es válida.

Interpretación del multiplicador de Lagrange

En el problema de la elección del consumidor del capítulo 12 obtuvimos el resultado de que el multiplicador de Lagrange λ representaba el cambio de valor de la función de Lagrange cuando cambia el presupuesto del consumidor. Interpretamos a λ como la utilidad marginal del ingreso. Ahora vamos a obtener una interpretación más general del multiplicador de Lagrange con la ayuda del teorema de la envolvente. Considere el problema

$$\begin{array}{ll} \text{Maximizar} & U = f(x, y) \\ \text{sujeto a} & g(x, y) = c \end{array}$$

donde c es una constante. La lagrangeana para este problema es

$$Z = f(x, y) + \lambda[c - g(x, y)] \tag{13.56}$$

Las condiciones de primer orden son

$$\begin{aligned} Z_x &= f_x(x, y) - \lambda g_x(x, y) = 0 \\ Z_y &= f_y(x, y) - \lambda g_y(x, y) = 0 \\ Z_\lambda &= c - g(x, y) = 0 \end{aligned} \tag{13.57}$$

A partir de las dos primeras ecuaciones de (13.57), obtenemos

$$\lambda = \frac{f_x}{g_x} = \frac{f_y}{g_y} \tag{13.58}$$

lo que nos da la condición de que la pendiente de la curva de nivel (curva de indiferencia) de la función objetivo debe ser igual a la pendiente de la restricción para el óptimo.

Las ecuaciones (13.57) definen implícitamente las soluciones

$$x^* = x^*(c) \qquad y^* = y^*(c) \qquad \lambda^* = \lambda^*(c) \tag{13.59}$$

Al sustituir (13.59) en la lagrangeana obtenemos la función de valor máximo,

$$V(c) = Z^*(c) = f(x^*(c), y^*(c)) + \lambda^*(c)[c - g(x_1^*(c), y^*(c))] \tag{13.60}$$

Al diferenciar respecto a c obtenemos

$$\begin{aligned} \frac{dV}{dc} = \frac{dZ^*}{dc} = f_x \frac{\partial x^*}{\partial c} + f_y \frac{\partial y^*}{\partial c} + [c - g(x^*(c), y^*(c))] \frac{\partial \lambda^*}{\partial c} \\ - \lambda^*(c) g_x \frac{\partial x^*}{\partial c} - \lambda^*(c) g_y \frac{\partial y^*}{\partial c} + \lambda^*(c) \frac{dc}{dc} \end{aligned}$$

Después de reordenar, obtenemos

$$\frac{dZ^*}{dc} = [f_x - \lambda^* g_x] \frac{\partial x^*}{\partial c} + [f_y - \lambda^* g_y] \frac{\partial y^*}{\partial c} + [c - g(x^*, y^*)] \frac{\partial \lambda^*}{\partial c} + \lambda^*$$

Mediante (13.57), los tres términos que están entre corchetes son iguales a cero. Por lo tanto, esta expresión se simplifica a

$$\frac{dV}{dc} = \frac{dZ^*}{dc} = \lambda^* \tag{13.61}$$

lo que demuestra que el valor óptimo λ^* mide la tasa de cambio del valor máximo de la función objetivo cuando cambia c, y por esta razón se le denomina el "precio sombra" de c.

Observe que en este caso, c interviene en el problema solamente a través de la restricción; no es un argumento de la función objetivo original.

13.6 La dualidad y el teorema de la envolvente

La función de desembolso de un consumidor y su función de utilidad indirecta ejemplifican las funciones de valor mínimo y máximo para los *problemas duales*.[5] Una función de desembolso especifica el desembolso mínimo requerido para obtener un nivel fijo de utilidad dados los precios de los bienes de consumo y la función de utilidad. Una función de utilidad indirecta especifica la utilidad máxima que puede obtenerse dados los precios, el ingreso y la función de utilidad.

El problema primal

Sea $U(x, y)$ una función de utilidad donde x y y son bienes de consumo. El consumidor tiene un presupuesto B y enfrenta precios de mercado P_x y P_y para los bienes x y y, respectivamente. Éste será considerado como el *problema primal*:

$$\begin{array}{lll} \text{Maximizar} & U = U(x, y) & \\ \text{sujeto a} & P_x x + P_y y = B & \end{array} \quad \text{[primal]} \qquad \textbf{(13.62)}$$

Para este problema, tenemos la lagrangeana conocida

$$Z = U(x, y) + \lambda(B - P_x x - P_y y)$$

Las condiciones de primer orden son

$$\begin{aligned} Z_x &= U_x - \lambda P_x = 0 \\ Z_y &= U_y - \lambda P_y = 0 \\ Z_\lambda &= B - P_x x - P_y y = 0 \end{aligned} \qquad \textbf{(13.63)}$$

Este sistema de ecuaciones define implícitamente una solución para x^m, y^m y λ^m como una función de las variables exógenas B, P_x, P_y:

$$\begin{aligned} x^m &= x^m(P_x, P_y, B) \\ y^m &= y^m(P_x, P_y, B) \\ \lambda^m &= \lambda^m(P_x, P_y, B) \end{aligned}$$

Las soluciones x^m y y^m son las funciones de demanda ordinaria del consumidor, llamadas algunas veces las funciones de demanda "marshallianas", de ahí el superíndice m.

Al sustituir las soluciones x^m y y^m en la función de utilidad obtenemos

$$U^* = U^*(x^m(P_x, P_y, B), y^m(P_x, P_y, B)) \equiv V(P_x, P_y, B) \qquad \textbf{(13.64)}$$

donde V es la función de utilidad directa, una función de valor máximo que muestra la utilidad máxima alcanzable en el problema (13.62). Regresaremos posteriormente a esta función.

[5] La dualidad en la teoría económica es la relación entre dos problemas de optimización restringida. Si uno de los problemas requiere la maximización restringida, el otro va a requerir la minimización restringida. La estructura y la solución de cada uno de los problemas pueden proveer información acerca de la estructura y la solución del otro problema.

El problema dual

Consideremos ahora un *problema dual* para el consumidor relacionado con el objetivo de minimizar el desembolso en x y y mientras que se mantiene un nivel fijo de utilidad U^* obtenido de (13.64) del problema primal:

$$\begin{array}{lll} \text{Minimizar} & E = P_x x + P_y y & \text{[dual]} \\ \text{sujeto a} & U(x, y) = U^* & \end{array} \tag{13.65}$$

Su lagrangeana es

$$Z^d = P_x x + P_y y + \mu\,[U^* - U(x, y)]$$

y las condiciones de primer orden son

$$\begin{aligned} Z_x^d &= P_x - \mu U_x = 0 \\ Z_y^d &= P_y - \mu U_y = 0 \\ Z_\lambda^d &= U^* - U(x, y) = 0 \end{aligned} \tag{13.66}$$

Este sistema de ecuaciones define implícitamente un conjunto de valores solución que debe etiquetarse como x^h, y^h y λ^h:

$$\begin{aligned} x^h &= x^h(P_x, P_y, U^*) \\ y^h &= y^h(P_x, P_y, U^*) \\ \mu^h &= \mu^h(P_x, P_y, U^*) \end{aligned}$$

Aquí x^h y y^h son las funciones de demanda compensadas (el "ingreso real" se mantiene constante). Normalmente se denominan funciones de demanda "hicksianas", de ahí el superíndice h.

Al sustituir x^h y y^h en la función objetivo del problema dual obtenemos

$$P_x x^h(P_x, P_y, U^*) + P_y y^h(P_x, P_y, U^*) \equiv E(P_x, P_y, U^*) \tag{13.67}$$

donde E es la función de desembolso, una función de valor mínimo que muestra el desembolso mínimo necesario para alcanzar el nivel de utilidad U^*.

Dualidad

Si tomamos las dos primeras ecuaciones de (13.63) y (13.64) y eliminamos los multiplicadores de Lagrange, podemos escribir

$$\frac{Px}{Py} = \frac{Ux}{Uy} \tag{13.68}$$

Ésta es la condición de tangencia para la cual el consumidor escoge el punto óptimo donde la pendiente de la curva de indiferencia es igual a la pendiente de la restricción presupuestaria. La condición de tangencia es idéntica para ambos problemas. De esta manera, cuando el nivel objetivo de utilidad en el problema de minimización se hace igual al valor U^* obtenido del problema de maximización, obtenemos

$$\begin{aligned} x^m(P_x, P_y, B) &= x^h(P_x, P_y, U^*) \\ y^m(P_x, P_y, B) &= y^h(P_x, P_y, U^*) \end{aligned} \tag{13.69}$$

es decir, las soluciones tanto para el problema de maximización como para el problema de minimización producen valores idénticos para x y y. Sin embargo, las soluciones son funciones de diferentes variables exógenas, de modo que los ejercicios estático-comparativos generalmente producirían resultados diferentes.

El hecho de que los valores solución para x y y en los problemas primal y dual se determinan por el punto de tangencia de la misma curva de indiferencia y la línea de restricción presupuestaria, significa que el desembolso minimizado en el problema dual es igual al presupuesto B dado del problema primal:

$$E(P_x, P_y, U^*) = B \tag{13.70}$$

Este resultado es paralelo al resultado de (13.64), lo que revela que el valor máximo de la utilidad V en el problema primal es igual al nivel de la función objetivo que es la función de de utilidad U^* dado en el problema dual.

Aun cuando los valores solución de x y y son idénticos en los dos problemas, no se puede decir lo mismo sobre los multiplicadores de Lagrange. De la primera ecuación de (13.63) y (13.66), podemos calcular $\lambda = U_x/P_x$, pero $\mu = P_x/U_x$. Entonces, los valores solución de λ y μ son recíprocos:

$$\lambda = \frac{1}{\mu} \qquad \text{o} \qquad \lambda^m = \frac{1}{\mu^h} \tag{13.71}$$

La identidad de Roy

Una aplicación del teorema de la envolvente es la obtención de la identidad de Roy. Ésta establece que la función de demanda marshalliana individual del consumidor es igual al negativo de la razón de dos derivadas parciales de la función de valor máximo.

La sustitución de los valores óptimos x^m, y^m y λ^m en la lagrangeana de (13.62) nos da

$$V(P_x, P_y, B) = U(x^m, y^m) + \lambda^m\left(B - P_x x^m - P_y y^m\right) \tag{13.72}$$

Cuando diferenciamos (13.72) respecto a P_x encontramos

$$\begin{aligned}\frac{\partial V}{\partial P_x} &= (U_x - \lambda^m P_x)\frac{\partial x^m}{\partial P_x} + (U_y - \lambda^m P_y)\frac{\partial y^m}{\partial P_x} \\ &\quad + (B - P_x x^m - P_y y^m)\frac{\partial \lambda^m}{\partial P_x} - \lambda^m x^m\end{aligned}$$

Para el óptimo, las condiciones de primer orden (13.63) nos permiten simplificar esto a

$$\frac{\partial V}{\partial P_x} = -\lambda^m x^m$$

Enseguida, diferenciamos la función valor respecto a B para obtener

$$\begin{aligned}\frac{\partial V}{\partial B} &= (U_x - \lambda^m P_x)\frac{\partial x^m}{\partial B} + (U_y - \lambda^m P_y)\frac{\partial y^m}{\partial B} \\ &\quad + (B - P_x x^m - P_y y^m)\frac{\partial \lambda^m}{\partial B} + \lambda^m\end{aligned}$$

Nuevamente para el óptimo, (13.63) nos permite simplificar esto a

$$\frac{\partial V}{\partial B} = \lambda^m$$

Al tomar la razón de estas dos derivadas parciales, encontramos que

$$\frac{\partial V/\partial P_x}{\partial V/\partial B} = -x^m \tag{13.73}$$

Este resultado, conocido como la *identidad de Roy*, muestra que la demanda marshalliana para el artículo x es el negativo de la razón de dos derivadas parciales de la función de valor máximo V respecto a P_x y B, respectivamente. En vista de la simetría entre x y y en el problema, también podemos escribir un resultado similar a (13.73) para y^m, la demanda marshalliana para y. Por supuesto, se pudo haber llegado a este resultado aplicando directamente el teorema de la envolvente.

El lema de Shephard

En la sección 13.5 obtuvimos el lema de Hotelling, que establece que las derivadas parciales del valor máximo de la función de ganancia suministran las funciones de demanda de insumos de la compañía y las funciones de oferta. Un enfoque similar aplicado a la función de desembolso proporciona el lema de Shephard.

Consideremos el problema de minimización del consumidor (13.65). La lagrangeana es

$$Z^d = P_x x + P_y y + \mu[U^* - U(x, y)]$$

A partir de las condiciones de primer orden, se definen implícitamente las siguientes soluciones

$$\begin{aligned} x^h &= x^h(P_x, P_y, U^*) \\ y^h &= y^h(P_x, P_y, U^*) \\ \mu^h &= \mu^h(P_x, P_y, U^*) \end{aligned}$$

La sustitución de estas soluciones en la lagrangeana conduce a la función de desembolso:

$$E(P_x, P_y, U^*) = P_x x^h + P_y y^h + \mu^h[U^* - U(x^h, y^h)]$$

Al tomar las derivadas parciales de esta función respecto a P_x y P_y y al evaluarlas para el óptimo, encontramos que $\partial E/\partial P_x$ y $\partial E/\partial P_y$ representan las demandas hicksianas del consumidor:

$$\begin{aligned} \frac{\partial E}{\partial P_x} &= (P_x - \mu^h U_x)\frac{\partial x^h}{\partial P_x} + (P_y - \mu^h U_y)\frac{\partial y^h}{\partial P_x} + [U^* - U(x^h, y^h)]\frac{\partial \mu^h}{\partial P_x} + x^h \\ &= (0)\frac{\partial x^h}{\partial P_x} + (0)\frac{\partial y^h}{\partial P_x} + (0)\frac{\partial \mu^h}{\partial P_x} + x^h = x^h \end{aligned} \tag{13.74}$$

y

$$\begin{aligned} \frac{\partial E}{\partial P_y} &= (P_x - \mu^h U_x)\frac{\partial x^h}{\partial P_y} + (P_y - \mu^h U_y)\frac{\partial y^h}{\partial P_y} + [U^* - U(x^h, y^h)]\frac{\partial \mu^h}{\partial P_y} + y^h \\ &= (0)\frac{\partial x^h}{\partial P_y} + (0)\frac{\partial y^h}{\partial P_y} + (0)\frac{\partial \mu^h}{\partial P_y} + y^h = y^h \end{aligned} \tag{13.74'}$$

Finalmente, al diferenciar E respecto a la restricción U^* nos da μ^h, el costo marginal de la restricción

$$\begin{aligned} \frac{\partial E}{\partial U^*} &= (P_x - \mu^h U_x)\frac{\partial x^h}{\partial U^*} + (P_y - \mu^h U_y)\frac{\partial y^h}{\partial P_y} \\ &\quad + [U^* - U(x^h, y^h)]\frac{\partial \mu^h}{\partial U^*} + \mu^h \\ &= (0)\frac{\partial x^h}{\partial U^*} + (0)\frac{\partial y^h}{\partial U^*} + (0)\frac{\partial \mu^h}{\partial U^*} + \mu^h = \mu^h \end{aligned} \tag{13.74''}$$

Juntas, las tres derivadas parciales (13.74), (13.74′) y (13.74″) se denominan *lema de Shephard.*

Ejemplo 1

Consideremos un consumidor con la función utilidad $U = xy$, quien enfrenta una restricción presupuestal de B y se le dan precios P_x y P_y.

El problema de elección es

$$\begin{aligned} &\text{Maximizar} && U = xy \\ &\text{sujeto a} && P_x x + P_y y = B \end{aligned}$$

La lagrangeana para este problema es

$$Z = xy + \lambda(B - P_x x - P_y y)$$

Las condiciones de primer orden son

$$\begin{aligned} Z_x &= y - \lambda P_x = 0 \\ Z_y &= x - \lambda P_y = 0 \\ Z_\lambda &= B - P_x x - P_y y = 0 \end{aligned}$$

La solución de las condiciones de primer orden aporta las siguientes soluciones:

$$x^m = \frac{B}{2P_x} \qquad y^m = \frac{B}{2P_y} \qquad \lambda^m = \frac{B}{2P_x P_y}$$

donde x^m y y^m son las funciones demanda marshalliana del consumidor. Para la condición de segundo orden, ya que el hessiano orlado es

$$|\overline{H}| = \begin{vmatrix} 0 & 1 & -P_x \\ 1 & 0 & -P_y \\ -P_x & -P_y & 0 \end{vmatrix} = 2P_x P_y > 0$$

La solución, ciertamente, representa un máximo.[6]

Ahora podemos deducir la función utilidad indirecta para este problema al sustituir x^m y y^m en la función utilidad:

$$V(P_x, P_y, B) = \left(\frac{B}{2P_x}\right)\left(\frac{B}{2P_y}\right) = \frac{B^2}{4P_x P_y} \qquad \textbf{(13.75)}$$

donde V se refiere a la utilidad maximizada. Puesto que V representa la utilidad maximizada, podemos hacer $V = U^*$ en (13.75) para obtener $B^2/4P_x P_y = U^*$, y luego reordenar términos para expresar a B como

$$B = (4P_x P_y U^*)^{1/2} = 2P_x^{1/2} P_y^{1/2} U^{*1/2}$$

Ahora, piense en el problema dual del consumidor que se refiere a la minimización de desembolso. En el problema dual, la función E de desembolso mínimo debe ser igual a la cantidad presupuestaria dada B del problema primal. Por lo tanto, de la ecuación anterior podemos concluir inmediatamente que

$$E(P_x, P_y, U^*) = B = 2P_x^{1/2} P_y^{1/2} U^{*1/2} \qquad \textbf{(13.76)}$$

[6] Tome en cuenta que el hessiano orlado se escribe aquí (y en el ejemplo 2 de la página 440) con las orlas en el tercer renglón y la tercera columna, en lugar de en el primer renglón y la primera columna, como en (12.19). Éste es el resultado de enumerar al multiplicador de Lagrange como la última en lugar de la primera variable, como lo hicimos en capítulos anteriores. El ejercicio 12.3-3 muestra que las dos expresiones alternas para el hessiano orlado son transformables entre sí mediante operaciones elementales de los renglones sin afectar su valor. Sin embargo, cuando aparecen en un problema más de dos variables de elección, es preferible usar el formato (12.19), porque eso facilita la escritura de los menores principales avanzados orlados.

Usemos ahora este ejemplo para verificar la identidad de Roy (13.73)

$$x^m = -\frac{\partial V/\partial P_x}{\partial V/\partial B}$$

Tomando las derivadas parciales relevantes de V, encontramos

$$\frac{\partial V}{\partial P_x} = -\frac{B^2}{4P_x^2 P_y}$$

y

$$\frac{\partial V}{\partial B} = \frac{B}{2P_x P_y}$$

El negativo de la razón de estas dos parciales es

$$-\frac{\dfrac{\partial V}{\partial P_x}}{\dfrac{\partial V}{\partial B}} = -\frac{\left(-\dfrac{B^2}{4P_x^2 P_y}\right)}{\left(\dfrac{B}{2P_x P_y}\right)} = \frac{B}{2P_x} = x^m$$

Entonces, encontramos que es válida la identidad de Roy.

Ejemplo 2

Consideremos ahora el problema dual de la minimización de costos dado un nivel fijo de utilidad relacionado con el ejemplo 1. Haciendo que U^* denote al nivel objetivo de utilidad, el problema es:

$$\begin{array}{ll} \text{Minimizar} & P_x x + P_y y \\ \text{sujeto a} & xy = U^* \end{array}$$

La lagrangeana para el problema es

$$Z^d = P_x x + P_y y + \mu(U^* - xy)$$

Las condiciones de primer orden son

$$\begin{aligned} Z_x^d &= P_x - \mu y = 0 \\ Z_y^d &= P_y - \mu x = 0 \\ Z_\mu^d &= U^* - xy = 0 \end{aligned}$$

La solución del sistema de ecuaciones para x, y y μ nos da

$$\begin{aligned} x^h &= \left(\frac{P_y U^*}{P_x}\right)^{\frac{1}{2}} \\ y^h &= \left(\frac{P_x U^*}{P_y}\right)^{\frac{1}{2}} \\ \mu^h &= \left(\frac{P_x P_y}{U^*}\right)^{\frac{1}{2}} \end{aligned} \tag{13.77}$$

donde x^h y y^h son las funciones de demanda compensadas del consumidor (hicksianas). Al verificar la condición de segundo orden para un mínimo encontramos

$$|\bar{H}| = \begin{vmatrix} 0 & -\mu & -y \\ -\mu & 0 & -x \\ -y & -x & 0 \end{vmatrix} = -2xy\mu < 0$$

Entonces, se cumple la condición suficiente para un mínimo.

La sustitución de x^h y y^h en la función objetivo original nos da la función de valor mínimo, o función de desembolso

$$
\begin{aligned}
E &= P_x x^h + P_y y^h = P_x\left(\frac{P_y U^*}{P_x}\right)^{1/2} + P_y\left(\frac{P_x U^*}{P_y}\right)^{1/2} \\
&= (P_x P_y U^*)^{1/2} + (P_x P_y U^*)^{1/2} \\
&= 2P_x^{1/2} P_y^{1/2} U^{*1/2} \qquad \textbf{(13.76')}
\end{aligned}
$$

Observe que este resultado es idéntico a (13.76) en el ejemplo 1. La única diferencia radica en el proceso usado para obtener el resultado. La ecuación (13.76′) se obtiene directamente de un problema de minimización de desembolso, mientras que (13.76) se deduce indirectamente, vía la relación de dualidad, de un problema de maximización de utilidad.

Usaremos ahora este ejemplo para probar la validez del lema de Shephard (13.74), (13.74′) y (13.74″). Diferenciando la función de desembolso de (13.76′) respecto a P_x, P_y y U^*, respectivamente, y relacionando las derivadas parciales resultantes, encontramos

$$
\begin{aligned}
\frac{\partial E(P_x, P_y, U^*)}{\partial P_x} &= \frac{P_y^{1/2} U^{*1/2}}{P_x^{1/2}} = x^h \\
\frac{\partial E(P_x, P_y, U^*)}{\partial P_y} &= \frac{P_x^{1/2} U^{*1/2}}{P_y^{1/2}} = y^h \\
\frac{\partial E(P_x, P_y, U^*)}{\partial U^*} &= \frac{P_x^{1/2} P_y^{1/2}}{U^{*1/2}} = \mu^h
\end{aligned}
$$

Entonces, el lema de Shephard es válido en este ejemplo.

EJERCICIO 13.6

1. Un consumidor tiene la siguiente función de utilidad: $U(x, y) = x(y + 1)$, donde x y y son cantidades de dos bienes de consumo cuyos precios son P_x y P_y, respectivamente. El consumidor también tiene un presupuesto de B. Por lo tanto, la lagrangeana del consumidor es

   $x(y + 1) + \lambda(B - P_x x - P_y y)$

   (*a*) Partiendo de las condiciones de primer orden, encuentre expresiones para las funciones de demanda. ¿Qué clase de bien es y? En particular, ¿qué ocurre cuando $P_y > B$?

   (*b*) Verifique que éste es un máximo revisando las condiciones de segundo orden. Al sustituir x^* y y^* en la función utilidad, encuentre una expresión para la función utilidad directa

   $U^* = U(P_x, P_y, B)$

   y obtenga una expresión para la función de desembolso

   $E = E(P_x, P_y, U^*)$

   (*c*) Este problema podría replantearse como el siguiente problema dual

   Minimizar $\quad P_x x + P_y y$

   sujeto a $\quad x(y + 1) = U^*$

   Encuentre los valores de x y y que resuelven este problema de minimización y demuestre que los valores de x y y son iguales a las derivadas parciales de la función de desembolso, $\partial E/\partial P_x$ y $\partial E/\partial P_y$, respectivamente.

13.7 Algunas observaciones finales

En esta parte del libro hemos cubierto las técnicas básicas de la optimización. La jornada, un poco ardua, nos ha llevado: (1) desde el caso de una sola variable de elección hasta el caso más general de n variables, (2) desde la función objetivo polinomial hasta la exponencial y logarítmica, y (3) desde la variedad de extremo sin restricciones hasta la restringida.

La mayor parte de esta discusión se basa en los métodos "clásicos" de optimización, con el cálculo diferencial como el pilar, y las derivadas de diferentes órdenes como herramientas primarias. Una debilidad del enfoque de la optimización, desde el punto de vista del cálculo, es su naturaleza esencialmente miope. Mientras que las condiciones de primer y segundo órdenes en términos de derivadas o diferenciales normalmente pueden localizar sin dificultad extremos relativos o locales, con frecuencia se requiere información adicional o investigación adicional para la identificación de extremos absolutos o globales. Nuestro estudio detallado de la concavidad, convexidad, cuasiconcavidad y cuasiconvexidad tiene como objetivo ser una piedra angular útil proveniente del reino de los extremos relativos al de los absolutos.

Una limitación más seria del enfoque del cálculo es su incapacidad para manejar las restricciones en forma de desigualdades. Por esta razón, la restricción presupuestaria en el modelo de maximización de utilidades, por ejemplo, se enuncia en una forma tal que el desembolso total sea exactamente *igual a* (y no "menor o igual que") una suma especificada. En otras palabras, la limitación del enfoque del cálculo hace necesario negar al consumidor la opción de ahorrar parte de los fondos disponibles. Y por la misma razón, el enfoque clásico no nos permite especificar explícitamente que las variables de elección deben ser no negativas, como es apropiado en la mayor parte del análisis económico.

Afortunadamente, nos liberamos de estas limitaciones cuando introducimos la técnica moderna de optimización conocida como programación lineal. Aquí podemos admitir abiertamente las restricciones de desigualdad, incluyendo restricciones no negativas sobre las variables de elección para el problema. Obviamente, esto representa un paso gigantesco hacia delante en el desarrollo de la metodología de la optimización.

Aun en la programación lineal, el marco de referencia analítico todavía permanece estático. El problema y su solución se relacionan únicamente con el estado óptimo para un instante de tiempo y no puede manejar la interrogante de cómo debe comportarse un agente de optimización para un periodo y bajo circunstancias dadas. Esta última interrogante pertenece al reino de la *optimización dinámica*, que no podremos manejar hasta haber aprendido los aspectos básicos del análisis dinámico: el análisis del movimiento de las variables en el tiempo. De hecho, aparte de su aplicación a la optimización dinámica, el análisis dinámico es en sí mismo una rama importante del análisis económico. Por esta razón, dirigiremos nuestra atención al tema del análisis dinámico en la parte 5.

Parte 5

Análisis dinámico

Capítulo 14

La dinámica económica y el cálculo integral

El término *dinámica*, tal como se aplica al análisis económico, ha tenido diferentes significados en distintas épocas y para diversos economistas.[1] Actualmente, este término describe el tipo de análisis cuyo objetivo es rastrear y estudiar las trayectorias específicas en el tiempo, de las variables, así como determinar si, dado un tiempo suficiente, estas variables tenderán a convergir a ciertos valores (equilibrio). Este tipo de información es importante porque llena una brecha considerable que ensombreció nuestro estudio de la estática y a la estática comparativa. En esta última, siempre hacemos la hipótesis arbitraria de que el proceso de ajuste económico conduce inevitablemente a un equilibrio. En un análisis dinámico, el aspecto de la "asequilibidad" debemos abordarlo frontalmente, en lugar de darlo como un hecho.

Una característica sobresaliente del análisis dinámico es la *afectación temporal* de las variables, lo que introduce la consideración explícita del *tiempo* en el escenario. Esto podemos hacerlo de dos maneras: considerar el tiempo como una variable *continua* o como una variable *discreta*. En el primer caso, a la variable le está ocurriendo algo en cada *instante* de tiempo (como en el interés compuesto continuo); en el último caso, por el contrario, la variable experimenta un cambio solamente una vez dentro de un *periodo* (por ejemplo, el interés se añade sólo al final de cada 6 meses). En ciertos contextos, uno de estos conceptos de tiempo puede ser más apropiado que el otro.

Estudiaremos primero el caso del tiempo continuo, para el cual son pertinentes las técnicas matemáticas del *cálculo integral* y las *ecuaciones diferenciales*. Posteriormente, en los capítulos 17 y 18, nos referiremos al caso del tiempo discreto, que utiliza los métodos de las *ecuaciones en diferencias*.

14.1 La dinámica y la integración

En un modelo estático, en términos generales, el problema es encontrar los valores de las variables endógenas que satisfacen alguna(s) condición(es) específicas de equilibrio. Aplicado al contexto de los modelos de optimización, la tarea es encontrar los valores de las variables de elección que maximizan (o minimizan) una función objetivo específica: con la condición de pri-

[1] Fritz Machlup, "Statics and Dynamics: Kaleidoscopic Words", *Southern Economic Journal*, octubre de 1959, pp. 91-110; reimpreso en Machlup, *Essays on Economic Semantics*, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1963, pp. 9-42.

mer orden que sirve como la condición de equilibrio. En contraste, en un modelo dinámico, el problema implica la determinación de la trayectoria en el tiempo de alguna variable, sobre la base de un patrón conocido de cambio (por ejemplo, una tasa instantánea de cambio dada).

Un ejemplo podrá aclarar esto. Suponga que conocemos que el tamaño de la población H cambia en el tiempo de acuerdo con la tasa

$$\frac{dH}{dt} = t^{-1/2} \tag{14.1}$$

entonces, tratamos de encontrar cuál(es) trayectoria(s) de tiempo de la población $H = H(t)$ puede(n) suministrar la tasa de cambio de (14.1).

Usted se dará cuenta de que si para empezar conocemos la función $H = H(t)$, podemos encontrar la derivada dH/dt por diferenciación. Pero en el problema que ahora afrontamos es la situación opuesta: debemos descubrir la función *primitiva* a partir de una función *derivada* dada. Matemáticamente, necesitamos el opuesto exacto del método de diferenciación, o del cálculo diferencial.

El método relevante, conocido como *integración*, o *cálculo integral*, lo estudiaremos en este capítulo. Por el momento, conformémonos con la observación de que la función $H(t) = 2t^{1/2}$ realmente tiene una derivada de la forma (14.1), por lo que en realidad califica como una solución a nuestro problema. La dificultad es que también hay funciones similares, tales como $H(t) = 2t^{1/2} + 15$ o $H(t) = 2t^{1/2} + 99$ o más generalmente

$$H(t) = 2t^{1/2} + c \qquad (c = \text{una constante arbitraria}) \tag{14.2}$$

las cuales tienen exactamente la misma derivada (14.1). Por lo tanto, no podemos determinar una trayectoria de tiempo única, a menos que podamos definir de alguna manera el valor de la constante c. Para lograr esto, debemos introducir información adicional en el modelo, generalmente en la forma que se conoce como una *condición inicial* o *condición de frontera*.

Si conocemos la población inicial $H(0)$ —es decir, el valor de H para $t = 0$, digamos, $H(0) = 100$— entonces podemos determinar el valor de la constante c. Haciendo $t = 0$ en (14.2), obtenemos

$$H(0) = 2(0)^{1/2} + c = c$$

Pero si $H(0) = 100$, entonces $c = 100$, y (14.2) se transforma en

$$H(t) = 2t^{1/2} + 100 \tag{14.2'}$$

donde la constante ya no es arbitraria. Más frecuentemente, para cualquier población inicial dada $H(0)$, la trayectoria de tiempo será

$$H(t) = 2t^{1/2} + H(0) \tag{14.2''}$$

Entonces, el tamaño de la población H para cualquier instante de tiempo consistirá, en el presente ejemplo, en la suma de la población inicial $H(0)$ más otro término que involucre a la variable de tiempo t. Una trayectoria de tiempo de este tipo realmente describe el itinerario completo de la variable H respecto al tiempo, por lo que constituye la solución de nuestro modelo dinámico. [La ecuación (14.1) también es una función de t. ¿Por qué *ésta* no puede considerarse también como una solución?]

Al ser tan sencillo, este ejemplo de la población ilustra la esencia de los problemas de la dinámica económica. Dado el patrón de comportamiento de una variable respecto al tiempo, deseamos encontrar una función que describa la trayectoria de tiempo de la variable. En el proceso, encontraremos una o más constantes arbitrarias, pero si poseemos suficiente información adicional en la forma de *condiciones iniciales*, podremos determinar el valor de estas constantes arbitrarias.

En los modelos más sencillos de problema, como el ya citado, podemos encontrar la solución mediante el método del cálculo integral, que trata del proceso de rastreo de una función de derivada dada hasta su función primitiva. En casos más complicados, también podemos recurrir a las técnicas conocidas de la rama de las matemáticas estrechamente relacionada conocida como *ecuaciones diferenciales*. Puesto que una ecuación diferencial se define como cualquier ecuación que contiene expresiones diferenciales o derivadas, seguramente (14.1) califica como una; en consecuencia, al encontrar su solución, de hecho ya hemos resuelto una ecuación diferencial, aunque muy simple.

Prosigamos ahora al estudio de los conceptos básicos del cálculo integral. Como estudiamos el cálculo diferencial con x (en lugar de t) como la variable independiente, usaremos x aquí también por razones de simetría. Sin embargo, en el presente estudio denotaremos, por conveniencia, las funciones primitivas con $F(x)$ y las derivadas con $f(x)$, en lugar de distinguirlas con el uso de una prima.

14.2 Integrales indefinidas

La naturaleza de las integrales

Hemos mencionado que la integración es el inverso de la diferenciación. Si la diferenciación de una función primitiva $F(x)$ dada conduce a la derivada $f(x)$, podemos "integrar" $f(x)$ para encontrar $F(x)$, siempre que dispongamos de información apropiada para determinar el valor de la constante arbitraria que va a surgir en el proceso de integración. La función $F(x)$ se define como la *integral* o *indefinida* (o *antiderivada*) de la función $f(x)$. Estos dos tipos de procesos pueden compararse con dos maneras de estudiar un árbol genealógico: la *integración* implica el rastreo del parentesco de la función $f(x)$, mientras que la *diferenciación* busca los descendientes de la función $F(x)$. Observe, no obstante, esta diferencia: mientras que la función primitiva $F(x)$ (diferenciable) produce invariablemente una descendencia solitaria, es decir, una derivada única $f(x)$, la función derivada $f(x)$ es rastreable hasta un número infinito de padres posibles a través de la integración, ya que si $F(x)$ es una integral de $f(x)$, entonces también debe serlo $F(x)$ más cualquier constante, como vimos en (14.2).

Necesitamos una notación especial para señalar la integración requerida de $f(x)$ respecto a x. La estándar es

$$\int f(x)\,dx$$

El símbolo de la izquierda —una S alargada (con la connotación de una suma, lo que se explica posteriormente)— se llama el *signo de integral*, mientras que la parte $f(x)$ se conoce como el *integrando* (la función que se va a integrar), y la parte dx —similar al dx del operador diferencial d/dx— nos recuerda que la operación debe realizarse respecto a la variable x. Sin embargo, usted también puede considerar a $f(x)\,dx$ como una entidad individual e interpretarla como la diferencial de la función primitiva $F(x)$ [es decir, $dF(x) = f(x)\,dx$]. Entonces, el signo integral que está enfrente puede verse como una instrucción para invertir el proceso de diferenciación que dio lugar a la diferencial. Con esta nueva notación podemos escribir que

$$\frac{d}{dx}F(x) = f(x) \quad \Rightarrow \quad \int f(x)\,dx = F(x) + c \tag{14.3}$$

donde la presencia de c, una *constante de integración* arbitraria, sirve para indicar al parentesco múltiple del integrando.

La integral $\int f(x)\,dx$ se conoce más específicamente como la *integral indefinida* de $f(x)$ (en contraste con la *integral definida,* que vamos a estudiar en la sección 14.2), porque no tiene un valor numérico definido. Como es igual a $F(x) + c$, su valor va a variar en general con el valor de x (aun si c se vuelve un número conocido). Entonces, al igual que en la derivada, una integral indefinida es en sí misma una función de la variable x.

Reglas básicas de la integración

Así como hay reglas de la derivación, también podemos desarrollar ciertas reglas para la integración. Como puede esperarse, estas últimas dependen en gran medida de las reglas de derivación con las que ya estamos familiarizados. Por ejemplo, a partir de la siguiente fórmula de derivada para una función potencia,

$$\frac{d}{dx}\left(\frac{x^{n+1}}{n+1}\right) = x^n \qquad (n \neq -1)$$

vemos que la expresión $x^{n+1}/(n+1)$ es la función primitiva para la función derivada x^n; entonces, al sustituirlas en $F(x)$ y $f(x)$ en (14.3), podemos enunciar el resultado como una regla de integración.

Regla I (la regla de la potencia)

$$\int x^n\,dx = \frac{1}{n+1}x^{n+1} + c \qquad (n \neq -1)$$

Ejemplo 1

Encuentre $\int x^3 dx$. Aquí, tenemos $n = 3$ y, por lo tanto,

$$\int x^3 dx = \frac{1}{4}x^4 + c$$

Ejemplo 2

Encuentre $\int x\,dx$. Como $n = 1$, tenemos

$$\int x\,dx = \frac{1}{2}x^2 + c$$

Ejemplo 3

¿Qué es $\int 1\,dx$? Para encontrar esta integral, recordamos que $x^0 = 1$; de modo que podemos hacer $n = 0$ en la regla de la potencia y obtener

$$\int 1\,dx = x + c$$

[$\int 1\,dx$ algunas veces se escribe sólo como $\int dx$, ya que $1\,dx = dx$.]

Ejemplo 4

Encuentre $\int \sqrt{x^3}\,dx$. Como $\sqrt{x^3} = x^{3/2}$, tenemos $n = \frac{3}{2}$; por lo tanto,

$$\int \sqrt{x^3}\,dx = \frac{x^{5/2}}{\frac{5}{2}} + c = \frac{2}{5}\sqrt{x^5} + c$$

Ejemplo 5

Encuentre $\int \frac{1}{x^4}dx$, $(x \neq 0)$. Como $1/x^4 = x^{-4}$, tenemos $n = -4$. Entonces, la integral es

$$\int \frac{1}{x^4}\,dx = \frac{x^{-4+1}}{-4+1} + c = -\frac{1}{3x^3} + c$$

Observe que los resultados de la integración pueden verificarse siempre mediante la diferenciación; si el proceso de integración es correcto, la derivada de la integral debe ser igual al integrando.

Se ha mostrado que las fórmulas de derivada para funciones simples exponencial y logarítmica son

$$\frac{d}{dx}e^x = e^x \qquad \text{y} \qquad \frac{d}{dx}\ln x = \frac{1}{x} \qquad (x > 0)$$

A partir de éstas, emergen otras dos reglas básicas de integración.

Regla II (la regla exponencial)

$$\int e^x\,dx = e^x + c$$

Regla III (la regla logarítmica)

$$\int \frac{1}{x}\,dx = \ln x + c \qquad (x > 0)$$

Llama la atención que el integrando implicado en la regla III sea $1/x = x^{-1}$, que es una forma especial de la función potencia x^n con $n = -1$. Este integrando específico es inadmisible bajo la regla de la potencia, pero ahora es apropiadamente tratada por la regla logarítmica.

Como está enunciada, la regla logarítmica se coloca bajo la restricción $x > 0$, ya que no existen los logaritmos para valores no positivos de x. Una formulación más general de la regla, que puede manejar valores negativos de x, es

$$\int \frac{1}{x}\,dx = \ln|x| + c \qquad (x \neq 0)$$

lo que también implica que $(d/dx)\ln|x| = 1/x$, así como $(d/dx)\ln x = 1/x$. Debe convencerse de que el reemplazo de x (con la restricción $x > 0$) por $|x|$ (con la restricción $x \neq 0$) de ninguna manera invalida la fórmula.

También como una cuestión de notación, debe señalarse que la integral $\int \frac{1}{x}\,dx$ algunas veces también se escribe como $\int \frac{dx}{x}$.

Como variantes de las reglas II y III, también tenemos las dos siguientes reglas.

Regla II*a*

$$\int f'(x)e^{f(x)}\,dx = e^{f(x)} + c$$

Regla III*a*

$$\int \frac{f'(x)}{f(x)}\,dx = \ln f(x) + c \qquad [f(x) > 0]$$

o sea

$$\ln|f(x)| + c \qquad [f(x) \neq 0]$$

Las bases para estas dos reglas pueden encontrarse en las reglas de las derivadas de (10.20).

Reglas de operación

Las tres reglas anteriores ilustran ampliamente el espíritu subyacente a todas las reglas de integración. También, siempre se añade una constante arbitraria al final (aun cuando se debe determinar su valor después mediante el uso de una condición inicial o de frontera) para indicar que una familia completa de funciones primitivas puede originar al integrando dado.

Sin embargo, para poder manejar integrandos más complicados, también nos serán útiles las dos siguientes reglas de operación respecto a las integrales.

Regla IV (la integral de una suma) La integral de la suma de un número finito de funciones es la suma de las integrales de esas funciones. Para el caso de dos funciones, esto significa que

$$\int [f(x) + g(x)]\, dx = \int f(x)\, dx + \int g(x)\, dx$$

Esta regla es una consecuencia natural del hecho de que

$$\underbrace{\frac{d}{dx}[F(x) + G(x)]}_{A} = \underbrace{\frac{d}{dx}F(x) + \frac{d}{dx}G(x)}_{B} = \underbrace{f(x) + g(x)}_{C}$$

Como $A = C$, basándonos en (14.3) podemos escribir

$$\int [f(x) + g(x)]\, dx = F(x) + G(x) + c \qquad \textbf{(14.4)}$$

Pero del hecho de que $B = C$, se sigue que

$$\int f(x)\, dx = F(x) + c_1 \qquad \text{y} \qquad \int g(x)\, dx = G(x) + c_2$$

Entonces, podemos obtener (por adición)

$$\int f(x)\, dx + \int g(x)\, dx = F(x) + G(x) + c_1 + c_2 \qquad \textbf{(14.5)}$$

Como las constantes c, c_1 y c_2 tienen valor arbitrario, podemos hacer $c = c_1 + c_2$. Entonces, las partes derechas de (14.4) y (14.5) se igualan y, como consecuencia, las partes izquierdas también deben igualarse. Esto prueba la Regla IV.

Ejemplo 6

Encuentre $\int (x^3 + x + 1)\, dx$. Por la regla IV, esta integral podemos expresarla como la suma de tres integrales: $\int x^3\, dx + \int x\, dx + \int 1\, dx$. Como los valores de estas tres integrales se han encontrado previamente en los ejemplos 1, 2 y 3, podemos simplemente combinar estos resultados para obtener

$$\int (x^3 + x + 1)\, dx = \left(\frac{x^4}{4} + c_1\right) + \left(\frac{x^2}{2} + c_2\right) + (x + c_3) = \frac{x^4}{4} + \frac{x^2}{2} + x + c$$

En la respuesta final hemos aglutinado las tres constantes con subíndice en una sola constante c.

Como práctica general, todas las constantes de integración arbitrarias aditivas que surgen durante el proceso pueden combinarse siempre en una sola constante arbitraria en la respuesta final.

Ejemplo 7

Encuentre $\int \left(2e^{2x} + \frac{14x}{7x^2 + 5}\right) dx$. Por la regla IV, podemos integrar por separado los dos términos aditivos en el integrando, y luego sumar los resultados. Como el término $2e^{2x}$ está en el formato de $f'(x)e^{f(x)}$ en la regla II*a*, con $f(x) = 2x$, la integral es $e^{2x} + c_1$. En forma similar, el otro término $14x/(7x^2+5)$, adopta la forma de $f'(x)/f(x)$, con $f(x)=7x^2+5>0$. Entonces, por la regla III*a*, la integral es $\ln(7x^2+5)+c_2$. Entonces, podemos escribir

$$\int\left(2e^{2x}+\frac{14x}{7x^2+5}\right)dx=e^{2x}+\ln(7x^2+5)+c$$

donde hemos combinado c_1 y c_2 en una sola constante arbitraria c.

Regla V (la integral de un múltiplo) La integral de k veces un integrando (siendo k una constante) es k veces la integral de ese integrando. En símbolos,

$$\int kf(x)\,dx=k\int f(x)\,dx$$

Operacionalmente, esta regla señala que una constante multiplicativa puede factorizarse fuera del signo de la integral. (*Advertencia:* ¡Un término *variable no puede* factorizarse hacia fuera de esta manera!) Para probar esta regla (para el caso en el cual k es un entero), recordemos que k veces $f(x)$ simplemente significa sumar $f(x)$ k veces. Por lo tanto, por la regla IV,

$$\begin{aligned}\int kf(x)\,dx&=\int\underbrace{[f(x)+f(x)+\cdots+f(x)]}_{k\text{ términos}}\,dx\\&=\underbrace{\int f(x)\,dx+\int f(x)\,dx+\cdots+\int f(x)\,dx}_{k\text{ términos}}=k\int f(x)\,dx\end{aligned}$$

Ejemplo 8

Encuentre $\int -f(x)\,dx$. Aquí $k=-1$, y entonces

$$\int -f(x)\,dx=-\int f(x)\,dx$$

es decir, la integral del negativo de una función es el negativo de la integral de esa función.

Ejemplo 9

Encuentre $\int 2x^2\,dx$. Factorizando el 2 y aplicando la regla I, tenemos

$$\int 2x^2\,dx=2\int x^2\,dx=2\left(\frac{x^3}{3}+c_1\right)=\frac{2}{3}x^3+c$$

Ejemplo 10

Encuentre $\int 3x^2\,dx$. En este caso, la factorización de la constante multiplicativa nos da

$$\int 3x^2\,dx=3\int x^2\,dx=3\left(\frac{x^3}{3}+c_1\right)=x^3+c$$

Observe que el término x^3 en la respuesta final, en contraste con el ejemplo anterior, no está multiplicado por ninguna expresión fraccionaria. Este resultado limpio se debe al hecho de que 3 (la constante multiplicativa del integrando) es igual a 2 (la potencia de la función) más 1. En referencia a la regla de potencias (regla I), vemos que la constante multiplicativa $(n+1)$ se va a cancelar con la fracción $1/(n+1)$, suministrando como la respuesta a $(x^{n+1}+c)$.

En general, siempre que tengamos una expresión $(n+1)x^n$ como el integrando, realmente no hay necesidad de factorizar la constante $(n+1)$ y luego integrar x^n; en vez de eso, podemos escribir $x^{n+1}+c$ inmediatamente como la respuesta.

Ejemplo 11 Encuentre $\int \left(5e^x - x^{-2} + \frac{3}{x}\right) dx, (x \neq 0)$. Este ejemplo ilustra las reglas IV y V; en realidad, también ilustra las tres primeras reglas:

$$\int \left(5e^x - \frac{1}{x^2} + \frac{3}{x}\right) dx = 5\int e^x\, dx - \int x^{-2}\, dx + 3\int \frac{1}{x}\, dx \quad \text{[mediante las reglas IV y V]}$$
$$= (5e^x + c_1) - \left(\frac{x^{-1}}{-1} + c_2\right) + (3\ln|x| + c_3)$$
$$= 5e^x + \frac{1}{x} + 3\ln|x| + c$$

Nuevamente puede verificarse la corrección del resultado mediante la diferenciación.

Reglas que incluyen la sustitución

Ahora introduciremos dos reglas más de integración que buscan simplificar el proceso de integración, cuando las circunstancias sean apropiadas, mediante una sustitución de la variable original de integración. Siempre que la variable de integración de nueva introducción haga más fácil el proceso de integración que para la variable original, estas reglas serán de utilidad.

Regla VI (regla de sustitución) La integral de $f(u)(du/dx)$ respecto a la variable x es la integral de $f(u)$ respecto a la variable u:

$$\int f(u)\frac{du}{dx}\, dx = \int f(u)\, du = F(u) + c$$

donde la operación $\int du$ se ha sustituido por la operación $\int dx$.

Esta regla, la contraparte en el cálculo integral de la regla de la cadena, puede probarse mediante la misma regla de la cadena. Dada una función $F(u)$, donde $u = u(x)$, la regla de la cadena establece que

$$\frac{d}{dx}F(u) = \frac{d}{du}F(u)\frac{du}{dx} = F'(u)\frac{du}{dx} = f(u)\frac{du}{dx}$$

Como $f(u)(du/dx)$ es la derivada de $F(u)$, de (14.3) se sigue que la integral (antiderivada) de la anterior debe ser

$$\int f(u)\frac{du}{dx}\, dx = F(u) + c$$

De hecho, puede observar que este resultado es también el de la *cancelación* de las dos expresiones dx a la izquierda.

Ejemplo 12 Encuentre $\int 2x(x^2 + 1)\, dx$. Podemos obtener la respuesta haciendo las multiplicaciones indicadas en el integrando:

$$\int 2x(x^2+1)\, dx = \int (2x^3 + 2x)dx = \frac{x^4}{2} + x^2 + c$$

Hagámoslo ahora por la regla de la sustitución. Sea $u = x^2 + 1$; entonces $du/dx = 2x$, o sea $dx = du/2x$. La sustitución de $du/2x$ en lugar de dx nos da

$$\int 2x(x^2+1)\, dx = \int 2xu\frac{du}{2x} = \int u\, du = \frac{u^2}{2} + c_1$$
$$= \frac{1}{2}(x^4 + 2x^2 + 1) + c_1 = \frac{1}{2}x^4 + x^2 + c$$

donde $c = \frac{1}{2} + c_1$. También podemos obtener la misma respuesta al sustituir du/dx en lugar de $2x$ (en vez de $du/2x$ en lugar de dx).

Ejemplo 13 Encuentre $\int 6x^2(x^3+2)^{99}\,dx$. Las multiplicaciones indicadas en el integrando de este ejemplo no se realizan tan fácilmente, por lo que ahora la regla de la sustitución tiene una mejor oportunidad de demostrar su efectividad. Sea $u = x^3 + 2$; entonces $du/dx = 3x^2$, de modo que

$$\int 6x^2(x^3+2)^{99}\,dx = \int \left(2\frac{du}{dx}\right) u^{99}\,dx = \int 2u^{99}\,du$$
$$= \frac{2}{100}u^{100} + c = \frac{1}{50}(x^3+2)^{100} + c$$

Ejemplo 14 Encuentre $\int 8e^{2x+3}\,dx$. Sea $u = 2x + 3$; entonces $ddu/dx = 2$, o sea $dx = du/2$. Entonces,

$$\int 8e^{2x+3}\,dx = \int 8e^u \frac{du}{2} = 4\int e^u\,du = 4e^u + c = 4e^{2x+3} + c$$

Como muestran estos ejemplos, esta regla es de ayuda siempre que podamos —mediante la elección juiciosa de una función $u = u(x)$— expresar el integrando (una función de x) como el producto de $f(u)$ (una función de u) y du/dx (la derivada de la función u que hemos escogido). Sin embargo, como lo ilustran los dos últimos ejemplos, esta regla también podemos usarla cuando el integrando original se puede transformar en una constante múltiplo de $f(u)(du/dx)$. Esto no afectaría la aplicabilidad porque la constante que multiplica puede factorizarse fuera del signo de la integral, lo que entonces dejaría un integrando de la forma $f(u)(du/dx)$, como se requiere en la regla de la sustitución. Cuando la sustitución de variables da por resultado una *variable* múltiplo de $f(u)(du/dx)$, digamos, x veces por esta última, no se permite la factorización y esta regla no nos va a ayudar. De hecho, no existe una fórmula general que dé la integral del producto de dos funciones en términos de las integrales separadas de estas funciones; ni tenemos una fórmula general que dé la integral de un cociente de dos funciones en términos de sus integrales separadas. Aquí radica la razón por la cual la integración, en general, es más difícil que la diferenciación y por la cual, con integrandos complicados, es más conveniente buscar la respuesta en tablas preparadas de las fórmulas de integración en lugar de intentar uno mismo la integración.

Regla VII (integración por partes) La integral de v respecto a u es igual a uv menos la integral de u respecto a v:

$$\int v\,du = uv - \int u\,dv$$

La esencia de esta regla es reemplazar la operación $\int du$ por la operación $\int dv$.

El razonamiento a que induce este resultado es relativamente sencillo. Primero, la regla del producto de las diferenciales nos da

$$d(uv) = v\,du + u\,dv$$

Si integramos ambos lados de la ecuación (es decir, integramos cada diferencial), obtenemos una nueva ecuación

$$\int d(uv) = \int v\,du + \int u\,dv$$

o sea, $$uv = \int v\,du + \int u\,dv \qquad \text{[no se necesita constante en la izquierda (¿por qué?)]}$$

Luego, al restar $\int u\,dv$ de ambos lados, surge el resultado previamente enunciado.

Ejemplo 15

Encuentre $\int x(x+1)^{1/2}\,dx$. A diferencia de los ejemplos 12 y 13, este ejemplo no es asequible al tipo de sustitución usado en la regla VI. (¿Por qué?) Sin embargo, podemos considerar que la integral dada está en la forma de $\int v\,du$, y aplicar la regla VII. Para lograr este objetivo, haremos $v = x$, lo que implica $dv = dx$, y también haremos $u = \frac{2}{3}(x+1)^{3/2}$, de modo que $du = (x+1)^{1/2}\,dx$. Entonces, podemos encontrar que la integral es

$$\begin{aligned}\int x(x+1)^{1/2}\,dx &= \int v\,du = uv - \int u\,dv \\ &= \frac{2}{3}(x+1)^{3/2}x - \int \frac{2}{3}(x+1)^{3/2}\,dx \\ &= \frac{2}{3}(x+1)^{3/2}x - \frac{4}{15}(x+1)^{5/2} + c\end{aligned}$$

Ejemplo 16

Encuentre $\int \ln x\,dx$, $(x > 0)$. No podemos aplicar aquí la regla logarítmica, porque esa regla maneja el integrando $1/x$, no ln x, ni la regla VI. Pero si hacemos $v = \ln x$, lo que implica que $dv = (1/x)\,dx$, y también hacemos $u = x$, de modo que $du = dx$, entonces la integración puede realizarse como:

$$\begin{aligned}\int \ln x\,dx &= \int v\,du = uv - \int u\,dv \\ &= x\ln x - \int dx = x\ln x - x + c = x(\ln x - 1) + c\end{aligned}$$

Ejemplo 17

Encuentre $\int xe^x\,dx$. En este caso, simplemente hacemos $v = x$ y $u = e^x$, de modo que $dv = dx$ y $du = e^x\,dx$. Aplicando la regla VII, tenemos

$$\begin{aligned}\int xe^x\,dx &= \int v\,du = uv - \int u\,dv \\ &= e^x x - \int e^x\,dx = e^x x - e^x + c = e^x(x-1) + c\end{aligned}$$

La validez de este resultado, como la de los ejemplos anteriores, podemos verificarla, por supuesto, por diferenciación.

EJERCICIO 14.2

1. Encuentre lo siguiente:

(*a*) $\int 16x^{-3}\,dx \quad (x \neq 0)$

(*b*) $\int 9x^8\,dx$

(*c*) $\int (x^5 - 3x)\,dx$

(*d*) $\int 2e^{-2x}\,dx$

(*e*) $\int \frac{4x}{x^2+1}\,dx$

(*f*) $\int (2ax + b)(ax^2 + bx)^7\,dx$

2. Encuentre:

(*a*) $\int 13e^x\,dx$

(*b*) $\int \left(3e^x + \frac{4}{x}\right)dx \quad (x > 0)$

(*c*) $\int \left(5e^x + \frac{3}{x^2}\right)dx \quad (x \neq 0)$

(*d*) $\int 3e^{-(2x+7)}\,dx$

(*e*) $\int 4xe^{x^2+3}\,dx$

(*f*) $\int xe^{x^2+9}\,dx$

3. Encuentre:

(*a*) $\int \frac{3dx}{x} \quad (x \neq 0)$

(*b*) $\int \frac{dx}{x-2} \quad (x \neq 2)$

(*c*) $\int \frac{2x}{x^2+3}\,dx$

(*d*) $\int \frac{x}{3x^2+5}\,dx$

4. Encuentre:

(*a*) $\int (x+3)(x+1)^{1/2}\,dx$

(*b*) $\int x \ln x\,dx \quad (x > 0)$

5. Dadas n constantes k_i (con $i = 1, 2, \ldots, n$) y n funciones $f_i(x)$, deduzca de las reglas IV y V que

$$\int \sum_{i=1}^{n} k_i f_i(x)\,dx = \sum_{i=1}^{n} k_i \int f_i(x)\,dx$$

14.3 Integrales definidas

Significado de las integrales definidas

Todas las integrales citadas en la sección 14.2 se les conoce como *indefinidas*: cada una es una función de una variable, por tanto, no posee un valor numérico definido. Ahora, para una integral indefinida dada de una función continua $f(x)$,

$$\int f(x)\,dx = F(x) + c$$

Si escogemos dos valores de x en el dominio, digamos, a y b $(a < b)$, los sustituimos sucesivamente en el lado derecho de la ecuación y formamos la diferencia

$$[F(b) + c] - [F(a) + c] = F(b) - F(a)$$

obtenemos un valor numérico específico, libre de la variable x, así como de la constante arbitraria c. Este valor se llama la *integral definida* de $f(x)$ de a a b. Nos referimos a a como el *límite inferior de integración* y a b como el *límite superior de integración.*

Con el fin de indicar los límites de integración, modificamos ahora el signo integral a la forma $\int_a^b$. La evaluación de la integral definida se simboliza entonces en los siguientes pasos:

$$\int_a^b f(x)\,dx = F(x)\Big]_a^b = F(b) - F(a) \qquad \textbf{(14.6)}$$

donde el símbolo $]_a^b$ (también escrito $|_a^b$ o $[\cdots]_a^b$) es una instrucción para sustituir a y b, sucesivamente, en lugar de x en el resultado de la integración para obtener $F(b)$ y $F(a)$, y luego tomar la diferencia, como se indica a la derecha de (14.6). Sin embargo, como primer paso, debemos encontrar la integral indefinida, aunque podemos omitir la constante c, ya que ésta se cancela de cualquier manera en el proceso de toma de diferencias.

Ejemplo 1 Evalúe $\int_1^5 3x^2\,dx$. Como la integral indefinida es $x^3 + c$, esta integral definida tiene el valor

$$\int_1^5 3x^2\,dx = x^3\Big]_1^5 = (5)^3 - (1)^3 = 125 - 1 = 124$$

Ejemplo 2 Evalúe $\int_a^b ke^x\,dx$. Aquí, los límites de integración se dan con símbolos; en consecuencia, el resultado de la integración también es en términos de estos símbolos:

$$\int_a^b ke^x\,dx = ke^x\Big]_a^b = k(e^b - e^a)$$

Ejemplo 3 Evalúe $\int_0^4 \left(\frac{1}{1+x} + 2x\right) dx$, $(x \neq -1)$. La integral indefinida es $\ln|1+x| + x^2 + c$; entonces, la respuesta es

$$\begin{aligned}\int_0^4 \left(\frac{1}{1+x} + 2x\right) dx &= \Big[\ln|1+x| + x^2\Big]_0^4 \\ &= (\ln 5 + 16) - (\ln 1 + 0) \\ &= \ln 5 + 16 \qquad [\text{ya que} \ln 1 = 0]\end{aligned}$$

Es importante darse cuenta de que los límites de integración a y b se refieren ambos a valores de la variable x. Si tuviéramos que usar la técnica de sustitución de variables (reglas VI y VII) durante la integración e introducimos una variable u, debemos tener cuidado de *no* considerar a a y b como los límites de u. El ejemplo 4 ilustra este punto.

Ejemplo 4 Evalúe $\int_1^2 (2x^3-1)^2(6x^2)\,dx$. Sea $u = 2x^3 - 1$; entonces $du/dx = 6x^2$, o sea $du = 6x^2\,dx$. Observe ahora que cuando $x = 1$, u será 1, pero cuando $x = 2$, u será 15; en otras palabras, los límites de integración en términos de la variable u deben ser 1 (inferior) y 15 (superior).

Por lo tanto, el reescribir la integral dada en u no nos dará $\int_1^2 u^2\,du$ sino

$$\int_1^{15} u^2\,du = \frac{1}{3}u^3\Big]_1^{15} = \frac{1}{3}(15^3 - 1^3) = 1\ 124\tfrac{2}{3}$$

Alternamente, primero podemos transformar u otra vez en x y luego usar los límites originales de 1 y 2 para obtener la respuesta idéntica:

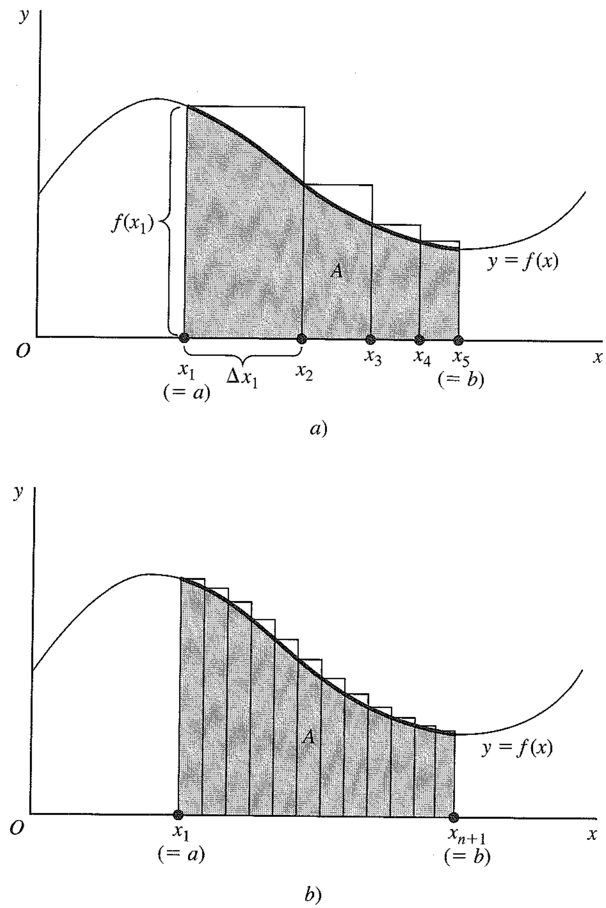
$$\left[\frac{1}{3}u^3\right]_{u=1}^{u=15} = \left[\frac{1}{3}(2x^3-1)^3\right]_{x=1}^{x=2} = \frac{1}{3}(15^3 - 1^3) = 1\ 124\tfrac{2}{3}$$

La integral definida como el área bajo la curva

Cada integral definida tiene un valor definido. Este valor puede interpretarse geométricamente como un área específica bajo una curva dada.

En la figura 14.1 se traza la gráfica de una función continua $y = f(x)$. Si deseamos medir el área A (sombreada) encerrada por la curva y el eje x entre los dos puntos a y b en el dominio, podemos proseguir de la siguiente manera. Primero, dividimos el intervalo $[a, b]$ en n subintervalos (no necesariamente de longitud igual). Cuatro de estos subintervalos se dibujan en la figura 14.1a —es decir, $n = 4$— siendo el primero $[x_1, x_2]$ y el último, $[x_4, x_5]$. Como cada uno representa un cambio de x, podemos denominarlos $\Delta x_1, \ldots, \Delta x_4$, respectivamente. Construyamos ahora cuatro bloques rectangulares en los subintervalos, de modo que la altura de cada bloque sea igual al mayor valor de la función alcanzado en ese bloque (lo que ocurre aquí en la frontera del lado izquierdo de cada rectángulo). Entonces, el primer bloque tiene

FIGURA 14.1



una altura $f(x_1)$ y un ancho Δx_1, y en general, el i-ésimo bloque tiene una altura $f(x_i)$ y un ancho Δx_i. El área total A^* de este conjunto de bloques es la suma

$$A^* = \sum_{i=1}^{n} f(x_i)\ \Delta x_i \qquad (n = 4 \text{ en la figura } 14.1a)$$

Sin embargo, es obvio que ésta *no* es el área bajo la curva que buscamos, sino solamente una aproximación muy rudimentaria.

Lo que hace que A^* se desvíe del valor verdadero de A es la porción no sombreada de los bloques rectangulares, que hacen que A^* sea una *sobre*estimación de A. Sin embargo, si podemos reducir de tamaño la parte no sombreada y hacemos que se aproxime a cero, el valor A^* de aproximación se acercará correspondientemente al valor verdadero A. Este resultado va a materializarse cuando intentemos una segmentación cada vez más fina del intervalo $[a, b]$, de modo que n aumente y Δx_i se acorte indefinidamente. Entonces, los bloques se harán más delgados (si son más numerosos) y va a disminuir la parte que está por arriba de la curva, como puede verse en la figura 14.1b. Llevada hasta el límite, esta operación de "adelgazamiento" obtiene el área exacta

$$\lim_{n\to\infty} \sum_{i=1}^{n} f(x_i)\ \Delta x_i = \lim_{n\to\infty} A^* = \text{área } A \tag{14.7}$$

siempre que este límite exista (existe en el presente caso). Realmente, esta ecuación constituye la definición formal del área bajo la curva.

La expresión sumatoria de (14.7), $\sum_{i=1}^{n} f(x_i)\ \Delta x_i$, tiene cierto parecido con la expresión de la integral definida $\int_a^b f(x)\,dx$. Realmente esta última expresión se basa en la anterior. El reemplazo de Δx_i por la diferencial dx se hace con la misma intención que en nuestro estudio anterior de la "aproximación" en la sección 8.1. Entonces, reescribimos $f(x_i)\ \Delta x_i$ como $f(x)\,dx$. ¿Qué pasa con el signo de la sumatoria? La notación $\sum_{i=1}^{n}$ representa la suma de un número *finito* de términos. Cuando hacemos $n \to \infty$, y tomamos el límite de esa suma, la notación regular para una operación de este tipo es más bien engorrosa. Entonces, se necesita un sustituto más sencillo. Ese sustituto es $\int_a^b$, donde el símbolo S alargado también indica una suma, y donde a y b (tal como $i = 1$ y n) sirven para especificar los límites inferior y superior de esta suma. En resumen, la integral definida es una versión abreviada para la expresión del límite de una suma en (14.7). Es decir,

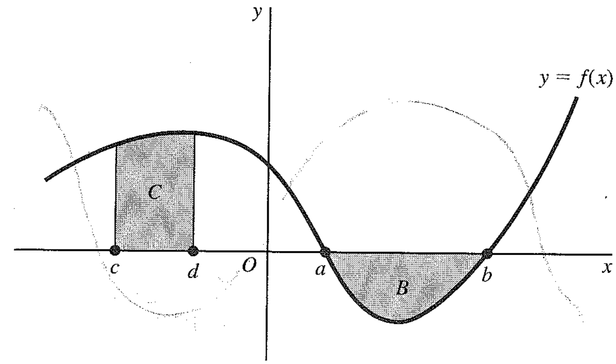
$$\int_a^b f(x)\,dx \equiv \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i)\ \Delta x_i = \text{área } A$$

Entonces, esta integral definida (denominada *integral de Riemann*) tiene ahora una connotación de *área,* así como una connotación de *suma*, porque $\int_a^b$ es la contraparte continua del concepto discreto de $\sum_{i=1}^{n}$.

En la figura 14.1 intentamos aproximar el área A al reducir sistemáticamente una sobreestimación A^* mediante una partición más fina del intervalo $[a, b]$. El límite resultante de la suma de las áreas de los rectángulos se llama la *integral superior*: una aproximación que viene desde arriba. También pudimos haber aproximado el área A desde abajo al formar bloques rectangulares inscritos en la curva en lugar de que sobresalgan (vea el ejercicio 14.3-3). El área total A^{**} de este nuevo conjunto de bloques va a *sub*estimar al área A, pero a medida que la partición de $[a, b]$ se hace más fina, nuevamente encontramos $\lim_{n\to\infty} A^{**} = A$. El límite citado al último de la suma de las áreas de los rectángulos se llama la *integral inferior*. La integral de Riemann existe si y sólo si la integral superior y la integral inferior tienen el mismo valor, $\int_a^b f(x)\,dx$, y se dice que la función $f(x)$ es *Riemann integrable*. Hay teoremas que especifican las condiciones bajo las cuales es integrable una función $f(x)$. De acuerdo con el teorema fundamental del cálculo, una función es integrable en $[a, b]$ si es continua en ese intervalo. Por lo tanto, siempre que trabajemos con funciones continuas no debemos preocuparnos respecto a esto.

Observemos otro punto. Aunque el área A en la figura 14.1 está situada enteramente bajo una porción decreciente de la curva $y = f(x)$, la igualdad conceptual de una integral definida con un área es válida también para porciones de la curva con pendiente hacia arriba. De hecho,

FIGURA 14.2



ambos tipos de pendiente pueden estar presentes simultáneamente; por ejemplo, podemos calcular $\int_0^b f(x)\,dx$ como el área bajo la curva en la figura 14.1 arriba de la línea Ob.

Observe que si calculamos el área B en la figura 14.2 mediante la integral definida $\int_a^b f(x)\,dx$, la respuesta será negativa, ya que la altura de cada bloque rectangular que interviene en esta área es negativa. Esto da lugar a la noción de un *área negativa*, la cual está situada *debajo* del eje x y *arriba* de una curva dada. En el caso de que estemos interesados en el valor numérico en lugar del valor algebraico de un área de este tipo, debemos tomar el valor absoluto de la integral definida relevante. Por otro lado, el área $C = \int_c^d f(x)\,dx$, tiene signo positivo aunque esté situada en la región negativa del eje x; esto se debe a que cada bloque rectangular tiene una altura positiva, así como un ancho positivo cuando nos movemos de c a d. A partir de esto, es clara la implicación de que el intercambio de los dos límites de integración alteraría el signo de Δx_i y de la integral definida al invertir la dirección del movimiento. Aplicada al área B, vemos que la integral definida $\int_b^a f(x)\,dx$ (de b a a) va a dar el negativo del área B; esto va a medir el valor numérico de esta área.

Algunas propiedades de las integrales definidas

La discusión del párrafo anterior nos conduce a la siguiente propiedad de las integrales definidas.

Propiedad I El intercambio de los límites de integración cambia el signo de la integral definida:

$$\int_b^a f(x)\,dx = -\int_a^b f(x)\,dx$$

Esto puede probarse de la siguiente manera:

$$\int_b^a f(x)\,dx = F(a) - F(b) = -[F(b) - F(a)] = -\int_a^b f(x)\,dx$$

Las integrales definidas también poseen otras propiedades interesantes.

Propiedad II Una integral definida tiene un valor de cero cuando los dos límites de integración son idénticos:

$$\int_a^a f(x)\,dx = F(a) - F(a) = 0$$

De acuerdo con la interpretación de "área", esto significa que el área (bajo la curva) no cuenta por arriba de cualquier *punto* individual en el dominio. Así es como debe ser, porque arriba de un punto en el eje x podemos dibujar solamente una *línea* (unidimensional), nunca un *área* (bidimensional).

Propiedad III Una integral definida puede expresarse como una suma de un número finito de subintegrales definidas como:

$$\int_a^d f(x)\,dx = \int_a^b f(x)\,dx + \int_b^c f(x)\,dx + \int_c^d f(x)\,dx \qquad (a < b < c < d)$$

Solamente se muestran tres subintegrales en esta ecuación, pero también es válida la generalización al caso de n subintegrales. Algunas veces esta propiedad se describe como la *propiedad de aditividad.*

En términos de área, esto significa que el área (bajo la curva) situada arriba del intervalo $[a, d]$ sobre el eje x puede obtenerse al sumar las áreas situadas arriba de los subintervalos del conjunto $\{[a, b], [b, c], [c, d]\}$. Ya que estamos manejando intervalos cerrados, observe que los puntos de frontera b y c se han incluido en *dos* áreas. ¿No es esto un doble conteo? En realidad lo es. Pero afortunadamente no se causa ningún daño, pues por la propiedad II el área de arriba de un punto individual es cero, de modo que el doble conteo no produce efecto sobre el cálculo. Pero no es necesario decir que nunca se permite el doble conteo de cualquier *intervalo.*

Anteriormente mencionamos que todas las funciones continuas son Riemann integrables. Ahora, por la propiedad III, también podemos encontrar las integrales definidas (áreas) de ciertas funciones discontinuas. Considere la función escalón en la figura 14.3*a*. A pesar de la discontinuidad del punto b en el intervalo $[a, c]$, podemos encontrar el área sombreada a partir de la suma

$$\int_a^b f(x)\,dx + \int_b^c f(x)\,dx$$

Lo mismo es aplicable también a la curva de la figura 14.3*b*.

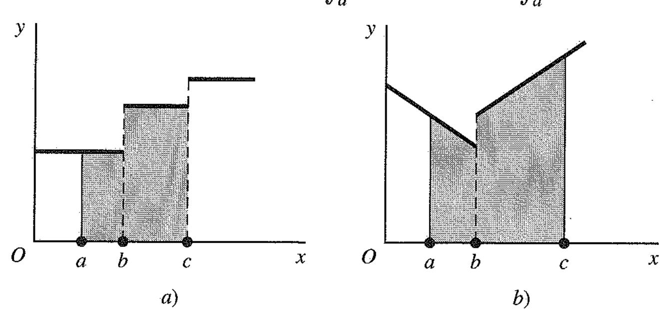
Propiedad IV

$$\int_a^b -f(x)\,dx = -\int_a^b f(x)\,dx$$

Propiedad V

$$\int_a^b kf(x)\,dx = k\int_a^b f(x)\,dx$$

FIGURA 14.3



Propiedad VI

$$\int_a^b [f(x) + g(x)]\, dx = \int_a^b f(x)\, dx + \int_a^b g(x)\, dx$$

Propiedad VII (integración por partes) Dados $u(x)$ y $v(x)$,

$$\int_{x=a}^{x=b} v\, du = uv\Big|_{x=a}^{x=b} - \int_{x=a}^{x=b} u\, dv$$

Estas últimas cuatro propiedades, todas prestadas de las reglas de la integración indefinida, no deben requerir mayor explicación.

Otra visión de la integral indefinida

Introducimos la integral definida mediante la asignación de dos límites de integración a una integral indefinida. Ahora que conocemos el significado de la integral definida, veamos cómo podemos regresar de lo último a la integral indefinida.

Suponga que, en lugar de fijar el límite superior de integración en b, permitimos que sea una variable, designada simplemente como x. Entonces, la integral adoptará la forma

$$\int_a^x f(x)\, dx = F(x) - F(a)$$

la cual, siendo ahora una función de x, denota un área *variable* bajo la curva de $f(x)$. Pero como el último término de la derecha es una constante, esta integral debe ser un miembro de la familia de funciones primitivas de $f(x)$, que anteriormente denotamos como $F(x) + c$. Si hacemos $c = -F(a)$, entonces la integral anterior se convierte exactamente en la integral indefinida $\int f(x)\, dx$.

Desde este punto de vista, podemos considerar que el símbolo $\int$ tiene el mismo significado que $\int_a^x$, siempre que se entienda que en esta última versión del símbolo el límite inferior de integración se relaciona con la constante de integración mediante la ecuación $c = -F(a)$.

EJERCICIO 14.3

1. Evalúe lo siguiente:

(*a*) $\int_1^3 \frac{1}{2}x^2\, dx$

(*b*) $\int_0^1 x(x^2 + 6)\, dx$

(*c*) $\int_1^3 3\sqrt{x}\, dx$

(*d*) $\int_2^4 (x^3 - 6x^2)\, dx$

(*e*) $\int_{-1}^1 (ax^2 + bx + c)\, dx$

(*f*) $\int_4^2 x^2\left(\frac{1}{3}x^3 + 1\right) dx$

2. Evalúe lo siguiente:

(*a*) $\int_1^2 e^{-2x}\, dx$

(*b*) $\int_{-1}^{e-2} \frac{dx}{x+2}$

(*c*) $\int_2^3 (e^{2x} + e^x)\, dx$

(*d*) $\int_e^6 \left(\frac{1}{x} + \frac{1}{1+x}\right) dx$

3. En la figura 14.1*a*, tome el valor más bajo de la función que se alcance en cada subintervalo como la altura del bloque rectangular, es decir, tome $f(x_2)$ en lugar de $f(x_1)$ como la altura del primer bloque, aunque todavía se retenga a Δx_1 como el ancho y haga lo mismo con los otros bloques.
   (*a*) Escriba una expresión sumatoria para el área total A^{**} de los nuevos rectángulos.
   (*b*) ¿A^{**} sobrestima o subestima al área deseada A?
   (*c*) ¿A^{**} tendería a aproximarse o a desviarse aún más de A si se introdujera una segmentación más fina de $[a, b]$? (*sugerencia*: intente con un diagrama).
   (*d*) En el límite, cuando el número n de subintervalos se aproxima a ∞, ¿se aproximaría el valor A^{**} al valor verdadero A, así como sucedió con el valor de aproximación A^*?
   (*e*) ¿Qué puede concluir de (*a*) a (*d*) acerca de la integrabilidad de Riemann de la función $f(x)$ en la figura 14.1*a*?
4. Se dice que la integral definida $\int_a^b f(x)\,dx$ representa un área bajo la curva. ¿Se refiere esta curva a la gráfica del integrando $f(x)$ o de la función primitiva $F(x)$? Si trazamos la gráfica de la función $F(x)$, ¿cómo podemos mostrar la integral definida dada en ella (por un área, un segmento de línea o un punto)?
5. Verifique que una constante c puede expresarse en forma equivalente como una integral definida:

   (*a*) $c \equiv \int_0^b \frac{c}{b}\,dx$ (*b*) $c \equiv \int_0^c 1\,dt$

14.4 Integrales impropias

Se dice que ciertas integrales son "impropias". Discutiremos brevemente dos tipos.

Límites infinitos de integración

Cuando tenemos integrales definidas de la forma

$$\int_a^\infty f(x)\,dx \qquad \text{y} \qquad \int_{-\infty}^b f(x)\,dx$$

con un límite de integración que es infinito, las llamamos *integrales impropias*. En estos casos, no se puede evaluar las integrales como, respectivamente,

$$F(\infty) - F(a) \qquad \text{y} \qquad F(b) - F(-\infty)$$

porque ∞ no es un número y, por lo tanto, no puede sustituir a x en la función $F(x)$. En lugar de esto, debemos recurrir una vez más al concepto de límites.

La primera integral impropia que citamos puede definirse como el límite de otra integral (propia) cuando el límite superior de integración de la última tiende a ∞, es decir,

$$\int_a^\infty f(x)\,dx \equiv \lim_{b\to\infty} \int_a^b f(x)\,dx \tag{14.8}$$

Si este límite existe, se dice que la integral impropia es convergente (o que converge), y el proceso de cálculo de los límites va a suministrar el valor de la integral. Si el límite no existe, se dice que la integral impropia es divergente y, de hecho, no tiene sentido. Por las mismas razones, podemos definir

$$\int_{-\infty}^b f(x)\,dx \equiv \lim_{a\to-\infty} \int_a^b f(x)\,dx \tag{14.8'}$$

con el mismo criterio de convergencia y divergencia.

Ejemplo 1 Evalúe $\int_1^\infty \frac{dx}{x^2}$. Primero observamos que

$$\int_1^b \frac{dx}{x^2} = \frac{-1}{x}\Big]_1^b = \frac{-1}{b} + 1$$

entonces, en línea con (14.8), la integral deseada es

$$\int_1^\infty \frac{dx}{x^2} = \lim_{b\to\infty} \int_1^b \frac{dx}{x^2} = \lim_{b\to\infty} \left(\frac{-1}{b} + 1\right) = 1$$

Esta integral impropia converge y tiene un valor de 1.

Como la expresión del límite es engorrosa para escribirla, algunas personas prefieren omitir la notación de "lím" y escribir simplemente

$$\int_1^\infty \frac{dx}{x^2} = \frac{-1}{x}\Big]_1^\infty = 0 + 1 = 1$$

Sin embargo, aun cuando esté escrita de esta forma, la integral impropia debemos interpretarla con el concepto de límite en mente.

Gráficamente, esta integral impropia todavía tiene la connotación de un área. Pero ya que permitimos que el límite superior de integración adopte valores crecientemente mayores en este caso, la frontera del lado derecho debemos prolongarla indefinidamente hacia la izquierda, como se muestra en la figura 14.4*a*. A pesar de esto, podemos considerar que el área tiene el valor definido (límite) de 1.

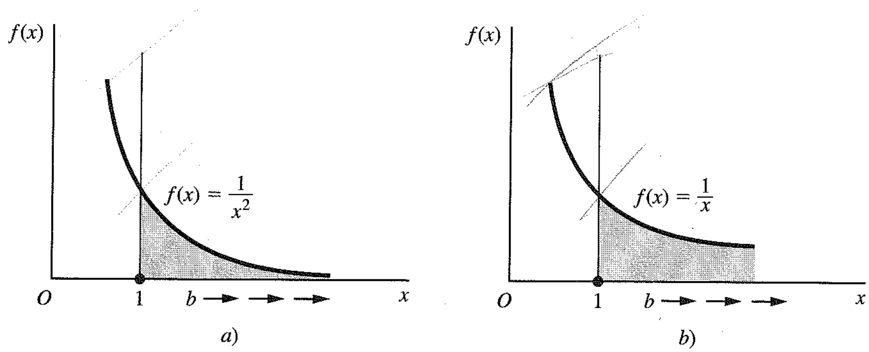
Ejemplo 2 Evalúe $\int_1^\infty \frac{dx}{x}$. Como antes, encontramos primero

$$\int_1^b \frac{dx}{x} = \ln x\Big]_1^b = \ln b - \ln 1 = \ln b$$

Cuando hacemos $b \to \infty$, por (10.16′) tenemos $\ln b \to \infty$. Entonces, la integral impropia dada es divergente.

La figura 14.4*b* muestra la gráfica de la función $1/x$, así como el área correspondiente a la integral dada. La prolongación indefinida hacia la izquierda de la frontera del lado derecho conducirá esta vez a un área infinita, aun cuando la forma de la gráfica exhibe una similitud superficial a la de la figura 14.4*a*.

FIGURA 14.4



¿Qué pasa si ambos límites de integración son infinitos? Una prolongación directa de (14.8) y (14.8′) sugeriría la definición

$$\int_{-\infty}^{\infty} f(x)\,dx = \lim_{\substack{b\to+\infty \\ a\to-\infty}} \int_a^b f(x)\,dx \qquad \textbf{(14.8'')}$$

Nuevamente, se dice que esta integral impropia converge si y sólo si existe el límite en cuestión.

Integrando infinito

Aún con límites finitos de integración, una integral todavía puede ser impropia si el integrando se hace infinito en alguna parte del intervalo de integración $[a, b]$. Para evaluar esta integral, nuevamente debemos hacer uso del concepto de límite.

Ejemplo 3 Evalúe $\int_0^1 \frac{1}{x}\,dx$. Esta integral es impropia debido a que, como lo muestra la figura 14.4*b*, el integrando es infinito para el límite inferior de integración ($1/x \to \infty$ cuando $x \to 0^+$). Por lo tanto, primero debemos hallar la integral

$$\int_a^1 \frac{1}{x}\,dx = \ln x\Big]_a^1 = \ln 1 - \ln a = -\ln a \qquad [\text{para } a > 0]$$

y luego evaluar su límite cuando $a \to 0^+$:

$$\int_0^1 \frac{1}{x} dx \equiv \lim_{a\to 0^+} \int_a^1 \frac{1}{x}\,dx = \lim_{a\to 0^+} (-\ln a)$$

Como este límite no existe (cuando $a \to 0^+$, $\ln a \to -\infty$), la integral dada es divergente.

Ejemplo 4 Evalúe $\int_0^9 x^{-1/2}\,dx$. Cuando $x \to 0^+$, el integrando $1/\sqrt{x}$ se hace infinito; la integral es impropia. Nuevamente podemos encontrar primero

$$\int_a^9 x^{-1/2}\,dx = 2x^{1/2}\Big]_a^9 = 6 - 2\sqrt{a}$$

El límite de esta expresión cuando $a \to 0^+$ es $6 - 0 = 6$. Entonces, la integral dada es convergente (a 6).

La situación en la cual el integrando se hace infinito para el límite *superior* de integración es similar. Sin embargo, es una proposición totalmente diferente, cuando un valor infinito del integrando se presenta en el intervalo abierto (a, b) en vez de a o b. En este caso, es necesario aprovechar la aditividad de las integrales definidas y descomponer primero la integral dada en subintegrales. Suponga que $f(x) \to \infty$ cuando $x \to p$, donde p es un punto en el intervalo (a, b); entonces, por la propiedad de aditividad, tenemos

$$\int_a^b f(x)\,dx = \int_a^p f(x)\,dx + \int_p^b f(x)\,dx$$

La integral dada a la izquierda puede considerarse como convergente si y sólo si cada subintegral tiene un límite.

Ejemplo 5 Evalúe $\int_{-1}^{1} \frac{1}{x^3}\, dx$. El integrando tiende a infinito cuando x se aproxima a cero; entonces, debemos escribir la integral dada como la suma

$$\int_{-1}^{1} x^{-3}\, dx = \int_{-1}^{0} x^{-3}\, dx + \int_{0}^{1} x^{-3}\, dx \text{ (digamos, } \equiv I_1 + I_2)$$

La integral I_1 es divergente porque

$$\lim_{b\to 0^-} \int_{-1}^{b} x^{-3}\, dx = \lim_{b\to 0^-} \left[\frac{-1}{2} x^{-2}\right]_{-1}^{b} = \lim_{b\to 0^-} \left(-\frac{1}{2b^2} + \frac{1}{2}\right) = -\infty$$

Entonces, podemos concluir inmediatamente, sin tener que evaluar I_2, que la integral dada es divergente.

EJERCICIO 14.4

1. Verifique las integrales definidas dadas en los ejercicios 14.3-1 y 14.3-2 para determinar si alguna es impropia. Si es impropia, indique qué variedad de integral impropia es.
2. ¿Cuál de las siguientes integrales es impropia y por qué?

   (*a*) $\int_{0}^{\infty} e^{-rt}\, dt$ (*d*) $\int_{-\infty}^{0} e^{rt}\, dt$

   (*b*) $\int_{2}^{3} x^4\, dx$ (*e*) $\int_{1}^{5} \frac{dx}{x-2}$

   (*c*) $\int_{0}^{1} x^{-2/3}\, dx$ (*f*) $\int_{-3}^{4} 6\, dx$
3. Evalúe todas las integrales *impropias* del problema 2.
4. Evalúe la integral I_2 del ejemplo 5 y muestre que también es divergente.
5. (*a*) Grafique la función $y = ce^{-t}$ para t no negativo, $(c > 0)$, y sombree el área que está bajo la curva.

   (*b*) Escriba una expresión matemática para esta área y determine si es finita.

14.5 Algunas aplicaciones de las integrales a la economía

Las integrales se usan de diferentes maneras en el análisis económico. En esta sección ilustraremos unas cuantas aplicaciones sencillas y luego mostraremos la aplicación al modelo de crecimiento de Domar en la sección 14.6.

Desde una función marginal a una función total

Dada una función total (por ejemplo, una función de costo total), el proceso de diferenciación puede proporcionarnos la función marginal (por ejemplo, la función de costo marginal). Debido a que el proceso de integración es el opuesto de la diferenciación, debe permitirnos, al contrario, inferir la función total a partir de una función marginal dada.

Ejemplo 1 Si el costo marginal (CM) de una compañía es la siguiente función de la cantidad producida, $C'(Q) = 2e^{0.2Q}$, y si el costo fijo es $C_F = 90$, encuentre la función de costo total $C(Q)$. Al integrar $C'(Q)$ respecto a Q, encontramos que

$$\int 2e^{0.2Q}\, dQ = 2\frac{1}{0.2}e^{0.2Q} + c = 10e^{0.2Q} + c \qquad \textbf{(14.9)}$$

Este resultado podemos tomarlo como la función deseada $C(Q)$ excepto que, en vista de la constante arbitraria c, la respuesta aparece como indeterminada. Afortunadamente, la información de que $C_F = 90$ podemos usarla como una condición inicial para determinar el valor de la constante. Cuando $Q = 0$, el costo total C va a consistir únicamente en C_F. Haciendo $Q = 0$ en el resultado de (14.9), debemos obtener un valor de 90; es decir, $10e^0 + c = 90$. Pero esto implicaría que $c = 90 - 10 = 80$. Entonces, la función de costo total es

$$C(Q) = 10e^{0.2Q} + 80$$

Observe que, a diferencia del caso de (14.2), donde la constante arbitraria c tiene el mismo valor que el valor inicial de la variable $H(0)$, en este ejemplo tenemos $c = 80$ pero $C(0) \equiv C_F = 90$, de modo que los dos toman valores diferentes. En general, no debemos suponer que la constante arbitraria c siempre será igual al valor inicial de la función total.

Ejemplo 2 Si la propensión marginal al ahorro (MPS) es la siguiente función de ingreso, $S'(Y) = 0.3 - 0.1Y^{-1/2}$, y si el ahorro agregado S es despreciable cuando el ingreso Y es 81, encuentre la función de ahorro $S(Y)$. Como la MPS es la derivada de la función S, este problema requiere la integración de $S'(Y)$:

$$S(Y) = \int (0.3 - 0.1Y^{-1/2})\, dY = 0.3Y - 0.2Y^{1/2} + c$$

El valor específico de la constante c puede deducirse del hecho de que $S = 0$ cuando $Y = 81$. Aun así, hablando estrictamente, ésta no es una condición *inicial* (no relacionada con $Y = 0$), la sustitución de esta información en la integral anterior servirá de cualquier modo para determinar a c. Como

$$0 = 0.3(81) - 0.2(9) + c \qquad \Rightarrow \qquad c = -22.5$$

la función de ahorro deseada es

$$S(Y) = 0.3Y - 0.2Y^{1/2} - 22.5$$

La técnica ilustrada en los ejemplos 1 y 2 podemos extenderla directamente a otros problemas que impliquen la búsqueda de funciones totales (tales como el ingreso total y el consumo total) a partir de funciones marginales dadas. Conviene reiterar que en problemas de este tipo la validez de la respuesta (una integral) siempre podemos verificarla por diferenciación.

La inversión y la formación de capital

La formación de capital es el proceso de aumentar un stock dado de capital. Considerando este proceso como continuo en el tiempo, podemos expresar al stock de capital como una función del tiempo, $K(t)$, y usar la derivada dK/dt para denotar la tasa de formación de capital.[2] Pero

[2] Acerca de la notación, la derivada de una variable respecto al tiempo con frecuencia se denota también con un punto colocado arriba de la variable, tal como $\dot{K} \equiv dK/dt$. En el análisis dinámico, donde las derivadas respecto al *tiempo* ocurren en abundancia, este símbolo más conciso puede contribuir en forma sustancial a la simplicidad en la notación. Sin embargo, un punto, por ser una marca imperceptible, se pierde de vista o se coloca erróneamente con facilidad; entonces, se requiere mucho cuidado al usar este símbolo.

la tasa de formación de capital para el tiempo t es idéntica a la tasa de flujo de *inversión neta* para el tiempo t, denotada por $I(t)$. Así, el stock de capital K y la inversión neta I se relacionan por las dos siguientes ecuaciones:

$$\frac{dK}{dt} \equiv I(t)$$

y
$$K(t) = \int I(t)\,dt = \int \frac{dK}{dt}\,dt = \int dK$$

La primera de estas ecuaciones es una identidad y muestra que la inversión neta y el incremento de capital significan lo mismo. Como $I(t)$ es la derivada de $K(t)$, la razón nos dice que $K(t)$ es la integral o antiderivada de $I(t)$, como se muestra en la segunda ocasión. La transformación del integrando en la última ecuación también es fácil de entender: el cambio de I a dK/dt es por definición, y la siguiente transformación es por cancelación de dos diferenciales idénticas, es decir, por la regla de la sustitución.

Algunas veces se usa en un modelo el concepto de *inversión bruta* junto con el de inversión neta. Si se denota la inversión bruta con I_g y la inversión neta con I, podemos relacionarlas entre sí mediante la ecuación

$$I_g = I + \delta K$$

donde δ representa la tasa de depreciación del capital y δK, la tasa de *inversión de reemplazo*.

Ejemplo 3 Suponga que el flujo de inversión neta lo describe la ecuación $I(t) = 3t^{1/2}$ y que el capital inicial, para el instante $t = 0$, es $K(0)$. ¿Cuál es la trayectoria de tiempo del capital K? Al integrar $I(t)$ respecto a t, obtenemos

$$K(t) = \int I(t)\,dt = \int 3t^{1/2}\,dt = 2t^{3/2} + c$$

Enseguida, si hacemos $t = 0$ en las expresiones de extrema izquierda y de extrema derecha, encontramos $K(0) = c$. Por lo tanto, la trayectoria de tiempo de K es

$$K(t) = 2t^{3/2} + K(0) \tag{14.10}$$

Observe la similitud básica que hay entre los resultados de (14.10) y (14.2″).

El concepto de integral definida entra en escena cuando deseamos encontrar la cantidad de formación de capital durante un intervalo (en vez de la trayectoria de tiempo de K). Como $\int I(t)\,dt = K(t)$, podemos escribir la integral definida

$$\int_a^b I(t)\,dt = K(t)\Big]_a^b = K(b) - K(a)$$

para indicar la acumulación total de capital durante el intervalo $[a, b]$. Por supuesto, esto también representa un área debajo de la curva $I(t)$. Sin embargo, debe observarse que en la gráfica de la función $K(t)$, esta integral definida aparecería como una distancia vertical: más específicamente como la diferencia entre las dos distancias verticales $K(b)$ y $K(a)$ (vea el ejercicio 14.3-4).

Para apreciar la diferencia entre $K(t)$ e $I(t)$ con mayor detalle, enfaticemos que el capital K es un concepto de *abasto*, mientras que la inversión I es un concepto de *flujo*. De acuerdo con esto, mientras que $K(t)$ nos habla de la *cantidad* de K existente para cada punto en el tiempo, $I(t)$ nos da la información acerca de la *tasa* de inversión (neta) por año (o por periodo) que prevalece para cada punto en el tiempo. Entonces, con objeto de calcular la *cantidad* de inversión neta considerada (capital acumulado), debemos especificar primero la longitud del intervalo que interviene. Este hecho también puede verse cuando reescribimos la identidad $dK/dt \equiv I(t)$ como $dK \equiv I(t)\, dt$, que establece que dK, el incremento en K, se basa no solamente en $I(t)$, la tasa de flujo, sino también en dt, el tiempo transcurrido. Esta necesidad de especificar el intervalo en la expresión $I(t)\, dt$ hace entrar en escena la integral definida, y hace surgir la representación de *área* bajo la curva $I(t)$ en contraste con la curva $K(t)$.

Ejemplo 4

Si la inversión neta es un flujo constante para $I(t) = 1\,000$ (dólares por año), ¿cuál será la inversión neta total (formación de capital) durante un año, de $t = 0$ a $t = 1$? Obviamente que la respuesta es $1 000; esto puede obtenerse formalmente de la siguiente manera:

$$\int_0^1 I(t)\, dt = \int_0^1 1\,000\, dt = 1\,000t\Big]_0^1 = 1\,000$$

Usted puede verificar que va a surgir la misma respuesta si, en lugar de lo anterior, el año considerado va de $t = 1$ a $t = 2$.

Ejemplo 5

Si $I(t) = 3t^{1/2}$ (miles de dólares por año) —un flujo no constante— ¿cuál será la formación de capital durante el intervalo [1, 4], es decir, durante el segundo, el tercero y el cuarto años? La respuesta radica en la integral definida

$$\int_1^4 3t^{1/2}\, dt = 2t^{3/2}\Big]_1^4 = 16 - 2 = 14$$

Basándonos en los ejemplos anteriores, podemos expresar la cantidad de acumulación de capital durante el intervalo $[0, t]$, para cualquier tasa de inversión $I(t)$, mediante la integral definida

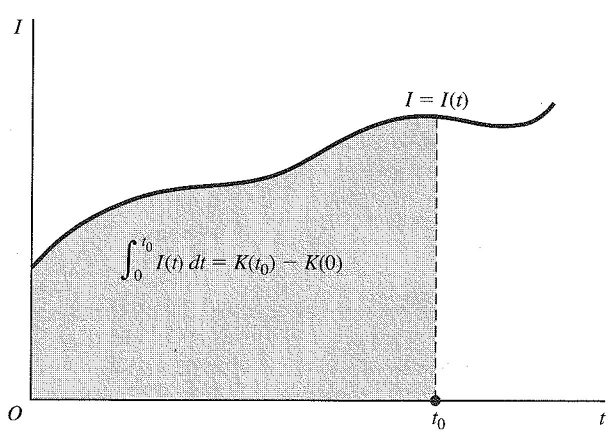
$$\int_0^t I(t)\, dt = K(t)\Big]_0^t = K(t) - K(0)$$

La figura 14.5 ilustra el caso del intervalo $[0, t_0]$. Desde un punto de vista diferente, la ecuación anterior aporta la siguiente expresión para la trayectoria de tiempo $K(t)$:

$$K(t) = K(0) + \int_0^t I(t)\, dt$$

La cantidad de K para cualquier instante t es el capital inicial más la acumulación total de capital que se ha dado desde entonces.

FIGURA 14.5



El valor presente de un flujo de efectivo

Nuestra discusión anterior del descuento y del valor presente, limitada al caso de un valor futuro V *individual*, nos condujo a las fórmulas de descuento

$$A = V(1+i)^{-t} \qquad \text{[caso } discreto\text{]}$$

y

$$A = Ve^{-rt} \qquad \text{[caso } continuo\text{]}$$

Suponga ahora que tenemos una corriente o un flujo de valores futuros: una serie de rentas cobrables en diferentes instantes o de desembolsos de costo pagaderos en diferentes instantes. ¿Cómo calculamos el valor presente de la corriente de efectivo, o flujo de efectivo completo?

En el caso *discreto*, si suponemos tres cifras de renta futuras R_t $(t = 1, 2, 3)$ disponibles al final del i-ésimo año y también suponemos una tasa de interés de i por año, los valores presentes de R_t serán, respectivamente,

$$R_1(1+i)^{-1} \qquad R_2(1+i)^{-2} \qquad R_3(1+i)^{-3}$$

De lo que se desprende que el valor presente total es la suma

$$\Pi = \sum_{t=1}^{3} R_t(1+i)^{-t} \tag{14.11}$$

(Π es la letra griega mayúscula pi, que aquí significa *presente*). Esto difiere de la fórmula de valor individual solamente por el reemplazo de V por R_t y por la inserción del signo Σ.

La idea de la suma trasciende rápidamente al caso de un flujo de efectivo continuo, pero en el último contexto el símbolo Σ debe dar paso, por supuesto, al signo de la integral definida. Considere una corriente continua de rentas a una tasa de $R(t)$ dólares por año. Esto significa que para $t = t_1$ la tasa de flujo es $R(t_1)$ dólares por año; pero para otro punto en el tiempo $t = t_2$ la tasa será $R(t_2)$ dólares por año, con t considerado como una variable continua. Para cualquier punto en el tiempo, la cantidad de renta durante el intervalo $[t, t + dt]$ puede escribirse como $R(t)\,dt$ [vea la discusión anterior de $dK \equiv I(t)\,dt$]. Cuando se descuente continuamente a una tasa de r por año, su valor presente deberá ser $R(t)e^{-rt}\,dt$. Si nuestro problema es encontrar el valor presente total de una corriente de 3 años, nuestra respuesta se encuentra en la siguiente integral definida:

$$\Pi = \int_0^3 R(t)e^{-rt}\,dt \tag{14.11'}$$

Esta expresión, la versión continua de la suma de (14.11), difiere de la fórmula de valor individual solamente en el reemplazo de V por $R(t)$ y en la adición del signo de la integral definida.[3]

[3] Puede observar que, mientras que el índice superior de la sumatoria y el límite superior de integración son idénticos a 3, el índice inferior de la sumatoria 1 difiere del límite inferior de integración 0. Esto se debe a que el primer ingreso en el flujo discreto, por hipótesis, no va a generarse hasta que $t = 1$ (al final del primer año), pero se supone que el flujo de ingreso en el caso continuo comienza inmediatamente después que $t = 0$.

Ejemplo 6 ¿Cuál es el valor presente de un flujo continuo de ingresos que dura y años para una tasa constante de D dólares por año y que se descuenta a una tasa de r por año? De acuerdo con (14.11′), tenemos

$$\begin{aligned}\Pi &= \int_0^y De^{-rt}\,dt = D\int_0^y e^{-rt}\,dt = D\left[\frac{-1}{r}e^{-rt}\right]_0^y \\ &= \frac{-D}{r}e^{-rt}\Big]_{t=0}^{t=y} = \frac{-D}{r}(e^{-ry}-1) = \frac{D}{r}(1-e^{-ry}) \qquad \textbf{(14.12)}\end{aligned}$$

Entonces, Π depende de D, r y y. Si $D = \$3\ 000$, $r = 0.06$ y $y = 2$, por ejemplo, tenemos

$$\Pi = \frac{3\ 000}{0.06}(1-e^{-0.12}) = 50\ 000(1-0.8869) = \$5\ 655 \qquad \text{[aproximadamente]}$$

Naturalmente que el valor de Π siempre es positivo; esto se infiere de la positividad de D y de r, así como $(1 - e^{-ry})$. (El número e elevado a cualquier potencia negativa siempre va a dar un valor positivo menor que la unidad, como puede verse en el segundo cuadrante en la figura 10.3*a*.)

Ejemplo 7 En el problema de almacenaje de vino de la sección 10.6 supusimos un costo cero de almacenaje. Esta hipótesis simplificadora se debió a nuestra ignorancia de una manera de calcular el valor presente de un flujo de costos. Sin embargo, ahora estamos listos para permitir al distribuidor de vinos que ejerza los costos de almacenaje.

Sea el costo de compra de una caja de vino una cantidad C, que se ejerce en tiempo presente. Su valor de venta (futuro), que varía con el tiempo, generalmente puede denotarse como $V(t)$, siendo su valor presente $V(t)e^{-rt}$. Mientras que el valor de venta representa un valor futuro individual (puede haber solamente una transacción de venta en este caso del vino), el costo de almacenaje es una corriente. Suponiendo que este costo sea una corriente constante para una tasa de s dólares por año, el valor presente total del costo de almacenaje ejercido para un total de t años es igual a

$$\int_0^t se^{-rt}\,dt = \frac{s}{r}(1-e^{-rt}) \qquad \text{[(vea (14.12)]}$$

Entonces, el valor presente *neto* —lo que el distribuidor buscaría maximizar— podemos expresarlo como

$$N(t) = V(t)e^{-rt} - \frac{s}{r}(1-e^{-rt}) - C = \left[V(t)+\frac{s}{r}\right]e^{-rt} - \frac{s}{r} - C$$

que es una función objetivo para una variable de elección individual t.

Para maximizar $N(t)$, debemos escoger el valor de t de modo que $N'(t) = 0$. Esta primera derivada es

$$\begin{aligned}N'(t) &= V'(t)e^{-rt} - r\left[V(t)+\frac{s}{r}\right]e^{-rt} \qquad \text{[regla del producto]} \\ &= [V'(t) - rV(t) - s]e^{-rt}\end{aligned}$$

y será cero si y sólo si

$$V'(t) = rV(t) + s$$

Entonces, esta última ecuación podemos tomarla como la condición de optimización necesaria para la elección del tiempo de venta t^*.

La interpretación económica de esta condición apela fácilmente al razonamiento intuitivo: $V'(t)$ representa la tasa de cambio del valor de venta, es decir, el incremento en V, si la venta se pospone por un año, mientras que los dos términos de la derecha indican, respectivamente, los incrementos en el costo del interés y en el costo de almacenaje por esa posposición de la venta (el ingreso y el costo son ambos considerados en el instante t^*). Entonces, la idea de igualar ambos lados para nosotros es simplemente algún "vino viejo en una botella nueva" ¡ya que no es otra cosa más que la misma condición MC = MR con una presentación diferente!

El valor presente de un flujo perpetuo

Si un flujo de efectivo fuera a persistir para siempre —una situación ejemplificada por el interés proveniente de un título perpetuo o de la renta de un bien de capital indestructible tal como un terreno— el valor presente del flujo sería

$$\Pi = \int_0^\infty R(t)e^{-rt}\,dt$$

lo cual es una integral impropia.

Ejemplo 8 Encuentre el valor presente de un flujo perpetuo de ingreso a una tasa uniforme de D dólares por año, si la tasa continua de descuento es r. Como al evaluar una integral impropia, simplemente tomamos el límite de una integral propia, el resultado de (14.12) todavía puede ser útil. Específicamente podemos escribir

$$\Pi = \int_0^\infty De^{-rt}\,dt = \lim_{y\to\infty}\int_0^y De^{-rt}\,dt = \lim_{y\to\infty}\frac{D}{r}(1-e^{-ry}) = \frac{D}{r}$$

Observe que el parámetro y (número de años) ha desaparecido de la respuesta final. Esto es como debe ser, ya que aquí manejamos un flujo *permanente*. También puede observar que nuestro resultado (valor presente = tasa de flujo de ingresos ÷ tasa de descuento) corresponde precisamente a la fórmula conocida para la llamada capitalización de un bien con un rendimiento perpetuo.

EJERCICIO 14.5

1. Dadas las siguientes funciones de ingreso marginal:
   (*a*) $R'(Q) = 28Q - e^{0.3Q}$ (*b*) $R'(Q) = 10(1+Q)^{-2}$
   encuentre en cada caso la función de ingreso total $R(Q)$. ¿Qué condición inicial puede introducir para determinar la constante de integración?
2. (*a*) Dada la propensión marginal a la importación $M'(Y) = 0.1$, y la información de que $M = 20$ cuando $Y = 0$, encuentre la función de importación $M(Y)$.
   (*b*) Dada la propensión marginal al consumo $C'(Y) = 0.8 + 0.1Y^{-1/2}$ y la información de que $C = Y$ cuando $Y = 100$, encuentre la función de consumo $C(Y)$.
3. Suponga que la tasa de inversión la describe la función $I(t) = 12t^{1/3}$ y que $K(0) = 25$:
   (*a*) Encuentre la trayectoria de tiempo del capital K.
   (*b*) Encuentre el monto de la acumulación de capital durante los intervalos [0, 1] y [1, 3], respectivamente.
4. Dada una corriente de ingreso continua a la tasa constante de \$1 000 por año:
   (*a*) ¿Cuál será el valor presente Π si si el flujo de ingreso dura 2 años y la tasa continua de descuento es 0.05 por año?
   (*b*) ¿Cual será el valor presente Π si la corriente de ingreso termina exactamente después de 3 años y la tasa de descuento es 0.04?
5. ¿Cuál es el valor presente de un flujo perpetuo de efectivo de:
   (*a*) \$1 450 por año, descontado a $r = 5\%$?
   (*b*) \$2 460 por año, descontado a $r = 8\%$?

14.6 El modelo de crecimiento de Domar

En el problema de crecimiento de la población de (14.1) y (14.2) y en el problema de la formación de capital de (14.10), el objetivo común es delinear una trayectoria de tiempo basándonos en algún patrón dado de cambio una variable. Por otro lado, en el modelo de crecimiento clásico del profesor Domar[4] la idea es estipular el tipo de trayectoria de tiempo que se requiere que prevalezca si debe satisfacerse una cierta condición de equilibrio de la economía.

Marco de análisis

Las premisas básicas del modelo de Domar son las siguientes:

1. Cualquier cambio del flujo de la tasa de inversión por año $I(t)$ va a producir un efecto doble: va a afectar a la demanda agregada, así como a la capacidad de producción de la economía.
2. Un cambio en $I(t)$ tiene un efecto en la demanda a través de un proceso multiplicador que se supone actúa instantáneamente. Entonces, un incremento en $I(t)$ elevará la tasa del flujo de ingreso por año $Y(t)$ por un múltiplo del incremento de $I(t)$. El multiplicador es $k = 1/s$, donde s representa la propensión marginal al ahorro (constante). Con la hipótesis de que $I(t)$ es el único flujo de desembolso (paramétrico) que influye en la tasa de flujo de ingreso, podemos afirmar que
$$\frac{dY}{dt} = \frac{dI}{dt}\frac{1}{s} \tag{14.13}$$
3. El efecto de capacidad de la inversión debe medirse por el cambio de la tasa de producción *potencial* que la economía es capaz degenerar. Suponiendo una relación constante de capacidad-capital, podemos escribir
$$\frac{\kappa}{K} \equiv \rho \qquad (= \text{constante})$$
donde κ (la letra griega minúscula kappa) representa la capacidad o el potencial del flujo de producción por año, y ρ (la letra griega minúscula rho: erre) denota la relación capacidad-capital. Esto implica que con un capital $K(t)$ la economía es potencialmente capaz de producir un producto o ingreso anual, igual a $\kappa \equiv \rho K$ dólares. Observe que, a partir de $\kappa \equiv \rho K$ (la función de producción), se sigue que $d\kappa = \rho\, dK$, y
$$\frac{d\kappa}{dt} = \rho\frac{dK}{dt} = \rho I \tag{14.14}$$

En el modelo de Domar, el equilibrio se define como una situación en la cual se emplea toda la capacidad de producción. Por lo tanto, el alcanzar un equilibrio requiere que la demanda agregada sea exactamente igual a la producción potencial que puede producirse en un año; es decir, $Y = \kappa$. Sin embargo, si comenzamos a partir de una situación de equilibrio, el requerimiento va a limitarse al balance de los respectivos *cambios* de la capacidad y de la demanda agregada; es decir,
$$\frac{dY}{dt} = \frac{d\kappa}{dt} \tag{14.15}$$

[4] Evsey D. Domar, "Capital Expansion, Rate of Growth, and Employment", *Econometrica*, abril de 1946, pp. 137-147; reimpreso en Domar, *Essays in the Theory of Economic Growth*, Oxford University Press, Fair Lawn, N. J., 1957, pp. 70-82.

¿Qué tipo de trayectoria de tiempo de inversión $I(t)$ puede satisfacer esta condición de equilibrio en todo momento?

Encontrando la solución

Para responder a esta pregunta, primero sustituimos (14.13) y (14.14) en la condición de equilibrio (14.15). El resultado es la siguiente ecuación diferencial:

$$\frac{dI}{dt}\frac{1}{s} = \rho I \qquad \text{o} \qquad \frac{1}{I}\frac{dI}{dt} = \rho s \tag{14.16}$$

Como (14.16) especifica un patrón de cambio definido para I, deberemos poder encontrar la trayectoria de inversión de equilibrio (o requerida) a partir de aquel.

En este caso sencillo, la solución se obtiene integrando directamente ambos lados de la segunda ecuación de (14.16) respecto a t. El hecho de que ambos lados sean idénticos para el equilibrio asegura la igualdad de sus integrales. Entonces,

$$\int \frac{1}{I}\frac{dI}{dt}\,dt = \int \rho s\,dt$$

Por la regla de la sustitución y la regla de los logaritmos, el lado izquierdo nos da

$$\int \frac{dI}{I} = \ln|I| + c_1 \qquad (I \neq 0)$$

mientras que el lado derecho aporta (siendo ρs una constante)

$$\int \rho s\,dt = \rho st + c_2$$

Al igualar los dos resultados y combinar las dos constantes, tenemos

$$\ln|I| = \rho st + c \tag{14.17}$$

Para obtener $|I|$ de $|I|$, realizamos una operación conocida como "tomar el antilog de ln $|I|$", que utiliza el hecho de que $e^{\ln x} = x$. Así, haciendo que cada lado de (14.17) se transforme en el exponente de la constante e, obtenemos

$$e^{\ln|I|} = e^{(\rho st + c)}$$

o sea $$|I| = e^{\rho st}e^{c} = Ae^{\rho st} \qquad \text{donde } A \equiv e^{c}$$

Si consideramos positiva la inversión, entonces $|I| = I$, de modo que el resultado anterior se transforma en $I(t) = Ae^{\rho st}$, donde A es arbitraria. Para eliminar esta constante arbitraria, hacemos $t = 0$ en la ecuación $I(t) = Ae^{\rho st}$, para obtener $I(0) = Ae^{0} = A$. Esto determina la constante A y nos permite expresar la solución —la trayectoria de inversión requerida— como

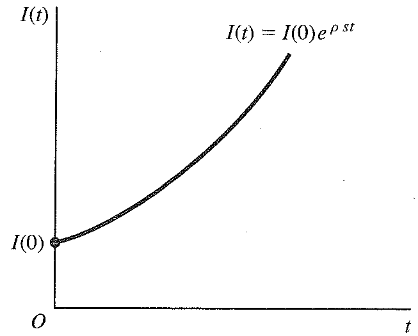
$$I(t) = I(0)e^{\rho st} \tag{14.18}$$

donde $I(0)$ denota la tasa inicial de inversión.[5]

Este resultado tiene un significado económico inquietante. Con objeto de conservar el balance entre la capacidad y la demanda en el tiempo, la tasa del flujo de inversión debe crecer precisamente según la tasa exponencial de ρs, a lo largo de una trayectoria tal como se ilustra en la figura 14.6. Es obvio que cuanto mayor sea la relación capacidad-capital o la propensión marginal al ahorro, mayor es la tasa requerida de crecimiento. Pero para cualquier tasa, una

[5] La solución (14.18) permanecerá válida aun si permitimos que la inversión sea negativa en el resultado $|I| = Ae^{\rho st}$. Vea el ejercicio 14.6-3.

FIGURA 14.6



vez que se conocen los valores de ρ y s, el crecimiento requerido de la trayectoria de inversión se establece en forma muy rígida.

El filo de la navaja

Resulta relevante preguntarnos en estos momentos qué va a suceder si la tasa *real* de crecimiento de la inversión —llamada tasa r— difiere de la tasa *requerida* ρs.

El enfoque de Domar es definir un *coeficiente de utilización*

$$u = \lim_{t\to\infty} \frac{Y(t)}{\kappa(t)} \qquad [u = 1 \text{ significa la utilización total de la capacidad}]$$

y muestra que $u = r/\rho s$, de modo que $u \gtreqless 1$ a medida que $r \gtreqless \rho s$. En otras palabras, si hay una discrepancia entre las tasas real y requerida ($r \neq \rho s$), encontraremos al final (a medida que $t \to \infty$) ya sea una escasez de la capacidad ($u > 1$) o un exceso de la capacidad ($u < 1$), dependiendo de si r es mayor o menor que ρs.

Sin embargo, podemos mostrar que la conclusión acerca de la escasez y el exceso de la capacidad realmente es aplicable a cualquier instante t, no solamente cuando $t \to \infty$. Para una tasa de crecimiento dada r implica que

$$I(t) = I(0)e^{rt} \qquad \text{y} \qquad \frac{dI}{dt} = rI(0)e^{rt}$$

Por lo tanto, mediante (14.13) y (14.14) tenemos

$$\frac{dY}{dt} = \frac{1}{s}\frac{dI}{dt} = \frac{r}{s}I(0)e^{rt}$$
$$\frac{d\kappa}{dt} = \rho I(t) = \rho I(0)e^{rt}$$

La relación entre estas dos derivadas,

$$\frac{dY/dt}{d\kappa/dt} = \frac{r}{\rho s}$$

deberá indicarnos la magnitud relativa del efecto de la creación de demanda y del efecto de la generación de capacidad de la inversión para cualquier instante t, bajo la tasa de crecimiento real de r. Si r (la tasa real) sobrepasa a ρs (la tasa requerida), entonces $dY/dt > d\kappa/dt$, y el efecto de la demanda va a sobrepasar al efecto de capacidad, causando una escasez de capacidad. Inversamente, si $r < \rho s$, habrá una deficiencia de la demanda agregada y, entonces, un exceso de capacidad.

Lo curioso acerca de esta conclusión es que si la inversión realmente crece a una tasa *mayor* que la requerida ($r > \rho s$), el resultado final será una *escasez* en vez de un exceso de capacidad. Es igualmente curioso que si el crecimiento real de la inversión está retrasado respecto a la tasa requerida ($r > \rho s$), encontraremos un *exceso* de capacidad en vez de una escasez. Debido a estos resultados paradójicos, si permitimos ahora que los empresarios ajusten la tasa real de crecimiento r (hasta ahora considerada como constante) de acuerdo con la situación prevaleciente de la capacidad, con seguridad van a hacer el tipo "equivocado" de ajuste. En el caso de $r > \rho s$, por ejemplo, la escasez emergente de capacidad va a motivar una tasa de inversión aún mayor. Pero esto significaría un incremento de r, en lugar de la reducción necesaria en estas circunstancias. En consecuencia, la discrepancia entre las dos tasas de crecimiento se intensificaría en lugar de reducirse.

La conclusión es que, dadas las constantes paramétricas ρ y s, la única manera de evitar tanto la escasez como el exceso de la capacidad de producción es guiar el flujo de inversión con mucho cuidado a lo largo de la trayectoria de equilibrio con una tasa de crecimiento $r^* = \rho s$. Y como hemos mostrado, cualquier desviación de esta trayectoria de tiempo del "filo de la navaja" acarreará una falla persistente para satisfacer la norma de la utilización completa que Domar concibió en este modelo. Tal vez éste no sea un prospecto muy alentador para tener en cuenta. Afortunadamente, se pueden lograr resultados más flexibles cuando se modifican ciertas hipótesis del modelo de Domar, como veremos en el modelo de crecimiento del profesor Solow, que vamos a estudiar en el capítulo 15.

EJERCICIO 14.6

1. ¿Cuántos factores de producción se consideran explícitamente en el modelo de Domar? ¿Qué implica este hecho respecto a la relación capital-mano de obra en la producción?
2. En la sección 10.2 aprendimos que la constante r en la función exponencial Ae^{rt} representa la tasa de crecimiento de la función. Aplique esto a (14.16) y deduzca (14.18) sin pasar por la integración.
3. Muestre que aun si permitimos que la inversión sea negativa en la ecuación $|I| = Ae^{\rho st}$, al determinar el valor de la constante arbitraria A de todas maneras terminamos en la solución (14.18).
4. Muestre que el resultado de (14.18) puede obtenerse en forma alterna al encontrar —e igualar— las integrales *definidas* de ambos lados de (14.16),
$$\frac{1}{I}\frac{dI}{dt} = \rho s$$
respecto a la variable t, con límites de integración $t = 0$ y $t = t$. Recuerde que cuando cambiamos la variable de integración de t a I, los límites de integración van a cambiar de $t = 0$ y $t = t$, respectivamente, a $I = I(0)$ e $I = I(t)$.

Capítulo 15

Tiempo continuo: ecuaciones diferenciales de primer orden

Con el modelo de crecimiento de Domar hemos resuelto una ecuación diferencial simple por integración directa. Para ecuaciones diferenciales más complicadas hay diferentes métodos de solución. Sin embargo, aun en los casos más complicados, la idea fundamental que está detrás de los métodos de solución es la de las técnicas del cálculo integral. Por esta razón, con frecuencia se denomina *integral* de esa ecuación a la solución de una ecuación diferencial.

En este capítulo sólo estudiaremos ecuaciones diferenciales de *primer orden*. En este contexto, el término *orden* se refiere al orden más alto de las derivadas (o diferenciales) que aparece en la ecuación diferencial; entonces, una ecuación diferencial de primer orden puede contener solamente la primera derivada, por decir, dy/dt.

15.1 Ecuaciones diferenciales lineales de primer orden con coeficientes constantes y términos constantes

Aunque la primera derivada dy/dt es la única que aparece en una ecuación diferencial de primer orden, también puede estar presente elevada a diversas potencias: dy/dt, $(dy/dt)^2$, o $(dy/dt)^3$. La potencia más alta que alcance la derivada en la ecuación se denomina *grado* de la ecuación diferencial. En el caso de que la derivada dy/dt aparezca sólo en primer grado, y así es con la variable dependiente y, además, no está presente ningún producto de la forma $y(dy/dt)$, entonces se dice que la ecuación es *lineal*. Así, una ecuación diferencial lineal de primer orden generalmente adoptará la forma[1]

$$\frac{dy}{dt} + u(t)y = w(t) \qquad \textbf{(15.1)}$$

[1] Observe que el término derivada dy/dt de (15.1) tiene un coeficiente unitario. Esto no implica que nunca pueda tener un coeficiente diferente de uno en la realidad, sino que cuando aparece un coeficiente de este tipo, siempre podemos "normalizar" la ecuación al dividir cada término entre dicho coeficiente. Por esta razón, la forma dada en (15.1) puede considerarse como una representación *general*.

donde u y w son dos funciones de t, así como y. Sin embargo, en contraste con dy/dt y y, no se impone ninguna restricción sobre la variable independiente t. Entonces, las funciones u y w podrían representar expresiones tales como t^2 y e^t o algunas funciones más complicadas de t; por otro lado, u y w también pueden ser constantes.

Este último punto nos conduce a una clasificación adicional. Cuando la función u (el coeficiente de la variable dependiente y) es una constante, y cuando la función w es un término aditivo constante, (15.1) se reduce al caso especial de una ecuación diferencial lineal de primer orden con *coeficiente constante y término constante*. En esta sección estudiaremos sólo esta variedad simple de ecuaciones diferenciales.

El caso homogéneo

Si u y w son funciones constantes y si w resulta ser idénticamente cero, (15.1) se transforma en

$$\frac{dy}{dt} + ay = 0 \tag{15.2}$$

donde a es alguna constante. Se dice que esta ecuación diferencial es *homogénea* debido al término constante cero (compare con los sistemas de ecuaciones homogéneas). La característica que define a una ecuación homogénea es que cuando todas las variables (aquí, dy/dt y y) se multiplican por una constante dada, la ecuación permanece válida. Esta característica es válida si el término constante es cero, pero se pierde si el término constante no es cero.

La ecuación (15.2) podemos escribirla en forma alterna como

$$\frac{1}{y}\frac{dy}{dt} = -a \tag{15.2$'$}$$

Pero usted reconocerá que la ecuación diferencial (14.16) que encontramos en el modelo de Domar es precisamente de esta forma. Por lo tanto, por analogía, podemos escribir la solución de (15.2) o (15.2′) como sigue:

$$y(t) = Ae^{-at} \qquad \text{[solución } \textit{general}\text{]} \tag{15.3}$$

o bien

$$y(t) = y(0)e^{-at} \qquad \text{[solución } \textit{definida por condición inicial}\text{]} \tag{15.3$'$}$$

En (15.3) aparece una constante arbitraria A; por lo tanto es una *solución general*. Cuando se sustituye A por algún valor específico, la solución se transforma en una *solución particular* de (15.2). Hay un número infinito de soluciones particulares, una para cada valor posible de A, incluyendo el valor $y(0)$. Sin embargo, este último valor tiene un significado especial: $y(0)$ es el único valor que puede hacer que la solución satisfaga la condición inicial. Como esto representa el resultado de determinar el valor de la constante arbitraria, nos referiremos a (15.3′) como la *solución definida por condición inicial* de la ecuación diferencial (15.2) o (15.2′).

Usted debe observar dos cosas acerca de la solución de una ecuación diferencial: (1) la solución no es un valor numérico, sino más bien una función $y(t)$ —una trayectoria de tiempo si t simboliza al tiempo— y (2) la solución $y(t)$ está libre de cualquier expresión de derivada o diferencial, de modo que tan pronto se sustituya un valor específico de t en ella, puede calcularse directamente un valor correspondiente de y.

El caso no homogéneo

Cuando una constante diferente de cero toma el lugar del cero en (15.2), tenemos una ecuación diferencial lineal *no homogénea*

$$\frac{dy}{dt} + ay = b \tag{15.4}$$

La solución de esta ecuación va a consistir en la suma de dos términos, uno de los cuales se llama la *función complementaria* (que denotamos como y_c), y el otro se conoce como la *integral particular* (denotada como y_p). Como mostraremos, cada uno de éstos tiene una interpretación económica importante. Presentaremos ahora el método de solución; su razonamiento se hará más claro posteriormente.

Aun cuando nuestro objetivo es resolver la ecuación *no* homogénea (15.4), con frecuencia tendremos que referirnos a su versión homogénea, como se muestra en (15.2). Para una referencia conveniente, a esta última la llamamos *ecuación reducida* de (15.4). La ecuación no homogénea (15.4) misma puede referirse, a su vez, como la *ecuación completa*. La función complementaria y_c no es más que la solución general de la ecuación reducida, mientras que la integral particular y_p es simplemente *cualquier* solución particular de la ecuación completa.

Nuestra discusión del caso homogéneo ya nos ha dado la solución general de la ecuación reducida, por lo tanto podemos escribir

$$y_c = Ae^{-at} \qquad \text{[mediante (15.3)]}$$

¿Qué podemos decir acerca de la integral particular? Como la integral particular es *cualquier* solución particular de la ecuación completa, podemos intentar primero el tipo más simple posible de solución, a saber, siendo y alguna constante ($y = k$). Si y es una constante, entonces se sigue que $dy/dt = 0$, y (15.4) se transforma en $ay = b$, con la solución $y = b/a$. Por lo tanto, la solución constante va a funcionar siempre que $a \neq 0$. En ese caso, tenemos

$$y_p = \frac{b}{a} \qquad (a \neq 0)$$

La suma de la función complementaria y de la integral particular constituye entonces la solución general de la ecuación completa (15.4):

$$y(t) = y_c + y_p = Ae^{-at} + \frac{b}{a} \qquad \text{[solución general, caso de } a \neq 0\text{]} \qquad \textbf{(15.5)}$$

Lo que la hace una solución general es la presencia de la constante arbitraria A. Por supuesto que podemos determinar el valor de esta constante mediante una condición inicial. Digamos que y toma el valor $y(0)$ cuando $t = 0$. Entonces, al hacer $t = 0$ en (15.5), encontramos que

$$y(0) = A + \frac{b}{a} \qquad \text{y} \qquad A = y(0) - \frac{b}{a}$$

Entonces podemos reescribir (15.5) como

$$y(t) = \left[y(0) - \frac{b}{a}\right]e^{-at} + \frac{b}{a} \qquad \left[\begin{array}{l}\text{solución definida por} \\ \text{condición inicial, caso de}\end{array}\; a \neq 0\right] \qquad \textbf{(15.5')}$$

Debemos observar que el uso de la condición inicial para determinar la constante arbitraria es —y debe ser— considerado como el paso *final*, después que hemos encontrado la solución general para la ecuación completa. Como los valores tanto de y_c como de y_p están relacionados con el valor de $y(0)$, ambos deben considerarse para hacer definitiva la constante A.

Ejemplo 1

Resuelva la ecuación $dy/dt + 2y = 6$, con la condición inicial $y(0) = 10$. Aquí, tenemos $a = 2$ y $b = 6$; entonces, mediante (15.5′), la solución es

$$y(t) = (10 - 3)e^{-2t} + 3 = 7e^{-2t} + 3$$

Ejemplo 2 Resuelva la ecuación $dy/dt + 4y = 0$, con la condición inicial $y(0) = 1$. Ya que $a = 4$ y $b = 0$, tenemos

$$y(t) = (1 - 0)e^{-4t} + 0 = e^{-4t}$$

Pudimos obtener la misma respuesta de (15.3′), la fórmula para el caso homogéneo. La ecuación homogénea (15.2) es meramente un caso especial de la ecuación no homogénea (15.4) cuando $b = 0$. En consecuencia, la fórmula (15.3′) es también un caso especial de la fórmula (15.5′) bajo la circunstancia de que $b = 0$.

¿Qué pasa si $a = 0$, de modo que la solución de (15.5′) es indefinida? En ese caso, la ecuación diferencial es de la forma extremadamente simple

$$\frac{dy}{dt} = b \tag{15.6}$$

Por integración directa, la solución general puede encontrarse rápidamente como

$$y(t) = bt + c \tag{15.7}$$

donde c es una constante arbitraria. De hecho, los dos términos componentes de (15.7) de nuevo pueden identificarse respectivamente como la función complementaria y como la integral particular de la ecuación diferencial dada. Como $a = 0$, la función complementaria puede expresarse simplemente como

$$y_c = Ae^{-at} = Ae^0 = A \qquad (A = \text{una constante arbitraria})$$

En cuanto a la integral particular, el hecho de que la solución constante $y = k$ no funciona en el presente caso de $a = 0$ sugiere que debemos intentar con una solución *no constante*. Consideremos el tipo más sencillo posible de esta última, a saber, $y = kt$. Si $y = kt$, entonces $dy/dt = k$, y la ecuación completa (15.6) se reduce a $k = b$, de modo que podemos escribir

$$y_p = bt \qquad (a = 0)$$

¡Nuestra nueva solución de prueba realmente funciona! Por lo tanto, la solución general de (15.6) es

$$y(t) = y_c + y_p = A + bt \qquad [\text{solución general, caso de } a = 0] \tag{15.7′}$$

que es idéntica al resultado de (15.7), ya que c y A no son más que notaciones alternas para una constante arbitraria. Observe, sin embargo, que en el caso presente y_c es una constante, mientras que y_p es una función del tiempo, el opuesto exacto de la situación de (15.5).

Al determinar la constante arbitraria, encontramos que la solución definida por condición inicial es

$$y(t) = y(0) + bt \qquad \begin{bmatrix} \text{solución definida por} \\ \text{condición inicial, caso de } a = 0 \end{bmatrix} \tag{15.7″}$$

Ejemplo 3 Resuelva la ecuación $dy/dt = 2$, con la condición inicial $y(0) = 5$. La solución, mediante (15.7″), es

$$y(t) = 5 + 2t$$

Verificación de la solución

La validez de todas las soluciones de las ecuaciones diferenciales siempre puede verificarse por diferenciación.

Si hacemos la prueba con la solución (15.5′), podemos obtener la derivada

$$\frac{dy}{dt} = -a\left[y(0) - \frac{b}{a}\right]e^{-at}$$

Cuando esta expresión para dy/dt y la expresión para $y(t)$ como se muestra en (15.5′) se sustituyen en el lado izquierdo de la ecuación diferencial (15.4), si la solución es correcta, ese lado debe reducirse exactamente al valor del término constante b en el lado derecho de (15.4). Al realizar esta sustitución, encontramos en verdad que

$$-a\left[y(0) - \frac{b}{a}\right]e^{-at} + a\left\{\left[y(0) - \frac{b}{a}\right]e^{-at} + \frac{b}{a}\right\} = b$$

Entonces, nuestra solución es correcta, pero siempre que también satisfaga la condición inicial. Para verificar esta última, hagamos $t = 0$ en la solución (15.5′). Como el resultado

$$y(0) = \left[y(0) - \frac{b}{a}\right] + \frac{b}{a} = y(0)$$

es una identidad, realmente se satisface la condición inicial.

Como paso final del proceso de solución de una ecuación diferencial, se recomienda que se habitúe usted a verificar la validez de su respuesta asegurándose que (1) la derivada de la trayectoria de tiempo $y(t)$ sea consistente con la ecuación diferencial dada y que (2) la solución definida satisfaga la condición inicial.

EJERCICIO 15.1

1. Encuentre y_c, y_p, la solución general, y la solución definida por la condición inicial, dados:

   (*a*) $\frac{dy}{dt} + 4y = 12;\ y(0) = 2$

   (*b*) $\frac{dy}{dt} - 2y = 0;\ y(0) = 9$

   (*c*) $\frac{dy}{dt} + 10y = 15;\ y(0) = 0$

   (*d*) $2\frac{dy}{dt} + 4y = 6;\ y(0) = 1\frac{1}{2}$

2. Verifique la validez de sus respuestas a los problemas anteriores.
3. Encuentre la solución de cada uno de los apartados siguientes usando una fórmula apropiada desarrollada en el texto:

   (*a*) $\frac{dy}{dt} + y = 4;\ y(0) = 0$

   (*b*) $\frac{dy}{dt} = 23;\ y(0) = 1$

   (*c*) $\frac{dy}{dt} - 5y = 0;\ y(0) = 6$

   (*d*) $\frac{dy}{dt} + 3y = 2;\ y(0) = 4$

   (*e*) $\frac{dy}{dt} - 7y = 7;\ y(0) = 7$

   (*f*) $3\frac{dy}{dt} + 6y = 5;\ y(0) = 0$

4. Verifique la validez de sus respuestas al problema 3.

15.2 La dinámica del precio de mercado

En el modelo de crecimiento de Domar (macro), encontramos una aplicación del caso *homogéneo* de las ecuaciones diferenciales lineales de primer orden. Para ilustrar el caso *no homogéneo*, presentemos un modelo dinámico (micro) del mercado.

El marco de referencia

Suponga que para un artículo específico las funciones de oferta y demanda son:

$$\begin{aligned} Q_d &= \alpha - \beta P \qquad (\alpha, \beta > 0) \\ Q_s &= -\gamma + \delta P \qquad (\gamma, \delta > 0) \end{aligned} \tag{15.8}$$

Entonces, de acuerdo con (3.4), el precio de equilibrio debe ser[2]

$$P^* = \frac{\alpha + \gamma}{\beta + \delta} \qquad (= \text{alguna constante positiva}) \tag{15.9}$$

Si el precio inicial $P(0)$ está precisamente al nivel de P^*, es claro que el mercado ya estará en equilibrio, y no se necesitará ningún análisis dinámico. El caso que nos interesa es $P(0) \neq P^*$, pero, se alcanza P^* (si es alcanzable) sólo después de un proceso de ajuste, durante el cual no solamente va a cambiar el precio con el tiempo sino que Q_d y Q_s, siendo funciones de P, también deben cambiar con el tiempo. A la luz de esto, las variables de precio y calidad pueden tomarse *todas* como *funciones del tiempo*.

Nuestra cuestión dinámica es ésta: dado suficiente tiempo para que el proceso de ajuste se lleve a cabo, ¿tiende el precio al nivel de equilibrio P^*?, es decir, ¿tiende la trayectoria de tiempo $P(t)$ a convergir a P^*, cuando $t \to \infty$?

La trayectoria de tiempo

Para responder la pregunta anterior debemos encontrar primero la trayectoria de tiempo $P(t)$. Pero eso, a su vez, requiere que primero se defina un patrón específico del cambio de precios. En general, los cambios de precios están controlados por la intensidad relativa de las fuerzas de la oferta y la demanda en el mercado. Supongamos por simplicidad que la tasa del cambio de precios (respecto al tiempo) en cualquier momento siempre es directamente proporcional a la *demanda excedente* $(Q_d - Q_s)$ que prevalece en ese momento. Este patrón de cambio puede expresarse simbólicamente como

$$\frac{dP}{dt} = j(Q_d - Q_s) \qquad (j > 0) \tag{15.10}$$

donde j representa un *coeficiente de ajuste* (constante). Con este patrón de cambio podemos tener $dP/dt = 0$ si y sólo si $Q_d = Q_s$. A este respecto, es provechoso observar que el término *precio de equilibrio* tiene dos sentidos: el sentido intertemporal (P es una constante en el tiempo) y el sentido de clarificación del mercado (el precio de equilibrio el iguala a Q_d con Q_s). En este modelo coinciden los dos sentidos, pero esto puede no ser verdad para todos los modelos.

En virtud de las funciones de oferta y demanda de (15.8), podemos expresar (15.10) específicamente en la forma

$$\frac{dP}{dt} = j(\alpha - \beta P + \gamma - \delta P) = j(\alpha + \gamma) - j(\beta + \delta)P$$

es decir,

$$\frac{dP}{dt} + j(\beta + \delta)P = j(\alpha + \gamma) \tag{15.10'}$$

[2] Hemos cambiado de los símbolos (a, b, c, d) de (3.4) a $(\alpha, \beta, \gamma, \delta)$ aquí para evitar cualquier confusión posible con el uso de *a* y *b* como parámetros en la ecuación diferencial (15.4) que vamos a aplicar al modelo de mercado.

Como ésta es precisamente la forma de la ecuación diferencial (15.4), y como el coeficiente de P es diferente de cero, podemos aplicar la fórmula solución (15.5′) y escribir la solución —la trayectoria de tiempo de los precios— como

$$\begin{aligned} P(t) &= \left[P(0) - \frac{\alpha+\gamma}{\beta+\delta}\right] e^{-j(\beta+\delta)t} + \frac{\alpha+\gamma}{\beta+\delta} \\ &= [P(0) - P^*]e^{-kt} + P^* \qquad [\text{por } (15.9);\ k \equiv j(\beta+\delta)] \qquad \textbf{(15.11)} \end{aligned}$$

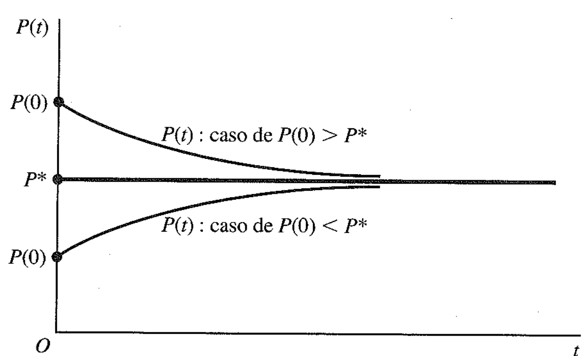
La estabilidad dinámica del equilibrio

Al final, la pregunta originalmente formulada: si $P(t) \to P^*$ cuando $t \to \infty$, equivale a la pregunta de si el primer término a la derecha de (15.11) va a tender a cero cuando $t \to \infty$. Como $P(0)$ y P^* son constantes, el factor clave será la expresión exponencial e^{-kt}. En vista del hecho de que $k > 0$, esa expresión tiende a cero cuando $t \to \infty$. En consecuencia, según la hipótesis de nuestro modelo, la trayectoria de tiempo va a conducir el precio hacia la posición de equilibrio. En una situación de este tipo, donde la trayectoria de tiempo de la variable relevante $P(t)$ *converge* al nivel P^* —interpretado aquí en su papel como el equilibrio intertemporal (en vez de clarificación de mercado)— se dice que el equilibrio es *dinámicamente estable.*

El concepto de estabilidad dinámica es importante. Examinémoslo más a fondo mediante un análisis detallado de (15.11). Dependiendo de las magnitudes relativas de $P(0)$ y P^*, la solución (15.11) comprende tres casos posibles. El primero es $P(0) = P^*$, que implica $P(t) = P^*$. En ese caso, la trayectoria de tiempo del precio puede dibujarse como la línea recta horizontal de la figura 15.1. Como mencionamos antes, la consecución del equilibrio en este caso es un hecho consumado. Segundo, podemos tener $P(0) > P^*$. En este caso, el primer término de la derecha de (15.11) es positivo, pero va a disminuir a medida que el incremento de t haga descender el valor de e^{-kt}. Entonces, la trayectoria de tiempo va a aproximarse al nivel de equilibrio P^* anterior, como lo ilustra la curva superior de la figura 15.1. Tercero, en el caso opuesto de $P(0) < P^*$, el nivel de equilibrio P^* va a aproximarse desde abajo, como lo ilustra la curva inferior de la misma figura. En general, para tener estabilidad dinámica, la *desviación* de la trayectoria de tiempo respecto al equilibrio debe ser o idénticamente cero (como en el caso 1) o decrecer uniformemente con el tiempo (como en los casos 2 y 3).

Una comparación de (15.11) con (15.5′) nos dice que el término P^*, la contraparte de b/a, no es otra cosa más que la integral particular y_p, mientras que el término exponencial es la función complementaria y_c (definida por condición inicial). Entonces, tenemos una interpretación económica de y_c y y_p; y_p representa al *nivel de equilibrio intertemporal* de la variable relevante, y y_c es la *desviación respecto al equilibrio.* La estabilidad dinámica requiere el desvanecimiento asintótico de la función complementaria cuando t se hace infinita.

FIGURA 15.1



En este modelo, la integral particular es una constante, de modo que tenemos un *equilibrio estacionario* en el sentido intertemporal, representado por P^*. Por otro lado, si la integral particular no es constante, como en (15.7′), podemos interpretarlo como un *equilibrio móvil*.

Un uso alterno del modelo

Lo que hemos hecho es analizar la estabilidad dinámica de equilibrio (la convergencia de la trayectoria de tiempo), dadas ciertas especificaciones del signo para los parámetros. Un tipo de investigación alterna es tratar de responder ¿qué restricciones específicas deben imponerse a los parámetros?, a fin de asegurar la estabilidad dinámica.

La respuesta está contenida en la solución (15.11). Si hacemos $P(0) \neq P^*$, vemos que el primer término y_c de (15.11) va a tender a cero cuando $t \to \infty$ si y sólo si $k > 0$, es decir, si y sólo si

$$j(\beta + \delta) > 0$$

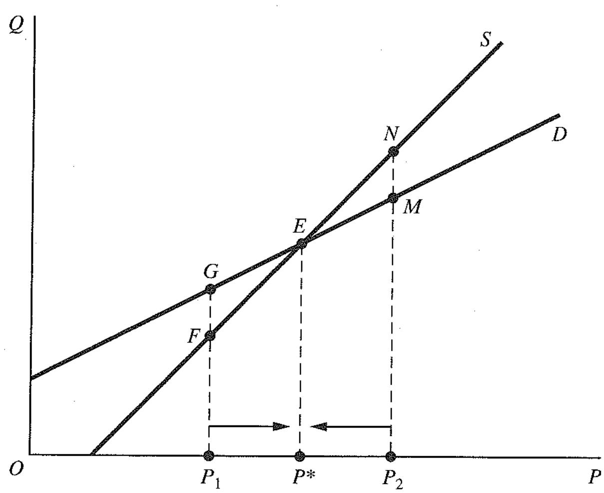
Entonces, podemos tomar esta última desigualdad como la restricción requerida para los parámetros j (el coeficiente de ajuste de los precios), β (el negativo de la pendiente de la curva de demanda, graficada con Q en el eje *vertical*) y δ (la pendiente de la curva de oferta, que se grafica en forma similar).

En el caso de que el ajuste de precio sea del tipo "normal", con $j > 0$, de modo que la demanda excedente impulse el precio hacia arriba en lugar de hacia abajo, entonces esta restricción se transforma en $(\beta + \delta) > 0$, o sea, en forma equivalente,

$$\delta > -\beta$$

En este caso, para tener estabilidad dinámica, la pendiente de la oferta debe sobrepasar a la pendiente de la demanda. Cuando tanto la oferta como la demanda tienen pendiente normal ($-\beta < 0,\ \delta > 0$), como en (15.8), obviamente se cumple este requerimiento. Pero aun si una de las curvas tiene pendiente "anómala", la condición todavía puede cumplirse, como cuando $\delta = 1$ y $-\beta = 1/2$ (demanda con pendiente positiva). Esta situación se ilustra en la figura 15.2, donde el precio de equilibrio P^* está determinado, como siempre, por el punto de intersección de las dos curvas. Si el precio inicial está en P_1, entonces Q_d (la distancia P_1G) va a sobrepasar a Q_s (la distancia P_1F), y la demanda excedente (FG) va a impulsar el precio

FIGURA 15.2



hacia arriba. Por otro lado, si el precio está inicialmente en P_2, entonces habrá una demanda excedente *negativa MN*, que va a impulsar el precio hacia abajo. Por lo tanto, como lo muestran las dos flechas de la figura, el ajuste de precios en este caso será *hacia* el equilibrio, sin importar de qué lado de P^* iniciemos. Sin embargo, debemos enfatizar que aun cuando estas flechas muestren la dirección, no tienen la capacidad de indicar la magnitud del cambio. Así, la figura 15.2 es básicamente de naturaleza estática, no dinámica, y puede servir sólo para ilustrar, no para reemplazar al análisis dinámico presentado.

EJERCICIO 15.2

1. Si tanto la oferta como la demanda de la figura 15.2 tienen pendiente negativa, ¿qué curva deberá tener mayor inclinación para que tenga estabilidad dinámica? ¿Concuerda su respuesta con el criterio $\delta > -\beta$?
2. Muestre que (15.10′) puede reescribirse como $dP/dt + k(P - P^*) = 0$. Si hacemos $P - P^* \equiv \Delta$ (lo que significa desviación), de modo que $d\Delta/dt = dP/dt$, la ecuación diferencial puede reescribirse adicionalmente como
$$\frac{d\Delta}{dt} + k\Delta = 0$$
Encuentre la trayectoria de tiempo $\Delta(t)$ y discuta la condición para estabilidad dinámica.
3. El modelo de mercado dinámico estudiado en esta sección sigue un patrón muy cercano al del estático de la sección 3.2. ¿Qué nueva característica específica es responsable de la transformación del modelo estático en el dinámico?
4. Sean la oferta y la demanda
$$Q_d = \alpha - \beta P + \sigma \frac{dP}{dt} \qquad Q_s = -\gamma + \delta P \qquad (\alpha, \beta, \gamma, \delta > 0)$$
(*a*) Suponiendo que la tasa de cambio de los precios respecto al tiempo es directamente proporcional a la demanda excedente, encuentre la trayectoria de tiempo $P(t)$ (solución general).
(*b*) ¿Cuál es el precio de equilibrio intertemporal? ¿Cuál es el precio de equilibrio de clarificación de mercado?
(*c*) ¿Qué restricción sobre el parámetro σ aseguraría la estabilidad dinámica?
5. Sean la oferta y la demanda
$$Q_d = \alpha - \beta P - \eta \frac{dP}{dt} \qquad Q_s = \delta P \qquad (\alpha, \beta, \eta, \delta > 0)$$
(*a*) Suponiendo que el mercado esté clarificado para cada instante de tiempo, encuentre la trayectoria de tiempo $P(t)$ (solución general).
(*b*) ¿Tiene este mercado un precio de equilibrio intertemporal dinámicamente estable?
(*c*) La hipótesis del presente modelo de que $Q_d = Q_s$ para todo t es idéntica con la del modelo del mercado estático en la sección 3.2; sin embargo, aquí todavía tenemos un modelo dinámico. ¿Cómo puede ser eso?

15.3 Coeficiente variable y término variable

En el caso más general de una ecuación diferencial lineal de primer orden

$$\frac{dy}{dt} + u(t)y = w(t) \qquad \textbf{(15.12)}$$

$u(t)$ y $w(t)$ representan, respectivamente, un coeficiente variable y un término variable. En este caso, ¿cómo encontramos la trayectoria de tiempo $y(t)$?

El caso homogéneo

Para el caso homogéneo, donde $w(t) = 0$, la solución es fácil de obtener. Como la ecuación diferencial tiene la forma

$$\frac{dy}{dt} + u(t)y = 0 \qquad \text{o} \qquad \frac{1}{y}\frac{dy}{dt} = -u(t) \tag{15.13}$$

al integrar ambos lados a su vez con respecto a t, tenemos

$$\text{Lado izquierdo} = \int \frac{1}{y}\frac{dy}{dt}\,dt = \int \frac{dy}{y} = \ln y + c \quad (\text{suponiendo } y > 0)$$

$$\text{Lado derecho} = \int -u(t)\,dt = -\int u(t)\,dt$$

En este último, el proceso de integración no puede desarrollarse más porque no se ha dado a $u(t)$ una forma específica; entonces, tenemos que aceptar sólo una expresión integral general. Cuando se igualan ambos lados, el resultado es

$$\ln y = -c - \int u(t)\,dt$$

Entonces, la trayectoria deseada y puede obtenerse tomando el antilog de $\ln y$:

$$y(t) = e^{\ln y} = e^{-c}e^{-\int u(t)\,dt} = Ae^{-\int u(t)\,dt} \qquad \text{donde } A \equiv e^{-c} \tag{15.14}$$

Ésta es la solución general de la ecuación diferencial (15.13).

Para subrayar la naturaleza variable del coeficiente $u(t)$, hasta ahora hemos descrito en forma explícita el argumento t. Sin embargo, por sencillez en la notación, a partir de ahora vamos a omitir el argumento y vamos a abreviar $u(t)$ como u.

En comparación con la solución general (15.3) para el caso de coeficientes constantes, la única modificación en (15.14) es el reemplazo de la expresión e^{-at} por la expresión más complicada $e^{-\int u\,dt}$. El razonamiento que subyace bajo este cambio puede entenderse mejor si interpretamos el término at en e^{-at} como una integral: $\int a\,dt = at$ (más una constante que el término A puede absorber, ya que e elevado a una potencia constante es nuevamente una constante). Ante esto, la diferencia entre las dos soluciones generales de hecho se convierte en una similitud. Como en ambos casos estamos tomando el coeficiente del término y en la ecuación diferencial —un término constante a en un caso, y un término variable u en el otro— e integrándolo respecto a t, y luego tomando el negativo de la integral resultante como el exponente de e.

Una vez que se llega a la solución general, es un asunto relativamente sencillo obtener la solución definida por condición inicial con ayuda de una condición inicial apropiada.

Ejemplo 1

Encuentre la solución general de la ecuación $\frac{dy}{dt} + 3t^2y = 0$. Aquí tenemos $u = 3t^2$, y $\int u\,dt = \int 3t^2\,dt = t^3 + c$. Por lo tanto, mediante (15.14), podemos escribir la solución como

$$y(t) = Ae^{-(t^3+c)} = Ae^{-t^3}e^{-c} = Be^{-t^3} \qquad \text{donde } B \equiv Ae^{-c}$$

Observe que si hubiéramos omitido la constante de integración c, no habríamos perdido información, porque habríamos obtenido $y(t) = Ae^{-t^3}$, que es realmente la solución idéntica, ya que A y B representan constantes arbitrarias. En otras palabras, la expresión e^{-c}, donde la constante c hace su única aparición, siempre puede incluirse bajo la otra constante A.

El caso no homogéneo

Para el caso no homogéneo, donde $w(t) \neq 0$, la solución no se obtiene tan fácilmente. Trataremos de encontrar esa solución vía el concepto de las ecuaciones diferenciales exactas, que estudiaremos en la sección 15.4. Sin embargo, no hay problema en enunciar el resultado primero aquí: dada la ecuación diferencial (15.12), la solución general es

$$y(t) = e^{-\int u\,dt}\left(A + \int w e^{\int u\,dt}\,dt\right) \qquad \textbf{(15.15)}$$

donde A es una constante arbitraria que se puede determinar si tenemos una condición inicial apropiada.

Es interesante que esta solución general, al igual que la solución en el caso de coeficiente constante término constante, nuevamente consista en dos componentes aditivas. Aún más, una de estas dos, $Ae^{-\int u\,dt}$, no es nada más que la solución general de la ecuación reducida (homogénea), obtenida anteriormente en (15.14), y está por lo tanto en la naturaleza de una función complementaria.

Ejemplo 2 Encuentre la solución general de la ecuación $\dfrac{dy}{dt} + 2ty = t$. Aquí tenemos

$$u = 2t \qquad w = t \qquad \text{y} \qquad \int u\,dt = t^2 + k \qquad (k \text{ arbitraria})$$

Entonces, mediante (15.15), tenemos

$$\begin{aligned}
y(t) &= e^{-(t^2+k)}\left(A + \int te^{t^2+k}\,dt\right) \\
&= e^{-t^2}e^{-k}\left(A + e^k \int te^{t^2}\,dt\right) \\
&= Ae^{-k}e^{-t^2} + e^{-t^2}\left(\frac{1}{2}e^{t^2} + c\right) \qquad [e^{-k}e^k = 1] \\
&= (Ae^{-k} + c)e^{-t^2} + \frac{1}{2} \\
&= Be^{-t^2} + \frac{1}{2} \qquad \text{donde } B \equiv Ae^{-k} + c \text{ es arbitraria}
\end{aligned}$$

La validez de esta solución puede verificarse nuevamente por diferenciación.

Es interesante observar que en este ejemplo pudimos haber omitido otra vez la constante de integración k, así como la constante de integración c, sin afectar el resultado final. Esto se debe a que tanto k como c pueden incluirse bajo la constante arbitraria B en la solución final. Se recomienda experimentar con el proceso más sencillo de aplicar (15.15), sin usar las constantes k y c, y verificar que surja la misma solución.

Ejemplo 3 Resuelva la ecuación $\dfrac{dy}{dt} + 4ty = 4t$. Esta vez vamos a omitir las constantes de integración. Como

$$u = 4t \qquad w = 4t \qquad \text{y} \qquad \int u\,dt = 2t^2 \qquad [\text{constante omitida}]$$

la solución general, mediante (15.15), es

$$\begin{aligned}
y(t) &= e^{-2t^2}\left(A + \int 4te^{2t^2}\,dt\right) = e^{-2t^2}\left(A + e^{2t^2}\right) \qquad [\text{constante omitida}] \\
&= Ae^{-2t^2} + 1
\end{aligned}$$

Como puede esperarse, la omisión de las constantes de integración sirve para simplificar el procedimiento en forma sustancial.

La ecuación diferencial $\frac{dy}{dt} + uy = w$ de (15.12) es más general que la ecuación $\frac{dy}{dt} + ay = b$ de (15.4), ya que u y w no son necesariamente constantes, como son a y b. De acuerdo con esto, la fórmula solución (15.15) es también más general que la fórmula solución (15.5). De hecho, cuando hacemos $u = a$ y $w = b$ (15.15) debe reducirse a (15.5). Éste es realmente el caso. Cuando tenemos

$$u = a \qquad w = b \qquad \text{y} \qquad \int u\,dt = at \qquad \text{[constante omitida]}$$

entonces, (15.15) se transforma en

$$y(t) = e^{-at}\left(A + \int be^{at}\,dt\right) = e^{-at}\left(A + \frac{b}{a}e^{at}\right) \qquad \text{[constante omitida]}$$

$$= Ae^{-at} + \frac{b}{a}$$

que es idéntica a (15.5).

EJERCICIO 15.3

Resuelva las siguientes ecuaciones diferenciales lineales de primer orden; si está dada una condición inicial, determine la constante arbitraria:

1. $\frac{dy}{dt} + 5y = 15$
2. $\frac{dy}{dt} + 2ty = 0$
3. $\frac{dy}{dt} + 2ty = t$; $y(0) = \frac{3}{2}$
4. $\frac{dy}{dt} + t^2y = 5t^2$; $y(0) = 6$
5. $2\frac{dy}{dt} + 12y + 2e^t = 0$; $y(0) = \frac{6}{7}$
6. $\frac{dy}{dt} + y = t$

15.4 Ecuaciones diferenciales exactas

Introduciremos ahora el concepto de las ecuaciones diferenciales exactas y usaremos el método de solución que emerge de este concepto para obtener la fórmula de la solución (15.15) de la ecuación diferencial (15.12). Aun cuando nuestro propósito inmediato es usarlo para resolver una ecuación diferencial *lineal*, una ecuación diferencial exacta puede ser lineal o no lineal en sí misma.

Ecuaciones diferenciales exactas

Dada una función de dos variables $F(y, t)$, su diferencial total es

$$dF(y, t) = \frac{\partial F}{\partial y}\,dy + \frac{\partial F}{\partial t}\,dt$$

Cuando esta diferencial se hace igual a cero, la ecuación resultante

$$\frac{\partial F}{\partial y}\,dy + \frac{\partial F}{\partial t}\,dt = 0$$

se conoce como una *ecuación diferencial exacta*, ya que el lado izquierdo es exactamente la diferencial de la función $F(y, t)$. Por ejemplo, dado

$$F(y, t) = y^2 t + k \qquad (k \text{ una constante})$$

la diferencial total es

$$dF = 2yt\,dy + y^2\,dt$$

Entonces, la ecuación diferencial

$$2yt\,dy + y^2\,dt = 0 \quad \text{o sea} \quad \frac{dy}{dt} + \frac{y^2}{2yt} = 0 \tag{15.16}$$

es exacta.

En general, una ecuación diferencial

$$M\,dy + N\,dt = 0 \tag{15.17}$$

es exacta si y sólo si existe una función $F(y, t)$ tal que $M = \partial F/\partial y$ y $N = \partial F/\partial t$. Sin embargo, por el teorema de Young, que establece que $\partial^2 F/\partial t\,\partial y = \partial^2 F/\partial y\,\partial t$, también podemos afirmar que (15.17) es exacta si y sólo si

$$\frac{\partial M}{\partial t} = \frac{\partial N}{\partial y} \tag{15.18}$$

Esta última ecuación nos da una prueba sencilla de la exactitud de una ecuación diferencial. Aplicada a (15.16), donde $M = 2yt$ y $N = y^2$, esta prueba suministra $\partial M/\partial t = 2y = \partial N/\partial y$; entonces, se verifica debidamente la exactitud de dicha ecuación diferencial.

Observe que no se han impuesto restricciones sobre los términos M y N respecto a la manera en que se presenta la variable y. Entonces, una ecuación diferencial exacta puede muy bien ser *no lineal* (en y). Sin embargo, siempre será de primer orden y de primer grado.

Al ser exacta, la ecuación diferencial simplemente dice

$$dF(y, t) = 0$$

Por lo tanto, la solución general debe ser claramente de la forma

$$F(y, t) = c$$

Debido a ello, la solución de una ecuación diferencial exacta es básicamente la búsqueda de la función $F(y, t)$ (primitiva) y luego hacerla igual a una constante arbitraria. Esbocemos un método para encontrarla para la ecuación $M\,dy + N\,dt = 0$.

Método de solución

Para comenzar, puesto que $M = \partial F/\partial y$, la función F debe contener la integral de M respecto a la variable y; por tanto, podemos escribir un resultado preliminar —de una forma aún no determinada— como sigue:

$$F(y, t) = \int M\,dy + \psi(t) \tag{15.19}$$

Aquí M, una derivada *parcial*, va a integrarse sólo respecto a y; es decir, t debe tratarse como una constante en el proceso de integración, así como se trató como una constante en la diferenciación parcial de $F(y, t)$ que condujo a $M = \partial F/\partial y$.[3] Como al diferenciar $F(y, t)$ parcialmente respecto a y, se cancelaría cualquier término aditivo que contenga sólo la variable t y/o algunas constantes (pero sin y), ahora debemos tener cuidado en restituir estos términos en el proceso de integración. Esto explica por qué introdujimos en (15.19) un término general $\psi(t)$, el cual, aunque no es exactamente igual a una constante de integración, tiene un papel que desempeñar, que es precisamente idéntico a este último. Es relativamente fácil obtener $\int M\,dy$; pero, ¿cómo identificamos la forma exacta de este término $\psi(t)$?

El truco es utilizar el hecho de que $N = \partial F/\partial t$. Pero el procedimiento se explica de la mejor manera con la ayuda de ejemplos específicos.

Ejemplo 1 Resuelva la ecuación diferencial exacta

$$2yt\,dy + y^2\,dt = 0 \qquad \text{[reproducido de (15.16)]}$$

En esta ecuación tenemos

$$M = 2yt \qquad \text{y} \qquad N = y^2$$

PASO 1 Mediante (15.19) podemos escribir primero el resultado preliminar

$$F(y, t) = \int 2yt\,dy + \psi(t) = y^2 t + \psi(t)$$

Observe que hemos omitido la constante de integración, porque puede fusionarse automáticamente con la expresión $\psi(t)$.

PASO 2 Si diferenciamos el resultado del Paso 1 parcialmente respecto a t, podemos obtener

$$\frac{\partial F}{\partial t} = y^2 + \psi'(t)$$

Pero como $N = \partial F/\partial t$, podemos igualar $N = y^2$ y $\partial F/\partial t = y^2 + \psi'(t)$, para obtener

$$\psi'(t) = 0$$

PASO 3 La integración del último resultado nos da

$$\psi(t) = \int \psi'(t)\,dt = \int 0\,dt = k$$

y ahora tenemos una forma específica de $\psi(t)$. En el presente caso resulta que $\psi(t)$ es simplemente una constante; más generalmente, puede ser una función no constante de t.

PASO 4 Los resultados de los Pasos 1 y 3 pueden combinarse para dar

$$F(y, t) = y^2 t + k$$

La solución de la ecuación diferencial exacta debe ser entonces $F(y, t) = c$. Pero como la constante k puede fusionarse con c, podemos escribir la solución simplemente como

$$y^2 t = c \qquad \text{o sea} \qquad y(t) = ct^{-1/2}$$

donde c es arbitraria.

[3] Algunos autores emplean el símbolo operador $\int(\cdots)\,\partial y$ para enfatizar que la integración se refiere sólo a y. Aquí usamos todavía el símbolo $\int(\cdots)\,dy$, ya que hay poca probabilidad de confusión.

Ejemplo 2 Resuelva la ecuación $(t+2y)\,dy+(y+3t^2)\,dt=0$. Primero verifiquemos si se trata de una ecuación diferencial exacta. Haciendo $M=t+2y$ y $N=y+3t^2$, encontramos que $\partial M/\partial t = 1 = \partial N/\partial y$. Entonces, la ecuación pasa la prueba de exactitud. Para encontrar la solución, nuevamente seguimos el procedimiento esbozado en el ejemplo 1.

PASO 1 Aplique (15.19) y escriba

$$F(y,t)=\int (t+2y)\,dy+\psi(t)=yt+y^2+\psi(t) \qquad [\text{constante fusionada con } \psi(t)]$$

PASO 2 Diferencie este resultado respecto a t, para obtener

$$\frac{\partial F}{\partial t}=y+\psi'(t)$$

Entonces, al igualar esto con $N=y+3t^2$, encontramos que

$$\psi'(t)=3t^2$$

PASO 3 Integre este último resultado para obtener

$$\psi(t)=\int 3t^2\,dt=t^3 \qquad [\text{la constante puede omitirse}]$$

PASO 4 Combine los resultados de los Pasos 1 y 3 para obtener la forma completa de la función $F(y,t)$:

$$F(y,t)=yt+y^2+t^3$$

lo que implica que la solución de la ecuación diferencial dada es

$$yt+y^2+t^3=c$$

Verifique que igualar a cero la diferencial total de esta ecuación producirá realmente la ecuación diferencial dada.

Este procedimiento de cuatro pasos podemos usarlo para resolver cualquier ecuación diferencial exacta. Es interesante que podamos aplicarlo aun cuando la ecuación dada *no sea* exacta. Sin embargo, para ver esto debemos introducir primero el concepto del factor de integración.

El factor de integración

Algunas veces una ecuación diferencial inexacta puede volverse exacta al multiplicar cada uno de sus términos por un factor común específico. Este factor se denomina *factor de integración.*

Ejemplo 3 La ecuación diferencial

$$2t\,dy+y\,dt=0$$

no es exacta, porque no satisface (15.18):

$$\frac{\partial M}{\partial t}=\frac{\partial}{\partial t}(2t)=2\neq\frac{\partial N}{\partial y}=\frac{\partial}{\partial y}(y)=1$$

Sin embargo, si multiplicamos cada término por y, la ecuación dada se transforma en (15.16), que se ha establecido que es exacta. Entonces y es un factor de integración para la ecuación diferencial en este ejemplo.

Mientras podamos encontrar un factor de integración para una ecuación diferencial inexacta, siempre podremos hacerla exacta y luego podremos usar rápidamente el procedimiento de solución de cuatro pasos.

Solución de las ecuaciones diferenciales lineales de primer orden

La ecuación diferencial general lineal de primer orden

$$\frac{dy}{dt} + uy = w$$

la cual, en el formato de (15.17), puede expresarse como

$$dy + (uy - w)\,dt = 0 \tag{15.20}$$

tiene el factor de integración

$$e^{\int u\,dt} \equiv \exp\left(\int u\,dt\right)$$

Este factor de integración, cuya forma de ninguna manera es intuitivamente obvia, puede "descubrirse" como sigue. Sea I el factor de integración (todavía desconocido). La multiplicación de (15.20) por I debería convertirlo en una ecuación diferencial exacta

$$\underbrace{I}_{M}\,dy + \underbrace{I(uy - w)}_{N}\,dt = 0 \tag{15.20'}$$

La prueba de exactitud exige que $\partial M/\partial t = \partial N/\partial y$. La inspección visual de las expresiones M y N sugieren que, como M consiste solamente en I, y como u y w sólo son funciones de t, la prueba de exactitud se reducirá a una condición muy simple si I sólo es también una función de t. Como entonces la prueba $\partial M/\partial t = \partial N/\partial y$ se transforma en

$$\frac{dI}{dt} = Iu \qquad \text{o sea} \qquad \frac{dI/dt}{I} = u$$

Así, la forma especial $I = I(t)$ puede realmente funcionar, siempre que tenga una tasa de crecimiento igual a u, o más explícitamente $u(t)$. De acuerdo con esto, $I(t)$ debe adoptar la forma específica

$$I(t) = Ae^{\int u\,dt} \qquad [\text{ver (15.13) y (15.14)}]$$

Sin embargo, como podemos verificar fácilmente, la constante A puede hacerse igual a 1 sin afectar la capacidad de $I(t)$ de cumplir la prueba de exactitud. Entonces, podemos usar la forma más simple $e^{\int u\,dt}$ como el factor de integración.

La sustitución de este factor de integración en (15.20′) suministra la ecuación diferencial exacta

$$e^{\int u\,dt}\,dy + e^{\int u\,dt}(uy - w)\,dt = 0 \tag{15.20''}$$

la cual puede resolverse entonces mediante el procedimiento de los cuatro pasos.

PASO 1 Primero aplicamos (15.19) para obtener

$$F(y, t) = \int e^{\int u\,dt}\,dy + \psi(t) = ye^{\int u\,dt} + \psi(t)$$

El resultado de la integración surge de esta forma sencilla porque el integrando es independiente de la variable y.

PASO 2 Enseguida diferenciamos el resultado del Paso 1 respecto a t para obtener

$$\frac{\partial F}{\partial t} = yue^{\int u\,dt} + \psi'(t) \qquad \text{[regla de la cadena]}$$

Y como esto puede igualarse a $N = e^{\int u\,dt}(uy - w)$, tenemos

$$\psi'(t) = -we^{\int u\,dt}$$

PASO 3 Ahora la integración directa suministra

$$\psi(t) = -\int we^{\int u\,dt}\,dt$$

Como no se ha dado forma específica a las funciones $u = u(t)$ y $w = w(t)$, no puede hacerse nada adicional acerca de esta integral, y debemos contentarnos con esta expresión más bien general para $\psi(t)$.

PASO 4 Sustituyendo esta expresión $\psi(t)$ en el resultado del Paso 1, encontramos que

$$F(y, t) = ye^{\int u\,dt} - \int we^{\int u\,dt}\,dt$$

Entonces, la solución general de la ecuación diferencial exacta (15.20″) —y de la ecuación diferencial lineal de primer orden (15.20) equivalente, aunque inexacta— es

$$ye^{\int u\,dt} - \int we^{\int u\,dt}\,dt = c$$

Al reordenar y sustituir el símbolo c por A (constante arbitraria), esto puede escribirse como

$$y(t) = e^{-\int u\,dt}\left(A + \int we^{\int u\,dt}\,dt\right) \tag{15.21}$$

el cual es exactamente el resultado dado anteriormente en (15.15).

EJERCICIO 15.4

1. Verifique que cada una de las siguientes ecuaciones diferenciales sea exacta y resuelva mediante el procedimiento de los cuatro pasos:
   (*a*) $2yt^3\,dy + 3y^2t^2\,dt = 0$
   (*b*) $3y^2t\,dy + (y^3 + 2t)\,dt = 0$
   (*c*) $t(1 + 2y)\,dy + y(1 + y)\,dt = 0$
   (*d*) $\dfrac{dy}{dt} + \dfrac{2y^4t + 3t^2}{4y^3t^2} = 0$ [*Sugerencia:* primero convierta a la forma de (15.17).]
2. ¿Son exactas las siguientes ecuaciones diferenciales? Si no, intente con t, y, y y^2 como posibles factores de integración.
   (*a*) $2(t^3 + 1)\,dy + 3yt^2\,dt = 0$
   (*b*) $4y^3t\,dy + (2y^4 + 3t)\,dt = 0$
3. Aplicando el procedimiento de los cuatro pasos a la ecuación general diferencial exacta $M\,dy + N\,dt = 0$, obtenga la siguiente fórmula para la solución general de una ecuación diferencial exacta:

$$\int M\,dy + \int N\,dt - \int\left(\frac{\partial}{\partial t}\int M\,dy\right)dt = c$$

15.5 Ecuaciones diferenciales no lineales de primer orden y primer grado

En una ecuación diferencial lineal restringimos al *primer grado* no sólo la derivada dy/dt, sino también la variable dependiente y, y no permitimos que aparezca el producto $y(dy/dt)$. Cuando y aparece con una potencia mayor que uno, la ecuación se transforma en *no lineal* aun si contiene sólo la derivada dy/dt en primer grado. En general, una ecuación de la forma

$$f(y, t)\, dy + g(y, t)\, dt = 0 \qquad \textbf{(15.22)}$$

o bien

$$\frac{dy}{dt} = h(y, t) \qquad \textbf{(15.22')}$$

donde no haya restricción para las potencias de y y t, constituirá una ecuación diferencial no lineal de primer orden y de primer grado, ya que dy/dt es una derivada de primer orden a la primera potencia. Ciertas variantes de estas ecuaciones pueden resolverse con relativa facilidad mediante procedimientos más o menos rutinarios. Discutiremos brevemente tres casos.

Ecuaciones diferenciales exactas

El primero es el conocido caso de las ecuaciones diferenciales exactas. Como señalamos anteriormente, la variable y puede aparecer en una ecuación exacta a una potencia elevada, como en (15.16) —$2yt\, dy + y^2\, dt = 0$— que debe compararse con (15.22). Es verdad que la cancelación del factor común y de ambos términos a la izquierda va a reducir la ecuación a una forma lineal, pero la propiedad de exactitud se pierde en ese caso. Por lo tanto, como una ecuación diferencial *exacta*, debe también considerase como no lineal.

Como ya hemos estudiado el método de solución para las ecuaciones diferenciales exactas, no es necesario hacer aquí ningún comentario adicional.

Variables separables

La ecuación diferencial de (15.22)

$$f(y, t)\, dy + g(y, t)\, dt = 0$$

puede ser que posea la conveniente propiedad de que la función f esté sólo en la variable y, mientras que la función g implica sólo la variable t, de modo que la ecuación se reduce a la forma especial

$$f(y)\, dy + g(t)\, dt = 0 \qquad \textbf{(15.23)}$$

En este caso, se dice que las variables son *separables*, porque los términos que incluyen y —agrupados en $f(y)$— pueden separarse matemáticamente de los términos que incluyen t, que se agrupan en $g(t)$. Para resolver este tipo especial de ecuación se requieren sólo técnicas de integración.

Ejemplo 1 Resuelva la ecuación $3y^2\, dy - t\, dt = 0$. Primero reescribamos la ecuación como

$$3y^2\, dy = t\, dt$$

Integrando los dos lados (cada uno de los cuales es un diferencial) e igualando los resultados, obtenemos

$$\int 3y^2\, dy = \int t\, dt \qquad \text{o} \qquad y^3 + c_1 = \frac{1}{2}t^2 + c_2$$

Entonces, la solución general puede escribirse como

$$y^3 = \frac{1}{2}t^2 + c \qquad \text{o} \qquad y(t) = \left(\frac{1}{2}t^2 + c\right)^{1/3}$$

Un aspecto notable es que la integración de cada término se realiza respecto a una variable diferente; esto es lo que hace que la ecuación de variables separables sea comparativamente fácil de manejar.

Ejemplo 2

Resuelva la ecuación $2t\,dy + y\,dt = 0$. A primera vista, esta ecuación diferencial no parece pertenecer a este tipo, porque no se amolda a la forma general de (15.23). Para ser específicos, se ve que los coeficientes de dy y dt incluyen las variables "equivocadas". Sin embargo, una transformación sencilla —dividir entre $2yt$ ($\neq 0$)— va a reducir la ecuación a la forma de variables separadas

$$\frac{1}{y}\,dy + \frac{1}{2t}\,dt = 0$$

Apoyándonos en nuestra experiencia con el ejemplo anterior podemos hacer manipulaciones en busca de la solución (sin trasponer un término primero) como sigue:[4]

$$\int \frac{1}{y}\,dy + \int \frac{1}{2t}\,dt = c$$

entonces

$$\ln y + \frac{1}{2}\ln t = c \qquad \text{o} \qquad \ln(yt^{1/2}) = c$$

Entonces, la solución es

$$yt^{1/2} = e^c = k \qquad \text{o} \qquad y(t) = kt^{-1/2}$$

donde k es una constante arbitraria, así como lo son los símbolos c y A empleados en otro lado.

Observe que, en lugar de resolver la ecuación en el ejemplo 2, como lo hicimos, también pudimos haberla transformado primero en una ecuación diferencial exacta (mediante el factor de integración y) y luego resolverla como tal. La solución, ya dada en el ejemplo 1 de la sección 15.4, debe ser idéntica a la recién obtenida por separación de variables. Lo importante es que una ecuación diferencial dada con frecuencia puede resolverse de más de una manera y, por lo tanto, podemos elegir el método de nuestra preferencia. En otros casos, una ecuación diferencial que no es asequible para un método específico puede hacerse asequible después de una transformación apropiada.

Ecuaciones reducibles a la forma lineal

Si la ecuación diferencial $dy/dt = h(y, t)$ adopta la forma no lineal específica

$$\frac{dy}{dt} + Ry = Ty^m \tag{15.24}$$

donde R y T son dos funciones de t, y m es cualquier número diferente de 0 y 1 (¿qué pasa si $m = 0$ o $m = 1$?), entonces la ecuación —denominada *ecuación de Bernoulli*— siempre puede reducirse a una ecuación diferencial lineal y resolverse como tal.

[4] Hablando estrictamente, en el resultado de la integración debimos haber escrito $\ln|y|$ y $\frac{1}{2}\ln|t|$. Si y y t pueden considerarse positivos, como es apropiado en la mayoría de los contextos en economía, entonces se presentará el resultado dado en el texto.

El procedimiento de reducción es relativamente sencillo. Primero, podemos dividir (15.24) entre y^m para obtener

$$y^{-m}\frac{dy}{dt} + Ry^{1-m} = T$$

Si adoptamos por comodidad una variable z como sigue:

$$z = y^{1-m} \qquad \left[\text{así que } \frac{dz}{dt} = \frac{dz}{dy}\,\frac{dy}{dt} = (1-m)y^{-m}\frac{dy}{dt}\right]$$

Entonces, la ecuación anterior puede escribirse como

$$\frac{1}{1-m}\,\frac{dz}{dt} + Rz = T$$

Aún más, después de multiplicar por $(1-m)\,dt$ y reordenar, podemos transformar la ecuación en

$$dz + [(1-m)Rz - (1-m)T]\,dt = 0 \qquad \textbf{(15.24')}$$

Se ve que ésta es una ecuación diferencial lineal de primer orden de la forma (15.20), en la cual la variable z ha ocupado el lugar de y.

Es evidente que podemos aplicar la formula (15.21) para encontrar su solución $z(t)$. Entonces, como paso final, podemos convertir z nuevamente en y mediante la sustitución inversa.

Ejemplo 3

Resuelva la ecuación $dy/dt + ty = 3ty^2$. Ésta es una ecuación de Bernoulli, con $m = 2$ (dándonos $z = y^{1-m} = y^{-1}$), $R = t$, y $T = 3t$. Entonces, mediante (15.24'), podemos escribir la ecuación diferencial linealizada como

$$dz + (-tz + 3t)\,dt = 0$$

Aplicando la fórmula (15.21), podemos encontrar que la solución es

$$z(t) = A\exp\left(\tfrac{1}{2}t^2\right) + 3$$

(Como ejercicio, rastree los pasos que conducen a esta solución.)

Como nuestro interés principal radica en la solución $y(t)$ en vez de $z(t)$, debemos realizar una transformación inversa usando la ecuación $z = y^{-1}$, o $y = z^{-1}$. Por lo tanto, al tomar el recíproco de $z(t)$, obtenemos

$$y(t) = \frac{1}{A\exp\left(\tfrac{1}{2}t^2\right) + 3}$$

como la solución deseada. Ésta es una solución general, porque está presente una constante A arbitraria.

Ejemplo 4

Resuelva la ecuación $dy/dt + (1/t)y = y^3$. Aquí tenemos $m = 3$ (por tanto, $z = y^{-2}$), $R = 1/t$, y $T = 1$; entonces, la ecuación puede linealizarse de forma

$$dz + \left(\frac{-2}{t}z + 2\right)dt = 0$$

Como podemos verificar, mediante el uso de la fórmula (15.21), la solución de esta ecuación diferencial es

$$z(t) = At^2 + 2t$$

Entonces, por la transformación inversa $y = z^{-1/2}$, se deduce que la solución general en la variable original debe escribirse como

$$y(t) = (At^2 + 2t)^{-1/2}$$

Como ejercicio, verifique la validez de las soluciones de estos dos últimos ejemplos por diferenciación.

EJERCICIO 15.5

1. Determine, para cada una de las siguientes ecuaciones diferenciales, (1) si las variables son separables y (2) si la ecuación es lineal o puede linealizarse:
   (*a*) $2t\,dy + 2y\,dt = 0$ (*c*) $\dfrac{dy}{dt} = -\dfrac{t}{y}$
   (*b*) $\dfrac{y}{y+t}\,dy + \dfrac{2t}{y+t}\,dt = 0$ (*d*) $\dfrac{dy}{dt} = 3y^2t$
2. Resuelva (*a*) y (*b*) en el problema 1 por separación de variables, considerando positivas y y t. Verifique sus respuestas por diferenciación.
3. Resuelva (*c*) en el problema 1 como una ecuación de variables separables y también como una ecuación de Bernoulli.
4. Resuelva (*d*) en el problema 1 como una ecuación de variables separables y también como una ecuación de Bernoulli.
5. Verifique la corrección de la solución intermedia $z(t) = At^2 + 2t$ en el ejemplo 4 mostrando que su derivada dz/dt es consistente con la ecuación diferencial linealizada.

15.6 El enfoque cualitativo gráfico

Los diversos casos de ecuaciones diferenciales no lineales anteriormente discutidos (ecuaciones diferenciales exactas, ecuaciones de variables separables y ecuaciones de Bernoulli) los hemos resuelto todos *cuantitativamente*. Es decir, en cada caso hemos buscado y encontrado una trayectoria de tiempo $y(t)$ la cual, para cada valor de t, nos dice el valor específico correspondiente de la variable y.

A veces, no podremos encontrar una solución cuantitativa a partir de una ecuación diferencial dada. Aun así, en estos casos tal vez se puedan evaluar las propiedades *cualitativas* de la trayectoria de tiempo —principalmente si $y(t)$ converge— observando de modo directo la ecuación diferencial misma o analizando su gráfica. Aun cuando dispongamos de soluciones cuantitativas, adicionalmente podemos también emplear las técnicas del análisis cualitativo si el aspecto cualitativo de la trayectoria de tiempo es nuestra preocupación principal o exclusiva.

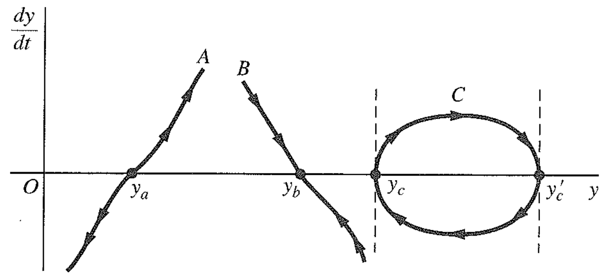
El diagrama de fases

Dada una ecuación diferencial de primer orden en la forma general

$$\frac{dy}{dt} = f(y)$$

ya sea lineal o no lineal en la variable y, podemos graficar dy/dt contra y como en la figura 15.3. Una representación geométrica de este tipo, factible siempre que dy/dt sea una función sólo de y, se llama *diagrama de fase*, y la gráfica que representa la función f, una *línea de fase*. (Una ecuación diferencial de esta forma —en la cual la variable t de tiempo no aparece como

FIGURA 15.3



un argumento separado de la función f— se dice que es una ecuación diferencial *autónoma*). Una vez que conocemos una línea de fase, su configuración va a impartir información cualitativa significativa en relación con la trayectoria de tiempo $y(t)$. La pista de esto radica en los dos siguientes comentarios generales:

1. En cualquier lado por *arriba* del eje horizontal (donde $dy/dt > 0$), y debe ser creciente con el tiempo y, por lo que toca al eje y, debe moverse de izquierda a derecha. Mediante un razonamiento análogo, cualquier punto que esté por *debajo* del eje horizontal debe asociarse con un movimiento hacia la izquierda de la variable y, ya que la negatividad de dy/dt significa que y disminuye con el tiempo. Estas tendencias direccionales explican por qué las cabezas de flecha de las líneas de fase ilustrativas de la figura 15.3 se dibujan tal como están. Por arriba del eje horizontal, las flechas apuntan uniformemente a la derecha —hacia el noreste o el sureste o hacia el este, dependiendo del caso. Lo opuesto es verdad por debajo del eje y. Aún más, estos resultados son independientes del signo algebraico de y; aun si la línea de fase A (o cualquier otra) se trasplanta a la izquierda del eje vertical, no se afecta la dirección de las flechas.
2. Si hay un nivel de equilibrio de y —en el sentido intertemporal del término—, puede presentarse sólo en el eje horizontal, donde $dy/dt = 0$ (y es estacionario con el tiempo). Por lo tanto, para encontrar un equilibrio es necesario sólo considerar la intersección de la línea de fase con el eje y.[5] Por otro lado, para probar la estabilidad dinámica del equilibrio, también debemos verificar si, independientemente de la posición inicial de y, la línea de fase siempre va a guiarlo hacia la posición de equilibrio en dicha intersección.

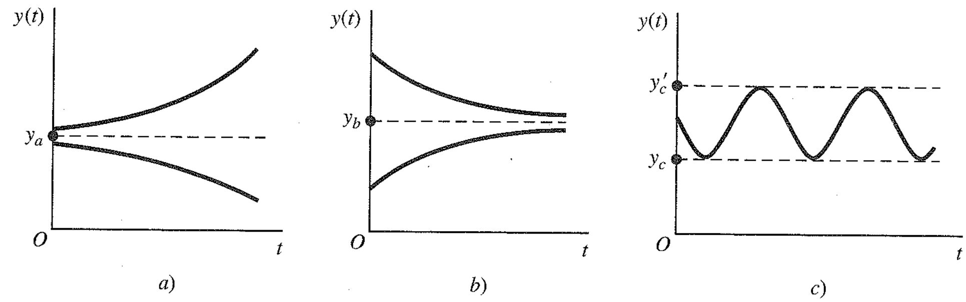
Tipos de trayectoria de tiempo

Basándonos en los comentarios generales anteriores, podemos observar tres tipos diferentes de trayectoria de tiempo a partir de las líneas de fase ilustrativas de la figura 15.3.

La línea de fase A tiene un equilibrio en el punto y_a; pero *arriba,* así como *debajo* de ese punto, las cabezas de flecha se alejan consistentemente del equilibrio. De esta manera, aunque el equilibrio puede lograrse si $y(0) = y_a$, el caso más común de $y(0) \neq y_a$ va a conducir a que y sea estrictamente creciente [si $y(0) > y_a$] o estrictamente decreciente [si $y(0) < y_a$]. Además, en este caso la desviación de y respecto a y_a tiende a crecer a un paso creciente porque, a medida que sigamos las cabezas de flecha en la línea de fase, nos desviaremos cada vez más del eje y, encontrando también por ello valores numéricos estrictamente crecientes de dy/dt. La trayectoria de tiempo $y(t)$ implicada por la línea de fase A puede representarse por lo tanto por las curvas mostradas en la figura 15.4*a*, donde y se grafica contra t (en vez de dy/dt contra y). El equilibrio y_a es dinámicamente inestable.

[5] Sin embargo, no todas las intersecciones representan posiciones de equilibrio. Veremos esto cuando estudiemos la línea de fase C en la figura 15.3.

FIGURA 15.4



En contraste, la línea de fase B implica un equilibrio estable para y_b. Si $y(0) = y_b$, el equilibrio prevalece inmediatamente. Pero la característica importante de la línea de fase B es que, aun si $y(0) \neq y_b$, el movimiento a lo largo de la línea de fase va a guiar a y hacia el nivel de y_b. La trayectoria de tiempo $y(t)$ que corresponde a este tipo de línea de fase debe ser por lo tanto de la forma que muestra la figura 15.4*b*, lo que nos recuerda al modelo dinámico del mercado.

El estudio anterior sugiere que, en general, es la pendiente de la línea de fase en el punto de intersección lo que tiene la clave de la estabilidad dinámica de equilibrio o la convergencia de la trayectoria de tiempo. Una pendiente *positiva* (finita), tal como en el punto y_a, produce una *inestabilidad* dinámica; mientras que una pendiente *negativa* (finita), tal como en y_b, implica *estabilidad* dinámica.

Esta generalización nos puede ayudar a obtener inferencias cualitativas acerca de las ecuaciones diferenciales dadas aun sin graficar las líneas de fase. Por ejemplo, considere la ecuación diferencial lineal de (15.4):

$$\frac{dy}{dt} + ay = b \qquad \text{o sea} \qquad \frac{dy}{dt} = -ay + b$$

Como la línea de fase tendrá la pendiente $-a$ (constante), como se supone diferente de cero, podemos inferir inmediatamente que (sin dibujar la línea)

$$a \gtrless 0 \quad \Leftrightarrow \quad y(t) \begin{Bmatrix} \text{converge a} \\ \text{diverge de} \end{Bmatrix} \text{equilibrio}$$

Como podemos esperar, este resultado coincide perfectamente con lo que nos muestra la solución cuantitativa de esta ecuación:

$$y(t) = \left[y(0) - \frac{b}{a}\right]e^{-at} + \frac{b}{a} \qquad \text{[de (15.5')]}$$

Hemos aprendido que, iniciando desde una posición de no equilibrio, la convergencia de $y(t)$ depende del prospecto de que $e^{-at} \to 0$ cuando $t \to \infty$. Esto puede suceder si y sólo si $a > 0$; si $a < 0$, entonces $e^{-at} \to \infty$ cuando $t \to \infty$, y $y(t)$ no puede convergir. Entonces, nuestra conclusión es la misma, ya sea que lleguemos a ella cuantitativa o cualitativamente.

Queda por discutir la línea de fase C, la cual, siendo un circuito cerrado que se sitúa a lo largo del eje horizontal, no califica como una *función* sino que muestra una *relación* entre dy/dt y y.[6] El nuevo elemento interesante que surge en este caso es la posibilidad de una trayectoria de tiempo que fluctúa periódicamente. Por la forma en que se dibuja la línea de fase C, encontraremos que y fluctúa entre los dos valores y_c y y'_c en un movimiento perpetuo. Con

[6] Esto puede surgir de una ecuación diferencial de segundo grado $(dy/dt)^2 = f(y)$.

objeto de generar la fluctuación periódica, el circuito debe atravesar el eje horizontal, de manera que dy/dt sea positiva y negativa en forma alterna. Además, en los dos puntos de intersección y_c y y'_c, la línea de fase debe tener una pendiente infinita; de otro modo, la intersección va a parecerse ya sea a y_a o a y_b, ninguno de los dos permite un flujo continuo de las cabezas de flecha. En la figura 15.4c se ilustra el tipo de trayectoria de tiempo $y(t)$ que corresponde a esta línea de fase de circuito. Observe que siempre que $y(t)$ toque el límite superior y'_c o el límite inferior y_c, tenemos $dy/dt = 0$ (extremos locales); pero estos valores no representan los valores de equilibrio de y. En términos de la figura 15.3, esto significa que no todas las intersecciones entre una línea de fase y el eje y son posiciones de equilibrio.

En resumen, para el estudio de la estabilidad dinámica de equilibrio (o la convergencia de la trayectoria de tiempo), tenemos la alternativa ya sea de encontrar la trayectoria de tiempo misma o simplemente dibujar la inferencia a partir de la línea de fase. Ilustraremos la aplicación de este último enfoque con el modelo de crecimiento de Solow. Luego, denotaremos el valor de equilibrio intertemporal de y mediante $\overline{y}$, siendo diferente de y^*.

EJERCICIO 15.6

1. Grafique la línea de fase para cada una de las siguientes ecuaciones y discuta sus implicaciones cualitativas:

   (*a*) $\dfrac{dy}{dt} = y - 7$ (*c*) $\dfrac{dy}{dt} = 4 - \dfrac{y}{2}$

   (*b*) $\dfrac{dy}{dt} = 1 - 5y$ (*d*) $\dfrac{dy}{dt} = 9y - 11$

2. Grafique la línea de fase para cada una de las siguientes ecuaciones e interprete:

   (*a*) $\dfrac{dy}{dt} = (y+1)^2 - 16 \qquad (y \geq 0)$

   (*b*) $\dfrac{dy}{dt} = \dfrac{1}{2}y - y^2 \qquad (y \geq 0)$

3. Dado $dy/dt = (y-3)(y-5) = y^2 - 8y + 15$:

   (*a*) Deduzca que hay dos niveles posibles de equilibrio de y, uno para $y = 3$ y el otro para $y = 5$.

   (*b*) Encuentre el signo de $\dfrac{d}{dy}\left(\dfrac{dy}{dt}\right)$ para $y = 3$ y $y = 5$, respectivamente. ¿Qué puede inferir a partir de esto?

15.7 El modelo de crecimiento de Solow

El modelo de crecimiento del profesor Robert Solow,[7] un laureado con el premio Nobel, tiene como objetivo mostrar, entre otras cosas, que la trayectoria de crecimiento del filo de la navaja de rasurar del modelo de Domar es principalmente el resultado de la hipótesis específica de producción-función adoptada en ese entonces y que, en otras circunstancias, puede no surgir la necesidad de un delicado balance.

El marco de referencia

En el modelo de Domar, el producto se enuncia explícitamente como una función solamente del capital: $\kappa = \rho K$ (la capacidad de producción, o producto potencial, es una constante múltiplo

[7] Robert M. Solow, "A Contribution to the Theory of Economic Growth", *Quarterly Journal of Economics*, febrero de 1956, pp. 65-94.

del capital). La ausencia del insumo de mano de obra en la función de producción conlleva la implicación de que la mano de obra siempre se combina con el capital en una proporción *fija*, de modo que es factible considerar explícitamente sólo uno de estos factores de la producción. En contraste, Solow busca analizar el caso en el cual el capital y la mano de obra pueden combinarse en proporciones *variables*. Entonces, la función de producción adopta la forma

$$Q = f(K, L) \qquad (K, L > 0)$$

donde Q es el producto (neto de depreciación), K es el capital y L es la mano de obra: todos se usan en el sentido *macro*. Se supone que f_K y f_L son positivos (productos marginales positivos), y f_{KK} y f_{LL} son negativos (retornos decrecientes para cada insumo). Aún más, la función de producción f se considera linealmente homogénea (retornos constantes a escala). En consecuencia, podemos escribir

$$Q = Lf\left(\frac{K}{L}, 1\right) = L\phi(k) \qquad \text{donde } k \equiv \frac{K}{L} \tag{15.25}$$

En vista de los signos supuestos de f_K y f_{KK}, la función ϕ de nueva introducción (la cual, fíjese bien, tiene sólo un argumento individual, k) debe caracterizarse por una primera derivada positiva y una segunda derivada negativa. Para verificar esta aseveración, de (12.94) recordamos primero que

$$f_K \equiv \text{MPP}_K = \phi'(k)$$

entonces, $f_K > 0$ significa automáticamente $\phi'(k) > 0$. Entonces, como

$$f_{KK} = \frac{\partial}{\partial K}\phi'(k) = \frac{d\phi'(k)}{dk}\frac{\partial k}{\partial K} = \phi''(k)\frac{1}{L} \qquad [\text{vea}\,(12.48)]$$

la hipótesis $f_{KK} < 0$ conduce directamente al resultado $\phi''(k) < 0$. Así la función ϕ —la cual, de acuerdo con (12.46), da el APP_L para cada relación capital— mano de obra es una que se incrementa con k para una tasa decreciente.

Dado que Q depende de K y L, ahora es necesario establecer cómo se determinan las últimas dos variables. Las hipótesis de Solow son:

$$\dot{K}\left(\equiv \frac{dK}{dt}\right) = sQ \quad [\text{se invierte una proporción constante de } Q] \tag{15.26}$$

$$\frac{\dot{L}}{L}\left(\equiv \frac{dL/dt}{L}\right) = \lambda \qquad (\lambda > 0) \;\; [\text{la fuerza laboral crece exponencialmente}] \tag{15.27}$$

El símbolo s representa una propensión al ahorro marginal (constante), y λ, una tasa (constante) de crecimiento de la mano de obra. Observe la naturaleza dinámica de estas hipótesis; no especifican cómo se determinan los *niveles* de K y L, sino cómo son las *tasas de cambio*.

Las ecuaciones (15.25) a (15.27) constituyen un modelo completo. Para resolver este modelo, lo condensaremos primero en una sola ecuación de una variable. Para comenzar, sustituya (15.25) en (15.26) para obtener

$$\dot{K} = sL\phi(k) \tag{15.28}$$

como $k \equiv K/L$ y $K \equiv kL$, sin embargo, podemos obtener otra expresión para $\dot{K}$ mediante la diferenciación de la última identidad:

$$\begin{aligned} \dot{K} &= L\dot{k} + k\dot{L} && [\text{regla del producto}] \\ &= L\dot{k} + k\lambda L && [\text{mediante } (15.27)] \end{aligned} \tag{15.29}$$

Cuando (15.29) se iguala con (15.28) y se elimina el factor común L, surge el resultado de que

$$\dot{k} = s\phi(k) - \lambda k \qquad \textbf{(15.30)}$$

Esta ecuación —una ecuación diferencial en la variable k, con dos parámetros s y λ— es la ecuación fundamental del modelo de crecimiento de Solow.

Análisis cualitativo-gráfico

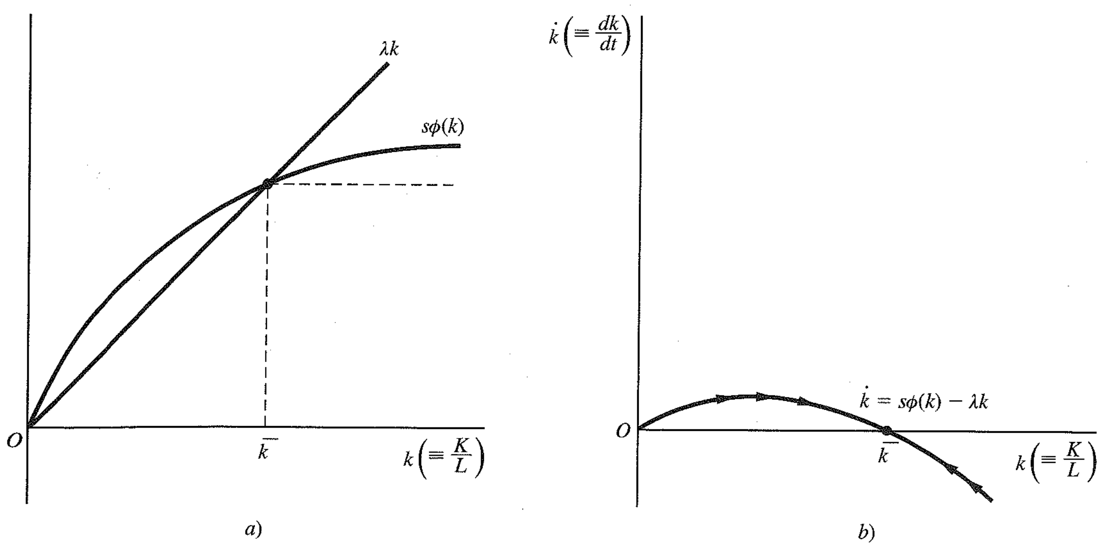
Aunque (15.30) se enuncia en forma de función general, no se dispone de una solución cuantitativa específica, pero podemos analizarla cualitativamente. Con este fin, debemos graficar una línea de fase, con $\dot{k}$ en el eje vertical y k en el horizontal.

Sin embargo, como (15.30) contiene dos términos a la derecha, primero grafiquemos éstos como dos curvas separadas. El término λk, una función lineal de k, obviamente va a hacerse presente en la figura 15.5*a* como una línea recta, con una intersección vertical cero y una pendiente igual a λ. Por otro lado, el término $s\phi(k)$ se grafica como una curva que se incrementa a una tasa decreciente, como $\phi(k)$, ya que $s\phi(k)$ es sólo una fracción constante de la curva $\phi(k)$. Si consideramos que K es un factor de producción indispensable, debemos iniciar la curva $s\phi(k)$ desde el punto de origen; esto es debido a que si $K = 0$ y entonces $k = 0$, Q también debe ser cero, como lo son $\phi(k)$ y $s\phi(k)$. La manera en que se dibuja la curva en realidad también refleja la hipótesis implícita de que existe un conjunto de valores k para los cuales $s\phi(k)$ sobrepasa a λk, de modo que las dos curvas se intersecan para algún valor positivo de k, a saber $\bar{k}$.

Basándose en estas dos curvas, el valor de $\dot{k}$ para cada valor de k puede medirse por la distancia vertical entre las dos curvas. Al graficar los valores de $\dot{k}$ contra k, como en la figura 15.5*b*, obtendremos entonces la línea de fase que necesitamos. Observe que, ya que las dos curvas de la figura 15.5*a* se intersecan cuando la relación capital-mano de obra es $\bar{k}$, la línea de fase de la figura 15.5*b* debe cruzar el eje horizontal para $\bar{k}$. Esto marca a $\bar{k}$ como la relación de capital-mano de obra de equilibrio intertemporal.

Siempre que la línea de fase tenga una pendiente negativa para $\bar{k}$, el equilibrio se identifica rápidamente como estable; dado cualquier valor inicial (positivo) de k, el movimiento dinámico del modelo debe conducirnos en forma convergente al nivel de equilibrio $\bar{k}$. El aspecto sig-

FIGURA 15.5



nificativo es que una vez que se alcanza este equilibrio —y entonces la relación capital— la mano de obra es (por definición) invariante con el tiempo: el capital debe crecer a partir de ese momento al mismo paso que la mano de obra, para la tasa idéntica λ. A su vez, esto implica que la inversión neta debe crecer a la tasa λ (vea el ejercicio 15.7-2). Observe sin embargo que la palabra *debe* se usa aquí no en el sentido de un requerimiento, sino con la implicación de la automaticidad. Entonces, el modelo de Solow sirve para mostrar que, dada una tasa de crecimiento de mano de obra λ, la economía por sí misma, y sin el delicado balance del modelo de Domar, puede finalmente alcanzar un estado de crecimiento uniforme en el cual la inversión va a crecer a la tasa λ, al igual que K y L. Aún más, con objeto de satisfacer (15.25), Q también debe crecer a la misma tasa ya que $\phi(\bar{k})$ es una constante cuando la relación capital-mano de obra permanece invariante para el nivel $\bar{k}$. Esta situación, en la cual las variables relevantes crecen todas a una tasa idéntica, se llama *estado uniforme* —una generalización del concepto del *estado estacionario*— (en el cual las variables relevantes permanecen constantes todas, o en otras palabras todas crecen a la tasa cero).

Observe que en el análisis anterior, por conveniencia se supone que la función de producción es invariante con el tiempo. Por otro lado, si se permite que mejore el estado de la tecnología, la función de producción deberá modificarse debidamente. Por ejemplo, ésta puede escribirse de forma

$$Q = T(t)f(K, L) \qquad \left(\frac{dT}{dt} > 0\right)$$

donde T, alguna medición de la tecnología, es una función creciente del tiempo. Debido al término multiplicativo creciente $T(t)$, una cantidad fija de K y L producirá un mayor producto para una fecha futura más que en el presente. En este caso, la curva $s\phi(k)$ de la figura 15.5 estará sujeta a un desplazamiento secular hacia arriba, lo que conduce a intersecciones sucesivamente más altas con el rayo λk y también a valores mayores de $\bar{k}$. Por lo tanto, con el mejoramiento tecnológico, será posible, en una sucesión de estados uniformes, tener una cantidad cada vez mayor de equipo capital disponible para cada trabajador representativo en la economía, con el aumento correspondiente en la productividad.

Una ilustración cuantitativa

El análisis anterior tuvo que ser cualitativo debido a la presencia de una función general $\phi(k)$ en el modelo. Pero si especificamos que la función de producción es una función de Cobb-Douglas linealmente homogénea, por ejemplo, entonces podemos encontrar también una solución cuantitativa.

Escribamos la función de producción como

$$Q = K^\alpha L^{1-\alpha} = L\left(\frac{K}{L}\right)^\alpha = Lk^\alpha$$

de modo que $\phi(k) = k^\alpha$. Entonces (15.30) se transforma en

$$\dot{k} = sk^\alpha - \lambda k \qquad \text{o sea} \qquad \dot{k} + \lambda k = sk^\alpha$$

que es una ecuación de Bernoulli en la variable k [vea (15.24)], con $R = \lambda$, $T = s$, y $m = \alpha$. Haciendo $z = k^{1-\alpha}$, obtenemos su versión linealizada

$$dz + [(1-\alpha)\lambda z - (1-\alpha)s]\,dt = 0$$

o sea

$$\frac{dz}{dt} + \underbrace{(1-\alpha)\lambda}_{a} z = \underbrace{(1-\alpha)s}_{b}$$

Ésta es una ecuación diferencial lineal con un coeficiente constante a y un término constante b. Entonces, por la fórmula (15.5′), tenemos

$$z(t) = \left[z(0) - \frac{s}{\lambda}\right]e^{-(1-\alpha)\lambda t} + \frac{s}{\lambda}$$

La sustitución de $z = k^{1-\alpha}$ va a suministrar entonces la solución final

$$k^{1-\alpha} = \left[k(0)^{1-\alpha} - \frac{s}{\lambda}\right]e^{-(1-\alpha)\lambda t} + \frac{s}{\lambda}$$

donde $k(0)$ es el valor inicial de la relación k capital-mano de obra.

Esta solución es lo que determina la trayectoria de tiempo de k. Recordando que $(1 - \alpha)$ y λ son ambos positivos, vemos que cuando $t \to \infty$ la expresión exponencial se aproxima a cero; en consecuencia,

$$k^{1-\alpha} \to \frac{s}{\lambda} \quad \text{o sea} \quad k \to \left(\frac{s}{\lambda}\right)^{1/(1-\alpha)} \quad \text{cuando } t \to \infty$$

Por lo tanto, la relación capital-mano de obra se aproxima a una constante para el valor de equilibrio. Este valor de equilibrio o valor de estado uniforme, $(s/\lambda)^{1/(1-\alpha)}$, varía directamente con la propensión al ahorro s, e inversamente con la tasa de crecimiento de la mano de obra λ.

EJERCICIO 15.7

1. Divida (15.30) entre k, e interprete la ecuación resultante en términos de las tasas de crecimiento de k, K y L.
2. Muestre que, si el capital está creciendo a la tasa λ (es decir, $K = Ae^{\lambda t}$), la inversión neta debe estar creciendo también a la tasa λ.
3. Las variables de insumo originales del modelo de Solow son K y L, pero la ecuación fundamental (15.30) se centra en la relación k de capital-mano de obra. ¿Qué hipótesis del modelo es (son) responsable(s) para este desplazamiento del enfoque (y lo hacen posible)? Explique.
4. Dibuje un diagrama de fase para cada una de las siguientes ecuaciones y discuta los aspectos cualitativos de la trayectoria de tiempo y/t):

   (*a*) $\dot{y} = 3 - y - \ln y$ (*b*) $\dot{y} = e^y - (y + 2)$f

Capítulo 16

Ecuaciones diferenciales de orden superior

En el capítulo 15 estudiamos los métodos de solución de una ecuación diferencial de *primer orden*, en la cual no hay derivadas (o diferenciales) de orden mayor que 1; sin embargo, a veces la especificación de un modelo puede incluir la segunda derivada o una derivada de orden aún mayor. Por ejemplo, se puede dar una función que describa "la tasa de cambio de la tasa de cambio" de la variable de ingreso, digamos Y,

$$\frac{d^2Y}{dt^2} = kY$$

a partir de la cual debemos encontrar la trayectoria de tiempo de Y. En este caso, la función dada constituye una ecuación diferencial de *segundo orden*, y la tarea de encontrar la trayectoria de tiempo $Y(t)$ es la de *resolver* la ecuación diferencial de segundo orden. Este capítulo aborda los métodos de solución y las aplicaciones económicas de estas ecuaciones diferenciales de orden superior, pero limitaremos nuestro estudio sólo al caso *lineal*.

Una variedad sencilla de las ecuaciones diferenciales lineales de orden n es de la siguiente manera:

$$\frac{d^ny}{dt^n} + a_1\frac{d^{n-1}y}{dt^{n-1}} + \cdots + a_{n-1}\frac{dy}{dt} + a_ny = b \qquad \textbf{(16.1)}$$

o con una notación alterna,

$$y^{(n)}(t) + a_1y^{(n-1)}(t) + \cdots + a_{n-1}y'(t) + a_ny = b \qquad \textbf{(16.1')}$$

Esta ecuación es de *orden n*, ya que la derivada n-ésima (el primer término a la izquierda) es la derivada más elevada presente. Es *lineal*, ya que todas las derivadas, así como la variable dependiente y, aparecen sólo en primer grado, y aún más, no hay términos de producto en los cuales se multipliquen y y cualquiera de sus derivadas. Además, puede observar que esta ecuación diferencial se caracteriza por *coeficientes constantes* (las a) y un *término constante* (b). Conservaremos en todo el capítulo la hipótesis de que los coeficientes son constantes. Por otro lado, el término b se supone constante como primera aproximación; posteriormente, en la sección 16.5 lo sustituiremos por un término variable.

16.1 Ecuaciones diferenciales lineales de segundo orden con coeficientes constantes y término constante

Por razones pedagógicas, estudiaremos primero el método de solución para el caso de *segundo orden* ($n = 2$). La ecuación diferencial relevante es entonces la ecuación sencilla

$$y''(t) + a_1 y'(t) + a_2 y = b \tag{16.2}$$

donde a_1, a_2 y b son todas constantes. Si el término b es idénticamente cero, tenemos una ecuación *homogénea*; pero si b es una constante diferente de cero, la ecuación es *no homogénea*. Nuestro estudio seguirá con la hipótesis de que (16.2) es no homogénea; al resolver la versión no homogénea de (16.2), la solución de la versión homogénea va a surgir automáticamente como un subproducto.

A este respecto, recordemos una proposición de la sección 15.1 que es igualmente aplicable aquí: si y_c es la *función complementaria*, es decir, la solución general (que contiene constantes arbitrarias) de la ecuación reducida de (16.2) y si y_p es la *integral particular*, es decir, cualquier solución particular (que no contiene constantes arbitrarias) de la ecuación completa (16.2), entonces $y(t) = y_c + y_p$ será la solución general de la ecuación completa. Como se explicó anteriormente, la componente y_p nos provee el valor de equilibrio de la variable y en el sentido intertemporal del término, mientras que la componente y_c revela, para cada instante de tiempo, la desviación de la trayectoria de tiempo $y(t)$ respecto al equilibrio.

La integral particular

Para el caso de coeficientes constantes y término constante, la integral particular es relativamente fácil de hallar. Como la integral particular puede ser *cualquier* solución de (16.2), es decir, cualquier valor de y que satisfaga esta ecuación no homogénea, debemos intentar siempre el tipo más sencillo posible: por ejemplo, y = una constante. Si y = una constante, se deduce que

$$y'(t) = y''(t) = 0$$

de modo que (16.2), en efecto, se transforma en $a_2 y = b$, con la solución $y = b/a_2$. Entonces, la integral particular deseada es

$$y_p = \frac{b}{a_2} \quad (\text{en caso de } a_2 \neq 0) \tag{16.3}$$

Como el proceso de encontrar el valor de y_p incluye la condición $y'(t) = 0$, el razonamiento para considerar ese valor como un equilibrio intertemporal se hace autoevidente.

Ejemplo 1 Encuentre la integral particular de la ecuación

$$y''(t) + y'(t) - 2y = -10$$

Los coeficientes relevantes aquí son $a_2 = -2$ y $b = -10$. Por lo tanto, la integral particular es $y_p = -10/(-2) = 5$.

¿Qué pasa si $a_2 = 0$, de modo que la expresión b/a_2 no está definida? En esta situación, como la solución constante para y_p no funciona, debemos intentar alguna forma *no constante* de solución. Tomando la posibilidad más sencilla, podemos intentar $y = kt$. Como $a_2 = 0$, la ecuación diferencial es ahora

$$y''(t) + a_1 y'(t) = b$$

pero si $y = kt$, lo que implica $y'(t) = k$ y $y''(t) = 0$, esta ecuación se reduce a $a_1 k = b$. Esto determina el valor de k como b/a_1, dándonos con ello la integral particular

$$y_p = \frac{b}{a_1}t \quad (\text{en caso de } a_2 = 0;\ a_1 \neq 0) \tag{16.3'}$$

Dado que y_p es en este caso una función no constante del tiempo, la consideraremos como un equilibrio móvil.

Ejemplo 2 Encuentre el y_p de la ecuación $y''(t) + y'(t) = -10$. Aquí tenemos $a_2 = 0$, $a_1 = 1$ y $b = -10$. Entonces, por (16.3′), podemos escribir

$$y_p = -10t$$

Si resulta que a_1 también es cero, entonces la forma de solución de $y = kt$ también se fragmenta, ya que la expresión bt/a_1 está ahora indefinida. Entonces debemos intentar una solución de la forma $y = kt^2$. Con $a_1 = a_2 = 0$, la ecuación diferencial se reduce ahora a la forma extremadamente simple

$$y''(t) = b$$

y si $y = kt^2$, lo que implica que $y'(t) = 2kt$ y $y''(t) = 2k$, la ecuación diferencial puede escribirse como $2k = b$. Entonces, encontramos $k = b/2$ y la integral particular es

$$y_p = \frac{b}{2}t^2 \quad (\text{en caso de } a_1 = a_2 = 0) \tag{16.3''}$$

El equilibrio representado por esta integral particular es nuevamente un equilibrio móvil.

Ejemplo 3 Encuentre el y_p de la ecuación $y''(t) = -10$. Como los coeficientes son $a_1 = a_2 = 0$ y $b = -10$, es aplicable la fórmula (16.3″). La respuesta deseada es $y_p = -5t^2$.

La función complementaria

La función complementaria de (16.2) se define como la solución general de su ecuación reducida (homogénea)

$$y''(t) + a_1 y'(t) + a_2 y = 0 \tag{16.4}$$

Ésta es la razón por la cual afirmamos que la solución de una ecuación homogénea siempre será un *subproducto* del proceso de solución de una ecuación completa.

Aun cuando nunca hemos enfrentado antes una ecuación de este tipo, nuestra experiencia con la función complementaria de las ecuaciones diferenciales de primer orden puede ofrecernos una sugerencia útil. A partir de las soluciones (15.3), (15.3′), (15.5) y (15.5′), es claro que las expresiones exponenciales de la forma Ae^{rt} destacan muy prominentemente en las funciones complementarias de las ecuaciones diferenciales de primer orden con coeficientes constantes. Ahora, ¿por qué no intentamos también una solución de la forma $y = Ae^{rt}$ en la ecuación de segundo orden?

Si adoptamos la solución de prueba $y = Ae^{rt}$, también debemos aceptar

$$y'(t) = rAe^{rt} \qquad \text{y} \qquad y''(t) = r^2 Ae^{rt}$$

como las derivadas de y. Basándose en estas expresiones para y, $y'(t)$ y $y''(t)$, la ecuación diferencial reducida (16.4) puede transformarse en

$$Ae^{rt}(r^2 + a_1 r + a_2) = 0 \tag{16.4'}$$

Siempre que escojamos los valores de A y r que satisfagan (16.4′), la solución de prueba $y = Ae^{rt}$ debería funcionar. Dado que e^{rt} nunca puede ser cero, debemos permitir $A = 0$ o ver que r satisfaga la ecuación

$$r^2 + a_1 r + a_2 = 0 \tag{16.4''}$$

Puesto que el valor de la constante A (arbitraria) debe hacerse definitivo mediante el uso de las condiciones iniciales del problema; sin embargo, no podemos simplemente hacer $A = 0$ a nuestro arbitrio. Por lo tanto, es esencial buscar valores de r que satisfagan a (16.4″).

La ecuación (16.4″) se conoce como la *ecuación característica* (o *ecuación auxiliar*) de la ecuación homogénea (16.4), o de la ecuación completa (16.2). Debido a que es una ecuación cuadrática en r, tiene dos raíces (soluciones), denominadas en el presente contexto *raíces características*, como sigue:[1]

$$r_1, r_2 = \frac{-a_1 \pm \sqrt{a_1^2 - 4a_2}}{2} \tag{16.5}$$

Estas dos raíces guardan entre sí una relación simple pero interesante, que puede servir como un medio conveniente para verificar nuestro cálculo: la *suma* de las dos raíces siempre es igual a $-a_1$ y su *producto* siempre es igual a a_2. La prueba de este enunciado es directa:

$$\begin{aligned} r_1 + r_2 &= \frac{-a_1 + \sqrt{a_1^2 - 4a_2}}{2} + \frac{-a_1 - \sqrt{a_1^2 - 4a_2}}{2} = \frac{-2a_1}{2} = -a_1 \\ r_1 r_2 &= \frac{(-a_1)^2 - (a_1^2 - 4a_2)}{4} = \frac{4a_2}{4} = a_2 \end{aligned} \tag{16.6}$$

Los valores de estas dos raíces son los únicos valores que podemos asignar a r en la solución $y = Ae^{rt}$. No obstante, esto significa que en efecto hay *dos* soluciones que van a funcionar,

$$y_1 = A_1 e^{r_1 t} \qquad \text{y} \qquad y_2 = A_2 e^{r_2 t}$$

donde A_1 y A_2 son dos constantes arbitrarias, y r_1 y r_2 son las raíces características encontradas de (16.5). Como queremos sólo *una* solución general, sin embargo, parece haber demasiadas. Ahora se nos presentan dos alternativas: (1) escoger aleatoriamente y_1 o y_2 o (2) combinarlas de alguna manera.

La primera alternativa, aunque más sencilla, es inaceptable. Existe sólo una constante arbitraria en y_1 o en y_2, pero para calificar como una solución general de una ecuación diferencial de *segundo orden* la expresión debe contener *dos* constantes arbitrarias. Este requerimiento proviene del hecho de que, al pasar de una función $y(t)$ a su segunda derivada $y''(t)$, "perdemos" dos constantes durante las dos rondas de diferenciación; por lo tanto, para revertir la función primitiva $y(t)$ de una ecuación diferencial de segundo orden debemos restituir dos constantes. Esto nos deja sólo la alternativa de combinar y_1 y y_2, a modo de incluir ambas

[1] Observe que la ecuación cuadrática (16.4″) está en la forma normalizada; el coeficiente del término r^2 es 1. Al aplicar la fórmula (16.5) para encontrar las raíces características de una ecuación diferencial, primero debemos asegurarnos de que la ecuación característica está realmente en la forma normalizada.

constantes A_1 y A_2. Tal como resulta, podemos simplemente tomar su *suma*, $y_1 + y_2$, como la solución general de (16.4). Demostremos que, si y_1 y y_2, respectivamente, satisfacen a (16.4), entonces la suma $(y_1 + y_2)$ también lo hará. Si y_1 y y_2 son realmente soluciones de (16.4), entonces al sustituir cada una de ellas en (16.4) debemos encontrar que son válidas las dos siguientes ecuaciones:

$$y_1''(t) + a_1 y_1'(t) + a_2 y_1 = 0$$
$$y_2''(t) + a_1 y_2'(t) + a_2 y_2 = 0$$

Sin embargo, al sumar estas ecuaciones encontramos que

$$\underbrace{[y_1''(t) + y_2''(t)]}_{=\frac{d^2}{dt^2}(y_1+y_2)} + a_1 \underbrace{[y_1'(t) + y_2'(t)]}_{=\frac{d}{dt}(y_1+y_2)} + a_2(y_1 + y_2) = 0$$

Entonces, al igual que y_1 o y_2, la suma $(y_1 + y_2)$ también satisface la ecuación (16.4). De acuerdo con esto, la solución general de la ecuación homogénea (16.4) o la función complementaria de la ecuación completa (16.2) pueden, en general, escribirse como $y_c = y_1 + y_2$.

Sin embargo, un examen más cuidadoso de la fórmula de raíces características (16.5) indica que por lo que toca a los valores de r_1 y r_2, pueden surgir tres casos posibles, algunos de los cuales tal vez requieran una modificación de nuestro resultado $y_c = y_1 + y_2$.

Caso 1 (raíces reales diferentes) Cuando $a_1^2 > 4a_2$, la raíz cuadrada de (16.5) es un número real, y las dos raíces r_1 y r_2 adoptan valores reales *diferentes*, ya que la raíz cuadrada se añade a $-a_1$ para r_1, pero se resta de $-a_1$ para r_2. En este caso, podemos escribir

$$y_c = y_1 + y_2 = A_1 e^{r_1 t} + A_2 e^{r_2 t} \qquad (r_1 \neq r_2) \qquad \textbf{(16.7)}$$

Debido a que las dos raíces son diferentes, las dos expresiones exponenciales deben ser linealmente independientes (ninguna de ellas es un múltiplo de la otra); en consecuencia, A_1 y A_2 siempre permanecerán como entidades separadas y proveerán dos constantes, como se requiere.

Ejemplo 4

Resuelva la ecuación diferencial

$$y''(t) + y'(t) - 2y = -10$$

Ya se ha encontrado que la integral particular de esta ecuación es $y_p = 5$, en el ejemplo 1. Encontremos la función complementaria. Dado que los coeficientes de la ecuación son $a_1 = 1$ y $a_2 = -2$, las raíces características son, por (16.5),

$$r_1, r_2 = \frac{-1 \pm \sqrt{1+8}}{2} = \frac{-1 \pm 3}{2} = 1, -2$$

(*Verifique:* $r_1 + r_2 = -1 = -a_1$; $r_1 r_2 = -2 = a_2$.) Como las raíces son números reales diferentes, la función complementaria es $y_c = A_1 e^t + A_2 e^{-2t}$. Por lo tanto, la solución general puede ser

$$y(t) = y_c + y_p = A_1 e^t + A_2 e^{-2t} + 5 \qquad \textbf{(16.8)}$$

Con objeto de determinar el valor de las constantes A_1 y A_2, ahora necesitamos dos condiciones iniciales. Sean estas condiciones $y(0) = 12$ y $y'(0) = -2$. Es decir, cuando $t = 0$, $y(t)$ y $y'(t)$ son, respectivamente, 12 y -2. Haciendo $t = 0$ en (16.8), encontramos que

$$y(0) = A_1 + A_2 + 5$$

Diferenciando (16.8) respecto a t y luego haciendo $t = 0$ en la derivada, encontramos que

$$y'(t) = A_1 e^t - 2A_2 e^{-2t} \qquad \text{y} \qquad y'(0) = A_1 - 2A_2$$

Por lo tanto, para satisfacer las dos condiciones iniciales, debemos hacer $y(0) = 12$ y $y'(0) = -2$, lo que conduce al siguiente par de ecuaciones simultáneas:

$$\begin{aligned} A_1 + A_2 &= 7 \\ A_1 - 2A_2 &= -2 \end{aligned}$$

con soluciones $A_1 = 4$ y $A_2 = 3$. Entonces la solución definida de la ecuación diferencial es

$$y(t) = 4e^t + 3e^{-2t} + 5 \tag{16.8'}$$

Como antes, podemos verificar la validez de esta solución por diferenciación. La primera y la segunda derivadas de (16.8′) son

$$y'(t) = 4e^t - 6e^{-2t} \qquad \text{y} \qquad y''(t) = 4e^t + 12e^{-2t}$$

Cuando esto se sustituye en la ecuación diferencial dada junto con (16.8′), el resultado es una identidad $-10 = -10$. Entonces la solución es correcta. Como usted puede verificar fácilmente, (16.8′) también satisface ambas condiciones iniciales.

Caso 2 (raíces reales repetidas) Cuando los coeficientes de la ecuación diferencial son tales que $a_1^2 = 4a_2$, se anula la raíz cuadrada en (16.5) y las dos raíces características toman un valor idéntico:

$$r(= r_1 = r_2) = -\frac{a_1}{2}$$

Las raíces de este tipo se conocen como *raíces repetidas*, o *raíces múltiples* (aquí son *dobles*).

Si intentamos escribir la función complementaria como $y_c = y_1 + y_2$, en este caso la suma va a colapsarse en una sola expresión

$$y_c = A_1 e^{rt} + A_2 e^{rt} = (A_1 + A_2)e^{rt} = A_3 e^{rt}$$

dejándonos sólo con una constante. Esto no es suficiente para conducirnos de una ecuación diferencial de segundo orden de vuelta a la función primitiva. La única salida es encontrar otro término componente elegible para la suma: un término que satisfaga a (16.4) y sea linealmente independiente del término $A_3 e^{rt}$, como para evitar este "colapso".

Una expresión que satisfaga estos requerimientos es $A_4 te^{rt}$. Puesto que la variable t ha ingresado en forma multiplicativa, es obvio que este término componente es linealmente independiente del término $A_3 e^{rt}$; entonces nos va a permitir introducir otra constante, A_4. Pero, ¿califica $A_4 te^{rt}$ como una solución de (16.4)? Si probamos $y = A_4 te^{rt}$, entonces por la regla del producto, podemos encontrar que sus derivadas primera y segunda son

$$y'(t) = (rt + 1)A_4 e^{rt} \qquad \text{y} \qquad y''(t) = (r^2 t + 2r)A_4 e^{rt}$$

Sustituyendo estas expresiones de y, y' y y'' en el lado izquierdo de (16.4), obtenemos la expresión

$$[(r^2 t + 2r) + a_1(rt + 1) + a_2 t]A_4 e^{rt}$$

Siempre que en el presente contexto tengamos $a_1^2 = 4a_2$ y $r = -a_1/2$, esta última expresión se anula idénticamente y entonces siempre es igual a la del lado derecho de (16.4); esto muestra que A_4te^{rt} ciertamente califica como solución.

Entonces, la función complementaria del caso de la doble raíz puede escribirse como

$$y_c = A_3e^{rt} + A_4te^{rt} \tag{16.9}$$

Ejemplo 5 Resuelva la ecuación diferencial

$$y''(t) + 6y'(t) + 9y = 27$$

Aquí los coeficientes son $a_1 = 6$ y $a_2 = 9$; como $a_1^2 = 4a_2$, las raíces se repiten. De acuerdo con la fórmula (16.5), tenemos $r = -a_1/2 = -3$. Entonces, en línea con el resultado de (16.9), la función complementaria puede escribirse como

$$y_c = A_3e^{-3t} + A_4te^{-3t}$$

La solución general de la ecuación diferencial dada es ahora también fácil de obtener. Si probamos una solución constante para la integral particular, obtenemos $y_p = 3$. De ahí que la solución general de la ecuación completa sea

$$y(t) = y_c + y_p = A_3e^{-3t} + A_4te^{-3t} + 3$$

Nuevamente las dos constantes arbitrarias pueden hacerse definitivas con dos condiciones iniciales. Suponga que las condiciones iniciales son $y(0) = 5$ y $y'(0) = -5$. Al hacer $t = 0$ en la solución general anterior, debemos encontrar $y(0) = 5$; es decir,

$$y(0) = A_3 + 3 = 5$$

Esto arroja $A_3 = 2$. Enseguida, al diferenciar la solución general y luego haciendo $t = 0$ y también $A_3 = 2$, debemos tener $y'(0) = -5$. Esto es,

$$y'(t) = -3A_3e^{-3t} - 3A_4te^{-3t} + A_4e^{-3t}$$

y

$$y'(0) = -6 + A_4 = -5$$

Esto arroja $A_4 = 1$. Finalmente podemos escribir la solución definida por las condiciones iniciales de la ecuación dada como

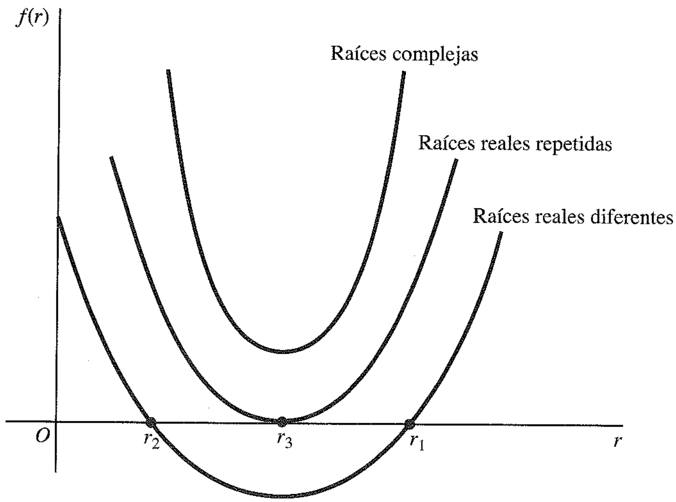
$$y(t) = 2e^{-3t} + te^{-3t} + 3$$

Caso 3 (raíces complejas) Queda por estudiar una tercera posibilidad en relación con la magnitud relativa de los coeficientes a_1 y a_2, a saber, $a_1^2 < 4a_2$. Cuando esto ocurra, la fórmula (16.5) va a incluir la raíz cuadrada de un número *negativo*, que no puede manipularse antes de haber introducido apropiadamente los conceptos de números *imaginarios* y *complejos*. Por lo tanto, por el momento estaremos satisfechos con la mera clasificación de este caso y dejaremos su estudio completo a las secciones 16.2 y 16.3.

Los tres casos citados pueden ilustrarse por las tres curvas de la figura 16.1, cada una de las cuales representa una versión diferente de la función cuadrática $f(r) = r^2 + a_1r + a_2$. Como aprendimos anteriormente, cuando una función de este tipo se hace igual a cero, el resultado es una *ecuación* cuadrática $f(r) = 0$ y la solución de esta última ecuación es simplemente "encontrar los ceros de la *función* cuadrática". Gráficamente, esto significa que las raíces de la ecuación deben encontrarse sobre el eje horizontal, cuando $f(r) = 0$.

La posición de la curva más baja en la figura 16.1 es tal que la curva interseca dos veces al eje horizontal; entonces podemos encontrar dos raíces diferentes r_1 y r_2, las cuales satisfacen la ecuación cuadrática $f(r) = 0$ y, por supuesto, tienen valor real. Así pues, la curva más baja

FIGURA 16.1



ilustra el caso 1. Si hacemos referencia a la curva de en medio, observamos que toca al eje horizontal sólo una vez, en r_3. Este último es el único valor de r que puede satisfacer la ecuación $f(r) = 0$. Por lo tanto, la curva de en medio ilustra el caso 2. Por último, observamos que la curva superior no toca para nada al eje horizontal, y que entonces no existe ninguna raíz con valor real para la ecuación $f(r) = 0$. Aun cuando no existan raíces reales en este caso, hay sin embargo dos números complejos que pueden satisfacer a la ecuación, como se mostrará en la sección 16.2.

La estabilidad dinámica del equilibrio

Para los casos 1 y 2, la condición de la estabilidad dinámica del equilibrio nuevamente depende de los signos algebraicos de las raíces características.

Para el caso 1, la función complementaria (16.7) consiste en las dos expresiones exponenciales $A_1e^{r_1t}$ y $A_2e^{r_2t}$. Los coeficientes A_1 y A_2 son constantes arbitrarias; su valor depende de las condiciones iniciales del problema. Entonces podemos estar seguros de un equilibrio dinámicamente estable ($y_c \to 0$ cuando $t \to \infty$), independientemente de cuáles sean las condiciones iniciales, si y solo si las raíces r_1 y r_2 son negativas *ambas*. Aquí enfatizamos la palabra *ambas* porque la condición de la estabilidad dinámica *no* permite ni siquiera que *una* de las raíces sea positiva o cero. Si $r_1 = 2$ y $r_2 = -5$, por ejemplo, puede parecer a primera vista que la segunda raíz, siendo mayor en valor absoluto, puede desplazar a la primera. Sin embargo, en la realidad es la raíz *positiva* la que finalmente debe dominar, ya que cuando t se incrementa, e^{2t} va a ser estrictamente creciente, pero e^{-5t} se aproxima a cero.

Para el caso 2, con raíces repetidas, la función complementaria (16.9) contiene no sólo la expresión familiar e^{rt}, sino también una expresión multiplicativa te^{rt}. Para que el primer término se aproxime a cero independientemente de las condiciones iniciales, es necesario y suficiente tener $r < 0$. ¿Pero eso garantizaría también la anulación de te^{rt}? Tal como resulta, la expresión te^{rt} (o en forma más general, t^ke^{rt}) posee el mismo tipo general de trayectoria de tiempo que e^{rt} ($r \neq 0$). Entonces la condición $r < 0$ es realmente necesaria y suficiente para que la función complementaria completa se aproxime a cero cuando $t \to \infty$, lo que conduce a un equilibrio intertemporal dinámicamente estable.

EJERCICIO 16.1

1. Encuentre La integral particular de cada ecuación:
   (*a*) $y''(t) - 2y'(t) + 5y = 2$ (*d*) $y''(t) + 2y'(t) - y = -4$
   (*b*) $y''(t) + y'(t) = 7$ (*e*) $y''(t) = 12$
   (*c*) $y''(t) + 3y = 9$
2. Encuentre la función complementaria de cada ecuación:
   (*a*) $y''(t) + 3y'(t) - 4y = 12$ (*c*) $y''(t) - 2y'(t) + y = 3$
   (*b*) $y''(t) + 6y'(t) + 5y = 10$ (*d*) $y''(t) + 8y'(t) + 16y = 0$
3. Encuentre la solución general de cada ecuación diferencial en el problema 2 y luego determine la solución que cumpla con las condiciones iniciales $y(0) = 4$ y $y'(0) = 2$.
4. ¿Son dinámicamente estables los equilibrios intertemporales encontrados en el problema 3?
5. Verifique que la solución definida por condiciones iniciales en el ejemplo 5 realmente (*a*) satisface las dos condiciones iniciales y (*b*) tiene primera y segunda derivadas que convierten en identidad a la ecuación diferencial dada.
6. Muestre que, cuando $t \to \infty$, el límite de te^{rt} es cero si $r < 0$, pero es infinita si $r \geq 0$.

16.2 Números complejos y funciones circulares

Cuando los coeficientes de una ecuación diferencial lineal de segundo orden, $y''(t) + a_1 y'(t) + a_2 y = b$, son tales que $a_1^2 < 4a_2$, la fórmula (16.5) de raíces características requeriría extraer la raíz cuadrada de un número *negativo*. Como el cuadrado de cualquier número real positivo o negativo es invariablemente positivo, mientras que el cuadrado de cero es cero, sólo un número real *no negativo* puede dar una raíz cuadrada de valor real. Entonces, si confinamos nuestra atención al sistema de los números reales, como hasta ahora, no disponemos de raíces características para este caso (caso 3). Este hecho nos motiva a considerar números fuera del sistema de los números reales.

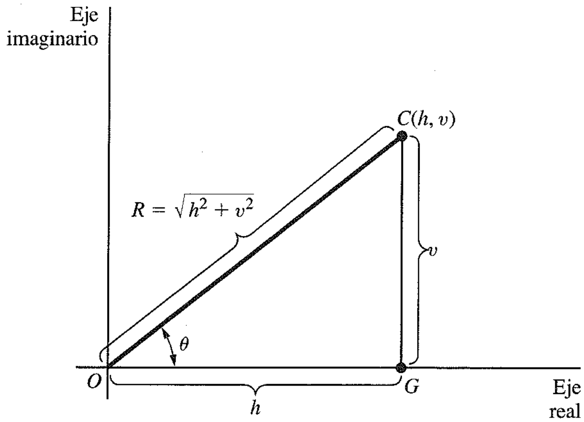
Números imaginarios y complejos

Conceptualmente, es posible definir un número $i \equiv \sqrt{-1}$, la cual, cuando se eleva al cuadrado, es igual a -1. Puesto que i es la raíz cuadrada de un número negativo, obviamente no tiene un valor real; por lo tanto, se denomina como *número imaginario*. Con esto a nuestra disposición, podemos escribir infinidad de números imaginarios, tales como $\sqrt{-9} = \sqrt{9}\sqrt{-1} = 3i$ y $\sqrt{-2} = \sqrt{2}i$.

Ampliando su aplicación un poco más, podemos construir aun otro tipo de número: uno que contenga una parte *real* así como una parte *imaginaria*, tal como $(8 + i)$ y $(3 + 5i)$. Conocidos como *números complejos*, éstos pueden representarse generalmente en la forma $(h + vi)$, donde h y v son dos números reales.[2] Desde luego, en el caso de $v = 0$, el número complejo se reduce a un número real, mientras que si $h = 0$, se transforma en un número imaginario. Entonces, el *conjunto de todos los números reales* (llamado **R**) constituye un subconjunto del *conjunto de todos los números complejos* (llamado **C**). En forma similar, el *conjunto de todos los números imaginarios* (llamado **I**) también constituye un subconjunto de **C**. Es decir, $\mathbf{R} \subset \mathbf{C}$ e $\mathbf{I} \subset \mathbf{C}$. Aún más, ya que los términos *real* e *imaginario* son mutuamente exclusivos, los conjuntos **R** e **I** deben ser disjuntos; es decir, $\mathbf{R} \cap \mathbf{I} = \varnothing$.

[2] Empleamos los símbolos *h* (para horizontal) y *v* (para vertical) en la notación general de los números complejos, ya que por último vamos a graficar los valores de *h* y *v*, respectivamente, en los ejes horizontal y vertical de un diagrama bidimensional.

FIGURA 16.2



Un número complejo $(h + vi)$ puede representarse gráficamente en lo que se denomina un *diagrama de Argand*, como se ilustra en la figura 16.2. Graficando h en sentido horizontal sobre el *eje real* y v en sentido vertical sobre el *eje imaginario*, el número $(h + vi)$ puede especificarse con el punto (h, v), el cual lo hemos rotulado alternativamente como C. Por supuesto que los valores de h y v tienen signo algebraico, de modo que si $h < 0$, el punto C estará a la izquierda del punto del origen; en forma similar, un v negativo implica una ubicación por debajo del eje horizontal.

Dados los valores de h y v, también podemos calcular la longitud de la línea OC aplicando el teorema de Pitágoras, que establece que el cuadrado de la hipotenusa de un triángulo rectángulo es la suma de los cuadrados de los otros dos lados. Si se denota la longitud de OC con R (que significa radio vector), tenemos

$$R^2 = h^2 + v^2 \qquad \text{y} \qquad R = \sqrt{h^2 + v^2} \tag{16.10}$$

donde la raíz cuadrada siempre se considera positiva. Algunas veces al valor de R se le denomina *valor absoluto*, o *módulo*, del número complejo $(h + vi)$. (Observe que el cambio de signos de h y v no va a producir ningún efecto sobre el valor absoluto del número complejo, R.) Entonces al igual que h y v, R tiene valor real, pero a diferencia de estos otros valores, R siempre es positivo. Encontraremos que el número R tiene mucha importancia en la siguiente discusión.

Raíces complejas

Mientras tanto, regresemos a la fórmula (16.5) y examinemos el caso de las raíces características complejas. Cuando los coeficientes de una ecuación diferencial de segundo orden son tales que $a_1^2 < 4a_2$, la expresión de raíz cuadrada de (16.5) puede escribirse como

$$\sqrt{a_1^2 - 4a_2} = \sqrt{4a_2 - a_1^2}\sqrt{-1} = \sqrt{4a_2 - a_1^2}\,i$$

Entonces, si adoptamos la notación

$$h = \frac{-a_1}{2} \qquad \text{y} \qquad v = \frac{\sqrt{4a_2 - a_1^2}}{2}$$

las dos raíces pueden denotarse como un par de *números complejos conjugados:*

$$r_1, r_2 = h \pm vi$$

Se dice que estas dos raíces complejas son conjugadas porque siempre aparecen juntas, siendo una la *suma* de h y vi, y siendo la otra la *diferencia* entre h y vi. Observe que comparten el mismo valor absoluto R.

Ejemplo 1

Encuentre las raíces de la ecuación característica $r^2 + r + 4 = 0$. Aplicando la fórmula conocida tenemos

$$r_1, r_2 = \frac{-1 \pm \sqrt{-15}}{2} = \frac{-1 \pm \sqrt{15}\sqrt{-1}}{2} = \frac{-1}{2} \pm \frac{\sqrt{15}}{2} i$$

lo que constituye un par de números complejos conjugados.

Como antes, podemos usar (16.6) para verificar nuestros cálculos. Si son correctos, debemos tener $r_1 + r_2 = -a_1 (= -1)$ y $r_1 r_2 = a_2 (= 4)$. Dado que ciertamente encontramos

$$\begin{aligned} r_1 + r_2 &= \left(\frac{-1}{2} + \frac{\sqrt{15}i}{2}\right) + \left(\frac{-1}{2} - \frac{\sqrt{15}i}{2}\right) \\ &= \frac{-1}{2} + \frac{-1}{2} = -1 \end{aligned}$$

y

$$\begin{aligned} r_1 r_2 &= \left(\frac{-1}{2} + \frac{\sqrt{15}i}{2}\right)\left(\frac{-1}{2} - \frac{\sqrt{15}i}{2}\right) \\ &= \left(\frac{-1}{2}\right)^2 - \left(\frac{\sqrt{15}i}{2}\right)^2 = \frac{1}{4} - \frac{-15}{4} = 4 \end{aligned}$$

nuestro cálculo está en verdad validado.

Aun en el caso de raíces complejas (caso 3), podemos expresar la función complementaria de una ecuación diferencial de acuerdo con (16.7), es decir,

$$y_c = A_1 e^{(h+vi)t} + A_2 e^{(h-vi)t} = e^{ht}(A_1 e^{vit} + A_2 e^{-vit}) \qquad \textbf{(16.11)}$$

Pero se ha introducido un nuevo aspecto: el número i aparece ahora en los exponentes de las dos expresiones que están entre paréntesis. ¿Cómo interpretamos estas funciones exponenciales imaginarias?

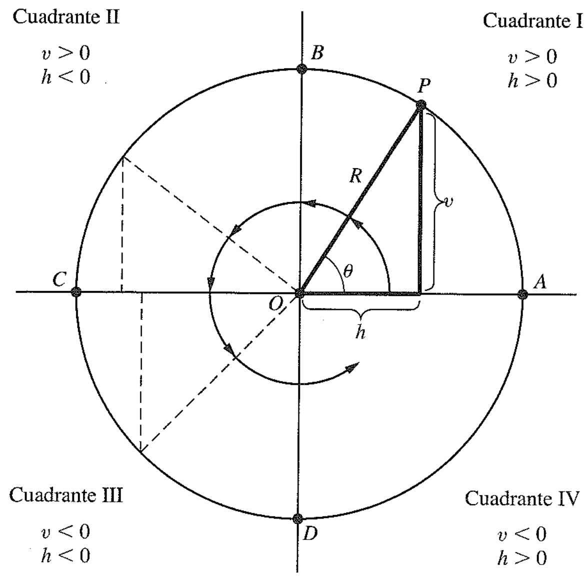
Para facilitar su interpretación, nos será de ayuda transformar primero estas expresiones en formas equivalentes de *función circular*. Como veremos más adelante, estas últimas funciones implican, como característica peculiar, las fluctuaciones periódicas de una variable. En consecuencia, la función complementaria (16.11), siendo transformable a las formas de la función circular, puede esperarse que genere una trayectoria de tiempo cíclica.

Funciones circulares

Considere un círculo con su centro en el punto de origen y con un radio de longitud R, como se muestra en la figura 16.3. Sea el radio, como una manecilla de reloj, que gira en sentido contrario a las manecillas del reloj. Comenzando en la posición OA, se va a mover de modo gradual hasta la posición OP, seguido sucesivamente de posiciones tales como OB, OC y OD; y al final de un ciclo, regresa a OA. A partir de ahí, el ciclo nada más se repite.

Cuando al estar en una posición específica —digamos OP— la manecilla del reloj forma un ángulo definido θ con la línea OA, y la punta de la manecilla (P) determina una distancia vertical v y una distancia horizontal h. A medida que cambia el ángulo θ durante el proceso de

FIGURA 16.3



rotación, v y h varían, aunque R no varía. Entonces, las relaciones v/R y h/R deben cambiar con θ; es decir, estas dos relaciones son funciones del ángulo θ. Específicamente, v/R y h/R se denominan, respectivamente, el *seno* (función) de θ y el *coseno* (función) de θ:

$$\operatorname{sen}\theta \equiv \frac{v}{R} \tag{16.12}$$

$$\cos\theta \equiv \frac{h}{R} \tag{16.13}$$

En vista de su relación con un círculo, estas funciones se denominan *funciones circulares*. Sin embargo, ya que también se asocian con un triángulo, se les denomina alternativamente *funciones trigonométricas*. Otro nombre para ellas (más deslumbrante) es *funciones sinusoidales*. Las funciones seno y coseno no son las únicas funciones circulares; otra que se encuentra con frecuencia es la función *tangente*, que se define como

$$\tan\theta = \frac{\operatorname{sen}\theta}{\cos\theta} = \frac{v}{h} \qquad (h \neq 0)$$

Sin embargo nuestra principal preocupación se relaciona con las funciones seno y coseno.

La variable independiente en una función circular es el ángulo θ, de modo que el mapeo que interviene aquí es de un *ángulo* a una *relación entre dos distancias*. Generalmente, los ángulos se miden en *grados* (por ejemplo, 30, 45 y 90°); sin embargo, en el trabajo analítico es más conveniente medir los ángulos en *radianes*. La ventaja de la medida en radianes proviene del hecho de que, cuando θ se mide de este modo, las derivadas de las funciones circulares van a conducir a expresiones más precisas, ya que la base e nos da derivadas más precisas para funciones exponenciales y logarítmicas. ¿Pero cuánto vale un radián? Para explicar esto, regresemos a la figura 16.3, donde se ha dibujado el punto P de modo que la longitud del *arco AP* sea exactamente igual al radio R. Entonces puede definirse al *radián* (abreviado como *rad*) como el tamaño del ángulo θ (en la figura 16.3) formado por un arco de longitud R. Puesto que la circunferencia de un círculo tiene una longitud total de $2\pi R$ (donde $\pi = 3.14159\ldots$), un círculo completo debe abarcar en su totalidad un ángulo de 2π rad. Sin embargo, en términos de grados, un círculo completo forma un ángulo de 360°; entonces, al igualar 360° con 2π rad, podemos llegar a la siguiente tabla de conversiones:

Grados	360	270	180	90	45	0
Radianes	2π	$\frac{3\pi}{2}$	π	$\frac{\pi}{2}$	$\frac{\pi}{4}$	0

Propiedades de las funciones seno y coseno

Dada la longitud de R, el valor del seno θ depende de la manera en que el valor de v cambia como respuesta a los cambios del ángulo θ. En la posición inicial OA, tenemos $v = 0$. A medida que la manecilla del reloj se mueve en sentido contrario, v comienza a tomar un valor creciente positivo, hasta culminar en el valor máximo de $v = R$ cuando la manecilla coincide con OB, es decir, cuando $\theta = \pi/2$ rad (= 90°). El movimiento adicional va a acortar gradualmente v, hasta que su valor se haga cero cuando la manecilla esté en la posición OC, es decir, cuando $\theta = \pi$ rad (= 180°). A medida que la manecilla entra en el tercer cuadrante, v comienza a tomar valores negativos; en la posición OD, tenemos $v = -R$. En el cuarto cuadrante, v todavía es negativo, pero va a incrementarse desde el valor de $-R$ hacia el valor de $v = 0$, el cual se alcanza cuando la manecilla regresa a OA: es decir, cuando $\theta = 2\pi$ rad (= 360°). Luego el ciclo se repite.

Cuando estos valores ilustrativos de v se sustituyen en (16.12), podemos obtener los resultados mostrados en el renglón de "seno θ" de la tabla 16.1. Sin embargo, para una descripción más completa de la función seno, vea la gráfica de la figura 16.4*a*, donde los valores de seno θ se grafican contra los de θ (expresado en radianes).

En contraste, el valor de cos θ depende de la manera en que h cambia en respuesta a los cambios de θ. En la posición inicial OA, tenemos $h = R$. Entonces, h disminuye gradualmente, hasta que $h = 0$ cuando $\theta = \pi/2$ (posición OB). En el segundo cuadrante, h se hace negativo, y cuando $\theta = \pi$ (posición OC), $h = -R$. El valor de h aumenta gradualmente desde $-R$ hasta cero en el tercer cuadrante, y cuando $\theta = 3\pi/2$ (posición OD), encontramos que $h = 0$. En el cuarto cuadrante, h se hace positivo nuevamente, y cuando la manecilla regresa a la posición OA ($\theta = 2\pi$), una vez más tenemos $h = R$. Luego el ciclo se repite.

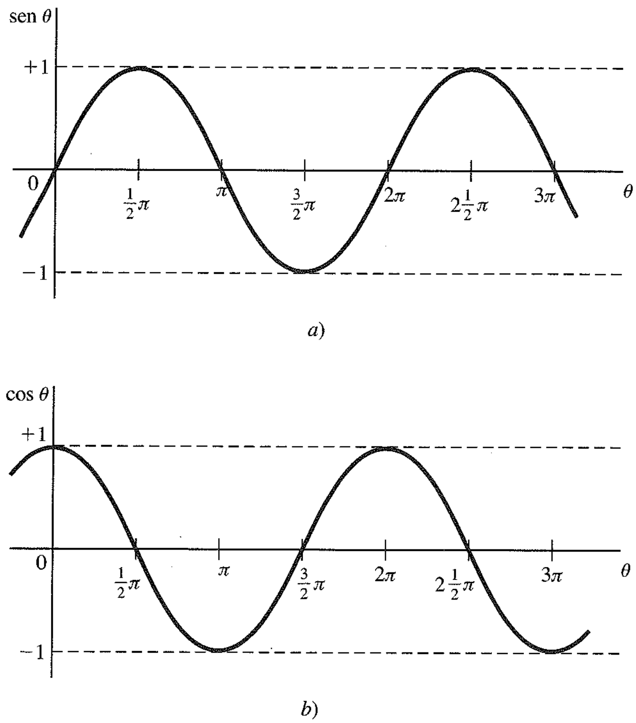
La sustitución de estos valores ilustrativos de h en (16.13) arroja los resultados del renglón inferior de la tabla 16.1, pero la figura 16.4*b* da una ilustración más completa de la función coseno.

Las funciones seno θ y cos θ comparten el mismo dominio, el conjunto de todos los números reales (las mediciones en radianes de θ). A este respecto, puede señalarse que un ángulo *negativo* simplemente se refiere a la rotación inversa de la manecilla del reloj; por ejemplo, un

TABLA 16.1

θ	0	$\frac{1}{2}\pi$	π	$\frac{3}{2}\pi$	2π
sen θ	0	1	0	−1	0
cos θ	1	0	−1	0	1

FIGURA 16.4



movimiento en el sentido de las manecillas del reloj de OA a OD en la figura 16.3 genera un ángulo de $-\pi/2$ rad $(= -90°)$. También existe una imagen común para las dos funciones, a saber, el intervalo cerrado $[-1, 1]$. Por esta razón, las gráficas de sen θ y cos θ se limitan en la figura 16.4 a una banda horizontal definida.

Una propiedad importante característica de las funciones seno y coseno es que ambas son *periódicas*; sus valores se repiten para cada 2π radianes (un círculo completo) que recorre el ángulo θ. Por lo tanto, se dice que cada función tiene un *periodo* de 2π. En vista de esta característica de periodicidad, son válidas las siguientes ecuaciones (para cualquier entero n):

$$\operatorname{sen}(\theta + 2n\pi) = \operatorname{sen}\theta \qquad \cos(\theta + 2n\pi) = \cos\theta$$

Es decir, la suma (o resta) de cualquier entero múltiplo de 2π a cualquier ángulo θ no va a afectar ni al valor de sen θ ni al del cos θ.

Las gráficas de las funciones seno y coseno indican un intervalo constante de fluctuación en cada periodo, a saber, ± 1. Algunas veces, esto se describe en forma alterna al decir que la *amplitud* de la fluctuación es 1. En virtud del periodo idéntico y de la amplitud idéntica, vemos que la curva de cos θ, si se desplaza hacia la derecha por $\pi/2$, va a coincidir exactamente con la curva sen θ. Por lo tanto se dice que estas dos curvas difieren sólo en la *fase*, es decir, difieren sólo en la localización de la cúspide para cada periodo. En forma simbólica, este hecho puede enunciarse por la ecuación

$$\cos\theta = \operatorname{sen}\left(\theta + \frac{\pi}{2}\right)$$

Las funciones seno y coseno obedecen a ciertas identidades. Entre éstas, las de uso más frecuente son

$$\begin{aligned} \operatorname{sen}(-\theta) &\equiv -\operatorname{sen}\theta \\ \cos(-\theta) &\equiv \cos\theta \end{aligned} \tag{16.14}$$

$$\operatorname{sen}^2\theta + \cos^2\theta \equiv 1 \qquad [\text{donde } \operatorname{sen}^2\theta \equiv (\operatorname{sen}\theta)^2, \text{ etc.}] \tag{16.15}$$

$$\begin{aligned} \operatorname{sen}(\theta_1 \pm \theta_2) &\equiv \operatorname{sen}\theta_1 \cos\theta_2 \pm \cos\theta_1 \operatorname{sen}\theta_2 \\ \cos(\theta_1 \pm \theta_2) &\equiv \cos\theta_1 \cos\theta_2 \mp \operatorname{sen}\theta_1 \operatorname{sen}\theta_2 \end{aligned} \tag{16.16}$$

El par de identidades (16.14) sirve para subrayar el hecho de que la función coseno es simétrica respecto al eje vertical (es decir, θ y $-\theta$ siempre suministran el mismo valor del coseno), mientas que la función seno no lo es. En (16.15) se muestra el hecho de que, para cualquier magnitud de θ, la suma de los cuadrados del seno y del coseno siempre es igual a la unidad. Y el conjunto de identidades de (16.16) da el seno y el coseno de la suma y de la diferencia de dos ángulos θ_1 y θ_2.

Finalmente, una palabra acerca de las derivadas. Siendo funciones continuas y suaves, tanto el sen θ como el cos θ son diferenciables. Las derivadas, $d(\operatorname{sen}\theta)/d\theta$ y $d(\cos\theta)/d\theta$, se obtienen tomando los límites, respectivamente, de los cocientes en diferencias $\Delta(\operatorname{sen}\theta)/\Delta\theta$ y $\Delta(\cos\theta)/\Delta\theta$ cuando $\Delta\theta \to 0$. Los resultados, enunciados aquí sin prueba, son

$$\frac{d}{d\theta}\operatorname{sen}\theta = \cos\theta \tag{16.17}$$

$$\frac{d}{d\theta}\cos\theta = -\operatorname{sen}\theta \tag{16.18}$$

Sin embargo, debemos enfatizar que estas fórmulas de derivadas son válidas sólo cuando θ se mide en radianes; si se mide en grados, por ejemplo, (16.17) se transforma en $d(\operatorname{sen}\theta)/d\theta = (\pi/180)\cos\theta$. Por la conveniencia de librarse del factor $(\pi/180)$, se prefieren las mediciones en radianes en vez de las mediciones en grados para el trabajo analítico.

Ejemplo 2 Encuentre la pendiente de la curva sen θ para $\theta = \pi/2$. La pendiente de la curva seno está dada por su derivada ($= \cos\theta$). Entonces, para $\theta = \pi/2$, la pendiente debería ser cos $(\pi/2) = 0$. Consulte la figura 16.4 para verificar este resultado.

Ejemplo 3 Encuentre la segunda derivada de sen θ. Por (16.17) sabemos que la primera derivada de sen θ es cos θ, por lo tanto la segunda derivada deseada es

$$\frac{d^2}{d\theta^2}\operatorname{sen}\theta = \frac{d}{d\theta}\cos\theta = -\operatorname{sen}\theta$$

Las relaciones de Euler

En la sección 9.5 mostramos que cualquier función que tenga derivadas finitas y continuas hasta el orden deseado puede expandirse a una función polinomial. Aún más, si el residuo R_n en la serie de Taylor que resulta (expansión para cualquier punto x_0) o en la serie de Maclaurin (expansión para $x_0 = 0$) se aproxima a cero a medida que el número de términos n se hace infinito, el polinomio puede escribirse como una serie infinita. Ahora vamos a expandir las funciones seno y coseno y luego intentaremos mostrar la forma en que las expresiones exponenciales imaginarias encontradas en (16.11) pueden transformarse en funciones circulares que tienen expansiones equivalentes.

Para la función seno, escriba $\phi(\theta) = \text{sen}\,\theta$; entonces se sigue que $\phi(0) = \text{sen}\,0 = 0$. Por derivación sucesiva obtenemos

$$\left.\begin{aligned} \phi'(\theta) &= \cos\theta \\ \phi''(\theta) &= -\text{sen}\theta \\ \phi'''(\theta) &= -\cos\theta \\ \phi^{(4)}(\theta) &= \text{sen}\,\theta \\ \phi^{(5)}(\theta) &= \cos\theta \\ \vdots \quad & \quad \vdots \end{aligned}\right\} \Rightarrow \left\{\begin{aligned} \phi'(0) &= \cos 0 = 1 \\ \phi''(0) &= -\text{sen}0 = 0 \\ \phi'''(0) &= -\cos 0 = -1 \\ \phi^{(4)}(0) &= \text{sen}\,0 = 0 \\ \phi^{(5)}(0) &= \cos 0 = 1 \\ \vdots \quad & \quad \vdots \end{aligned}\right.$$

Cuando se sustituye en (9.14), donde ahora θ reemplaza a x, esto nos dará la siguiente serie de Maclaurin con residuo:

$$\text{sen}\,\theta = 0 + \theta + 0 - \frac{\theta^3}{3!} + 0 + \frac{\theta^5}{5!} + \cdots + \frac{\phi^{(n+1)}(p)}{(n+1)!}\theta^{n+1}$$

Ahora, la expresión $\phi^{(n+1)}(p)$ del último término (residuo), que representa a la derivada $(n+1)$ evaluada para $\theta = p$ puede sólo adoptar la forma de $\pm \cos p$ o $\pm$ sen p y como tales pueden tomar sólo un valor en el intervalo $[-1, 1]$, independientemente de lo grande que sea n. Por otro lado, $(n+1)!$ va a crecer rápidamente cuando $n \to \infty$: de hecho mucho más rápidamente que θ^{n+1} a medida que n aumenta. Entonces, el residuo va a aproximarse a cero cuando $n \to \infty$ y, por lo tanto, podemos expresar a la serie de Maclaurin como una serie infinita:

$$\text{sen}\,\theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \frac{\theta^7}{7!} + \cdots \qquad \textbf{(16.19)}$$

En forma similar, si escribimos $\psi(\theta) = \cos\theta$, entonces $\psi(0) = \cos 0 = 1$, y las derivadas sucesivas serán

$$\left.\begin{aligned} \psi'(\theta) &= -\text{sen}\theta \\ \psi''(\theta) &= -\cos\theta \\ \psi'''(\theta) &= \text{sen}\theta \\ \psi^{(4)}(\theta) &= \cos\theta \\ \psi^{(5)}(\theta) &= -\text{sen}\theta \\ \vdots \quad & \quad \vdots \end{aligned}\right\} \Rightarrow \left\{\begin{aligned} \psi'(0) &= -\text{sen}0 = 0 \\ \psi''(0) &= -\cos 0 = -1 \\ \psi'''(0) &= \text{sen}0 = 0 \\ \psi^{(4)}(0) &= \cos 0 = 1 \\ \psi^{(5)}(0) &= -\text{sen}0 = 0 \\ \vdots \quad & \quad \vdots \end{aligned}\right.$$

Basándonos en estas derivadas, podemos expandir $\cos\theta$ como sigue:

$$\cos\theta = 1 + 0 - \frac{\theta^2}{2!} + 0 + \frac{\theta^4}{4!} + \cdots + \frac{\psi^{(n+1)}(p)}{(n+1)!}\theta^{n+1}$$

Dado que el residuo nuevamente tiende a cero cuando $n \to \infty$, la función coseno también puede expresarse como una serie infinita, como sigue:

$$\cos\theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \cdots \qquad \textbf{(16.20)}$$

Probablemente observó que con (16.19) y (16.20) a la mano, ahora podemos construir una tabla de los valores de seno y coseno para todos los valores posibles de θ (en radianes). Sin embargo, nuestro interés inmediato radica en encontrar la relación entre las expresiones exponenciales imaginarias y las funciones circulares. Con este fin, expandamos ahora las dos expresiones exponenciales $e^{i\theta}$ y $e^{-i\theta}$. El lector reconocerá que éstos son sólo casos especiales de la expresión e^x, que se ha demostrado anteriormente en (10.6) que tiene la expansión

$$e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + \cdots$$

Por lo tanto, haciendo $x = i\theta$, podemos obtener inmediatamente

$$\begin{aligned} e^{i\theta} &= 1 + i\theta + \frac{(i\theta)^2}{2!} + \frac{(i\theta)^3}{3!} + \frac{(i\theta)^4}{4!} + \frac{(i\theta)^5}{5!} + \cdots \\ &= 1 + i\theta - \frac{\theta^2}{2!} - \frac{i\theta^3}{3!} + \frac{\theta^4}{4!} + \frac{i\theta^5}{5!} - \cdots \\ &= \left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \cdots\right) + i\left(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \cdots\right) \end{aligned}$$

En forma similar, haciendo $x = -i\theta$, surge el siguiente resultado:

$$\begin{aligned} e^{-i\theta} &= 1 - i\theta + \frac{(-i\theta)^2}{2!} + \frac{(-i\theta)^3}{3!} + \frac{(-i\theta)^4}{4!} + \frac{(-i\theta)^5}{5!} + \cdots \\ &= 1 - i\theta - \frac{\theta^2}{2!} + \frac{i\theta^3}{3!} + \frac{\theta^4}{4!} - \frac{i\theta^5}{5!} - \cdots \\ &= \left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \cdots\right) - i\left(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \cdots\right) \end{aligned}$$

Al sustituir (16.19) y (16.20) en estos dos resultados, podemos establecer rápidamente el siguiente par de identidades, que se conocen como *relaciones de Euler*:

$$e^{i\theta} \equiv \cos\theta + i\,\text{sen}\,\theta \qquad \textbf{(16.21)}$$

$$e^{-i\theta} \equiv \cos\theta - i\,\text{sen}\,\theta \qquad \textbf{(16.21')}$$

Esto nos va a permitir transformar cualquier función exponencial imaginaria en una combinación lineal equivalente de las funciones seno y coseno y viceversa.

Ejemplo 4 Encuentre el valor de $e^{i\pi}$. Primero convirtamos esta expresión en una expresión trigonométrica. Haciendo $\theta = \pi$ en (16.21), se encuentra que $e^{i\pi} = \cos\pi + i\,\text{sen}\,\pi$. Como $\cos\pi = -1$ y $\text{sen}\,\pi = 0$, se sigue que $e^{i\pi} = -1$.

Ejemplo 5 Muestre que $e^{-i\pi/2} = -i$. Haciendo $\theta = \pi/2$ en (16.21'), tenemos

$$e^{-i\pi/2} = \cos\frac{\pi}{2} - i\,\text{sen}\,\frac{\pi}{2} = 0 - i(1) = -i$$

Representaciones alternas de los números complejos

Hasta ahora hemos representado un par de números complejos conjugados en la forma general $(h \pm vi)$. Como h y v se refieren a la abscisa y a la ordenada en el sistema coordenado Cartesiano de un diagrama de Argand, la expresión $(h \pm vi)$ representa la *forma Cartesiana* de un par de números complejos conjugados. Como un subproducto del estudio de las funciones circulares y de las relaciones de Euler, podemos expresar ahora $(h \pm vi)$ de otras dos maneras.

Haciendo referencia a la figura 16.2, vemos que en cuanto se especifican h y v, el ángulo θ y el valor de R también quedan determinados. Puesto que un θ dado y un R dado pueden identificar juntos un único punto en el diagrama de Argand, podemos emplear θ y R para especificar el par particular de números complejos. Al reescribir las definiciones de las funciones seno y coseno en (16.12) y (16.13), como

$$v = R\,\text{sen}\,\theta \quad \text{y} \quad h = R\cos\theta \tag{16.22}$$

los números complejos conjugados $(h \pm vi)$ pueden transformarse como sigue:

$$h \pm vi = R\cos\theta \pm Ri\,\text{sen}\,\theta = R(\cos\theta \pm i\,\text{sen}\,\theta)$$

Al hacerlo, hemos cambiado de las coordenadas Cartesianas de los números complejos (h y v) a lo que se llama sus *coordenadas polares* (R y θ). De acuerdo con esto, la expresión de la derecha en la ecuación anterior ejemplifica la *forma polar* de un par de números complejos conjugados.

Aún más, en vista de las relaciones de Euler, la forma polar también puede reescribirse en *forma exponencial* como sigue: $R(\cos\theta \pm i\,\text{sen}\,\theta) = Re^{\pm i\theta}$. Entonces tenemos un total de tres representaciones alternas de los números complejos conjugados:

$$h \pm vi = R(\cos\theta \pm i\,\text{sen}\,\theta) = Re^{\pm i\theta} \tag{16.23}$$

Si nos dan los valores de R y θ, la transformación a h y v es directa: usamos las dos ecuaciones de (16.22). ¿Qué pasa con la transformación inversa? Con valores dados de h y v, no hay dificultad para encontrar el valor correspondiente de R, que es igual a $\sqrt{h^2 + v^2}$. Pero surge una ligera ambigüedad respecto a θ: el valor deseado de θ (en radianes) es aquel que satisface las dos condiciones $\cos\theta = h/R$ y $\text{sen}\,\theta = v/R$; pero para valores dados de h y v, ¡θ no es único! ¿Por qué? Afortunadamente, el problema no es grave, ya que al limitar nuestra atención al intervalo $[0, 2\pi)$ en el dominio, la indeterminación se resuelve rápidamente.

Ejemplo 6

Encuentre la forma Cartesiana del número complejo $5e^{3i\pi/2}$. Aquí tenemos $R = 5$ y $\theta = 3\pi/2$; entonces, por (16.22) y la tabla 16.1,

$$h = 5\cos\frac{3\pi}{2} = 0 \qquad \text{y} \qquad v = 5\,\text{sen}\,\frac{3\pi}{2} = -5$$

Entonces, la forma Cartesiana es simplemente $h + vi = -5i$.

Ejemplo 7

Encuentre las formas polar y exponencial de $(1 + \sqrt{3}i)$. En este caso, tenemos $h = 1$ y $v = \sqrt{3}$; entonces $R = \sqrt{1+3} = 2$. La tabla 16.1 no nos sirve esta vez para encontrar el valor de θ, pero la tabla 16.2, la cual lista algunos valores seleccionados adicionales de sen θ y cos θ, nos va a ser de ayuda. Específicamente, buscamos un valor de θ tal que $\cos\theta = h/R = 1/2$ y

TABLA 16.2

θ	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{3\pi}{4}$
sen θ	$\frac{1}{2}$	$\frac{1}{\sqrt{2}}\left(=\frac{\sqrt{2}}{2}\right)$	$\frac{\sqrt{3}}{2}$	$\frac{1}{\sqrt{2}}\left(=\frac{\sqrt{2}}{2}\right)$
cos θ	$\frac{\sqrt{3}}{2}$	$\frac{1}{\sqrt{2}}\left(=\frac{\sqrt{2}}{2}\right)$	$\frac{1}{2}$	$\frac{-1}{\sqrt{2}}\left(=\frac{-\sqrt{2}}{2}\right)$

$\operatorname{sen}\theta = v/R = \sqrt{3}/2$. El valor $\theta = \pi/3$ cumple con los requerimientos. Entonces, de acuerdo con (16.23), la transformación deseada es

$$1 + \sqrt{3}i = 2\left(\cos\frac{\pi}{3} + i\operatorname{sen}\frac{\pi}{3}\right) = 2e^{i\pi/3}$$

Antes de dejar este tema, observemos una extensión importante del resultado de (16.23). Suponiendo que tenemos la n-ésima potencia de un número complejo —digamos, $(h + vi)^n$— ¿cómo escribimos sus formas polar y exponencial? La forma exponencial es la más fácil de obtener. Como $h + vi = Re^{i\theta}$, se sigue que

$$(h + vi)^n = (Re^{i\theta})^n = R^n e^{in\theta}$$

En forma similar podemos escribir

$$(h - vi)^n = (Re^{-i\theta})^n = R^n e^{-in\theta}$$

Observe que la potencia n ha ocasionado dos cambios: (1) ahora R se transforma en R^n, y (2) ahora θ se transforma en $n\theta$. Cuando estos dos cambios se insertan en la forma polar en (16.23), encontramos que

$$(h \pm vi)^n = R^n(\cos n\theta \pm i\operatorname{sen} n\theta) \qquad \textbf{(16.23')}$$

Es decir,

$$[R(\cos\theta \pm i\operatorname{sen}\theta)]^n = R^n(\cos n\theta \pm i\operatorname{sen} n\theta)$$

Conocido como el *teorema de De Moivre*, este resultado indica que, para elevar un número complejo a la n-ésima potencia, debemos modificar simplemente sus coordenadas polares al elevar a R a la n-ésima potencia y al multiplicar θ por n.

EJERCICIO 16.2

1. Encuentre las raíces de las siguientes ecuaciones cuadráticas:
   (*a*) $r^2 - 3r + 9 = 0$ (*c*) $2x^2 + x + 8 = 0$
   (*b*) $r^2 + 2r + 17 = 0$ (*d*) $2x^2 - x + 1 = 0$
2. (*a*) ¿Cuántos grados hay en un radián?
   (*b*) ¿Cuántos radianes hay en un grado?
3. Haciendo referencia a la figura 16.3, y usando el teorema de Pitágoras, pruebe que
   (*a*) $\operatorname{sen}^2\theta + \cos^2\theta \equiv 1$ (*b*) $\operatorname{sen}\frac{\pi}{4} = \cos\frac{\pi}{4} = \frac{1}{\sqrt{2}}$
4. Mediante las identidades (16.14), (16.15) y (16.16), muestre que
   (*a*) $\operatorname{sen} 2\theta \equiv 2\operatorname{sen}\theta\cos\theta$
   (*b*) $\cos 2\theta \equiv 1 - 2\operatorname{sen}^2\theta$
   (*c*) $\operatorname{sen}(\theta_1 + \theta_2) + \operatorname{sen}(\theta_1 - \theta_2) \equiv 2\operatorname{sen}\theta_1\cos\theta_2$
   (*d*) $1 + \tan^2\theta \equiv \dfrac{1}{\cos^2\theta}$
   (*e*) $\operatorname{sen}\left(\frac{\pi}{2} - \theta\right) \equiv \cos\theta$ (*f*) $\cos\left(\frac{\pi}{2} - \theta\right) \equiv \operatorname{sen}\theta$
5. Aplicando la regla de la cadena:
   (*a*) Desarrolle las fórmulas de derivación para $\frac{d}{d\theta}\operatorname{sen} f(\theta)$ y $\frac{d}{d\theta}\cos f(\theta)$, donde $f(\theta)$ es una función de θ.
   (*b*) Encuentre las derivadas de $\cos\theta^3$, $\operatorname{sen}(\theta^2 + 3\theta)$, $\cos e^\theta$ y $\operatorname{sen}(1/\theta)$.

6. De las relaciones de Euler, deduzca que:

   (*a*) $e^{-i\pi} = -1$ (*c*) $e^{i\pi/4} = \frac{\sqrt{2}}{2}(1+i)$

   (*b*) $e^{i\pi/3} = \frac{1}{2}(1+\sqrt{3}i)$ (*d*) $e^{-3i\pi/4} = -\frac{\sqrt{2}}{2}(1+i)$

7. Encuentre la forma Cartesiana de cada número complejo:

   (*a*) $2\left(\cos\frac{\pi}{6} + i\,\text{sen}\frac{\pi}{6}\right)$ (*b*) $4e^{i\pi/3}$ (*c*) $\sqrt{2}e^{-i\pi/4}$

8. Encuentre las formas polar y exponencial de los siguientes números complejos:

   (*a*) $\frac{3}{2} + \frac{3\sqrt{3}}{2}i$ (*b*) $4(\sqrt{3}+i)$

16.3 Análisis del caso de las raíces complejas

Con los conceptos de números complejos y funciones circulares a nuestra disposición, ahora estamos preparados para examinar el caso de las raíces complejas (caso 3), estudiado en la sección 16.1. Recordará que la clasificación de los tres casos, de acuerdo con la naturaleza de las raíces características, alude sólo a la función complementaria de una ecuación diferencial. Entonces podemos continuar centrando nuestra atención en la ecuación reducida

$$y''(t) + a_1y'(t) + a_2y = 0 \qquad \text{[reproducido de (16.4)]}$$

La función complementaria

Cuando los valores de los coeficientes a_1 y a_2 son tales que $a_1^2 < 4a_2$, las raíces características serán el par de números complejos conjugados

$$r_1, r_2 = h \pm vi$$

donde

$$h = -\frac{1}{2}a_1 \qquad \text{y} \qquad v = \frac{1}{2}\sqrt{4a_2 - a_1^2}$$

La función complementaria, como se ha visto anteriormente, estará en la forma

$$y_c = e^{ht}(A_1e^{vit} + A_2e^{-vit}) \qquad \text{[reproducido de (16.11)]}$$

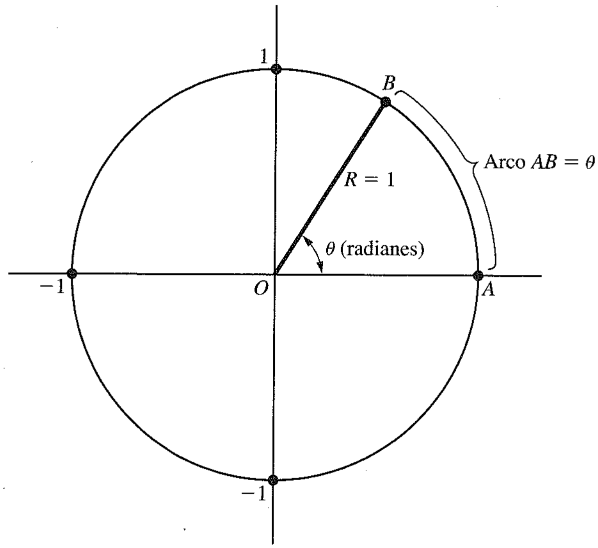
Transformemos primero las expresiones exponenciales imaginarias que están entre paréntesis a expresiones trigonométricas equivalentes, de modo que podamos interpretar la función complementaria como una función circular. Esto puede lograrse con el uso de las relaciones de Euler. Haciendo $\theta = vt$ en (16.21) y (16.21′), encontramos que

$$e^{vit} = \cos vt + i\,\text{sen}\,vt \qquad \text{y} \qquad e^{-vit} = \cos vt - i\,\text{sen}\,vt$$

De éstos, se sigue que la función complementaria de (16.11) puede reescribirse como

$$\begin{aligned} y_c &= e^{ht}[A_1(\cos vt + i\,\text{sen}\,vt) + A_2(\cos vt - i\,\text{sen}\,vt)] \\ &= e^{ht}[(A_1 + A_2)\cos vt + (A_1 - A_2)i\,\text{sen}\,vt] \end{aligned} \qquad \textbf{(16.24)}$$

FIGURA 16.5



Aún más, si empleamos los símbolos

$$A_5 \equiv A_1 + A_2 \quad \text{y} \quad A_6 \equiv (A_1 - A_2)i$$

es posible simplificar (16.24) a[3]

$$y_c = e^{ht}(A_5 \cos vt + A_6 \operatorname{sen} vt) \tag{16.24'}$$

donde las nuevas constantes arbitrarias A_5 y A_6 se van a determinar posteriormente.

Si usted es escrupuloso, puede sentirse un poco incómodo acerca de la sustitución de θ por vt en el procedimiento anterior. La variable θ mide un ángulo, pero vt es una magnitud en unidades de t (en nuestro contexto es el tiempo). Por lo tanto, ¿cómo podemos hacer la sustitución $\theta = vt$? La respuesta a esta pregunta puede explicarse de la mejor manera haciendo referencia al *círculo unitario* (un círculo con radio $R = 1$) de la figura 16.5. Es verdad que hemos usado θ para designar un ángulo; pero como el ángulo se mide en unidades de radianes, el valor de θ siempre es la relación de la longitud del arco AB entre el radio R. Cuando $R = 1$, tenemos específicamente

$$\theta \equiv \frac{\text{arc } AB}{R} \equiv \frac{\text{arc } AB}{1} \equiv \text{arc } AB$$

En otras palabras, θ es no sólo la medida en radianes del ángulo, sino también la longitud del arco AB, el cual es un número más que un ángulo. Si se grafica el paso del tiempo en la circunferencia del círculo unitario (en el sentido contrario de las manecillas del reloj), en vez de en una línea recta como lo hacemos al graficar una serie de tiempo, realmente no hay ninguna

[3] El hecho de que al definir A_6 incluyamos el número imaginario i no es de ninguna manera un intento para "esconder la basura bajo la alfombra". Dado que A_6 es una constante arbitraria, puede adoptar tanto un valor imaginario como un valor real. Tampoco es cierto que, tal como se define, A_6 necesariamente será imaginaria. En realidad, si A_1 y A_2 son un par de números complejos conjugados, digamos $m \pm ni$, entonces A_5 y A_6 serán ambos reales: $A_5 = A_1 + A_2 = (m + ni) + (m - ni) = 2m$ y $A_6 = (A_1 - A_2)i = [(m + ni) - (m - ni)]i = (2ni)i = -2n$.

diferencia si consideramos un periodo como un incremento en la medición en radianes del ángulo θ o como un alargamiento del arco AB. Aún si $R \neq 1$, adicionalmente, podemos aplicar la misma línea de razonamiento, excepto que en ese caso θ será igual a (arco AB)/R; es decir, el ángulo θ y el arco AB deben guardar entre sí una proporción fija, en lugar de ser iguales. Entonces, la sustitución $\theta = vt$ es en verdad legítima.

Un ejemplo de solución

Encontremos la solución de la ecuación diferencial

$$y''(t) + 2y'(t) + 17y = 34$$

con las condiciones iniciales $y(0) = 3$ y $y'(0) = 11$.

Como $a_1 = 2$, $a_2 = 17$ y $b = 34$, podemos encontrar inmediatamente que la integral particular es

$$y_p = \frac{b}{a_2} = \frac{34}{17} = 2 \qquad \text{[mediante (16.3)]}$$

Aún mas, puesto que $a_1^2 = 4 < 4a_2 = 68$, las raíces características serán el par de números complejos conjugados $(h \pm vi)$, donde

$$h = -\frac{1}{2}a_1 = -1 \qquad \text{y} \qquad v = \frac{1}{2}\sqrt{4a_2 - a_1^2} = \frac{1}{2}\sqrt{64} = 4$$

Entonces mediante (16.24′), la función complementaria es

$$y_c = e^{-t}(A_5 \cos 4t + A_6 \operatorname{sen} 4t)$$

Al combinar y_c y y_p, la solución general puede expresarse como

$$y(t) = e^{-t}(A_5 \cos 4t + A_6 \operatorname{sen} 4t) + 2$$

Para determinar las constantes A_5 y A_6 utilizamos las dos condiciones iniciales. Primero al hacer $t = 0$ en la solución general, encontramos que

$$\begin{aligned} y(0) &= e^0(A_5 \cos 0 + A_6 \operatorname{sen} 0) + 2 \\ &= (A_5 + 0) + 2 = A_5 + 2 \qquad [\cos 0 = 1; \operatorname{sen} 0 = 0] \end{aligned}$$

Mediante la condición inicial $y(0) = 3$, podemos especificar entonces $A_5 = 1$. Enseguida, diferenciemos la solución general respecto a t—usando la regla del producto y las fórmulas de derivadas (16.17) y (16.18) mientras se tiene en mente la regla de la cadena [ejercicio 16.2-5]— para encontrar $y'(t)$ y luego $y'(0)$:

$$y'(t) = -e^{-t}(A_5 \cos 4t + A_6 \operatorname{sen} 4t) + e^{-t}[A_5(-4 \operatorname{sen} 4t) + 4A_6 \cos 4t]$$

de modo que

$$\begin{aligned} y'(0) &= -(A_5 \cos 0 + A_6 \operatorname{sen} 0) + (-4A_5 \operatorname{sen} 0 + 4A_6 \cos 0) \\ &= -(A_5 + 0) + (0 + 4A_6) = 4A_6 - A_5 \end{aligned}$$

Mediante la segunda condición inicial $y'(0) = 11$, y en vista de que $A_5 = 1$, entonces se hace claro que $A_6 = 3$.[4] Por lo tanto, la solución definida por las condiciones iniciales es

$$y(t) = e^{-t}(\cos 4t + 3 \operatorname{sen} 4t) + 2 \qquad \textbf{(16.25)}$$

[4] Observe que aquí, A_6 realmente es un número real, aun cuando hemos incluido al número imaginario i en su definición.

Como antes, la componente y_p $(= 2)$ puede interpretarse como el nivel de equilibrio intertemporal de y, mientras que la componente y_c representa la desviación respecto al equilibrio. Debido a la presencia de las funciones circulares en y_c puede esperarse que la trayectoria de tiempo (16.25) exhiba un patrón de fluctuación. ¿Pero qué patrón específico será incluido?

La trayectoria de tiempo

Ya conocemos las trayectorias de una función simple seno o coseno, como se muestra en la figura 16.4. Ahora debemos estudiar las trayectorias de ciertas variantes y combinaciones de las funciones seno y coseno de modo que podamos interpretar en general la función complementaria (16.24′).

$$y_c = e^{ht}(A_5 \cos vt + A_6 \operatorname{sen} vt)$$

y en particular la componente y_c de (16.25).

Examinemos primero el término $(A_5 \cos vt)$. Por sí misma, la expresión $(\cos vt)$ es una función circular de (vt), con periodo 2π $(= 6.2832)$ y amplitud 1. El periodo de 2π significa que la gráfica va a repetir su configuración cada vez que (vt) se incremente por 2π. Sin embargo, cuando t sola se toma como la variable independiente, la repetición ocurre cada vez que t se incrementa por $2\pi/v$, de modo que con referencia a t —como es apropiado en el análisis económico dinámico— consideraremos que el periodo de $(\cos vt)$ es de $2\pi/v$ (sin embargo, la amplitud permanece como 1). Ahora, cuando se agrega una constante multiplicativa A_5 a $(\cos vt)$, esto hace que la imagen de fluctuación cambie de ± 1 a $\pm A_5$. Entonces la amplitud se transforma ahora en A_5, aun cuando el periodo no esté afectado por esta constante. En resumen, $(A_5 \cos vt)$ es una función coseno de t, con periodo $2\pi/v$ y amplitud A_5. Por la misma razón, $(A_6 \operatorname{sen} vt)$ es una función seno de t, con periodo $2\pi/v$ y amplitud A_6.

Habiendo un periodo común, la suma $(A_5 \cos vt + A_6 \operatorname{sen} vt)$ también va a exhibir un ciclo repetitivo cada vez que t se incremente por $2\pi/v$. Para mostrar esto con más rigor, observemos que para valores dados de A_5 y A_6 siempre podemos encontrar dos constantes A y ε tales que

$$A_5 = A \cos \varepsilon \qquad \text{y} \qquad A_6 = -A \operatorname{sen} \varepsilon$$

Entonces podemos expresar dicha suma como

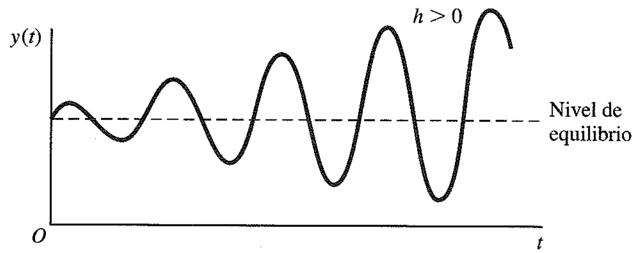
$$\begin{aligned} A_5 \cos vt + A_6 \operatorname{sen} vt &= A \cos \varepsilon \cos vt - A \operatorname{sen} \varepsilon \operatorname{sen} vt \\ &= A(\cos vt \cos \varepsilon - \operatorname{sen} vt \operatorname{sen} \varepsilon) \\ &= A \cos(vt + \varepsilon) \qquad [\text{por } (16.16)] \end{aligned}$$

Ésta es una función coseno de t modificada, con amplitud A y periodo $2\pi/v$, ya que cada vez que t se incrementa por $2\pi/v$, $(vt + \varepsilon)$ va a incrementarse por 2π, con lo cual se completa un ciclo de la curva coseno.

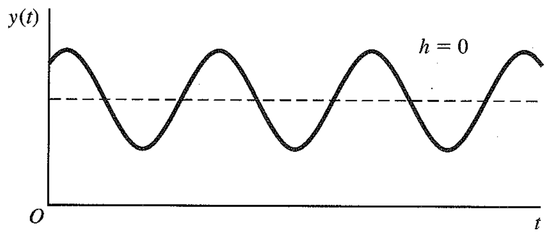
Si y_c hubiera consistido sólo en la expresión $(A_5 \cos vt + A_6 \operatorname{sen} vt)$, la implicación habría sido que la trayectoria de tiempo de y sería una fluctuación que nunca termina de amplitud constante alrededor del valor de equilibrio de y, como lo representa y_p. Pero de hecho también hay que considerar al término multiplicativo e^{ht}. Este último término es de mucha importancia ya que, como veremos, contiene la clave para responder la pregunta sobre la convergencia de la trayectoria en el tiempo.

Si $h > 0$, el valor de e^{ht} va a incrementarse continuamente a medida que t se incrementa. Esto va a producir un efecto amplificador sobre la amplitud de $(A_5 \cos vt + A_6 \operatorname{sen} vt)$ y causa desviaciones cada vez mayores respecto al equilibrio para cada ciclo sucesivo. Como se ilustra en la figura 16.6*a*, la trayectoria de tiempo va a caracterizarse en este caso por una *fluc-*

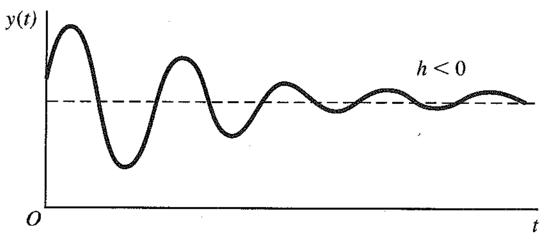
FIGURA 16.6



a)



b)



c)

tuación explosiva. Si $h = 0$, por otro lado, entonces $e^{ht} = 1$, y la función complementaria será simplemente $(A_5 \cos vt + A_6 \operatorname{sen} vt)$, de la cual se ha mostrado que tiene una amplitud constante. En este segundo caso, cada ciclo va a exhibir un patrón uniforme de desviación respecto al equilibrio como se ilustra en la trayectoria de tiempo de la figura 16.6*b*. Ésta es una trayectoria de tiempo con *fluctuación uniforme*. Por último, si $h < 0$, el término e^{ht} va a disminuir continuamente a medida que t se incrementa, y cada ciclo sucesivo tendrá una amplitud menor que el anterior, de la misma manera que una ola va desvaneciéndose. Este caso se ilustra en la figura 16.6*c*, donde la trayectoria de tiempo se caracteriza por una *fluctuación amortiguada*. La solución en (16.25), con $h = -1$, ejemplifica este último caso. Debe ser claro que sólo el caso de la fluctuación amortiguada puede producir una trayectoria de tiempo *convergente*; en los otros dos casos, la trayectoria de tiempo es *no convergente* o *divergente*.[5]

En los tres diagramas de la figura 16.6, se supone que el equilibrio intertemporal es estacionario. Si es móvil, los tres tipos ilustrados de trayectoria de tiempo van a fluctuar alrededor de aquél, pero puesto que un equilibrio móvil generalmente se grafica como una curva en vez de

[5] Usaremos las dos palabras *no convergente* y *divergente* en forma intercambiable, aunque esta última es aplicable más estrictamente a la variedad explosiva de no convergencia en vez de la fluctuación uniforme.

una línea recta horizontal, la fluctuación va a adoptar la naturaleza de, digamos, una serie de ciclos de negocios alrededor de una tendencia secular.

La estabilidad dinámica del equilibrio

El concepto de convergencia de la trayectoria de tiempo de una variable está inextricablemente unido al concepto de estabilidad dinámica del equilibrio intertemporal de esa variable. De modo específico, el equilibrio es dinámicamente estable si y sólo si la trayectoria de tiempo es convergente. La condición de convergencia de la trayectoria $y(t)$, es decir, $h < 0$ (figura 16.6*c*), es también por lo tanto la condición de estabilidad dinámica del equilibrio intertemporal de y.

Usted recordará que para los casos 1 y 2 en los cuales las raíces características son reales, la condición de estabilidad dinámica de equilibrio es que cada raíz característica sea negativa. En el caso presente (caso 3), con raíces complejas, la condición parece ser más especializada; estipula sólo que la parte real (h) de las raíces complejas ($h \pm vi$) sea negativa. Sin embargo, es posible unificar los tres casos y consolidar las condiciones aparentemente diferentes en una sola, que en general es aplicable. Simplemente interprete cualquier raíz real r como una raíz compleja cuya parte imaginaria es cero ($v = 0$). Entonces la condición "la parte *real* de toda raíz característica debe ser negativa" se hace claramente aplicable a los tres casos y surge como la única condición que necesitamos.

EJERCICIO 16.3

Encuentre y_p y y_c, la solución general, y la solución definida por las condiciones iniciales de cada una de las siguientes ecuaciones diferenciales:

1. $y''(t) - 4y'(t) + 8y = 0$; $y(0) = 3$, $y'(0) = 7$
2. $y''(t) + 4y'(t) + 8y = 2$; $y(0) = 2\frac{1}{4}$, $y'(0) = 4$
3. $y''(t) + 3y'(t) + 4y = 12$; $y(0) = 2$, $y'(0) = 2$
4. $y''(t) - 2y'(t) + 10y = 5$; $y(0) = 6$, $y'(0) = 8\frac{1}{2}$
5. $y''(t) + 9y = 3$; $y(0) = 1$, $y'(0) = 3$
6. $2y''(t) - 12y'(t) + 20y = 40$; $y(0) = 4$, $y'(0) = 5$
7. ¿Cuáles de las ecuaciones diferenciales de los problemas 1 a 6 proporcionan trayectorias de tiempo con (*a*) fluctuación amortiguada; (*b*) fluctuación uniforme; (*c*) fluctuación explosiva?

16.4 Un modelo de mercado con expectativas de precio

En la formulación anterior del modelo dinámico de mercado, tanto Q_d como Q_s se consideran funciones sólo del precio presente P. Pero algunas veces los compradores y los vendedores pueden basar su comportamiento de mercado no sólo en el precio presente sino también en la *tendencia* de precios que prevalece en el tiempo, ya que es probable que la tendencia de precios los conduzca a ciertas *expectativas* respecto al nivel de precios en el futuro, y estas expectativas pueden a su vez influir en sus decisiones de oferta y demanda.

La tendencia de precios y las expectativas de precios

En el contexto de tiempo continuo, la información de la tendencia de precios va a encontrarse principalmente en las dos derivadas dP/dt (si es que el precio está subiendo) y d^2P/dt^2 (si es que está subiendo a una tasa creciente). Para considerar la tendencia de precios, incluyamos estas derivadas como argumentos adicionales en las funciones de oferta y demanda:

$$Q_d = D[P(t), P'(t), P''(t)]$$
$$Q_s = S[P(t), P'(t), P''(t)]$$

Si nos limitamos a la versión lineal de estas funciones y simplificamos la notación para las variables independientes a P, P' y P'', podemos escribir

$$\begin{aligned} Q_d &= \alpha - \beta P + mP' + nP'' \qquad (\alpha, \beta > 0) \\ Q_s &= -\gamma + \delta P + uP' + wP'' \qquad (\gamma, \delta > 0) \end{aligned} \tag{16.26}$$

donde los parámetros α, β, γ y δ provienen de los modelos de mercado anteriores, pero m, n, u y w son nuevos.

Los cuatro parámetros nuevos, cuyos signos no se han restringido, agrupan las expectativas de precio de los compradores y los vendedores. Si $m > 0$, por ejemplo, un aumento de precio va a causar el incremento de Q_d. Esto sugeriría que los compradores esperan que el precio *continúe* subiendo y, por lo tanto, prefieren incrementar sus compras ahora, cuando el precio todavía es relativamente bajo. Por otro lado, el signo opuesto de m implicaría la expectativa de una rápida inversión de la tendencia de precios, de modo que los compradores preferirían hacer recortes en las compras presentes y esperar que se imponga después un precio más bajo. La inclusión del parámetro n hace que el comportamiento de los compradores dependa también de la tasa de cambio de dP/dt. Entonces, los nuevos parámetros m y n inyectan un elemento sustancial de especulación de precios en el modelo. Los parámetros u y w conllevan una implicación similar para el lado de los vendedores.

Un modelo simplificado

Por sencillez, supondremos que sólo la función de demanda contiene expectativas de precios. Específicamente, hacemos m y n diferentes de cero, pero hacemos $u = w = 0$ en (16.26). Supongamos aún más que el mercado está en ceros para todo instante de tiempo. Entonces podemos igualar las funciones de oferta y demanda para obtener (después de la normalización) la ecuación diferencial

$$P'' + \frac{m}{n}P' - \frac{\beta + \delta}{n}P = -\frac{\alpha + \gamma}{n} \tag{16.27}$$

Esta ecuación está en la forma de (16.2) con las siguientes sustituciones:

$$y = P \qquad a_1 = \frac{m}{n} \qquad a_2 = -\frac{\beta + \delta}{n} \qquad b = -\frac{\alpha + \gamma}{n}$$

Como el patrón de cambio de P incluye la segunda derivada P'', así como la primera derivada P', este modelo ciertamente es diferente del modelo dinámico de mercado presentado en la sección 15.2.

Observe, sin embargo, que este modelo difiere del anterior de otra manera. En la sección 15.2 está presente un mecanismo de ajuste dinámico, $dP/dt = j(Q_d - Q_s)$. Puesto que esa ecuación implica que $dP/dt = 0$ si y sólo si $Q_d = Q_s$, el sentido intertemporal y el sentido de equilibrio de clarificación del mercado coinciden en ese modelo. En contraste, este modelo supone que el mercado está en ceros en todo instante de tiempo. Entonces todo precio alcanzado en el mercado es un precio de equilibrio en el sentido de que el mercado está en ceros, aunque tal vez no califique como precio de equilibrio intertemporal. En otras palabras, los dos sentidos del equilibrio son ahora desiguales. Observe también que el mecanismo de ajuste $dP/dt = j(Q_d - Q_s)$, que contiene una derivada, es lo que hace dinámico al anterior modelo de mercado.

En este modelo, sin mecanismo de ajuste, la naturaleza dinámica del modelo emana de los términos de expectativa mP' y nP''.

La trayectoria de tiempo de los precios

El precio de equilibrio intertemporal de este modelo —la integral particular P_p (anteriormente y_p)— se localiza muy fácil a través de (16.3). Es

$$P_p = \frac{b}{a_2} = \frac{\alpha + \gamma}{\beta + \delta}$$

Debido a que ésta es una constante (positiva), representa un equilibrio estacionario.

En cuanto a la función complementaria P_c (anteriormente y_c), hay tres casos posibles.

Caso 1 (raíces reales diferentes)

$$\left(\frac{m}{n}\right)^2 > -4\left(\frac{\beta + \delta}{n}\right)$$

La función complementaria para este caso es, mediante (16.7),

$$P_c = A_1 e^{r_1 t} + A_2 e^{r_2 t}$$

donde

$$r_1, r_2 = \frac{1}{2}\left[-\frac{m}{n} \pm \sqrt{\left(\frac{m}{n}\right)^2 + 4\left(\frac{\beta + \delta}{n}\right)}\right] \qquad \textbf{(16.28)}$$

De acuerdo con esto, la solución general es

$$P(t) = P_c + P_p = A_1 e^{r_1 t} + A_2 e^{r_2 t} + \frac{\alpha + \gamma}{\beta + \delta} \qquad \textbf{(16.29)}$$

Caso 2 (raíces reales dobles)

$$\left(\frac{m}{n}\right)^2 = -4\left(\frac{\beta + \delta}{n}\right)$$

En este caso, las raíces características adoptan el valor individual

$$r = -\frac{m}{2n}$$

Entonces, mediante (16.9), la solución general puede escribirse como

$$P(t) = A_3 e^{-mt/2n} + A_4 t e^{-mt/2n} + \frac{\alpha + \gamma}{\beta + \delta} \qquad \textbf{(16.29')}$$

Caso 3 (raíces complejas)

$$\left(\frac{m}{n}\right)^2 < -4\left(\frac{\beta + \delta}{n}\right)$$

Para este tercero y último caso, las raíces características son el par de números complejos conjugados

$$r_1, r_2 = h \pm vi$$

donde

$$h = -\frac{m}{2n} \qquad \text{y} \qquad v = \frac{1}{2}\sqrt{-4\left(\frac{\beta+\delta}{n}\right) - \left(\frac{m}{n}\right)^2}$$

Por lo tanto, mediante (16.24′), tenemos la solución general

$$P(t) = e^{-mt/2n}(A_5 \cos\, vt + A_6 \operatorname{sen} vt) + \frac{\alpha+\gamma}{\beta+\delta} \qquad \textbf{(16.29″)}$$

De estos resultados pueden deducirse un par de conclusiones generales. Primero, si $n > 0$, entonces $-4(\beta+\delta)/n$ debe ser negativo y, por lo tanto, menor que $(m/n)^2$. Entonces los casos 2 y 3 pueden descartarse inmediatamente. Aún más, con n positivo (tal como β y δ), la expresión bajo el radical de (16.28) necesariamente sobrepasa a $(m/n)^2$, y entonces la raíz cuadrada debe ser mayor que $|m/n|$. Entonces el signo $\pm$ de (16.28) produciría una raíz positiva (r_1) y una raíz negativa (r_2). En consecuencia, el equilibrio intertemporal es dinámicamente inestable, a menos que el valor de la constante A_1 determinado por las condiciones iniciales resulte ser cero en (16.29).

Segundo, si $n < 0$, entonces los tres casos se hacen factibles. Para el caso 1, podemos estar seguros de que ambas raíces serán negativas si m es negativo (¿por qué?). Es interesante que la raíz repetida del caso 2 también sea negativa si m es negativo. Aún más, ya que h, la parte real de las raíces complejas del caso 3, adopta el mismo valor que la raíz repetida r del caso 2, la negatividad de m también va a garantizar que h es negativo. En resumen, para los tres casos, la estabilidad dinámica de equilibrio se asegura cuando los parámetros m y n son negativos.

Ejemplo 1

Sean las funciones de oferta y demanda

$$Q_d = 42 - 4P - 4P' + P''$$
$$Q_s = -6 + 8P$$

Con condiciones iniciales $P(0) = 6$ y $P'(0) = 4$. Suponiendo que el mercado está en ceros en todos los instantes de tiempo, encuentre la trayectoria de tiempo $P(t)$.

En este ejemplo, los valores de los parámetros son

$$\alpha = 42 \qquad \beta = 4 \qquad \gamma = 6 \qquad \delta = 8 \qquad m = -4 \qquad n = 1$$

Como n es positivo, nuestro estudio anterior sugiere que sólo puede presentarse el caso 1, y que las dos raíces (reales) r_1 y r_2 van a tomar signos opuestos. La sustitución de los valores de los parámetros de (16.28) en verdad confirma esto, ya que

$$r_1, r_2 = \frac{1}{2}(4 \pm \sqrt{16+48}) = \frac{1}{2}(4 \pm 8) = 6, -2$$

La solución general es, entonces, mediante (16.29),

$$P(t) = A_1 e^{6t} + A_2 e^{-2t} + 4$$

Aún más, al considerar las condiciones iniciales, encontramos que $A_1 = A_2 = 1$, de modo que la solución definida es

$$P(t) = e^{6t} + e^{-2t} + 4$$

En vista de la raíz positiva $r_1 = 6$, el equilibrio intertemporal ($P_p = 4$) es dinámicamente inestable.

La solución anterior se encuentra a través del uso de las fórmulas (16.28) y (16.29). En forma alterna, podemos igualar primero las funciones dadas de oferta y demanda para obtener la ecuación diferencial

$$P'' - 4P' - 12P = -48$$

y luego resolver esta ecuación como un caso específico de (16.2).

Ejemplo 2 Dadas las funciones de oferta y demanda

$$Q_d = 40 - 2P - 2P' - P''$$
$$Q_s = -5 + 3P$$

Con $P(0) = 12$ y $P'(0) = 1$, encuentre $P(t)$ con la hipótesis de que el mercado siempre está en ceros.

Aquí, los parámetros m y n son negativos. Por lo tanto, de acuerdo con nuestra discusión general anterior, el equilibrio intertemporal debe ser dinámicamente estable. Para encontrar la solución específica, podemos igualar primero Q_d y Q_s para obtener la ecuación diferencial (después de multiplicar todo por -1)

$$P'' + 2P' + 5P = 45$$

El equilibrio intertemporal está dado por la integral particular

$$P_p = \frac{45}{5} = 9$$

De la ecuación característica de la ecuación diferencial,

$$r^2 + 2r + 5 = 0$$

encontramos que las raíces son complejas:

$$r_1, r_2 = \frac{1}{2}(-2 \pm \sqrt{4-20}) = \frac{1}{2}(-2 \pm 4i) = -1 \pm 2i$$

Esto significa que $h = -1$ y $v = 2$, de modo que la solución general es

$$P(t) = e^{-t}(A_5 \cos 2t + A_6 \operatorname{sen} 2t) + 9$$

Para determinar las constantes arbitrarias A_5 y A_6, hacemos $t = 0$ en la solución general, para obtener

$$P(0) = e^0(A_5 \cos 0 + A_6 \operatorname{sen} 0) + 9 = A_5 + 9 \qquad [\cos 0 = 1; \operatorname{sen} 0 = 0]$$

Aún más, al diferenciar la solución general y luego hacer $t = 0$, encontramos que

$$P'(t) = -e^{-t}(A_5 \cos 2t + A_6 \operatorname{sen} 2t) + e^{-t}(-2A_5 \operatorname{sen} 2t + 2A_6 \cos 2t)$$

[regla del producto y regla de la cadena]

y
$$\begin{aligned} P'(0) &= -e^0(A_5 \cos 0 + A_6 \operatorname{sen} 0) + e^0(-2A_5 \operatorname{sen} 0 + 2A_6 \cos 0) \\ &= -(A_5 + 0) + (0 + 2A_6) = -A_5 + 2A_6 \end{aligned}$$

Entonces, en virtud de las condiciones iniciales $P(0) = 12$ y $P'(0) = 1$, tenemos $A_5 = 3$ y $A_6 = 2$. En consecuencia, la solución definida por las condiciones iniciales es

$$P(t) = e^{-t}(3 \cos 2t + 2 \operatorname{sen} 2t) + 9$$

Obviamente que esta trayectoria de tiempo tiene fluctuación periódica; el periodo es $2\pi/v = \pi$. Es decir, se da un ciclo completo cada vez que t se incrementa por $\pi = 3.14159 \ldots$ En vista del término multiplicativo e^{-t}, la fluctuación se amortigua. La trayectoria de tiempo, que inicia desde el precio inicial $P(0) = 12$, converge hasta el precio de equilibrio intertemporal $P_p = 9$ de una manera cíclica.

EJERCICIO 16.4

1. Sean los parámetros m, n, u y w de (16.26) todos diferentes de cero.
   (*a*) Suponiendo que el mercado está en ceros para todo instante de tiempo, escriba la nueva ecuación diferencial del modelo.
   (*b*) Encuentre el precio de equilibrio intertemporal.
   (*c*) ¿Bajo qué circunstancias puede descartarse la fluctuación periódica?
2. Sean las funciones de oferta y demanda como en (16.26), pero con $u = w = 0$ como en la discusión del texto.
   (*a*) Si el mercado no siempre está en ceros, sino que se ajusta de acuerdo a
   $$\frac{dP}{dt} = j(Q_d - Q_s) \qquad (j > 0)$$
   escriba la nueva ecuación diferencial apropiada.
   (*b*) Encuentre el precio de equilibrio intertemporal $\overline{P}$ y el precio de equilibrio para el mercado en ceros P^*.
   (*c*) Establezca la condición para tener una trayectoria de precio fluctuante. ¿Puede presentarse la fluctuación si $n > 0$?
3. Sean la oferta y la demanda
   $$Q_d = 9 - P + P' + 3P'' \qquad Q_s = -1 + 4P - P' + 5P''$$
   con $P(0) = 4$ y $P'(0) = 4$.
   (*a*) Encuentre la trayectoria de precio, suponiendo que el mercado está en ceros para todo instante de tiempo.
   (*b*) ¿Es convergente la trayectoria de tiempo? ¿Con fluctuación?

16.5 La interacción de la inflación y el desempleo

En esta sección ilustramos el uso de una ecuación diferencial de segundo orden con un modelo macro que trata el problema de la inflación y el desempleo.

La relación de Phillips

Uno de los conceptos más ampliamente usados en el análisis moderno del problema de la inflación y del desempleo es la relación de Phillips.[6] En su formulación original, ilustra una relación negativa con una base empírica entre la tasa de crecimiento del salario en dinero y la tasa de desempleo:

$$w = f(U) \qquad [f'(U) < 0] \qquad \textbf{(16.30)}$$

[6] A. W. Phillips, "The Relationship Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861-1957", *Economica*, noviembre de 1958, pp. 283-299.

donde la letra minúscula w denota la tasa de crecimiento del salario monetario W (es decir, $w \equiv \dot{W}/W$) y U es la tasa de desempleo. Por esto sólo pertenece al mercado del trabajo. Sin embargo, su uso posterior de la relación de Phillips se ha adaptado a una función que enlaza la *tasa de inflación* (en lugar de w) con la tasa de desempleo. Esta adaptación puede justificarse al argumentar que los precios al alza son de uso generalizado, de modo que una w positiva, que refleja el costo creciente del salario monetario, necesariamente conlleva implicaciones inflacionarias. Y esto es lo que hace que la tasa de inflación, al igual que w, sea una función de U. Sin embargo, la presión inflacionaria de una w positiva puede ser compensada por un incremento en la productividad laboral, que se supone que es exógena, y aquí se denota por T. Específicamente, el efecto inflacionario puede materializarse sólo hasta el momento en el cual el salario monetario crezca con mayor rapidez que la productividad. Si se denota la tasa de inflación —es decir, la tasa de crecimiento del nivel de precios P— por la letra minúscula p, ($p \equiv \dot{P}/P$), podemos escribir

$$p = w - T \qquad \textbf{(16.31)}$$

Al combinar (16.30) y (16.31), y al adoptar la versión lineal de la función $f(U)$, obtenemos entonces una relación adaptada de Phillips

$$p = \alpha - T - \beta U \qquad (\alpha, \beta > 0) \qquad \textbf{(16.32)}$$

La relación de Phillips aumentada con expectativas

Más recientemente, los economistas han preferido usar la versión *aumentada con expectativas* de la relación de Phillips

$$w = f(U) + g\pi \qquad (0 < g \leq 1) \qquad \textbf{(16.30')}$$

donde π denota la tasa esperada de inflación. La idea subyacente de (16.30′), tal como fue propuesta por el profesor Friedman (premio Nobel)[7] es que si una tendencia inflacionaria ha estado en vigor por suficiente tiempo, las personas tienden a formar ciertas expectativas de inflación que luego intentan incorporar a sus demandas de salario monetario. Entonces w debe ser una función creciente de π. Transpuesta a (16.32), esta idea conduce a la ecuación

$$p = \alpha - T - \beta U + g\pi \qquad (0 < g \leq 1) \qquad \textbf{(16.33)}$$

Con la introducción de una nueva variable para denotar la tasa esperada de inflación, se hace necesario formular una hipótesis acerca de cómo se forman específicamente las expectativas de inflación.[8] Aquí adoptamos la hipótesis de las *expectativas adaptativas*

$$\frac{d\pi}{dt} = j(p - \pi) \qquad (0 < j \leq 1) \qquad \textbf{(16.34)}$$

Observe que, más que explicar la magnitud absoluta de π esta ecuación describe en vez de eso su patrón de cambio con el tiempo. Si la tasa real de inflación p sobrepasa a la tasa esperada π esta última, que ahora se ha comprobado que es demasiado baja, se revisa hacia arriba ($d\pi/dt > 0$). Inversamente, si p decae respecto a π, entonces π se revisa en dirección hacia

[7] Milton Friedman, "The Role of Monetary Policy", *American Economic Review*, marzo de 1968, pp. 1-17.

[8] Esto contrasta con la sección 16.4, donde las expectativas de precio se estudiaron sin introducir una nueva variable que represente al precio esperado. Como resultado, las hipótesis relativas a la formación de las expectativas estuvieron sólo implícitamente incorporadas en los parámetros m, n, u y w, en (16.26).

abajo. En el formato, (16.34) se parece mucho al mecanismo de ajuste $dP/dt = j(Q_d - Q_s)$ del modelo de mercado. Pero aquí la fuerza de impulso que hay detrás del ajuste es la discrepancia entre las tasas de inflación *real* y *esperada*, en vez de Q_d y Q_s.

La retroalimentación de la inflación hacia el desempleo

Se puede considerar que (16.33) y (16.34) constituyen un modelo completo. Puesto que hay tres variables en un sistema de dos ecuaciones, sin embargo una de las variables tiene que considerarse como exógena. Por ejemplo, si se consideran a π y p como endógenas, entonces U debe tratarse como exógena. Una alternativa mas satisfactoria es introducir una tercera ecuación para explicar la variable U, de modo que el modelo va a enriquecerse en sus características de comportamiento. Tiene más importancia el que esto nos ofrece una oportunidad para considerar el efecto de retroalimentación de la inflación sobre el desempleo. La ecuación (16.33) nos dice la forma en que U afecta a p: principalmente desde el punto de vista del abastecimiento en la economía. Pero con toda seguridad que p puede afectar a su vez a U. Por ejemplo, la tasa de inflación suele influir en las decisiones del público en cuanto al ahorro de consumo, y por lo tanto también en la demanda agregada de la producción nacional, y esta última a su vez va a afectar a la tasa de desempleo. Inclusive la tasa de inflación puede afectar la efectividad de las políticas de gobierno respecto al manejo de la demanda. Dependiendo de la tasa de inflación, un nivel dado de gasto monetario (política fiscal) podría trascender a niveles variables de gasto real, y en forma similar, una tasa dada de expansión nominal monetaria (política monetaria) podría conducir a tasas variables de expansión monetaria real. Éstas, a su vez, implicarían efectos diferenciados sobre el producto y el desempleo.

Por simplicidad, sólo tomaremos en consideración la retroalimentación a través de la conducción de la política monetaria. Si denotamos el balance nominal del dinero como M y su tasa de crecimiento como $m \equiv \dot{M}/M$, postulemos que[9]

$$\frac{dU}{dt} = -k(m - p) \qquad (k > 0) \tag{16.35}$$

Recordando (10.25), y aplicándola hacia atrás, vemos que la expresión $(m - p)$ representa la tasa de crecimiento del dinero *real*:

$$m - p = \frac{\dot{M}}{M} - \frac{\dot{P}}{P} = r_M - r_P = r_{(M/P)}$$

Entonces (16.35) estipula que dU/dt está relacionado en sentido negativo con la tasa de crecimiento del balance monetario real. Como ahora la variable p entra en la determinación de dU/dt, el modelo contiene una retroalimentación de la inflación hacia el desempleo.

La trayectoria de tiempo de π

Juntas, (16.33) a (16.35) constituyen un modelo cerrado en las tres variables π, p y U. Sin embargo, al eliminar dos de las tres variables, podemos condensar el modelo en una sola ecuación diferencial con una sola variable. Suponga que hacemos que esta variable individual sea π. Entonces podemos sustituir primero (16.33) en (16.34) para obtener

$$\frac{d\pi}{dt} = j(\alpha - T - \beta U) - j(1 - g)\pi \tag{16.36}$$

[9] En una discusión anterior, denotamos a la oferta monetario como M_s, para diferenciarlo de la demanda monetaria M_d. Aquí podemos usar simplemente la letra M sin subíndice, ya que no hay posibilidad de confusión.

Si esta ecuación contuviera la expresión dU/dt en vez de a U, podríamos haber sustituido directamente (16.35) en (16.36). Pero tal como está (16.36), debemos crear primero en forma deliberada un término dU/dt al diferenciar (16.36) respecto a t, con el resultado

$$\frac{d^2\pi}{dt^2} = -j\beta\frac{dU}{dt} - j(1-g)\frac{d\pi}{dt} \tag{16.37}$$

Entonces, la sustitución de (16.35) en esto arroja

$$\frac{d^2\pi}{dt^2} = j\beta km - j\beta kp - j(1-g)\frac{d\pi}{dt} \tag{16.37'}$$

Todavía hay una variable p que debe eliminarse. Para lograrlo, observamos que (16.34) implica

$$p = \frac{1}{j}\frac{d\pi}{dt} + \pi \tag{16.38}$$

Con el uso de este resultado en (16.37′) y simplificando, obtenemos finalmente la ecuación diferencial deseada sólo en la variable π:

$$\frac{d^2\pi}{dt^2} + \underbrace{[\beta k + j(1-g)]}_{a_1}\frac{d\pi}{dt} + \underbrace{(j\beta k)}_{a_2}\pi = \underbrace{j\beta km}_{b} \tag{16.37''}$$

La integral particular de esta ecuación es simplemente

$$\pi_p = \frac{b}{a_2} = m$$

Entonces, en este modelo el valor de equilibrio intertemporal de la tasa esperada de inflación depende exclusivamente de la tasa de crecimiento del dinero nominal.

Para la función complementaria, las dos raíces son, como antes,

$$r_1, r_2 = \frac{1}{2}\left(-a_1 \pm \sqrt{a_1^2 - 4a_2}\right) \tag{16.39}$$

donde, como puede observarse de (16.37″), tanto a_1 como a_2 son positivas. Basándonos en un criterio *a priori*, no es posible determinar si a_1^2 sobrepasaría, sería igual o sería menor que $4a_2$. Entonces, es concebible que surjan los tres casos de raíces características: raíces reales diferentes, raíces reales repetidas, o raíces complejas. Sin embargo, cualquiera que sea el caso que se presente, el equilibrio intertemporal va a probar ser dinámicamente estable en el presente modelo. Esto puede explicarse como sigue: Suponga primero que prevalece el caso 1, con $a_1^2 > 4a_2$. Entonces la raíz cuadrada de (16.39) arroja un número real. Como a_2 es positiva, $\sqrt{a_1^2 - 4a_2}$ es necesariamente menor que $\sqrt{a_1^2} = a_1$. Se sigue que r_1 es negativo, al igual que r_2, lo que implica un equilibrio dinámicamente estable. ¿Qué pasa si $a_1^2 = 4a_2$ (caso 2)? En ese caso, la raíz cuadrada es cero, de modo que $r_1 = r_2 = -a_1/2 < 0$. Y la negatividad de las raíces repetidas implica nuevamente estabilidad dinámica. Finalmente, para el caso 3, la parte real de las raíces complejas es $h = -a_1/2$. Como ésta tiene el mismo valor que las raíces repetidas del caso 2, se puede aplicar una conclusión idéntica en relación con la estabilidad dinámica.

Aunque sólo hemos estudiado la trayectoria de tiempo de π, el modelo puede también suministrar información sobre las otras variables. Para encontrar la trayectoria de tiempo de, digamos, la variable U, podemos *ya sea* iniciar mediante la condensación del modelo en una ecuación diferencial en U en lugar de en π (vea el ejercicio 16.5-2) o deducir la trayectoria de U a partir de la trayectoria de π ya encontrada (vea el ejemplo 1).

Ejemplo 1 Sean las tres ecuaciones del modelo que adoptan la forma específica

$$p = \frac{1}{6} - 3U + \pi \tag{16.40}$$

$$\frac{d\pi}{dt} = \frac{3}{4}(p - \pi) \tag{16.41}$$

$$\frac{dU}{dt} = -\frac{1}{2}(m - p) \tag{16.42}$$

Entonces tenemos los valores de los parámetros $\beta = 3$, $h = 1$, $j = \frac{3}{4}$ y $k = \frac{1}{2}$; así, con referencia a (16.37″), encontramos

$$a_1 = \beta k + j(1 - g) = \frac{3}{2} \qquad a_2 = j\beta k = \frac{9}{8} \qquad \text{y} \qquad b = j\beta km = \frac{9}{8}m$$

La integral particular es $b/a_2 = m$. Con $a_1^2 < 4a_2$, las raíces características son complejas:

$$r_1, r_2 = \frac{1}{2}\left(-\frac{3}{2} \pm \sqrt{\frac{9}{4} - \frac{9}{2}}\right) = \frac{1}{2}\left(-\frac{3}{2} \pm \frac{3}{2}i\right) = -\frac{3}{4} \pm \frac{3}{4}i$$

Es decir, $h = -\frac{3}{4}$ y $v = \frac{3}{4}$. En consecuencia, la solución general para la tasa esperada de inflación es

$$\pi(t) = e^{-3t/4}\left(A_5 \cos\frac{3}{4}t + A_6 \operatorname{sen}\frac{3}{4}t\right) + m \tag{16.43}$$

que describe una trayectoria de tiempo con fluctuación amortiguada alrededor del valor de equilibrio m.

A partir de esto, también podemos deducir las trayectorias de tiempo para las variables p y U. De acuerdo con (16.41), p puede expresarse en términos de π y $d\pi/dt$ mediante la ecuación

$$p = \frac{4}{3}\frac{d\pi}{dt} + \pi$$

La trayectoria π en la solución general (16.43) implica la derivada

$$\begin{aligned}\frac{d\pi}{dt} &= -\frac{3}{4}e^{-3t/4}\left(A_5 \cos\frac{3}{4}t + A_6 \operatorname{sen}\frac{3}{4}t\right) \\ &\quad + e^{-3t/4}\left(-\frac{3}{4}A_5 \operatorname{sen}\frac{3}{4}t + \frac{3}{4}A_6 \cos\frac{3}{4}t\right) \quad \text{[regla del producto y regla de la cadena]}\end{aligned}$$

Con el uso de la solución (16.43) y su derivada, entonces tenemos

$$p(t) = e^{-3t/4}\left(A_6 \cos\frac{3}{4}t - A_5 \operatorname{sen}\frac{3}{4}t\right) + m \tag{16.44}$$

Al igual que la tasa *esperada* de inflación π, la tasa *real* de inflación p tiene una trayectoria fluctuante de tiempo que converge hacia el valor de equilibrio m.

En cuanto a la variable U, (16.40) nos dice que puede expresarse en términos de π y p como sigue:

$$U = \frac{1}{3}(\pi - p) + \frac{1}{18}$$

Por lo tanto, en virtud de las soluciones (16.43) y (16.44), podemos escribir la trayectoria de tiempo de la tasa de desempleo como

$$U(t) = \frac{1}{3}e^{-3t/4}\left[(A_5 - A_6)\cos\frac{3}{4}t + (A_5 + A_6)\operatorname{sen}\frac{3}{4}t\right] + \frac{1}{18} \tag{16.45}$$

Nuevamente, esta trayectoria tiene fluctuación amortiguada, con $\frac{1}{18}$ como $\overline{U}$, el valor de U de equilibrio intertemporal dinámicamente estable.

Debido a que los valores de equilibrio intertemporal de π y p son ambos iguales al parámetro de la política monetaria m, el valor de m —la tasa de crecimiento del dinero nominal— suministra el eje alrededor del cual fluctúan las trayectorias de tiempo de π y p. Si ocurre un cambio en m, un nuevo valor de equilibrio de π y p va a reemplazar inmediatamente al anterior, y cualquier valor que las variables π y p lleguen a adoptar en el momento de un cambio en la política monetaria se transformará en los valores iniciales de los cuales emanen las nuevas trayectorias de π y p.

En contraste, el valor de equilibrio intertemporal $\overline{U}$ no depende de m. De acuerdo con (16.45), U converge hacia la constante $\frac{1}{18}$ independientemente de la tasa de crecimiento del dinero nominal, y por lo tanto independientemente de la tasa de inflación de equilibrio. A este valor de equilibrio constante de U se le denomina la *tasa natural de desempleo*. El hecho de que la tasa natural de desempleo es consistente con cualquier tasa de inflación de equilibrio puede representarse en el espacio Up mediante una línea recta vertical paralela al eje p. Esta línea vertical que relaciona entre sí los valores de equilibrio de U y p, se conoce como la *curva de Phillips a largo plazo*. Sin embargo, la forma vertical de esta curva depende del valor especial de un parámetro adoptado en este ejemplo. Cuando se altera ese valor, como en el ejercicio 16.5-4, la curva de Phillips a largo plazo puede ya no ser vertical.

EJERCICIO 16.5

1. En el modelo de inflación-desempleo retenga (16.33) pero borre (16.35) y permita que U sea exógeno.
   (*a*) ¿Qué tipo de ecuación diferencial va a surgir?
   (*b*) ¿Cuántas raíces características puede obtener? ¿Es posible tener una fluctuación periódica en la función complementaria?
2. En la discusión del texto, condensamos el modelo de inflación-desempleo en una ecuación diferencial en la variable π. Muestre que el modelo puede condensarse en forma alterna en una ecuación diferencial de segundo orden en la variable U, con los mismos coeficientes a_1 y a_2 que en (16.37″), pero un término constante diferente $b = kj[\alpha - T - (1 - g)m]$.
3. Reemplacemos la hipótesis de expectativas adaptativas (16.34) por la así llamada hipótesis perfecta de previsión $\pi = p$, pero conserve (16.33) y (16.35).
   (*a*) Obtenga una ecuación diferencial en la variable p.
   (*b*) Obtenga una ecuación diferencial en la variable U.
   (*c*) ¿Cómo difieren estas ecuaciones fundamentalmente de la que obtuvimos bajo la hipótesis de expectativas adaptativas?
   (*d*) ¿Qué cambio en la restricción de parámetros es necesario para que las nuevas ecuaciones diferenciales tengan significado?
4. En el ejemplo 1, conserve (16.41) y (16.42) pero reemplace (16.40) por
   $$p = \frac{1}{6} - 3U + \frac{1}{3}\pi$$
   (*a*) Encuentre $p(t)$, $\pi(t)$ y $U(t)$.
   (*b*) ¿Todavía fluctúan las trayectorias de tiempo? ¿Todavía son convergentes?
   (*c*) ¿Qué son $\overline{p}$ y $\overline{U}$, los valores de equilibrio intertemporal de p y U?
   (*d*) ¿Es todavía verdad que $\overline{U}$ no está funcionalmente relacionada con $\overline{p}$? Si enlazamos estos dos valores de equilibrio entre sí en una curva de Phillips a largo plazo, ¿podemos obtener todavía una curva vertical? ¿Qué hipótesis del ejemplo 1 es entonces crucial para obtener una curva de Phillips vertical a largo plazo?

16.6 Ecuaciones diferenciales con un término variable

En las ecuaciones diferenciales consideradas en la sección 16.1,

$$y''(t) + a_1 y'(t) + a_2 y = b$$

el término b del miembro derecho es una constante. ¿Qué pasa si en lugar de b tenemos a la derecha un *término variable*, es decir, alguna función de t tal como bt^2, e^{bt} o b sen t? La respuesta es que entonces debemos modificar nuestra integral particular y_p. Afortunadamente, la función complementaria no se afecta por la presencia de un término variable, ya que y_c tiene que ver sólo con la ecuación reducida, cuyo miembro derecho siempre es cero.

Método de los coeficientes indeterminados

Explicaremos un método para encontrar y_p, conocido como el *método de los coeficientes indeterminados*, que es aplicable a las ecuaciones diferenciales de coeficiente constante y término variable, siempre que el término variable junto con sus derivadas sucesivas contengan sólo un número *finito* de diferentes tipos de expresión (aparte de las constantes multiplicativas). La explicación de este método puede desarrollarse de la mejor manera con una ilustración concreta.

Ejemplo 1 Encuentre la integral particular de

$$y''(t) + 5y'(t) + 3y = 6t^2 - t - 1 \tag{16.46}$$

Por definición, la integral particular es un valor de y que satisface la ecuación dada, es decir, un valor de y que haga al miembro izquierdo idénticamente igual al miembro derecho independientemente del valor de t. Dado que el miembro izquierdo contiene la función $y(t)$ y las derivadas $y'(t)$ y $y''(t)$ —mientras que el miembro derecho contiene múltiplos de las expresiones t^2, t y una constante— preguntamos: ¿Qué forma de función general de $y(t)$, junto con la primera y la segunda derivadas, nos darán los tres tipos de expresión t^2, t y una constante? La respuesta obvia es una función de la forma $B_1t^2 + B_2t + B_3$ (donde B_i son coeficientes que deben determinarse), ya que si escribimos la integral particular como

$$y(t) = B_1t^2 + B_2t + B_3$$

podemos derivar

$$y'(t) = 2B_1t + B_2 \qquad \text{y} \qquad y''(t) = 2B_1 \tag{16.47}$$

Y estas tres ecuaciones están compuestas en verdad con los tipos mencionados de expresión. Al sustituirlos en (16.46) y agrupando términos, obtenemos

$$\text{Miembro izquierdo} = (3B_1)t^2 + (10B_1 + 3B_2)t + (2B_1 + 5B_2 + 3B_3)$$

Y cuando esto se iguala término a término con el miembro derecho, podemos determinar los coeficientes B_i como sigue:

$$\left.\begin{aligned} 3B_1 &= 6 \\ 10B_1 + 3B_2 &= -1 \\ 2B_1 + 5B_2 + 3B_3 &= -1 \end{aligned}\right\} \quad \Rightarrow \quad \left\{\begin{aligned} B_1 &= 2 \\ B_2 &= -7 \\ B_3 &= 10 \end{aligned}\right.$$

Entonces, la integral particular deseada puede escribirse como

$$y_p = 2t^2 - 7t + 10$$

Este método puede funcionar sólo cuando el número de tipos de expresión es finito (vea el ejercicio 16.6-1). En general, cuando se cumpla este prerrequisito, la integral particular puede tomarse como que está en la forma de una combinación lineal de todos los diferentes tipos de expresión contenidos en el término variable dado, así como en todas sus derivadas. Observe en particular que una expresión constante debe incluirse en la integral particular, si el término variable original o cualquiera de las derivadas sucesivas contienen un término constante.

Ejemplo 2 Como una ilustración adicional, encontremos la forma general para la integral particular adecuada para el término variable (b sen t). La diferenciación repetida arroja en este caso las derivadas sucesivas (b cos t), ($-b$ sen t), ($-b$ cos t), (b sen t), etc., que incluyen sólo dos tipos diferentes de expresión. Por lo tanto podemos ensayar una integral particular de la forma $(B_1 \operatorname{sen} t + B_2 \cos t)$.

Una modificación

En ciertos casos surge una complicación al aplicar el método. Cuando el coeficiente del término y en la ecuación diferencial dada es cero, tal como en

$$y''(t) + 5y'(t) = 6t^2 - t - 1$$

la forma de prueba previamente utilizada para la y_p, es decir, $B_1t^2 + B_2t + B_3$, no va a funcionar. La causa de esta falla es que, ya que el término $y(t)$ está fuera de la escena y que sólo las derivadas $y'(t)$ y $y''(t)$ como se muestra en (16.47) van a sustituirse en el miembro izquierdo, ningún término B_1t^2 va a aparecer jamás a la izquierda que sea igual al término $6t^2$ a la derecha. La forma de salir de esta dificultad es usar la solución de prueba $t(B_1t^2 + B_2t + B_3)$; o si esto también falla (por ejemplo, dada la ecuación $y''(t) = 6t^2 - t - 1$), usar $t^2(B_1t^2 + B_2t + B_3)$, etcétera.

En verdad, podemos emplear el mismo truco aun en otra circunstancia difícil, como se ilustra en el ejemplo 3.

Ejemplo 3 Encuentre la integral particular de

$$y''(t) + 3y'(t) - 4y = 2e^{-4t} \tag{16.48}$$

Aquí, el término variable está en la forma de e^{-4t}, pero todas las derivadas sucesivas (a saber, $-8e^{-4t}$, $32e^{-4t}$, $-128e^{-4t}$–$128e^{-4t}$, etc.) adoptan la misma forma. Si probamos la solución

$$y(t) = Be^{-4t} \qquad [\text{con } y'(t) = -4Be^{-4t} \text{ y } y''(t) = 16Be^{-4t}]$$

y sustituimos esto en (16.48), obtenemos el resultado desfavorable

$$\text{Miembro izquierdo} = (16 - 12 - 4)Be^{-4t} = 0 \tag{16.49}$$

que obviamente no puede igualarse con el término $2e^{-4t}$ en el miembro derecho.

La causa de esto es el hecho de que el coeficiente exponencial en el término variable (-4) resulta ser igual a una de las raíces de la ecuación característica de (16.48):

$$r^2 + 3r - 4 = 0 \qquad (\text{raíces } r_1, r_2 = 1, -4)$$

Se recordará que la ecuación característica se obtiene a través de un proceso de diferenciación;[10] pero la expresión $(16 - 12 - 4)$ de (16.49) se deriva mediante el mismo proceso. No es sorprendente, por lo tanto, que $(16 - 12 - 4)$ sea meramente una versión específica de $r^2 + 3r - 4$ con r igual a -4. Ya que -4 resulta ser una ecuación característica, la expresión cuadrática

$$r^2 + 3r - 4 = 16 - 12 - 4$$

debe necesariamente ser idénticamente cero.

[10] Vea la discusión en el texto que conduce a (16.4'').

Para enfrentar esta situación, probemos a su vez la solución

$$y(t) = Bte^{-4t}$$

con derivadas

$$y'(t) = (1 - 4t)Be^{-4t} \qquad y \qquad y''(t) = (-8 + 16t)Be^{-4t}$$

La sustitución de esto en (16.48) nos da ahora: miembro izquierdo $= -5\ Be^{-4t}$. Cuando esto se iguala con el miembro derecho, determinamos que el coeficiente es $B = -2/5$. En consecuencia, la integral particular deseada de (16.48) puede escribirse como

$$y_p = \frac{-2}{5}te^{-4t}$$

EJERCICIO 16.6

1. Muestre que el método de los coeficientes indeterminados es inaplicable a la ecuación diferencial $y''(t) + ay'(t) + by = t^{-1}$.
2. Encuentre la integral particular de cada una de las siguientes ecuaciones por el método de los coeficientes indeterminados:
   (*a*) $y''(t) + 2y'(t) + y = t$ (*c*) $y''(t) + y'(t) + 2y = e^t$
   (*b*) $y''(t) + 4y'(t) + y = 2t^2$ (*d*) $y''(t) + y'(t) + 3y = \operatorname{sen} t$

16.7 Ecuaciones diferenciales lineales de orden superior

Los métodos de solución introducidos en las secciones anteriores se extienden rápidamente a una ecuación diferencial lineal de orden n. Con coeficientes constantes y un término constante, esta ecuación puede escribirse en forma general como

$$y^{(n)}(t) + a_1y^{(n-1)}(t) + \cdots + a_{n-1}y'(t) + a_ny = b \qquad \textbf{(16.50)}$$

Cómo encontrar la solución

En este caso de coeficientes constantes y término constante, la presencia de las derivadas superiores no afecta al método de búsqueda estudiado anteriormente para encontrar la integral particular.

Si probamos el tipo más sencillo posible de solución, $y = k$, podemos ver que todas las derivadas desde $y'(t)$ hasta $y^{(n)}(t)$ serán cero; por lo tanto (16.50) se reduce a $a_nk = b$ y podemos escribir

$$y_p = k = \frac{b}{a_n} \qquad (a_n \neq 0) \qquad [\text{vea } (16.3)]$$

Sin embargo, para el caso $a_n = 0$, debemos probar una solución de la forma $y = kt$. Entonces, ya que $y'(t) = k$, se anulan todas las derivadas superiores, (16.50) se reduce a $a_{n-1}k = b$, de donde se obtiene la integral particular

$$y_p = kt = \frac{b}{a_{n-1}}t \qquad (a_n = 0;\ a_{n-1} \neq 0) \qquad [\text{vea } (16.3')]$$

Si resulta que $a_n = a_{n-1} = 0$, entonces también va a fallar esta última solución; en lugar de ello, debe probarse una solución de la forma $y = kt^2$. Las adaptaciones adicionales de este procedimiento deben ser obvias.

En cuanto a la función complementaria, la inclusión de las derivadas de orden superior en la ecuación diferencial tiene el efecto de elevar el grado de la ecuación característica. La función complementaria se define como la solución general de la ecuación reducida

$$y^{(n)}(t) + a_1 y^{(n-1)}(t) + \cdots + a_{n-1} y'(t) + a_n y = 0 \qquad \textbf{(16.51)}$$

Probando $y = Ae^{rt}$ $(\neq 0)$ como una solución y utilizando el conocimiento de que esto implica $y'(t) = rAe^{rt}$, $y''(t) = r^2 Ae^{rt}, \ldots, y^{(n)}(t) = r^n Ae^{rt}$, podemos reescribir (16.51) como

$$Ae^{rt}(r^n + a_1 r^{n-1} + \cdots + a_{n-1} r + a_n) = 0$$

Esta ecuación se resuelve con cualquier valor de r que satisfaga la siguiente ecuación característica (polinomio de grado n)

$$r^n + a_1 r^{n-1} + \cdots + a_{n-1} r + a_n = 0 \qquad \textbf{(16.51')}$$

Por supuesto que habrá n raíces para este polinomio, y cada una de ellas debe incluirse en la solución general de (16.51). Entonces, nuestra función complementaria debe ser en general de la forma

$$y_c = A_1 e^{r_1 t} + A_2 e^{r_2 t} + \cdots + A_n e^{r_n t} \qquad \left(= \sum_{i=1}^{n} A_i e^{r_i t} \right)$$

Sin embargo, como antes, debemos hacer algunas modificaciones en el caso de que las n raíces no sean todas reales y distintas. Primero, supongamos que hay raíces repetidas, digamos $r_1 = r_2 = r_3$. Entonces, para evitar el "colapso", debemos escribir los tres primeros términos de las soluciones como $A_1 e^{r_1 t} + A_2 t e^{r_1 t} + A_3 t^2 e^{r_1 t}$ [vea (16.9)]. En caso de tener también $r_4 = r_1$, el cuarto término debe alterarse como $A_4 t^3 e^{r_1 t}$, etcétera.

Segundo, suponga que dos de las raíces son complejas, es decir,

$$r_5, r_6 = h \pm vi$$

entonces, los términos quinto y sexto de la solución anterior deben combinarse en la siguiente expresión:

$$e^{ht}(A_5 \cos vt + A_6 \operatorname{sen} vt) \qquad [\text{vea } (16.24')]$$

Por la misma razón, si se encuentran dos pares *distintos* de raíces complejas, debe haber dos expresiones trigonométricas de este tipo (con un conjunto diferente de valores de h, v y dos constantes arbitrarias para cada uno).[11] Como una posibilidad adicional, si resulta que hay dos pares de raíces complejas *repetidas*, entonces debemos usar e^{ht} como el término multiplicativo para uno, pero usar te^{ht} para el otro. También, aun cuando h y v tengan valores idénticos en las raíces complejas repetidas, ahora debemos asignar a cada una un par diferente de constantes arbitrarias.

Una vez que encontramos y_p y y_c, la solución general de la ecuación completa (16.50) es fácil. Como antes, es simplemente la suma de la función complementaria más la integral particular: $y(t) = y_c + y_p$. En esta solución general podemos contar un total de n constantes arbitrarias. Entonces, para determinar la solución, requerimos n condiciones iniciales.

[11] Es interesante observar que, como las raíces complejas siempre se presentan en pares conjugados, podemos estar seguros de tener *cuando menos* una raíz real si la ecuación diferencial es de orden *impar*, es decir, cuando n es un número impar.

Ejemplo 1 Encuentre la solución general de

$$y^{(4)}(t) + 6y'''(t) + 14y''(t) + 16y'(t) + 8y = 24$$

La integral particular de esta ecuación de cuarto orden es simplemente

$$y_p = \frac{24}{8} = 3$$

Su ecuación característica es, por (16.51′),

$$r^4 + 6r^3 + 14r^2 + 16r + 8 = 0$$

que puede factorizarse en la forma

$$(r+2)(r+2)(r^2 + 2r + 2) = 0$$

De las primeras dos expresiones que están entre paréntesis, podemos obtener las raíces dobles $r_1 = r_2 = -2$, pero la última expresión (cuadrática) arroja el par de raíces complejas $r_3, r_4 = -1 \pm i$, con $h = -1$ y $v = 1$. En consecuencia, la función complementaria es

$$y_c = A_1 e^{-2t} + A_2 t e^{-2t} + e^{-t}(A_3 \cos t + A_4 \operatorname{sen} t)$$

y la solución general es

$$y(t) = A_1 e^{-2t} + A_2 t e^{-2t} + e^{-t}(A_3 \cos t + A_4 \operatorname{sen} t) + 3$$

Las cuatro constantes A_1, A_2, A_3 y A_4 podemos determinarlas, por supuesto, si se dan cuatro condiciones iniciales.

Observe que todas las raíces características en este ejemplo son ya sea reales o negativas o son complejas y con una parte real negativa. Por lo tanto, la trayectoria de tiempo debe ser convergente y el equilibrio intertemporal es dinámicamente estable.

La convergencia y el teorema de Routh

La solución de una ecuación característica de orden superior no siempre es una tarea sencilla. Por esta razón, es una gran ayuda encontrar una manera de evaluar la convergencia o la divergencia de una trayectoria de tiempo sin tener que encontrar las raíces características. Afortunadamente, existe un método de este tipo, el cual puede suministrar un análisis cualitativo (aunque no es gráfico) de una ecuación diferencial.

Este método se encuentra en el *teorema de Routh*,[12] el cual establece que

> Las partes reales de todas las raíces de la ecuación polinomial de grado n
>
> $$a_0 r^n + a_1 r^{n-1} + \cdots + a_{n-1} r + a_n = 0$$
>
> son negativas si y sólo si los n primeros de la siguiente secuencia de determinantes
>
> $$|a_1|; \quad \begin{vmatrix} a_1 & a_3 \\ a_0 & a_2 \end{vmatrix}; \quad \begin{vmatrix} a_1 & a_3 & a_5 \\ a_0 & a_2 & a_4 \\ 0 & a_1 & a_3 \end{vmatrix}; \quad \begin{vmatrix} a_1 & a_3 & a_5 & a_7 \\ a_0 & a_2 & a_4 & a_6 \\ 0 & a_1 & a_3 & a_5 \\ 0 & a_0 & a_2 & a_4 \end{vmatrix}; \quad \ldots$$
>
> son todos positivos.

Al aplicar este teorema, debemos recordar que $|a_1| \equiv a_1$. Aún más, debemos entender que tenemos que tomar $a_m = 0$ para todo $m > n$. Por ejemplo, dada una ecuación polinomial de

[12] Para un estudio de este teorema y un esbozo de su demostración, vea Paul A. Samuelson, *Foundations of Economic Analysis*, Harvard University Press, 1947, pp. 429-435, y las referencias citadas ahí.

tercer grado ($n = 3$), necesitamos examinar los signos de los primeros *tres* determinantes listados en el teorema de Routh; para este propósito, debemos hacer $a_4 = a_5 = 0$.

La relevancia de este teorema para el problema de la convergencia debe hacerse autoevidente cuando recordamos que, con objeto de que la trayectoria de tiempo $y(t)$ converja independientemente de cuáles sean las condiciones iniciales, todas las raíces características de la ecuación diferencial deben tener partes reales negativas. Como la ecuación característica (16.51′) es una ecuación polinomial de grado n, con $a_0 = 1$, el teorema de Routh puede ser una ayuda directa en la prueba de la convergencia. De hecho, observamos que los coeficientes de la ecuación característica (16.51′) son totalmente idénticos con los de la ecuación diferencial dada (16.51), entonces es perfectamente aceptable sustituir los coeficientes de (16.51) directamente en la secuencia de determinantes mostrada en el teorema de Routh para la prueba, siempre que tomemos $a_0 = 1$. Siempre que a la condición citada en el teorema se le adjudique la naturaleza de "si y sólo si", constituye obviamente una condición necesaria y suficiente.

Ejemplo 2

Pruebe por el teorema de Routh si la ecuación diferencial del ejemplo 1 tiene una trayectoria de tiempo convergente. Esta ecuación es de cuarto orden, de modo que $n = 4$. Los coeficientes son $a_0 = 1$, $a_1 = 6$, $a_2 = 14$, $a_3 = 16$, $a_4 = 8$, y $a_5 = a_6 = a_7 = 0$. Al sustituirlos en los primeros cuatro determinantes, encontramos que sus valores son 6, 68, 800 y 6 400 respectivamente. Puesto que todos son positivos, podemos concluir que la trayectoria de tiempo es convergente.

EJERCICIO 16.7

1. Encuentre la integral particular de cada una de las siguientes ecuaciones:
   (*a*) $y'''(t) + 2y''(t) + y'(t) + 2y = 8$
   (*b*) $y'''(t) + y''(t) + 3y'(t) = 1$
   (*c*) $3y'''(t) + 9y''(t) = 1$
   (*d*) $y^{(4)}(t) + y''(t) = 4$
2. Encuentre y_p y y_c (y la solución general) de:
   (*a*) $y'''(t) - 2y''(t) - y'(t) + 2y = 4$
   [*Sugerencia:* $r^3 - 2r^2 - r + 2 = (r-1)(r+1)(r-2)$]
   (*b*) $y'''(t) + 7y''(t) + 15y'(t) + 9y = 0$
   [*Sugerencia:* $r^3 + 7r^2 + 15r + 9 = (r+1)(r^2 + 6r + 9)$]
   (*c*) $y'''(t) + 6y''(t) + 10y'(t) + 8y = 8$
   [*Sugerencia:* $r^3 + 6r^2 + 10r + 8 = (r+4)(r^2 + 2r + 2)$]
3. Basándose en los signos de las raíces características obtenidas en el problema 2, analice la estabilidad dinámica de equilibrio. Luego verifique su respuesta por el teorema de Routh.
4. Sin encontrar sus raíces características, determine si las siguientes ecuaciones diferenciales van a dar lugar a trayectorias de tiempo convergentes:
   (*a*) $y'''(t) - 10y''(t) + 27y'(t) - 18y = 3$
   (*b*) $y'''(t) + 11y''(t) + 34y'(t) + 24y = 5$
   (*c*) $y'''(t) + 4y''(t) + 5y'(t) - 2y = -2$
5. Deduzca del teorema de Routh que, para la ecuación diferencial lineal de segundo orden $y''(t) + a_1 y'(t) + a_2 y = b$, la trayectoria solución será convergente independientemente de las condiciones iniciales si y sólo si los coeficientes a_1 y a_2 son ambos positivos.

Capítulo 17

Tiempo discreto: ecuaciones en diferencias de primer orden

En el contexto de tiempo continuo, el patrón de cambio de una variable y está englobado en las derivadas $y'(t)$, $y''(t)$, etcétera. El cambio de tiempo contemplado en éstas ocurre en forma continua. A su vez, cuando el tiempo se considera una variable *discreta*, de modo que la variable t puede adoptar sólo valores enteros, es obvio que el concepto de la derivada ya no es apropiado. Entonces, como veremos, el patrón de cambio de la variable y debe describirse mediante las así llamadas diferencias, en lugar de derivadas o diferenciales de $y(t)$. De acuerdo con esto, las técnicas de las ecuaciones diferenciales van a dar paso a las técnicas de las *ecuaciones en diferencias*.

Cuando manejamos tiempo discreto, el valor de la variable y cambia sólo cuando la variable t cambia de un valor entero al siguiente, tal como de $t = 1$ a $t = 2$. Mientras tanto, no debe ocurrir nada con y. A la luz de esto, se hace más conveniente interpretar los valores de t refiriéndose a ellos como *periodos* —en vez de *puntos*— con $t = 1$ que denota al periodo 1 y $t = 2$ que denota al periodo 2, etcétera. Entonces podemos considerar simplemente que y tiene un valor único para cada periodo. En vista de esta interpretación, la versión de tiempo discreto de la dinámica económica generalmente se denomina *análisis de periodos*. Sin embargo, debemos enfatizar que "periodo" lo usamos aquí no en el sentido del calendario sino en el sentido analítico. Entonces, un periodo puede incluir un alcance de tiempo de calendario en un modelo económico específico y otro totalmente diferente en otro. Adicionalmente, aún para el mismo modelo, no es necesario interpretar que cada periodo sucesivo sea equivalente a un tiempo de calendario igual. En el sentido analítico, un periodo es simplemente un lapso que transcurre antes de que la variable y experimente un cambio.

17.1 Tiempo discreto, diferencias y ecuaciones en diferencias

El cambio de tiempo continuo a tiempo discreto no produce ningún efecto sobre la naturaleza fundamental del análisis dinámico, aunque debe cambiarse la formulación del problema. Básicamente, nuestro problema dinámico todavía es encontrar una trayectoria de tiempo a partir de un patrón dado del cambio de una variable y con el tiempo. Pero ahora el patrón de cambio debemos representarlo por el cociente en diferencias $\Delta y/\Delta t$, que es la contraparte en tiempo discreto de la derivada dy/dt. Sin embargo, recuerde que ahora t puede adoptar sólo valores enteros; entonces, cuando comparamos los valores de y en dos periodos consecutivos, debemos tener $\Delta t = 1$. Por esta razón, el cociente en diferencias $\Delta y/\Delta t$ puede simplificarse a la expresión Δy; esto se llama la *primera diferencia* de y. El símbolo Δ, que significa diferencia, puede interpretarse de acuerdo con esto como una instrucción para tomar la primera diferencia de (y). Como tal, constituye la contraparte en tiempo discreto del símbolo operador d/dt.

Por supuesto que la expresión Δy puede adoptar diferentes valores, dependiendo de qué par de periodos consecutivos intervienen en la toma de las diferencias (o "diferendos"). Para evitar ambigüedades, añadamos un subíndice de tiempo a y y definamos en forma más específica la primera diferencia como sigue:

$$\Delta y_t \equiv y_{t+1} - y_t \tag{17.1}$$

donde y_t representa el valor de y en el t-ésimo periodo, y y_{t+1} es su valor en el periodo que sigue inmediatamente al periodo t-ésimo. Con esta simbología, podemos describir el patrón de cambio de y por una ecuación tal como

$$\Delta y_t = 2 \tag{17.2}$$

o bien

$$\Delta y_t = -0.1 y_t \tag{17.3}$$

Las ecuaciones de este tipo se llaman *ecuaciones en diferencias*. Observe la notable semejanza entre estas dos últimas ecuaciones, por un lado, y las ecuaciones en diferencias $dy/dt = 2$ y $dy/dt = -0.1y$ por el otro.

Aun cuando las ecuaciones en diferencias derivan su nombre de las expresiones en diferencias tal como Δy_t, hay formas equivalentes alternas de las ecuaciones de este tipo que están completamente libres de expresiones en Δ y que son de uso más conveniente. En virtud de (17.1), podemos reescribir (17.2) como

$$y_{t+1} - y_t = 2 \tag{17.2'}$$

o

$$y_{t+1} = y_t + 2 \tag{17.2''}$$

Para (17.3), las formas equivalentes alternas correspondientes son

$$y_{t+1} - 0.9 y_t = 0 \tag{17.3'}$$

o bien

$$y_{t+1} = 0.9 y_t \tag{17.3''}$$

Las versiones con los números con doble prima son convenientes cuando calculamos un valor de y a partir de un valor conocido de y del periodo anterior. Sin embargo, en discusiones posteriores vamos a emplear principalmente las versiones con números con una sola prima, es decir, los de (17.2′) y (17.3′).

Es importante observar que la elección de los subíndices de tiempo en una ecuación en diferencias es un poco arbitraria. Por ejemplo, sin ningún cambio en el significado, (17.2′) puede reescribirse como $y_t - y_{t-1} = 2$, donde $(t - 1)$ se refiere al periodo que precede inmediatamente al t-ésimo. O podemos expresarlo en forma equivalente como $y_{t+2} - y_{t+1} = 2$.

También podemos señalar que, aunque hemos usado en forma consistente el símbolo y con subíndice, también es aceptable usar $y(t)$, $y(t+1)$ y $y(t-1)$ en su lugar. Sin embargo, con objeto de evitar el uso de la notación $y(t)$ para ambos casos de tiempo continuo y tiempo discreto, vamos a seguir el mecanismo del subíndice en el estudio del análisis de periodos.

En forma análoga a las ecuaciones diferenciales, las ecuaciones en diferencias pueden ser lineales o no lineales, homogéneas o no homogéneas, y de primero o segundo orden (o también orden superior). Considere por ejemplo (17.2′). Ésta puede clasificarse como: (1) lineal, ya que ningún término y (para cualquier periodo) está elevado a la segunda potencia (o mayor) ni está multiplicado por un término y de otro periodo; (2) no homogénea, ya que el miembro derecho (donde no hay término en y) es diferente de cero, y (3) de primer orden, ya que existe sólo una *primera diferencia* Δy_t, que incluye sólo un retraso de tiempo de un periodo. (En contraste, una ecuación en diferencias de segundo orden, que vamos a estudiar en el capítulo 18, incluye un retraso de dos periodos y, por lo tanto, intervienen tres términos en y: y_{t+2}, y_{t+1}, así como y_t.)

En realidad, (17.2′) también puede caracterizarse como que tiene coeficientes constantes y un término constante ($= 2$). Dado que el caso de coeficientes constantes es el único que vamos a considerar, en lo sucesivo esta caracterización la supondremos como implícita. En el presente capítulo también vamos a conservar el aspecto del término constante y discutiremos en el capítulo 18 el caso del término variable.

Verifique que la ecuación (17.3′) es también lineal y de primer orden; pero a diferencia de (17.2′), es homogénea.

17.2 Solución de una ecuación en diferencias de primer orden

Al resolver una ecuación diferencial, nuestro objetivo era encontrar una trayectoria de tiempo $y(t)$. Como sabemos, esta trayectoria de tiempo es una función del tiempo que está totalmente libre de cualquier expresión con derivadas (o diferenciales) y que es perfectamente consistente con la ecuación diferencial dada, así como con las condiciones iniciales. La trayectoria de tiempo que buscamos a partir de una ecuación de diferencias es de naturaleza similar. Nuevamente, debe ser una función de t —una fórmula que defina los valores de y para cada periodo— lo cual es consistente con la ecuación en diferencias dada, así como con las condiciones iniciales. Además, no debe contener ninguna expresión en diferencias tal como Δy_t (o expresiones como $y_{t+1} - y_t$).

La solución de las ecuaciones diferenciales es una cuestión de integración en su análisis final. ¿Cómo resolvemos una ecuación en diferencias?

Método iterativo

Antes de desarrollar un método general para afrontarlo, expliquemos primero un método relativamente corriente, el *método iterativo* —el cual, aunque elemental, va a ser muy revelador de la naturaleza esencial de una solución.

En este capítulo nos interesa sólo el caso de primer orden; por ello la ecuación en diferencia describe el patrón de cambio de y entre sólo *dos* periodos consecutivos. Una vez que se especifica un patrón de este tipo, tal como en (17.2″), y que se nos da un valor inicial y_0, no hay problema para encontrar y_1 a partir de la ecuación. En forma similar, una vez encontrado y_1, y_2 la podremos obtener inmediatamente, y así, mediante la aplicación repetida (iteración) del patrón de cambio especificado en la ecuación de diferencias. Entonces, los resultados de la interacción nos van a permitir inferir una trayectoria de tiempo.

Ejemplo 1 Encuentre la solución de la ecuación en diferencias (17.2), suponiendo un valor inicial de $y_0 = 15$. Para desarrollar el proceso iterativo es más conveniente usar la forma alternativa de la ecuación en diferencias (17.2″), es decir, $y_{t+1} = y_t + 2$, con $y_0 = 15$. A partir de esta ecuación, podemos deducir paso por paso que

$$\begin{aligned} y_1 &= y_0 + 2 \\ y_2 &= y_1 + 2 = (y_0 + 2) + 2 = y_0 + 2(2) \\ y_3 &= y_2 + 2 = [y_0 + 2(2)] + 2 = y_0 + 3(2) \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned}$$

y, en general, para cualquier periodo t,

$$y_t = y_0 + t(2) = 15 + 2t \tag{17.4}$$

Esta última ecuación indica el valor de y para cualquier periodo (incluyendo el periodo inicial $t = 0$); por lo tanto constituye la solución de (17.2).

El proceso de iteración es elemental, es análogo a la solución de ecuaciones diferenciales simples por integración directa, pero sirve para señalar claramente la manera en que se genera una trayectoria de tiempo. En general, el valor de y_t depende de una manera específica del valor de y en el periodo inmediatamente anterior (y_{t-1}); entonces, un valor inicial y_0 va a conducirnos sucesivamente a $y_1, y_2, \ldots$, vía el patrón de cambio prescrito.

Ejemplo 2 Resuelva la ecuación en diferencias (17.3); esta vez, con un valor inicial no especificado y denotado simplemente como y_0. Nuevamente es más conveniente trabajar con la versión alternativa de (17.3″), es decir, $y_{t+1} = 0.9y_t$. Por iteración, tenemos

$$\begin{aligned} y_1 &= 0.9y_0 \\ y_2 &= 0.9y_1 = 0.9(0.9y_0) = (0.9)^2 y_0 \\ y_3 &= 0.9y_2 = 0.9(0.9)^2 y_0 = (0.9)^3 y_0 \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned}$$

Éstos pueden resumirse en la solución

$$y_t = (0.9)^t y_0 \tag{17.5}$$

Para aumentar el interés, a este ejemplo podemos impartirle algún contenido económico. En el análisis del multiplicador simple, un solo gasto de inversión en el periodo 0 va a atraer rondas sucesivas de gastos, lo cual, a su vez, originará cantidades variables del incremento del ingreso para periodos sucesivos. Usando y para denotar el *incremento de ingreso*, tenemos $y_0 =$ el monto de la inversión en el periodo 0; pero los incrementos de ingreso subsiguientes dependerán de la propensión marginal al consumo (PMC). Si PMC = 0.9 y si el ingreso de cada periodo se consume sólo en el siguiente periodo, entonces 90 por ciento de y_0 será consumido en el periodo 1, lo que conduce a un incremento del ingreso en el periodo 1 de $y_1 = 0.9y_0$. Mediante un razonamiento similar, podemos encontrar $y_2 = 0.9y_1$, etcétera. Vemos que éstos son precisamente los resultados del proceso iterativo citado anteriormente. En otras palabras, la generación del proceso de ingreso del multiplicador puede describirse con una ecuación en diferencias tal como (17.3″), y una solución como (17.5) nos va a decir cuál debe ser la magnitud del incremento del ingreso para cualquier periodo t.

Ejemplo 3 Resuelva la ecuación diferencial homogénea

$$my_{t+1} - ny_t = 0$$

Al normalizar y transponer, esto puede escribirse como

$$y_{t+1} = \left(\frac{n}{m}\right) y_t$$

que es lo mismo que (17.3″) en el ejemplo 2 excepto por el reemplazo de 0.9 por n/m. Entonces, por analogía, la solución debe ser

$$y_t = \left(\frac{n}{m}\right)^t y_0$$

Observe el término $\left(\frac{n}{m}\right)^t$. A través de este término es como los diferentes valores de t van a conducir hacia los valores correspondientes de y. Por lo tanto, corresponde a la expresión e^{rt} en las soluciones de las ecuaciones diferenciales. Si la escribimos en forma más general como b^t (b representa una base) y añadimos la constante multiplicativa A más general (en lugar de y_0), vemos que la solución de la ecuación en diferencias homogénea general del ejemplo 3 va a ser de la forma

$$y_t = Ab^t$$

Encontraremos que esta expresión Ab^t desempeñará el mismo papel importante en la ecuación en diferencias que la expresión Ae^{rt} en las ecuaciones diferenciales.[1] Sin embargo, aun cuando ambas son expresiones exponenciales, la primera es de base b, mientras que la última es de base e. Es razonable que, así como el tipo de la trayectoria continua en el tiempo $y(t)$ depende principalmente del valor de r, la trayectoria discreta en el tiempo y_t depende principalmente del valor de b.

Método general

Por ahora, usted debe haber quedado muy impresionado por las diferentes similitudes entre las ecuaciones diferenciales y las ecuaciones en diferencias. Como puede conjeturarse, el método general de solución que finalmente explicaremos será paralelo con el de las ecuaciones diferenciales.

Suponga que estamos buscando la solución para la ecuación en diferencias de primer orden

$$y_{t+1} + ay_t = c \tag{17.6}$$

donde a y c son dos constantes. La solución general consistirá en la suma de dos componentes: una *solución particular* y_p, que es *cualquier* solución de la ecuación no homogénea completa (17.6), y una *función complementaria* y_c, que es la solución general de la ecuación reducida de (17.6):

$$y_{t+1} + ay_t = 0 \tag{17.7}$$

La componente y_p representa nuevamente el nivel de equilibrio intertemporal de y, y la componente y_c, las desviaciones de la trayectoria de tiempo respecto a ese equilibrio. La suma de y_c y y_p constituye la solución *general*, debido a la presencia de una constante arbitraria. Como antes, con objeto de establecer la solución que cumpla con la condición inicial, por supuesto es necesaria una condición inicial.

Veamos primero la función complementaria. Nuestra experiencia con el ejemplo 3 sugiere que podemos intentar una solución de la forma $y_t = Ab^t$ (con $Ab^t \neq 0$, ya que de otra mane-

[1] Usted puede objetar esta afirmación al señalar que la solución (17.4) del ejemplo 1 no contiene un término de la forma Ab^t. Sin embargo, este último hecho surge sólo porque en el ejemplo 1 tenemos $b = n/m = 1/1 = 1$, de modo que el término Ab^t se reduce a una constante.

ra y_t resultará ser simplemente una línea recta horizontal situada en el eje t); en ese caso, también tenemos $y_{t+1} = Ab^{t+1}$. Si son válidos estos valores de y_t y y_{t+1}, la ecuación homogénea (17.7) se transformará en

$$Ab^{t+1} + aAb^t = 0$$

la cual, al cancelarse el factor común diferente de cero Ab^t, arroja

$$b + a = 0 \qquad \text{o bien} \qquad b = -a$$

Esto significa que, para la solución de prueba, debemos hacer $b = -a$; entonces, la función complementaria debe escribirse como

$$y_c(= Ab^t) = A(-a)^t$$

Busquemos ahora la solución particular, que tiene que ver con la ecuación completa (17.6). A este respecto, el ejemplo 3 no nos ayuda en nada, porque ese ejemplo se relaciona sólo con una ecuación homogénea. Sin embargo, observamos que para y_p podemos escoger *cualquier* solución de (17.6); entonces si puede funcionar una solución de prueba de la forma más simple $y_t = k$ (una constante), no se encontrará ninguna dificultad real. Ahora, si $y_t = k$, entonces y va a conservar el mismo valor constante con el tiempo, y debemos tener también $y_{t+1} = k$. La sustitución de estos valores en (17.6) arroja

$$k + ak = c \qquad \text{y} \qquad k = \frac{c}{1+a}$$

Dado que este valor particular de k satisface a la ecuación, la solución particular puede escribirse como

$$y_p(= k) = \frac{c}{1+a} \qquad (a \neq -1)$$

Siendo esto una constante, se indica en este caso un equilibrio estacionario.

Si resulta que $a = -1$, como en el ejemplo 1, la solución particular $c/(1+a)$ queda indefinida y debe buscarse alguna otra solución de la ecuación no homogénea (17.6). En este caso, empleamos el truco que ya nos es conocido de intentar una solución de la forma $y_t = kt$. Esto implica, por supuesto, que $y_{t+1} = k(t+1)$. Sustituyendo esto en (17.6), encontramos

$$k(t+1) + akt = c \qquad \text{y} \qquad k = \frac{c}{t+1+at} = c \qquad [\text{ya que } a = -1]$$

entonces $$y_p(= kt) = ct$$

Esta forma de la solución particular es una función no constante de t; por lo tanto, representa un equilibrio móvil.

Sumando y_c y y_p podemos escribir ahora la solución general de una de las dos formas siguientes:

$$y_t = A(-a)^t + \frac{c}{1+a} \qquad [\text{solución general, caso de } a \neq -1] \qquad \textbf{(17.8)}$$

$$y_t = A(-a)^t + ct = A + ct \qquad [\text{solución general, caso de } a = -1] \qquad \textbf{(17.9)}$$

Ninguna de estas soluciones está completamente determinada por la presencia de la constante arbitraria A. Para eliminar esta constante arbitraria, aprovechamos la condición inicial $y_t = y_0$ cuando $t = 0$. Haciendo $t = 0$ en (17.8), tenemos

$$y_0 = A + \frac{c}{1+a} \qquad \text{y} \qquad A = y_0 - \frac{c}{1+a}$$

En consecuencia, la versión de (17.8) que cumple con la condición inicial es

$$y_t = \left(y_0 - \frac{c}{1+a}\right)(-a)^t + \frac{c}{1+a} \qquad \text{[solución definida, caso de } a \neq -1] \qquad \textbf{(17.8')}$$

Haciendo $t = 0$ en (17.9), por otro lado, encontramos $y_0 = A$, de modo que la versión de (17.9) que cumple la condición inicial es

$$y_t = y_0 + ct \qquad \text{[solución definida, caso de } a = -1] \qquad \textbf{(17.9')}$$

Si este último resultado se aplica al ejemplo 1, la solución que surge es exactamente la misma que la solución iterativa (17.4).

Usted puede verificar la validez de cada una de estas soluciones mediante los dos siguientes pasos. Primero, haciendo $t = 0$ en (17.8′), vea que la última ecuación se reduce a la identidad $y_0 = y_0$, lo que implica la satisfacción de la condición inicial. Segundo, al sustituir la fórmula para y_t (17.8′) y una fórmula similar para y_{t+1} —obtenida al reemplazar t por $(t + 1)$ en (17.8′)— en (17.6), vea que la última se reduce a la identidad $c = c$, lo que implica que la trayectoria de tiempo es consistente con la ecuación en diferencias dada. La verificación de la validez de la solución (17.9′) es análoga.

Ejemplo 4

Resuelva la ecuación en diferencias de primer orden

$$y_{t+1} - 5y_t = 1 \qquad \left(y_0 = \tfrac{7}{4}\right)$$

Siguiendo el procedimiento usado al obtener (17.8′), podemos encontrar y_c intentando una solución $y_t = Ab^t$ (lo que implica $y_{t+1} = Ab^{t+1}$). Al sustituir estos valores en la versión homogénea $y_{t+1} - 5y_t = 0$ y al cancelar el factor común Ab^t, obtenemos $b = 5$. Entonces,

$$y_c = A(5)^t$$

Para encontrar y_p, intente la solución $y_t = k$, que implica $y_{t+1} = k$. Al sustituir esto en la ecuación en diferencias completa, encontramos $k = -\frac{1}{4}$. Por lo tanto,

$$y_p = -\tfrac{1}{4}$$

Se sigue que la solución general es

$$y_t = y_c + y_p = A(5)^t - \tfrac{1}{4}$$

Haciendo $t = 0$ aquí y utilizando la condición inicial $y_0 = \frac{7}{4}$, obtenemos $A = 2$. Entonces, la solución definida puede escribirse finalmente como

$$y_t = 2(5)^t - \tfrac{1}{4}$$

Como la ecuación en diferencias dada en este ejemplo es un caso especial de (17.6), con $a = -5$, $c = 1$ y $y_0 = \frac{7}{4}$, y como (17.8′) es la "fórmula" de la solución para este tipo de ecuación en diferencias, pudimos haber encontrado nuestra solución al insertar los valores específicos de los parámetros en (17.8′), con el resultado de que

$$y_t = \left(\frac{7}{4} - \frac{1}{1-5}\right)(5)^t + \frac{1}{1-5} = 2(5)^t - \frac{1}{4}$$

que concuerda perfectamente con la respuesta anterior.

Observe que el término y_{t+1} de (17.6) tiene un coeficiente unitario. Si una ecuación en diferencias dada tiene un coeficiente no unitario para este término, debe normalizarse antes de usar la fórmula solución (17.8′).

EJERCICIO 17.2

1. Convierta las siguientes ecuaciones en diferencias a la forma de (17.2″):
(*a*) $\Delta y_t = 7$
(*b*) $\Delta y_t = 0.3 y_t$
(*c*) $\Delta y_t = 2y_t - 9$
2. Resuelva las siguientes ecuaciones en diferencias por iteración:
(*a*) $y_{t+1} = y_t + 1$ $\quad (y_0 = 10)$
(*b*) $y_{t+1} = \alpha y_t$ $\quad (y_0 = \beta)$
(*c*) $y_{t+1} = \alpha y_t - \beta$ $\quad (y_t = y_0$ cuando $t = 0)$
3. Exprese las ecuaciones del problema 2 en la forma de (17.6), y resuélvalas aplicando la fórmula (17.8′) o la (17.9′), la que sea más apropiada. ¿Concuerdan sus respuestas con las obtenidas con el método iterativo?
4. Para cada una de las siguientes ecuaciones en diferencias, use el procedimiento ilustrado en la obtención de (17.8′) y (17.9′) para encontrar y_c, y_p y la solución definida:
(*a*) $y_{t+1} + 3y_t = 4$ $\quad (y_0 = 4)$
(*b*) $2y_{t+1} - y_t = 6$ $\quad (y_0 = 7)$
(*c*) $y_{t+1} = 0.2y_t + 4$ $\quad (y_0 = 4)$

17.3 La estabilidad dinámica del equilibrio

Para el caso continuo del tiempo, la estabilidad dinámica de equilibrio depende del término Ae^{rt} en la función complementaria. En el análisis de periodos, el papel correspondiente lo desempeña el término Ab^t en la función complementaria. Dado que su interpretación es un poco más complicada que Ae^{rt}, intentemos aclararla antes de seguir adelante.

La importancia de b

Que el equilibrio sea dinámicamente estable es una cuestión de si la función complementaria va a tender a cero cuando $t \to \infty$. Básicamente, debemos analizar la trayectoria del término Ab^t a medida que t crece de modo indefinido. Es obvio que el valor de b (la base de este término exponencial) es de importancia crucial en este aspecto. Consideremos primero su importancia individual, ignorando el coeficiente A (suponiendo que $A = 1$).

Para propósitos analíticos, podemos dividir la imagen de los valores posibles de b, $(-\infty, +\infty)$, en siete regiones distintas, tal como se establece en las dos primeras columnas de la tabla 17.1, arreglada en orden descendente de magnitud de b. Estas regiones también están identificadas en la figura 17.1 en una escala vertical de b, con los puntos $+1$, 0 y -1 como los puntos de demarcación. De hecho, estos tres últimos puntos en sí mismos constituyen las regiones II, IV y VI. Por otro lado, las regiones III y V corresponden al conjunto de todas las fracciones positivas y al conjunto de todas las fracciones negativas, respectivamente. Las dos regiones restantes, I y VII, son aquellas en las cuales el valor numérico de b sobrepasa la unidad.

Para cada región, la expresión exponencial b^t genera un tipo diferente de trayectoria de tiempo. Éstas se ejemplifican en la tabla 17.1 y se ilustran en la figura 17.1. En la región I (donde $b > 1$), b^t debe incrementarse con t a un ritmo creciente. La configuración general de la trayectoria de tiempo va a asumir, por lo tanto, la forma de la gráfica superior de la figura

TABLA 17.1
Una clasificación de los valores de b

Región	Valor de b		Valor de b^t	Valor de b^t para diferentes periodos				
				$t=0$	$t=1$	$t=2$	$t=3$	$t=4\cdots$
I	$b>1$	$(\lvert b\rvert>1)$	e.g., $(2)^t$	1	2	4	8	16
II	$b=1$	$(\lvert b\rvert=1)$	$(1)^t$	1	1	1	1	1
III	$0<b<1$	$(\lvert b\rvert<1)$	e.g., $\left(\frac{1}{2}\right)^t$	1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$
IV	$b=0$	$(\lvert b\rvert=0)$	$(0)^t$	0	0	0	0	0
V	$-1<b<0$	$(\lvert b\rvert<1)$	e.g., $\left(-\frac{1}{2}\right)^t$	1	$-\frac{1}{2}$	$\frac{1}{4}$	$-\frac{1}{8}$	$\frac{1}{16}$
VI	$b=-1$	$(\lvert b\rvert=1)$	$(-1)^t$	1	-1	1	-1	1
VII	$b<-1$	$(\lvert b\rvert>1)$	e.g., $(-2)^t$	1	-2	4	-8	16

17.1. Observe que esta gráfica se muestra como una función escalón en vez de una curva continua; esto se debe a que estamos tratando con el análisis de los periodos. En la región II ($b = 1$), b^t permanecerá como la unidad para todos los valores de t. Su gráfica será entonces una línea recta horizontal. Enseguida en la región III, b^t representa una fracción positiva elevada a potencias enteras. A medida que aumenta la potencia, b^t debe disminuir, aunque siempre permanezca positiva. El siguiente caso, el de $b = 0$ de la región IV, es muy similar al caso de $b = 1$; pero aquí tenemos $b^t = 0$ en vez de $b^t = 1$, de modo que su gráfica va a coincidir con el eje horizontal. Sin embargo, este caso es sólo de interés secundario, ya que anteriormente hemos adoptado la hipótesis de que $Ab^t \neq 0$.

Cuando nos movemos a las regiones negativas, ocurre un fenómeno nuevo interesante: ¡El valor de b^t se *alterna* entre valores positivos y negativos de periodo a periodo! Este hecho se destaca claramente en los últimos tres renglones de la tabla 17.1 y en las últimas tres gráficas de la figura 17.1. En la región V, donde b es una fracción negativa, la trayectoria de tiempo alterna tiende a acercarse al eje horizontal (vea la región III de fracciones positivas). En contraste, cuando $b = -1$ (región VI), se da una alternancia perpetua entre $+1$ y -1. Y finalmente, cuando $b < -1$ (región VII), la trayectoria de tiempo alterna va a desviarse cada vez más del eje horizontal.

Lo que llama la atención es que, mientras que el fenómeno de una trayectoria de tiempo fluctuante no puede surgir posiblemente de un solo término Ae^{rt} (el caso de raíces complejas de la ecuación diferencial de segundo orden requiere un *par* de raíces complejas), la fluctuación puede generarse por un solo término b^t (o Ab^t). Observe, sin embargo, que la naturaleza de la fluctuación es algo diferente; a diferencia del patrón de las funciones circulares, la fluctuación ilustrada en la figura 17.1 no es suave. Por esta razón, emplearemos la palabra *oscilación* para denotar al nuevo tipo de fluctuación no continuo, aun cuando muchos autores usan los términos fluctuación y oscilaciones en forma intercambiable.

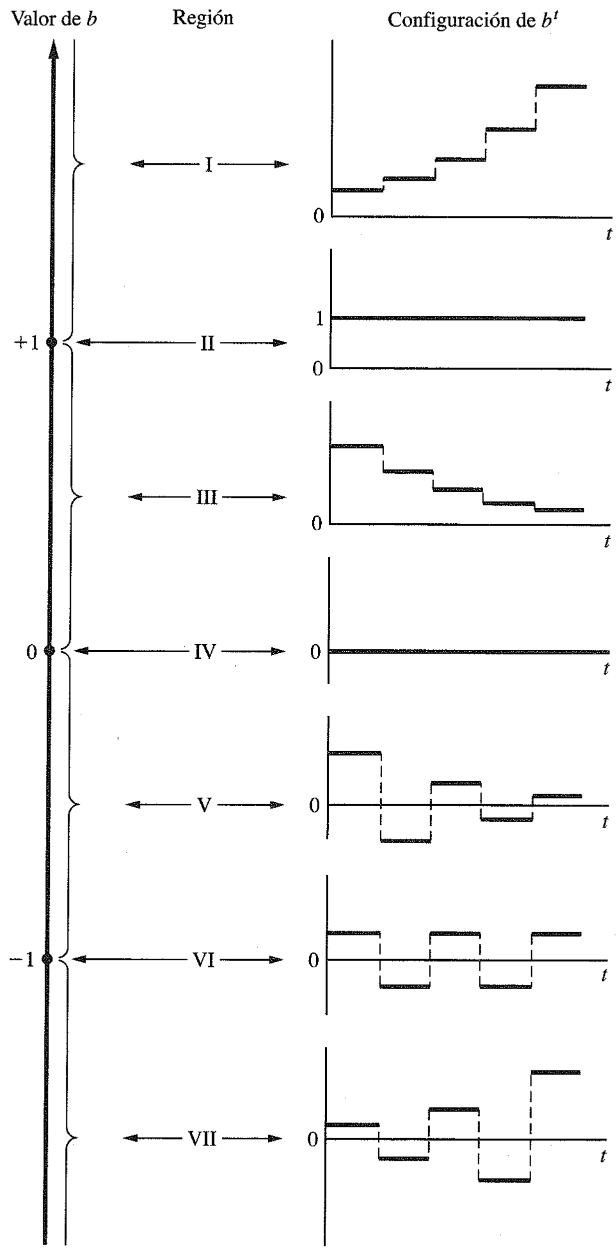
La esencia de la discusión anterior puede trasmitirse en el siguiente enunciado general: la trayectoria de tiempo de b^t ($b \neq 0$) será

$$\left.\begin{array}{l}\text{No oscilatorio}\\ \text{Oscilatorio}\end{array}\right\}\quad \text{si} \quad \left\{\begin{array}{l} b > 0\\ b < 0\end{array}\right.$$

$$\left.\begin{array}{l}\text{Divergente}\\ \text{Convergente}\end{array}\right\}\quad \text{si} \quad \left\{\begin{array}{l} |b| > 1\\ |b| < 1\end{array}\right.$$

Es importante observar que, mientras que la convergencia de la expresión e^{rt} depende del *signo* de r, la convergencia de la expresión b^t depende, a su vez, del *valor absoluto* de b.

FIGURA 17.1



La función de A

Hasta ahora hemos dejado fuera deliberadamente la constante multiplicativa A. Pero sus efectos —de los cuales hay dos— son relativamente fáciles de considerar. Primero, la *magnitud* de A puede servir para "aumentar" (si por ejemplo $A = 3$) o para "disminuir" (si por ejemplo $A = \frac{1}{5}$) los valores de b^t. Es decir, puede producir un *efecto de escala* sin cambiar la configuración básica de la trayectoria de tiempo. Por otro lado, el *signo* de A afecta a la forma de la trayectoria ya que, si b^t se multiplica por $A = -1$, entonces cada trayectoria de tiempo mostrada en la figura 17.1 será reemplazada por su propia imagen espejo con referencia al eje horizontal. Entonces, un A negativo puede producir un *efecto de espejo,* así como un efecto de escala.

Convergencia al equilibrio

La discusión anterior presenta la interpretación del término Ab^t en la función complementaria, la cual, como recordamos, representa las desviaciones respecto a algún nivel de equilibrio intertemporal. Si se añade un término (digamos) $y_p = 5$ al término Ab^t, la trayectoria de tiempo debe desplazarse verticalmente hacia arriba por un valor constante de 5. Esto no va a afectar de ninguna manera a la convergencia o la divergencia de la trayectoria de tiempo, pero va a alterar el nivel respecto al cual se mide la convergencia o la divergencia. Lo que ilustra la figura 17.1 es la convergencia (o la falta de ella) de la expresión Ab^t a cero. Cuando se incluye y_p, el problema se transforma en la determinación de la convergencia de la trayectoria de tiempo $y_t = y_c + y_p$ al nivel de equilibrio y_p.

A este respecto, demos alguna explicación para el caso especial de $b = 1$ (región II). Una trayectoria de tiempo tal que

$$y_t = A(1)^t + y_p = A + y_p$$

da la impresión de que converge, ya que el término multiplicativo $(1)^t = 1$ no produce ningún efecto explosivo. Observe, sin embargo, que y_t adopta ahora el valor $(A + y_p)$ en vez del valor de equilibrio y_p; de hecho, nunca puede alcanzar a y_p (a menos que $A = 0$). Como una ilustración de este tipo de situación, podemos citar la trayectoria de tiempo de (17.9), en la cual interviene un equilibrio móvil $y_p = ct$. Esta trayectoria de tiempo debe considerarse divergente, no por la aparición de t en la solución particular sino porque, con A diferente de cero, habrá una desviación constante respecto al equilibrio móvil. Así, al estipular la condición de convergencia de la trayectoria de tiempo y_t hacia el equilibrio y_p, debemos descartar el caso de $b = 1$.

En suma, la solución

$$y_t = Ab^t + y_p$$

es una trayectoria convergente si y sólo si $|b| < 1$.

Ejemplo 1 ¿Qué tipo de trayectoria de tiempo representa $y_t = 2(-\frac{4}{5})^t + 9$? Como $b = -\frac{4}{5} < 0$, la trayectoria de tiempo es oscilatoria. Pero como $|b| = \frac{4}{5} < 1$, la oscilación se amortigua, y la trayectoria de tiempo converge al nivel de equilibrio de 9.

Debe tener cuidado de no confundir $2(-\frac{4}{5})^t$ con $-2(\frac{4}{5})^t$, ya que representan configuraciones totalmente diferentes de trayectorias de tiempo.

Ejemplo 2 ¿Cómo caracteriza a la trayectoria de tiempo $y_t = 3(2)^t + 4$? Puesto que $b = 2 > 0$, no va a haber oscilación. Pero ya que $|b| = 2 > 1$, la trayectoria de tiempo va a divergir del nivel de equilibrio de 4.

EJERCICIO 17.3

1. Discuta la naturaleza de las siguientes trayectorias de tiempo:

(*a*) $y_t = 3^t + 1$ (*c*) $y_t = 5\left(-\frac{1}{10}\right)^t + 3$

(*b*) $y_t = 2\left(\frac{1}{3}\right)^t$ (*d*) $y_t = -3\left(\frac{1}{4}\right)^t + 2$

2. ¿Cuál es la naturaleza de la trayectoria de tiempo obtenida de cada una de las ecuaciones en diferencias en el ejercicio 17.2-4?
3. Encuentre las soluciones de lo siguiente y determine si las trayectorias de tiempo son oscilatorias y convergentes:
   (*a*) $y_{t+1} - \frac{1}{3}y_t = 6 \qquad (y_0 = 1)$
   (*b*) $y_{t+1} + 2y_t = 9 \qquad (y_0 = 4)$
   (*c*) $y_{t+1} + \frac{1}{4}y_t = 5 \qquad (y_0 = 2)$
   (*d*) $y_{t+1} - y_t = 3 \qquad (y_0 = 5)$

17.4 El modelo de la telaraña

Para ilustrar el uso de las ecuaciones en diferencias de primer orden en el análisis económico, citaremos dos variantes del modelo de mercado para un solo artículo. La primera variante, conocida como el *modelo de la telaraña*, difiere de nuestros modelos de mercado anteriores en que trata a Q_s como una función no del precio de lista sino del precio del periodo anterior.

El modelo

Considere una situación en la cual el fabricante toma decisiones sobre la producción con un periodo de anticipación de la venta real —tal como en la producción agrícola, donde la siembra debe anteceder con mucho tiempo a la cosecha y venta del producto. Supongamos que la decisión del productor en el periodo t se basa en el precio P_t prevaleciente entonces. Como este producto no va a estar disponible para la venta hasta el periodo $(t+1)$, sin embargo, P_t va a determinar no a Q_{st}, sino a $Q_{s,t+1}$. Entonces ahora tenemos una función de oferta "retrasada".[2]

$$Q_{s,t+1} = S(P_t)$$

o en forma equivalente, al desplazar hacia atrás en un periodo los índices de tiempo,

$$Q_{st} = S(P_{t-1})$$

Cuando una función de oferta de este tipo interactúa con una función de demanda de la forma

$$Q_{dt} = D(P_t)$$

aparecen interesantes patrones en la dinámica de precio.

Considerando las versiones lineales de estas funciones de oferta (retrasada) y de demanda (sin retraso), y suponiendo que para cada periodo el precio de mercado siempre se establece a un nivel que pone al clarificar al mercado, tenemos un modelo de mercado con las siguientes tres ecuaciones:

$$\begin{aligned} Q_{dt} &= Q_{st} \\ Q_{dt} &= \alpha - \beta P_t \qquad (\alpha, \beta > 0) \\ Q_{st} &= -\gamma + \delta P_{t-1} \qquad (\gamma, \delta > 0) \end{aligned} \tag{17.10}$$

[2] Aquí estamos haciendo la hipótesis implícita de que la producción completa de un periodo se va a colocar en el mercado, sin ninguna parte conservada en el almacenamiento. Esta hipótesis es apropiada cuando el artículo en cuestión es perecedero o cuando no se lleva ningún inventario. En la sección 17.5 vamos a considerar un modelo con inventario.

Sin embargo, al sustituir las últimas dos ecuaciones en la primera, el modelo puede reducirse a una ecuación en diferencias individual de primer orden como sigue:

$$\beta P_t + \delta P_{t-1} = \alpha + \gamma$$

Con objeto de resolver esta ecuación, primero es conveniente normalizarla y desplazar los subíndices de tiempo hacia delante un periodo [alterar t a $(t+1)$, etc.]. El resultado,

$$P_{t+1} + \frac{\delta}{\beta} P_t = \frac{\alpha + \gamma}{\beta} \qquad \textbf{(17.11)}$$

será entonces una réplica de (17.6), con las sustituciones

$$y = P \qquad a = \frac{\delta}{\beta} \qquad \text{y} \qquad c = \frac{\alpha + \gamma}{\beta}$$

Como δ y β son ambas positivas, se sigue que $a \neq -1$. En consecuencia, podemos aplicar la fórmula (17.8′) para obtener la trayectoria de tiempo

$$P_t = \left(P_0 - \frac{\alpha + \gamma}{\beta + \delta}\right)\left(-\frac{\delta}{\beta}\right)^t + \frac{\alpha + \gamma}{\beta + \delta} \qquad \textbf{(17.12)}$$

donde P_0 representa el precio inicial.

Las telarañas

Podemos observar tres puntos respecto a esta trayectoria de tiempo. En primer lugar, la expresión $(\alpha + \gamma)/(\beta + \delta)$, que constituye la solución particular de la ecuación en diferencias, puede tomarse como el precio de equilibrio intertemporal del modelo:[3]

$$\overline{P} = \frac{\alpha + \gamma}{\beta + \delta}$$

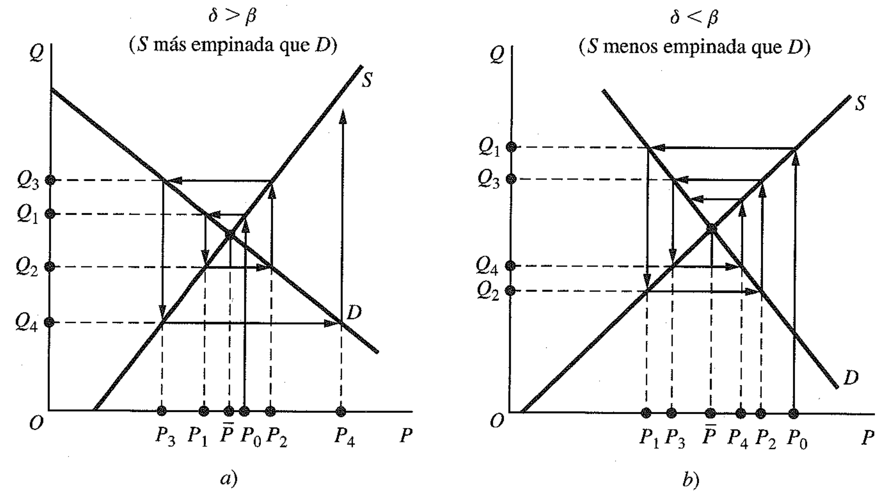
Dado que ésta es una constante, éste es un equilibrio estacionario. Al sustituir $\overline{P}$ en nuestra solución, podemos expresar la trayectoria de tiempo P_t alternativamente en la forma

$$P_t = (P_0 - \overline{P})\left(-\frac{\delta}{\beta}\right)^t + \overline{P} \qquad \textbf{(17.12')}$$

Esto nos conduce al segundo punto, es decir, a la importancia de la expresión $(P_0 - \overline{P})$. Como esto corresponde a la constante A en el término Ab^t, el signo va a proyectarse sobre la interrogante de si la trayectoria de tiempo va a iniciar arriba o debajo del equilibrio (efecto de espejo), mientras que su magnitud va a decidir qué tan arriba o abajo del equilibrio (efecto de escala). Por último, está la expresión $(-\delta/\beta)$, que corresponde a la componente b de Ab^t. A partir de nuestra especificación de modelo de que $\beta, \delta > 0$, podemos deducir una trayectoria de tiempo oscilatoria. Este hecho da lugar al fenómeno de la telaraña, como finalmente

[3] Por lo que toca al sentido de equilibrio del agotamiento del mercado, el precio alcanzado en cada periodo es un precio de equilibrio, ya que hemos supuesto $Q_{dt} - Q_{st}$ para todo t.

FIGURA 17.2



veremos. Por supuesto que pueden surgir *tres* variedades posibles de patrones de oscilación en el modelo. De acuerdo con la tabla 17.1, o la figura 17.1, la oscilación es

$$\left.\begin{array}{l}\text{Explosiva}\\ \text{Uniforme}\\ \text{Amortiguada}\end{array}\right\} \quad \text{si} \quad \delta \gtreqless \beta$$

donde el término *oscilación uniforme* se refiere al tipo de trayectoria de la región VI.

Con objeto de visualizar las telarañas, ilustremos el modelo (17.10) en la figura 17.2. La segunda ecuación de (17.10) se grafica como una curva de demanda lineal con pendiente hacia abajo, con su pendiente numéricamente igual a β. En forma similar, podemos trazar a partir de la tercera ecuación una curva lineal de oferta con una pendiente igual a δ, si hacemos que el eje Q represente en este caso a una cantidad ofertada *retrasada*. Los casos de $\delta > \beta$ (S más empinado que D) y $\delta < \beta$ (S más llano que D) se ilustran en la figura 17.2*a* y *b*, respectivamente. Sin embargo, en cada uno de los casos la intersección de D y S conducirá al precio $\overline{P}$ de equilibrio intertemporal.

Cuando $\delta > \beta$, como en la figura 17.2*a*, la interacción de la oferta y la demanda producirá una oscilación explosiva como sigue. Dado un precio inicial P_0 (aquí supuesto por arriba de $\overline{P}$), podemos seguir la punta de la flecha y leer sobre la curva S que el nivel de la oferta en el siguiente periodo (periodo 1) será de Q_1. Con objeto de poner al mercado en ceros, el nivel de la demanda en el periodo 1 debe ser también Q_1, lo que es posible si y sólo si el precio se coloca al nivel de P_1 (vea la flecha hacia abajo). Ahora, vía la curva S, el precio P_1 conducirá a Q_2 como la cantidad ofertada en el periodo 2, y para poner al mercado en ceros para el último periodo, el precio debe colocarse al nivel de P_2 de acuerdo con la curva de demanda. Repitiendo este razonamiento, podemos rastrear los precios y las cantidades para periodos subsiguientes siguiendo simplemente las cabezas de flecha en el diagrama, con lo cual se teje una "telaraña" alrededor de las curvas de oferta y demanda. Al comparar los niveles de precio, $P_0,\ P_1,\ P_2, \ldots$, observamos en este caso no sólo un patrón oscilatorio de cambio sino también una tendencia a que el precio amplíe su desviación respecto a $\overline{P}$ a medida que pasa el tiempo. Con la telaraña que se teje de dentro hacia fuera, la trayectoria de tiempo es divergente y la oscilación es explosiva.

En comparación, en el caso de la figura 17.2*b*, donde $\delta < \beta$, el proceso de tejido crea una telaraña que es centrípeta. A partir de P_0, si seguimos las cabezas de flecha vamos a acercarnos cada vez más a la intersección de las curvas de oferta y demanda, donde está $\overline{P}$. Aun cuando todavía es oscilatorio, esta trayectoria de precio es convergente.

En la figura 17.2 no hemos mostrado una tercera posibilidad, es decir, $\delta = \beta$. Sin embargo, el procedimiento de análisis gráfico que interviene es perfectamente análogo a los otros dos casos. Por lo tanto se le deja a usted como un ejercicio.

La discusión anterior ha tratado sólo de la trayectoria de tiempo de P (es decir, P_t); sin embargo, después de encontrar P_t falta sólo un corto paso para llegar a la trayectoria de tiempo de Q. La segunda ecuación de (17.10) relaciona a Q_{dt} con P_t, de modo que si (17.12) o (17.12′) se sustituyen en la ecuación de la demanda, podemos obtener inmediatamente la trayectoria de tiempo Q_{dt}. Aún más, puesto que Q_{dt} debe ser igual a Q_{st} para cada periodo (con mercados que se vacían), podemos simplemente referirnos a la trayectoria de tiempo como Q_t en vez de Q_{dt}. Basándonos en la figura 17.2 se ve fácilmente el razonamiento de esta sustitución. Cada punto de la curva D relaciona un P_i con un Q_i que pertenecen al mismo periodo; por lo tanto, la función de demanda puede servir para mapear la trayectoria de tiempo del precio sobre la trayectoria de tiempo de la cantidad.

Usted debe observar que la técnica gráfica de la figura 17.2 es aplicable aun cuando las curvas D y S sean no lineales.

EJERCICIO 17.4

1. Basándose en (17.10), encuentre la trayectoria de tiempo de Q y analice la condición de su convergencia.
2. Dibuje un diagrama similar a los de la figura 17.2 para mostrar que, para el caso de $\delta = \beta$, el precio oscilará uniformemente sin amortiguamiento ni explosión.
3. Dadas la oferta y la demanda para el modelo de telaraña que sigue, encuentre el precio de equilibrio intertemporal, y determine si el equilibrio es estable:
   (*a*) $Q_{dt} = 18 - 3P_t \qquad Q_{st} = -3 + 4P_{t-1}$
   (*b*) $Q_{dt} = 22 - 3P_t \qquad Q_{st} = -2 + P_{t-1}$
   (*c*) $Q_{dt} = 19 - 6P_t \qquad Q_{st} = 6P_{t-1} - 5$
4. Para el modelo (17.10), sean la condición $Q_{dt} = Q_{st}$ y la función de demanda que permanecen como son, pero cambie la función de oferta a
   $$Q_{st} = -\gamma + \delta P_t^*$$
   donde P_t^* denota al precio esperado para el periodo t. Aún más, suponga que los vendedores tienen el tipo "adaptativo" de expectativa de precio:[4]
   $$P_t^* = P_{t-1}^* + \eta(P_{t-1} - P_{t-1}^*) \qquad (0 < \eta \leq 1)$$
   donde η (la letra griega eta) es un coeficiente de ajuste de expectativa.
   (*a*) Dé una interpretación económica a la ecuación anterior. ¿En qué aspectos es similar y diferente de la ecuación de expectativas adaptativas (16.34)?
   (*b*) ¿Qué ocurre si η adapta su valor máximo? ¿Podemos considerar el modelo de la telaraña como un caso especial del presente modelo?

[4] Vea Marc Nerlove, "Adaptive Expectations and Cobweb Phenomena", *Quarterly Journal of Economics,* mayo de 1958, pp. 227-240.

(*c*) Muestre que el nuevo modelo puede representarse mediante la ecuación en diferencias de primer orden

$$P_{t+1} - \left(1 - \eta - \frac{\eta\delta}{\beta}\right) P_t = \frac{\eta(\alpha + \gamma)}{\beta}$$

(*Sugerencia*: Resuelva la función de oferta para P_t^* y luego use la información de que $Q_{st} = Q_{dt} = \alpha - \beta P_t$.)

(*d*) Encuentre la trayectoria de tiempo del precio. ¿Es esta trayectoria necesariamente oscilatoria? ¿Puede ser oscilatoria? ¿Bajo qué circunstancias?

(*e*) Muestre que la trayectoria de tiempo P_t, si es oscilatoria, convergirá sólo si $1 - 2/\eta < -\delta/\beta$. En comparación con la solución de la telaraña (17.12) o (17.12′), ¿tiene el nuevo modelo un rango más amplio o más estrecho para los valores de $-\delta/\beta$ que inducen a la estabilidad?

5. El modelo de telaraña, al igual que los modelos de mercado dinámico analizados previamente, se basa esencialmente en el modelo de mercado estático de la sección 3.2. ¿Qué hipótesis económica introduce la parte dinámica en el presente caso? Explique.

17.5 Un modelo de mercado con inventario

En el modelo anterior, se supone que el precio se establece de manera tal que el inventario se vacía en cada periodo. La implicación de esta hipótesis es que ya sea que el artículo es perecedero y no puede almacenarse o que, aun cuando sea almacenable, no se lleva ningún inventario. Ahora vamos a construir un modelo en el cual los vendedores llevan un inventario del artículo.

El modelo

Supongamos lo siguiente:

1. Tanto la cantidad demandada, Q_{dt}, como la cantidad producida al presente, Q_{st}, son funciones lineales sin retraso del precio P_t.
2. El ajuste de precio se efectúa no a través de la clarificación del mercado para cada periodo, sino a través de un proceso de fijación de precios por los vendedores. Al inicio de cada periodo, los vendedores establecen un precio para ese periodo después de considerar la situación del inventario. Si como resultado del precio del periodo anterior se acumuló el inventario, el precio del periodo presente se establece a un nivel más bajo que antes, con objeto de "mover" la mercancía; pero si en lugar de eso disminuyó el inventario, el precio presente se fija más alto que antes.
3. El ajuste de precios que se hace de periodo en periodo es inversamente proporcional al cambio observado en el inventario (existencias).

Con estas hipótesis, podemos escribir las siguientes ecuaciones:

$$\begin{aligned} Q_{dt} &= \alpha - \beta P_t & &(\alpha, \beta > 0) \\ Q_{st} &= -\gamma + \delta P_t & &(\gamma, \delta > 0) \qquad \textbf{(17.13)} \\ P_{t+1} &= P_t - \sigma(Q_{st} - Q_{dt}) & &(\sigma > 0) \end{aligned}$$

donde σ denota al coeficiente de *ajuste de precios inducidos por las existencias*. Observe que (17.13) no es nada más que la contraparte discreta en el tiempo del modelo de mercado de la sección 15.2, aunque ahora hemos conducido al proceso de ajuste de precios en términos del *inventario* $(Q_{st} - Q_{dt})$ en vez de la demanda excesiva $(Q_{dt} - Q_{st})$. Sin embargo, los resultados analíticos son muy diferentes; simplemente porque para el tiempo discreto podemos encontrar el fenómeno de las oscilaciones. Obtengamos y analicemos la trayectoria de tiempo P_t.

La trayectoria de tiempo

Al sustituir las dos primeras ecuaciones en la tercera, el modelo puede condensarse en una sola ecuación diferencial:

$$P_{t+1} - [1 - \sigma(\beta + \delta)]P_t = \sigma(\alpha + \gamma) \tag{17.14}$$

y su solución está dada por (17.8′):

$$\begin{aligned} P_t &= \left(P_0 - \frac{\alpha + \gamma}{\beta + \delta}\right)[1 - \sigma(\beta + \delta)]^t + \frac{\alpha + \gamma}{\beta + \delta} \\ &= (P_0 - \overline{P})[1 - \sigma(\beta + \delta)]^t + \overline{P} \end{aligned} \tag{17.15}$$

Por lo tanto, obviamente que la estabilidad dinámica del modelo dependerá de la expresión $1 - \sigma(\beta + \delta)$; por comodidad, nos referiremos a esta expresión como b.

Con referencia a la tabla 17.1, vemos que al analizar la expresión exponencial b^t podemos definir siete regiones diferentes de valores de b. Sin embargo, ya que nuestras especificaciones del modelo $(\sigma, \beta, \delta > 0)$ han descartado de hecho las dos primeras regiones, quedan sólo cinco casos posibles, como se lista en la tabla 17.2. Para cada una de estas regiones, la especificación b de la segunda columna puede traducirse en una especificación σ equivalente, como se muestra en la tercera columna. Por ejemplo, para la región III, la especificación b es $0 < b < 1$; por lo tanto podemos escribir

$$0 < 1 - \sigma(\beta + \delta) < 1$$

$$-1 < -\sigma(\beta + \delta) < 0 \qquad \text{[restando 1 de las tres partes]}$$

y $$\frac{1}{\beta + \delta} > \sigma > 0 \qquad \text{[dividiendo por } -(\beta + \delta)]$$

TABLA 17.2
Tipos de trayectoria de tiempo

Región	Valor de $b \equiv 1 - \sigma(\beta + \delta)$	Valor de σ	Natureza de la trayectoria de tiempo P_t
III	$0 < b < 1$	$0 < \sigma < \frac{1}{\beta + \delta}$	No oscilatoria y convergente
IV	$b = 0$	$\sigma = \frac{1}{\beta + \delta}$	Permanece en equilibrio[5]
V	$-1 < b < 0$	$\frac{1}{\beta + \delta} < \sigma < \frac{2}{\beta + \delta}$	Con oscilación amortiguada
VI	$b = -1$	$\sigma = \frac{2}{\beta + \delta}$	Con oscilación uniforme
VII	$b < -1$	$\sigma > \frac{2}{\beta + \delta}$	Con oscilación explosiva

[5] El hecho de que el precio va a permanecer en equilibrio en este caso también puede verse directamente en (17.14). Con $\sigma = 1/(\beta + \delta)$, el coeficiente de P_t se hace cero, y (17.14) se reduce a $P_{t+1} = \sigma(\alpha + \gamma) = (\alpha + \gamma)/(\beta + \delta) = \overline{P}$.

Esto último nos da la especificación deseada equivalente de σ para la región III. La transferencia para las otras regiones puede desarrollarse en forma análoga. Como el tipo de trayectoria de tiempo que pertenece a cada región ya se conoce de la figura 17.1, la especificación de σ nos permite discernir a partir de los valores dados de σ, β y δ la naturaleza general de la trayectoria de tiempo P_t, como se esboza en la última columna de la tabla 17.2.

Ejemplo 1

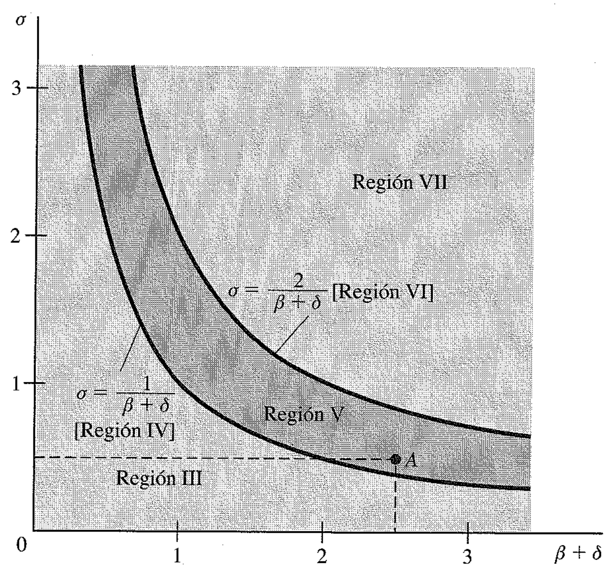
Si los vendedores de nuestro modelo incrementan siempre (disminuyen) el precio en 10 por ciento de la cantidad de la disminución (incremento) del inventario, y si la curva de demanda tiene una pendiente de -1 y la curva de oferta una pendiente de 15 (ambas pendientes respecto al eje de los precios), ¿qué tipo de trayectoria de tiempo P_t vamos a encontrar?

Aquí tenemos $\sigma = 0.1$, $\beta = 1$ y $\delta = 15$. Como $1/(\beta + \delta) = \frac{1}{16}$ y $2/(\beta + \delta) = \frac{1}{8}$, el valor de σ $(= \frac{1}{10})$ está situado entre los dos primeros valores; por lo tanto es un caso de la región V. La trayectoria de tiempo P_t va a caracterizarse por la oscilación amortiguada.

Resumen gráfico de los resultados

La tabla 17.2, que contiene cinco casos posibles de la especificación σ, puede hacerse mucho más fácil de comprender si los resultados se presentan gráficamente. Como la especificación σ incluye esencialmente una comparación de las magnitudes relativas de los parámetros σ frente a $(\beta + \delta)$, grafiquemos σ contra $(\beta + \delta)$, como en la figura 17.3. Observe que necesitamos preocuparnos sólo del cuadrante positivo ya que, por la especificación del modelo, σ y $(\beta + \delta)$ son ambas positivas. De la tabla 17.2, es evidente que las regiones IV y VI se especifican mediante las ecuaciones $\sigma = 1/(\beta + \delta)$ y $\sigma = 2/(\beta + \delta)$, respectivamente. Puesto que cada una de éstas exhibe la gráfica de una hipérbola rectangular, las dos regiones se representan gráficamente por las dos hipérbolas de la figura 17.3. Aún más, una vez que tenemos las dos hipérbolas, las otras tres regiones quedan inmediatamente en su lugar. Por ejemplo, la región III es simplemente el conjunto de puntos situados debajo de la hipérbola inferior, donde tenemos a σ menor que $1/(\beta + \delta)$. En forma similar, la región V está representada por el conjunto de puntos situados entre las dos hipérbolas, mientras que todos los puntos ubicados arriba de la hipérbola más alta pertenecen a la región VII.

FIGURA 17.3



Ejemplo 2 Si $\sigma = \frac{1}{2}$, $\beta = 1$ y $\delta = \frac{3}{2}$, ¿arrojará nuestro modelo (17.13) una trayectoria de tiempo convergente P_t? Los valores paramétricos dados corresponden al punto A de la figura 17.3. Puesto que está situado dentro de la región V, la trayectoria de tiempo es convergente, aunque oscilatoria.

Observará que en los dos modelos recién presentados nuestros resultados analíticos se enuncian en cada caso como un conjunto de casos alternativos posibles: tres tipos de trayectoria oscilatoria para las telarañas, y cinco tipos de trayectoria de tiempo en el modelo del inventario. Esta riqueza de resultados analíticos proviene por supuesto de la formulación paramétrica de los modelos. El hecho de que nuestro resultado no puede enunciarse en una respuesta individual inequívoca es, por supuesto, un mérito más bien que una debilidad.

EJERCICIO 17.5

1. Al resolver (17.14), ¿por qué debe usarse la fórmula (17.8′) en vez de (17.9′)?
2. Basándose en la tabla 17.2, verifique la validez de la translación de la especificación b a la especificación σ para las regiones IV a VII.
3. Si el modelo (17.13) tiene la siguiente forma numérica:
$$Q_{dt} = 21 - 2P_t$$
$$Q_{st} = -3 + 6P_t$$
$$P_{t+1} = P_t - 0.3(Q_{st} - Q_{dt})$$
encuentre la trayectoria de tiempo P_t y determine si es convergente.
4. Suponga que en el modelo (17.13) la oferta en cada periodo es una cantidad fija, digamos, $Q_{st} = k$, en vez de una función de precios. Analice el comportamiento de los precios respecto al tiempo. ¿Qué restricción debe imponerse sobre k para hacer que la solución tenga un significado económico?

17.6 Ecuaciones en diferencias no lineales. Método gráfico cualitativo

Hasta ahora hemos utilizado en nuestros modelos sólo ecuaciones en diferencias *lineales*; pero los hechos de la vida económica no siempre se ajustan de manera conveniente a la linealidad. Afortunadamente, cuando se presenta la no linealidad en el caso de los modelos de ecuaciones en diferencias de primer orden, existe un método de análisis sencillo que es aplicable bajo condiciones bastante generales. Este método, que es de naturaleza gráfica, se parece mucho al del análisis cualitativo de las ecuaciones en diferencias de primer orden presentado en la sección 15.6.

Diagrama de fase

Las ecuaciones en diferencias no lineales en las cuales intervienen sólo las variables y_{t+1} y y_t, tales como

$$y_{t+1} + y_t^3 = 5 \qquad \text{o} \qquad y_{t+1} + \operatorname{sen} y_t - \ln y_t = 3$$

pueden representarse en forma categórica mediante la ecuación

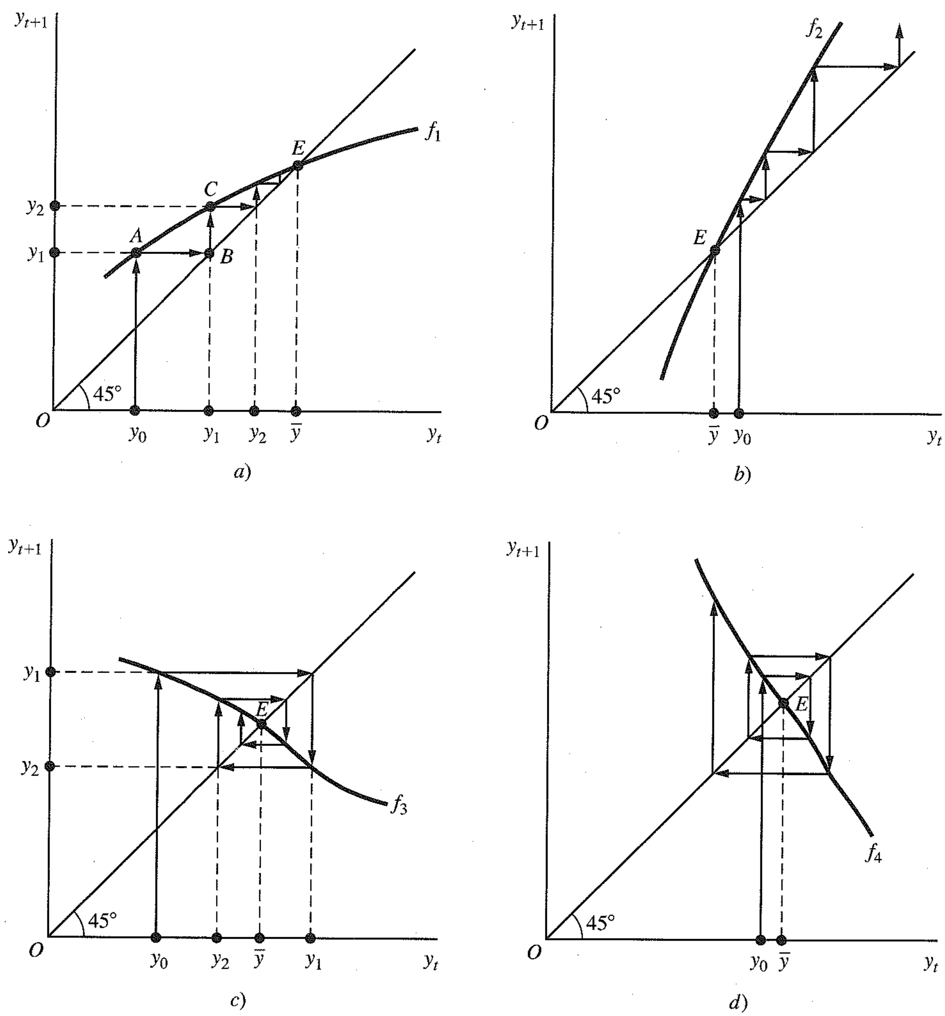
$$y_{t+1} = f(y_t) \tag{17.16}$$

donde f puede ser una función de cualquier grado de complejidad, siempre que sea una función de y_t sólo sin t como argumento. Cuando se grafican entre sí las dos variables y_{t+1} y y_t en un plano coordenado Cartesiano, el diagrama resultante constituye un *diagrama de fase*, y la curva correspondiente a f es una *línea de fase*. A partir de éstos, es posible analizar la trayectoria de tiempo de la variable mediante el proceso de iteraciones.

Los términos diagrama de fase y línea de fase se usan aquí análogamente al caso de las ecuaciones diferenciales; pero observe una diferencia en el trazo del diagrama. En el caso de las ecuaciones diferenciales, graficamos dy/dt contra y como en la figura 15.3, de modo que, con objeto de obtener una analogía perfecta en el presente caso, debemos tener Δy_t en el eje vertical y y_t en el horizontal. Esto no es imposible de lograr, pero es mucho más conveniente a su vez colocar y_{t+1} en el eje vertical, como lo hemos hecho en la figura 17.4, donde se usa la misma escala en ambos ejes. Observe la presencia de una línea de 45° en cada diagrama de la figura 17.4; esta línea nos va a prestar un gran servicio al desarrollar nuestro análisis gráfico.

Ilustremos el procedimiento involucrado mediante la figura 17.4a, donde hemos dibujado una línea de fase (rotulada f_1) que representa una ecuación en diferencias específica $y_{t+1} = f_1(y_t)$.

FIGURA 17.4



Si se nos da un valor inicial y_0 (graficado en el eje horizontal), mediante iteraciones podemos trazar todos los valores subsiguientes de y como sigue. Primero, como la línea de fase f_1 mapea el valor inicial y_0 sobre y_1 de acuerdo con la ecuación

$$y_1 = f_1(y_0)$$

podemos subir directamente desde y_0 a la línea de fase hasta encontrar el punto A y leer su altura sobre el eje vertical como el valor de y_1. Enseguida, buscamos mapear y_1 sobre y_2 de acuerdo con la ecuación

$$y_2 = f_1(y_1)$$

Para este propósito, debemos graficar primero y_1 en el eje horizontal, en forma similar a y_0 durante el primer mapeo. Esta transposición requerida de y_1 desde el eje vertical al horizontal se logra de la manera más sencilla mediante el uso de la línea de 45°, la cual, con una pendiente igual a $+1$, es el lugar geométrico de los puntos con abscisas y ordenadas idénticas, tales como $(2, 2)$ y $(5, 5)$. Entonces, para transponer y_1 desde el eje vertical, podemos simplemente trazar una referencia hasta la línea de 45°, alcanzar el punto B, y luego trazar una referencia hacia abajo hasta el eje horizontal para localizar el punto y_1. Mediante la repetición de este proceso, podemos mapear y_1 sobre y_2 vía el punto C sobre la línea de fase, y luego usar la línea de 45° para la transposición de y_2, etcétera.

Ahora que está clara la naturaleza de la iteración, podemos observar que la iteración deseada puede alcanzarse simplemente siguiendo las cabezas de flecha desde y_0 hasta A (sobre la línea de fase), hasta B (sobre la línea de 45°), hasta C (sobre la línea de fase), etcétera —alternando siempre entre las dos figuras— sin que nunca sea necesario hacer referencia a los ejes nuevamente.

Tipos de trayectoria de tiempo

Las iteraciones gráficas esbozadas hasta aquí son igualmente aplicables a los otros tres diagramas de la figura 17.4. En realidad, estos cuatro diagramas sirven para ilustrar cuatro variedades básicas de líneas de fase, cada una implica un tipo diferente de trayectoria de tiempo. Las dos primeras líneas de fase, f_1 y f_2, se caracterizan por pendientes positivas, con una pendiente menor que la unidad y la otra mayor que la unidad:

$$0 < f_1'(y_t) < 1 \qquad \text{y} \qquad f_2'(y_t) > 1$$

Por otro lado, las dos restantes tienen pendiente negativa; específicamente tenemos

$$-1 < f_3'(y_t) < 0 \qquad \text{y} \qquad f_4'(y_t) < -1$$

Para cada diagrama de la figura 17.4, el valor de equilibrio intertemporal de y (es decir, $\overline{y}$) se localiza en la intersección de la línea de fase con la línea de 45°, que hemos rotulado como E. Esto es así debido a que el punto E sobre la línea de fase, es simultáneamente un punto sobre la línea de 45°, por tanto va a mapear un y_t sobre un y_{t+1} de valor idéntico; y cuando $y_{t+1} = y_t$, por definición y debe estar en equilibrio intertemporal. Nuestra tarea principal es determinar si, dado un valor inicial $y_0 \neq \overline{y}$, el patrón de cambio implicado por la línea de fase va a conducirnos consistentemente hacia $\overline{y}$ (convergente) o a alejarse de éste (divergente).

Para la línea de fase f_1, el proceso iterativo conduce desde y_0 hasta $\overline{y}$ en una trayectoria uniforme, sin oscilaciones. Usted puede verificar que, si y_0 se coloca a la derecha de $\overline{y}$, también habrá un movimiento uniforme hacia $\overline{y}$, aunque será en dirección a la izquierda. Estas trayectorias de tiempo son convergentes hasta el equilibrio, y su configuración general sería del mismo tipo que se muestra en la región III de la figura 17.1.

Dada la línea de fase f_2, cuya pendiente sobrepasa la unidad, surge una trayectoria de tiempo divergente. A partir de un valor inicial y_0 mayor que $\overline{y}$, las cabezas de flecha nos conducen constantemente lejos del equilibrio a valores cada vez más altos de y. Como usted puede verificar, un valor inicial más bajo que $\overline{y}$ a un movimiento divergente uniforme similar, aunque en dirección opuesta.

Cuando la línea de fase tiene una inclinación negativa, como en f_3 y f_4, el movimiento da lugar a oscilaciones, y aparece ahora el fenómeno de *estar por encima* de la marca de equilibrio. En el diagrama c, y_0 conduce a y_1, lo que sobrepasa a $\overline{y}$, sólo para ser seguido por y_2, el cual casi le atina a $\overline{y}$, etcétera. En estos casos, la convergencia de la trayectoria de tiempo va a depender de que la pendiente de la línea de fase sea menor que 1 en valor absoluto. Éste es el caso de la línea de fase f_3, donde la amplitud de estar por encima tiende a disminuir para periodos sucesivos. Para la línea de fase f_4, cuya pendiente sobrepasa numéricamente 1, por otro lado, prevalece la tendencia opuesta, lo que conduce a una trayectoria de tiempo divergente.

Las trayectorias de tiempo oscilatorias generadas por las líneas de fase f_3 y f_4 nos recuerdan a las telarañas de la figura 17.2. Sin embargo, en la figura 17.4*c* o *d*, la telaraña se teje alrededor de una línea de fase (que contiene un retraso) y de la línea de 45°, en vez de alrededor de una curva de demanda y una curva de oferta (retrasada). Aquí se usa una línea de 45° como una ayuda mecánica para la transposición de un valor de y, mientras que en la figura 17.2, la curva D (que juega un papel similar al de la línea de 45° en la figura 17.4) es una parte integral del modelo mismo. Específicamente, una vez que se determina Q_{st} en la curva de oferta, hacemos que las cabezas de flecha alcancen a la curva D con objeto de encontrar un precio que "agote al mercado", como era la regla del juego en el modelo de la telaraña. En consecuencia, existe una diferencia básica en el rotulado de los ejes: en la figura 17.2 existen dos variables totalmente diferentes, P y Q, pero en la figura 17.4 los ejes representan los valores de la misma variable y para dos periodos consecutivos. Observe, sin embargo, que si analizamos la gráfica de la ecuación en diferencias (17.11) que resume al modelo de la telaraña, en vez de las funciones separadas de oferta y demanda en (17.10), entonces el diagrama resultante será una línea de fase tal como se muestra en la figura 17.4. En otras palabras, existen realmente dos formas alternativas de analizar gráficamente el modelo de la telaraña, lo que va a suministrar un resultado idéntico.

La regla básica que surge de la consideración anterior de la línea de fase es que el *signo algebraico* de la pendiente determina si va a haber *oscilación*, y el *valor absoluto* de la pendiente gobierna el aspecto de la *convergencia*. Si resulta que la línea de fase contiene segmentos con pendiente tanto positiva como negativa, y si el valor absoluto de la pendiente es mayor que 1 en algunos puntos, y es menor que 1 para todos los demás, es natural que la trayectoria de tiempo se haga más complicada. Sin embargo, aun en estos casos, podemos emplear con igual facilidad el análisis gráfico-iterativo. Por supuesto que debe dársenos un valor inicial antes de iniciar la iteración apropiadamente. En verdad que para estos casos más complicados, un valor inicial diferente puede conducirnos a una trayectoria de tiempo de una clase totalmente diferente (vea los ejercicios 17.6-2 y 17.6-3).

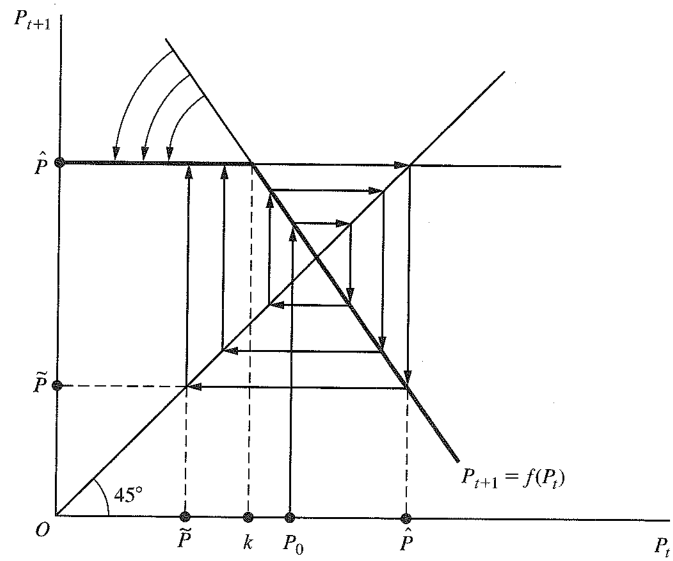
Un mercado con precio máximo

Vamos a citar ahora un ejemplo en economía de una ecuación no lineal en diferencias. En la figura 17.4, las cuatro líneas de fase no lineales resultan ser, todas, de la variedad uniforme; en el presente ejemplo, mostraremos una línea de fase no uniforme.

Como punto de partida, consideremos la ecuación en diferencias *lineal* (17.11) del modelo de telaraña y expresémosla como

$$P_{t+1} = \frac{\alpha + \gamma}{\beta} - \frac{\delta}{\beta} P_t \qquad \left(\frac{\delta}{\beta} > 0\right) \tag{17.17}$$

FIGURA 17.5



Esta ecuación tiene la forma $P_{t+1} = f(P_t)$, con $f'(P_t) = -\delta/\beta < 0$. Hemos graficado esta línea de fase lineal en la figura 17.5 con la hipótesis de que la pendiente es mayor que 1 en valor absoluto, lo que implica oscilación *explosiva.*

Ahora impongamos precio máximo legal de precios $\hat{P}$ (lea: "P con gorro"). Esto puede mostrarse en la figura 17.5 como una línea recta horizontal ya que, independientemente del nivel de P_t, ahora P_{t+1} está imposibilitado de sobrepasar el nivel de $\hat{P}$. Lo que hace esto es invalidar la parte de la línea de fase situada arriba de $\hat{P}$, o visto de manera diferente, doblegar la parte superior de la línea de fase hasta el nivel de $\hat{P}$, lo que conduce a una línea de fase acotada.[6] En vista del acotamiento, la nueva línea de fase (línea gruesa) es no lineal y también es irregular. Al igual que una función escalón, esta línea irregular va a requerir más de una ecuación para expresarla algebraicamente:

$$P_{t+1} = \begin{cases} \hat{P} & (\text{para } P_t \leq k) \\ \dfrac{\alpha + \gamma}{\beta} - \dfrac{\delta}{\beta} P_t & (\text{para } P_t > k) \end{cases} \qquad \textbf{(17.17')}$$

donde k denota el valor de P_t en el vértice en donde cambia de dirección.

Suponiendo un precio inicial P_0, tracemos en forma iterativa la trayectoria de tiempo de los precios. Durante la primera etapa de la iteración, cuando surte efecto el segmento de la línea de fase que tiene pendiente hacia abajo, la tendencia oscilatoria explosiva se hace claramente manifiesta. Sin embargo, después de unos cuantos periodos, las cabezas de flecha comienzan a alcanzar al precio máximo, y a partir de ahí la trayectoria de tiempo va a evolucionar hasta un movimiento cíclico perpetuo entre $\hat{P}$ y un *piso de precios* efectivo $\tilde{P}$ (lea: "P con tilde"). Entonces, en virtud del techo de precios, se contiene efectivamente la tendencia explosiva intrínseca del modelo, y ahora la oscilación que tiende a ensancharse se detiene hasta convertirse en una oscilación uniforme que produce un así llamado ciclo límite.

[6] Hablando estrictamente, también debemos "doblar" la parte de la línea de fase situada a la derecha del punto $\hat{P}$ en el eje horizontal. Pero no hace ningún daño el dejarlo como está, siempre que el otro extremo ya haya sido doblado, ya que la transposición de P_{t+1} al eje horizontal va a desplazar automáticamente al límite superior de $\hat{P}$ automáticamente al eje P_t.

Lo que es importante acerca de este resultado es que, mientras que en el caso de una línea de fase lineal puede producirse una trayectoria oscilatoria uniformemente si y sólo si la pendiente de la línea de fase es -1, ahora después de la introducción de la no linealidad puede surgir el mismo resultado analítico aun cuando la línea de fase tenga una pendiente diferente de -1. La implicación económica de esto es de considerable importancia. Si observamos la oscilación más o menos uniforme de una variable en la trayectoria de tiempo real e intentamos explicarlo mediante un modelo *lineal*, nos veremos obligados a depender de la especificación de modelo más bien especial —y no plausible— de que la pendiente de la línea de fase es exactamente -1. Pero si se introduce la no linealidad, ya sea de la variedad regular o irregular, entonces podemos usar una gran cantidad de hipótesis más razonables, cada una de las cuales puede explicar igualmente la característica observada de la oscilación uniforme.

EJERCICIO 17.6

1. En los modelos de ecuaciones en diferencias, la variable t puede adoptar sólo valores enteros. ¿Implica esto que en los diagramas de fase de la figura 17.4 las variables y_t y y_{t+1} deben considerarse como variables discretas?
2. Use la mitad izquierda de una curva con forma de U invertida como línea de fase que alcance la línea de 45° en dos puntos L (izquierdo) y R (derecho).
   (*a*) ¿Es éste un caso de equilibrios múltiples?
   (*b*) Si el valor inicial y_0 está situado a la izquierda de L, ¿qué tipo de trayectoria de tiempo va a obtenerse?
   (*c*) ¿Qué pasa si el valor inicial está situado entre L y R?
   (*d*) ¿Qué pasa si el valor inicial está situado a la derecha de R?
   (*e*) ¿Qué puede concluir usted acerca de la estabilidad dinámica de equilibrio en L y R, respectivamente?
3. Use una curva inversa con forma de U como línea de fase. Haga que el segmento con pendiente hacia arriba alcance la línea de 45° en el punto L, y haga que el segmento con pendiente hacia abajo llegue a la línea de 45° en el punto R. Responda las mismas cinco preguntas propuestas en el problema 2 (*Nota:* su respuesta va a depender de la forma específica en que se dibuje la línea de fase; explore diferentes posibilidades).
4. En la figura 17.5 elimine el precio legal máximo e imponga en su lugar un precio mínimo P_m.
   (*a*) ¿Cómo va a cambiar la línea de fase?
   (*b*) ¿Tendrá un vértice? ¿Será no lineal?
   (*c*) ¿Va a desarrollarse también un movimiento uniformemente oscilatorio en los precios?
5. Con referencia a (17.17′) y la figura 17.5, muestre que la constante k puede expresarse como

$$k = \frac{\alpha + \gamma}{\delta} - \frac{\beta}{\delta}\hat{P}$$

Capítulo 18

Ecuaciones en diferencias de orden superior

Los modelos económicos del capítulo 17 incluyen ecuaciones en diferencias que relacionan entre sí a P_t con P_{t-1}. Como el valor de P en un periodo puede determinar en forma única al valor de P en el siguiente, la trayectoria de tiempo de P se hace totalmente determinada una vez que se especifica un valor inicial de P_0. Sin embargo, Suele ocurrir que el valor de una variable económica para el periodo t (digamos y_t) depende no sólo de y_{t-1} sino también de y_{t-2}. Esta situación va a dar lugar a una ecuación en diferencias de segundo orden.

Hablando estrictamente, una *ecuación en diferencias de segundo orden* es aquella que implica una expresión $\Delta^2 y_t$, llamada la *segunda diferencia* de y_t, pero no contiene diferencias de orden mayores de 2. El símbolo Δ^2, la contraparte en tiempo discreto del símbolo d^2/dt^2, es una instrucción para "tomar la segunda diferencia" como sigue:

$$\begin{aligned}
\Delta^2 y_t &= \Delta(\Delta y_t) = \Delta(y_{t+1} - y_t) && \text{[por (17.1)]} \\
&= (y_{t+2} - y_{t+1}) - (y_{t+1} - y_t) && \text{[nuevamente por (17.1)]}^1 \\
&= y_{t+2} - 2y_{t+1} + y_t
\end{aligned}$$

Entonces una segunda diferencia de y_t es transformable en una suma de términos que incluyen un retraso de dos periodos. Dado que las expresiones como $\Delta^2 y_t$ y Δy_t son bastante complicadas para trabajar con ellas, simplemente vamos a redefinir una ecuación en diferencias de segundo orden como la que incluye un retraso de tiempo de dos periodos en la variable. En forma similar, una ecuación en diferencias de tercer orden es aquella que incluye un retraso de tiempo de tres periodos, etcétera.

Concentrémonos primero en el método de solución de una ecuación en diferencias de segundo orden, dejando la generalización a las ecuaciones de orden superior para la sección 18.4. Para conservar manejable el alcance de la discusión, en el presente capítulo trataremos sólo las ecuaciones diferenciales lineales con coeficientes constantes. Sin embargo, examinaremos tanto las ecuaciones con término constante como las de término variable.

[1] Es decir, primero movemos hacia delante los subíndices en la expresión $(y_{t+1} - y_t)$, para obtener una nueva expresión $(y_{t+2} - y_{t+1})$, y luego restamos de ésta la expresión original. Observe que, como la diferencia resultante puede escribirse como $\Delta y_{t+1} - \Delta y_t$, podemos inferir la siguiente regla de operación:

$$\Delta(y_{t+1} - y_t) = \Delta y_{t+1} - \Delta y_t$$

Esto nos recuerda de la regla aplicable a la derivada de una suma o de una diferencia.

18.1 Ecuaciones en diferencias lineales de segundo orden con coeficientes constantes y término constante

Una variedad simple de las ecuaciones en diferencias de segundo orden adopta la forma

$$y_{t+2} + a_1 y_{t+1} + a_2 y_t = c \qquad \textbf{(18.1)}$$

Usted va a reconocer que esta ecuación es lineal, no homogénea y con coeficientes constantes (a_1, a_2) y término constante c.

La solución particular

Como antes, podemos esperar que la solución de (18.1) tenga dos componentes: una solución particular y_p que representa el nivel de equilibrio intertemporal de y, y una función complementaria y_c que especifica para cada periodo la desviación respecto al equilibrio. La solución particular, que se define como cualquier solución de la ecuación completa, puede algunas veces encontrarse simplemente al intentar una solución de la forma $y_t = k$. Al sustituir este valor constante de y en (18.1), obtenemos

$$k + a_1 k + a_2 k = c \qquad \text{y} \qquad k = \frac{c}{1 + a_1 + a_2}$$

Entonces, siempre que $(1 + a_1 + a_2) \neq 0$, la integral particular sea

$$y_p(= k) = \frac{c}{1 + a_1 + a_2} \quad (\text{en caso de } a_1 + a_2 \neq -1) \qquad \textbf{(18.2)}$$

Ejemplo 1 Encuentre la integral particular de $y_{t+2} - 3y_{t+1} + 4y_t = 6$. Aquí tenemos $a_1 = -3$, $a_2 = 4$ y $c = 6$. Como $a_1 + a_2 \neq -1$, la solución particular puede obtenerse a partir de (18.2) como sigue:

$$y_p = \frac{6}{1 - 3 + 4} = 3$$

En caso de que $a_1 + a_2 = -1$, entonces la solución de prueba $y_t = k$ falla y podemos intentar en su lugar $y_t = kt$. Al sustituir esta última en (18.1) y teniendo en mente que ahora tenemos $y_{t+1} = k(t + 1)$ y $y_{t+2} = k(t + 2)$, encontramos que

$$k(t + 2) + a_1 k(t + 1) + a_2 kt = c$$

$$\text{y} \qquad k = \frac{c}{(1 + a_1 + a_2)t + a_1 + 2} = \frac{c}{a_1 + 2} \qquad [\text{ya que } a_1 + a_2 = -1]$$

Entonces podemos escribir la solución particular como

$$y_p(= kt) = \frac{c}{a_1 + 2} t \quad (\text{en caso de } a_1 + a_2 = -1;\ a_1 \neq -2) \qquad \textbf{(18.2')}$$

Ejemplo 2 Encuentre la solución particular de $y_{t+2} + y_{t+1} - 2y_t = 12$. Aquí, $a_1 = 1$, $a_2 = -2$ y $c = 12$. Obviamente, la fórmula (18.2) no es aplicable, pero (18.2′) sí lo es. Entonces,

$$y_p = \frac{12}{1 + 2} t = 4t$$

Esta solución particular representa un equilibrio móvil.

Si $a_1 + a_2 = -1$, pero al mismo tiempo $a_1 = -2$ (es decir, si $a_1 = -2$ y $a_2 = 1$), entonces podemos adoptar una solución de prueba de la forma $y_t = kt^2$, lo que implica que $y_{t+1} = k(t+1)^2$, etcétera. Como el lector puede verificar, en este caso la solución particular resulta ser

$$y_p = kt^2 = \frac{c}{2}t^2 \quad \text{(en caso de } a_1 = -2;\ a_2 = 1) \tag{18.2''}$$

Sin embargo, dado que esta fórmula es aplicable sólo al caso único de la ecuación en diferencias $y_{t+2} - 2y_{t+1} + y_t = c$, su utilidad es más bien limitada.

La función complementaria

Para encontrar la función complementaria debemos concentrarnos en la ecuación reducida

$$y_{t+2} + a_1 y_{t+1} + a_2 y_t = 0 \tag{18.3}$$

Nuestra experiencia con las ecuaciones diferenciales de primer orden nos ha enseñado que la expresión Ab^t juega un papel prominente en la solución general de esta ecuación. Por lo tanto intentemos una solución de la forma $y_t = Ab^t$, lo que implica naturalmente que $y_{t+1} = Ab^{t+1}$, etc. Ahora nuestra tarea es determinar los valores de A y b.

Al sustituir la solución de prueba en (18.3), la ecuación se transforma en

$$Ab^{t+2} + a_1 Ab^{t+1} + a_2 Ab^t = 0$$

o después de cancelar el factor común (diferente de cero) Ab^t,

$$b^2 + a_1 b + a_2 = 0 \tag{18.3'}$$

Esta ecuación cuadrática —la *ecuación característica* de (18.3) o de (18.1)— que es comparable con (16.4''), posee las dos *raíces características*

$$b_1, b_2 = \frac{-a_1 \pm \sqrt{a_1^2 - 4a_2}}{2} \tag{18.4}$$

cada una de las cuales es aceptable en la solución Ab^t. De hecho, *tanto* b_1 *como* b_2 deben aparecer en la solución general de la ecuación en diferencia homogénea (18.3) porque, al igual que en el caso de las ecuaciones diferenciales, esta solución general debe consistir de dos partes *linealmente independientes*, cada una con su propia constante arbitraria multiplicativa.

Podemos encontrar tres situaciones posibles respecto a las raíces características, dependiendo de la expresión con raíz cuadrada de (18.4). Usted descubrirá que esto se parece mucho al análisis de las ecuaciones diferenciales de segundo orden de la sección 16.1.

Caso 1 (raíces reales diferentes) Cuando $a_1^2 > 4a_2$, la raíz cuadrada en (18.4) es un número real, y b_1 y b_2 son reales y diferentes. En ese caso, b_1^t y b_2^t son linealmente independientes, y la función complementaria puede escribirse simplemente como una combinación lineal de estas expresiones; es decir,

$$y_c = A_1 b_1^t + A_2 b_2^t \tag{18.5}$$

Compare esto con (16.7).

Ejemplo 3 Encuentre la solución de $y_{t+2} + y_{t+1} - 2y_t = 12$. Esta ecuación tiene los coeficientes $a_1 = 1$ y $a_2 = -2$; de (18.4) se encuentra que las raíces características son $b_1, b_2 = 1, -2$. Entonces, la función complementaria es

$$y_c = A_1(1)^t + A_2(-2)^t = A_1 + A_2(-2)^t$$

Dado que en el ejemplo 2 ya vimos que la solución particular de la ecuación en diferencias dada es $y_p = 4t$, podemos escribir la solución general como

$$y_t = y_c + y_p = A_1 + A_2(-2)^t + 4t$$

Todavía hay dos constantes arbitrarias, A_1 y A_2, que deben determinarse; para lograr esto, son necesarias *dos* condiciones iniciales. Suponga que se nos da $y_0 = 4$ y $y_1 = 5$. Entonces, puesto que al hacer $t = 0$ y $t = 1$ sucesivamente en la solución general, encontramos

$$y_0 = A_1 + A_2 \qquad (= 4 \text{ por la primera condición inicial})$$
$$y_1 = A_1 - 2A_2 + 4 \qquad (= 5 \text{ por la segunda condición inicial})$$

las constantes arbitrarias se determinan como $A_1 = 3$ y $A_2 = 1$. Entonces, la solución definida puede escribirse finalmente como

$$y_t = 3 + (-2)^t + 4t$$

Caso 2 (raíces reales repetidas) Cuando $a_1^2 = 4a_2$ se anula la raíz cuadrada de (18.4) y las raíces características se repiten:

$$b(= b_1 = b_2) = -\frac{a_1}{2}$$

Ahora, si expresamos la función complementaria en la forma de (18.5), los dos componentes se colapsan en un solo término:

$$A_1b_1^t + A_2b_2^t = (A_1 + A_2)b^t \equiv A_3b^t$$

Esto no va a funcionar, pues ahora nos falta una constante.

Para suministrar la componente que falta —la cual nosotros recordamos que debe ser linealmente independiente del término A_3b^t— el viejo truco de multiplicar b^t por la variable t va a funcionar de nueva cuenta. Por lo tanto, el nuevo término componente debe adoptar la forma A_4tb^t. Debe ser obvio que esto es linealmente independiente de A_3b^t, ya que nunca podremos obtener la expresión A_4tb^t al agregar un coeficiente constante a A_3b^t. Podemos verificar con facilidad que A_4tb^t en realidad califica como una solución de la ecuación homogénea (18.3) al igual que A_3b^t, al sustituir $y_t = A_4tb^t$ [y $y_{t+1} = A_4(t+1)b^{t+1}$, etc.] en (18.3)[2] y ver que esta última va a reducirse a una identidad $0 = 0$.

Por lo tanto, la función complementaria para el caso de raíces repetidas es

$$y_c = A_3b^t + A_4tb^t \tag{18.6}$$

lo que debe compararse con (16.9).

Ejemplo 4 Encuentre la función complementaria de $y_{t+2} + 6y_{t+1} + 9y_t = 4$. Siendo los coeficientes $a_1 = 6$ y $a_2 = 9$, se ve que las raíces características son $b_1 = b_2 = -3$. Por lo tanto tenemos

$$y_c = A_3(-3)^t + A_4t(-3)^t$$

Si avanzamos un paso más, podemos encontrar fácilmente $y_p = \frac{1}{4}$, de modo que la solución general de la ecuación en diferencias dada es

$$y_t = A_3(-3)^t + A_4t(-3)^t + \tfrac{1}{4}$$

Dadas dos condiciones iniciales, nuevamente pueden asignarse valores definidos a A_3 y A_4.

[2] Para esta sustitución debe tenerse en mente que tenemos en el presente caso $a_1^2 = 4a_2$ y $b = -a_1/2$.

Caso 3 (raíces complejas) Bajo la posibilidad restante de $a_1^2 < 4a_2$, las raíces características son complejas conjugadas. Específicamente, están en la forma

$$b_1, b_2 = h \pm vi$$

donde

$$h = -\frac{a_1}{2} \qquad \text{y} \qquad v = \frac{\sqrt{4a_2 - a_1^2}}{2} \tag{18.7}$$

Entonces la función complementaria misma se transforma en

$$y_c = A_1 b_1^t + A_2 b_2^t = A_1(h + vi)^t + A_2(h - vi)^t$$

Tal como está, y_c no se interpreta con facilidad. Pero afortunadamente, gracias al teorema de De Moivre, dado en (16.23′), esta función complementaria puede transformarse sin problemas en términos trigonométricos, que ya hemos aprendido a interpretar.

De acuerdo con dicho teorema, podemos escribir

$$(h \pm vi)^t = R^t(\cos\theta t \pm i \operatorname{sen}\theta t)$$

donde el valor de R (siempre considerado positivo) es, por (16.10),

$$R = \sqrt{h^2 + v^2} = \sqrt{\frac{a_1^2 + 4a_2 - a_1^2}{4}} = \sqrt{a_2} \tag{18.8}$$

y θ es la medida en radianes del ángulo en el intervalo $[0, 2\pi)$, que satisface las condiciones

$$\cos\theta = \frac{h}{R} = \frac{-a_1}{2\sqrt{a_2}} \qquad \text{y} \qquad \operatorname{sen}\theta = \frac{v}{R} = \sqrt{1 - \frac{a_1^2}{4a_2}} \tag{18.9}$$

Por lo tanto, la función complementaria puede transformase como sigue:

$$\begin{aligned} y_c &= A_1 R^t(\cos\theta t + i \operatorname{sen}\theta t) + A_2 R^t(\cos\theta t - i \operatorname{sen}\theta t) \\ &= R^t[(A_1 + A_2)\cos\theta t + (A_1 - A_2) i \operatorname{sen}\theta t] \\ &= R^t(A_5 \cos\theta t + A_6 \operatorname{sen}\theta t) \end{aligned} \tag{18.10}$$

donde hemos adoptado los símbolos

$$A_5 \equiv A_1 + A_2 \qquad \text{y} \qquad A_6 \equiv (A_1 - A_2)i$$

La función complementaria (18.10) difiere de la contraparte en las ecuaciones diferenciales (16.24′) en dos aspectos importantes. Primero, las expresiones $\cos\theta t$ y sen θt han reemplazado a los previamente utilizados $\cos vt$ y sen vt. Segundo, el factor multiplicativo R^t (una exponencial con base R) ha reemplazado a la expresión exponencial natural e^{ht}. En resumen, hemos cambiado de las coordenadas cartesianas (h y v) de las raíces complejas a las coordenadas polares (R y θ). Los valores de R y θ pueden determinarse a partir de (18.8) y (18.9) una vez que h y v se conocen. También se pueden calcular R y θ directamente de los valores de los parámetros a_1 y a_2 (18.8) y (18.9), siempre que nos aseguremos primero que $a_1^2 < 4a_2$, y que las raíces son realmente complejas.

Ejemplo 5 Encuentre la solución general de $y_{t+2} + \frac{1}{4}y_t = 5$. Con coeficientes $a_1 = 0$ y $a_2 = \frac{1}{4}$, esto constituye una ilustración del caso de raíces complejas de $a_1^2 < 4a_2$. Por (18.7), las partes real e imaginaria de las raíces son $h = 0$ y $v = \frac{1}{2}$. Se sigue de (18.8) que

$$R = \sqrt{0 + \left(\frac{1}{2}\right)^2} = \frac{1}{2}$$

Puesto que el valor de θ es aquel que satisface las dos ecuaciones

$$\cos\theta = \frac{h}{R} = 0 \qquad \text{y} \qquad \operatorname{sen}\theta = \frac{v}{R} = 1$$

puede concluirse de la tabla 16.1 que

$$\theta = \frac{\pi}{2}$$

En consecuencia, la función complementaria es

$$y_c = \left(\frac{1}{2}\right)^t \left(A_5 \cos\frac{\pi}{2}t + A_6 \operatorname{sen}\frac{\pi}{2}t\right)$$

Para encontrar y_p, intentemos una solución constante $y_t = k$ en la ecuación completa. Esto suministra $k = 4$; entonces, $y_p = 4$, y la solución general puede escribirse como

$$y_t = \left(\frac{1}{2}\right)^t \left(A_5 \cos\frac{\pi}{2}t + A_6 \operatorname{sen}\frac{\pi}{2}t\right) + 4 \tag{18.11}$$

Ejemplo 6 Encuentre la solución general de $y_{t+2} - 4y_{t+1} + 16y_t = 0$. En primer lugar, la solución particular se halla fácilmente como $y_p = 0$. Esto significa que la solución general $y_t\ (= y_c + y_p)$ es idéntica a y_c. Para encontrar esta última, observamos que los coeficientes $a_1 = -4$ y $a_2 = 16$ producen raíces complejas. Entonces podemos sustituir los valores de a_1 y a_2 directamente en (18.8) y (18.9) para obtener

$$R = \sqrt{16} = 4$$

$$\cos\theta = \frac{4}{2 \cdot 4} = \frac{1}{2} \qquad \text{y} \qquad \operatorname{sen}\theta = \sqrt{1 - \frac{16}{4 \cdot 16}} = \sqrt{\frac{3}{4}} = \frac{\sqrt{3}}{2}$$

Las dos últimas ecuaciones nos permiten encontrar de la tabla 16.2 que

$$\theta = \frac{\pi}{3}$$

Se sigue que la función complementaria —que también sirve aquí como la solución general— es

$$y_c(= y_t) = 4^t \left(A_5 \cos\frac{\pi}{3}t + A_6 \operatorname{sen}\frac{\pi}{3}t\right) \tag{18.12}$$

La convergencia de la trayectoria de tiempo

Como en el caso de las ecuaciones diferenciales de primer orden, la convergencia de la trayectoria de tiempo y_t depende sólo de si y_c tiende a cero cuando $t \to \infty$. Lo que aprendimos acerca de las diferentes configuraciones de la expresión b^t, de la figura 17.1, es por lo tanto aplicable todavía, aunque en el presente contexto tendremos que considerar *dos* raíces características en lugar de una.

Considere primero el caso de las raíces reales diferentes: $b_1 \neq b_2$. Si $|b_1| > 1$ y $|b_2| > 1$, entonces ambos términos componentes en la función complementaria (18.5) $A_1b_1^t$ y $A_2b_2^t$ serán explosivos y, por lo tanto, y_c debe ser divergente. En el caso opuesto de $|b_1| < 1$ y $|b_2| < 1$, ambos términos de y_c van a convergir a cero a medida que t aumenta indefinidamente, así como también lo hará y_c. ¿Qué pasa si $|b_1| > 1$ pero $|b_2| < 1$? En este caso intermedio, es evidente que el término $A_2 b_2^t$ tiende a "menguar", mientras que el otro término tiende a desviarse cada vez más de cero. Se sigue que el término $A_1 b_1^t$ debe dominar finalmente la escena y hacer que la trayectoria sea divergente.

A la raíz con el valor *absoluto* más alto llamémosla *raíz dominante*. Entonces resulta que es la raíz dominante b_1 la que realmente imparte el tono de la trayectoria de tiempo, cuando menos respecto a su convergencia o divergencia últimas. Tal es realmente el caso. Por lo tanto podemos afirmar que *una trayectoria de tiempo será convergente —independientemente de cuáles sean las condiciones iniciales— si y sólo si la raíz dominante es menor que 1 en valor absoluto*. El lector puede verificar que este enunciado es válido para los casos en los cuales ambas raíces son mayores que o menores que 1 en valor absoluto (lo que se discutió anteriormente), y cuando una raíz tiene un valor absoluto de exactamente 1 (lo que *no* se discutió anteriormente). Observe, sin embargo, que aun cuando la convergencia final depende sólo de la raíz dominante, la raíz *no* dominante también va a ejercer una influencia definida sobre la trayectoria de tiempo, cuando menos en los periodos iniciales. Por lo tanto, la configuración exacta de y_t depende todavía de ambas raíces.

Si regresamos al caso de las raíces repetidas, encontramos que la función complementaria consiste en los términos $A_3 b^t$ y $A_4 t b^t$, como se muestra en (18.6). El primero ya nos es conocido, pero se necesita alguna explicación para el segundo, el cual incluye un factor t. Si $|b| > 1$, el término b^t será explosivo, y el factor t va a servir simplemente para intensificar la explosividad a medida que t se incrementa. Si $|b| < 1$, por otro lado, la parte b^t (que tiende a cero a medida que t se incrementa) y la parte t sigue una evolución contraria; es decir, el valor de t va a contrarrestar a b^t en lugar de reforzarlo. ¿Qué fuerza va a ser mayor? La respuesta es que la fuerza de amortiguamiento de b^t siempre va a aventajar a la fuerza explosiva de t. Por esta razón, el requerimiento básico para la convergencia en el caso de las raíces repetidas es todavía que la raíz sea menor de 1 en valor absoluto.

Ejemplo 7

Analice la convergencia de las soluciones de los ejemplos 3 y 4. Para el ejemplo 3, la solución es

$$y_t = 3 + (-2)^t + 4t$$

donde las raíces son 1 y -2, respectivamente $[3(1)^t = 3]$, y donde hay un equilibrio móvil $4t$. Siendo la raíz dominante -2, la trayectoria de tiempo es divergente.

Para el ejemplo 4, donde la solución es

$$y_t = A_3(-3)^t + A_4 t(-3)^t + \frac{1}{4}$$

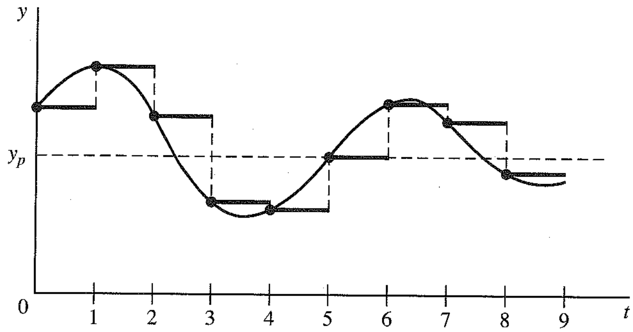
y donde $|b| = 3$, también tenemos divergencia.

Consideremos ahora el caso de las raíces complejas. A partir de la forma general de la función complementaria de (18.10),

$$y_c = R^t(A_5 \cos\theta t + A_6 \operatorname{sen}\theta t)$$

es claro que la expresión que está entre paréntesis, como la de (16.24′), va a producir un patrón de fluctuación de naturaleza periódica. Sin embargo, como la variable t sólo puede adoptar valores enteros 0, 1, 2, ... en el presente contexto, vamos a atrapar y utilizar sólo un subconjunto de los puntos de la gráfica de una función circular. El valor de y para cada uno de estos puntos siempre va a prevalecer para un periodo completo, hasta que se alcance el siguiente punto relevante. Como se ilustra en la figura 18.1, la trayectoria resultante no es ni el tipo

FIGURA 18.1



oscilatorio común (que no se alterna entre los valores arriba y abajo de y_p para periodos consecutivos), ni el tipo fluctuante común (no uniforme); en vez de esto, exhibe una especie de fluctuación *escalonada.* Sin embargo, por lo que toca a la convergencia, el factor decisivo es realmente el término R^t, el cual, al igual que el término e^{ht} de (16.24′), va a determinar si la fluctuación escalonada se intensifica o mitiga a medida que t se incremente. En el presente caso, la fluctuación puede estrecharse gradualmente si y sólo si $R < 1$. Ya que R es por definición el valor absoluto de las raíces complejas conjugadas $(h \pm vi)$, la condición de convergencia es nuevamente que las raíces características sean menores que la unidad en valor absoluto.

En resumen: para los tres casos de las raíces características, la trayectoria de tiempo va a converger en el equilibrio intertemporal (estacionaria o móvil) si y sólo si el valor absoluto de cada una de las raíces es menor de 1, independientemente de cuáles puedan ser las condiciones iniciales.

Ejemplo 8 ¿Son convergentes las trayectorias de tiempo (18.11) y (18.12)? En (18.11) tenemos $R = \frac{1}{2}$; por lo tanto la trayectoria de tiempo va a convergir al equilibrio estacionario (= 4). Por otro lado, en (18.12) tenemos $R = 4$, de modo que la trayectoria de tiempo no va a convergir al equilibrio (= 0).

EJERCICIO 18.1

1. Escriba la ecuación característica para cada una de las siguientes ecuaciones en diferencias, y encuentre las raíces características:
   (*a*) $y_{t+2} - y_{t+1} + \frac{1}{2}y_t = 2$ (*c*) $y_{t+2} + \frac{1}{2}y_{t+1} - \frac{1}{2}y_t = 5$
   (*b*) $y_{t+2} - 4y_{t+1} + 4y_t = 7$ (*d*) $y_{t+2} - 2y_{t+1} + 3y_t = 4$
2. Para cada una de las ecuaciones en diferencias del problema 1 enuncie, basándose en las raíces características, si la trayectoria de tiempo incluye la oscilación o la fluctuación escalonada, y si es explosiva.
3. Encuentre las soluciones particulares de las ecuaciones del problema 1. ¿Representan éstas un equilibrio estacionario o un equilibrio móvil?
4. Resuelva las siguientes ecuaciones en diferencias:
   (*a*) $y_{t+2} + 3y_{t+1} - \frac{7}{4}y_t = 9$ $(y_0 = 6;\ y_1 = 3)$
   (*b*) $y_{t+2} - 2y_{t+1} + 2y_t = 1$ $(y_0 = 3;\ y_1 = 4)$
   (*c*) $y_{t+2} - y_{t+1} + \frac{1}{4}y_t = 2$ $(y_0 = 4;\ y_1 = 7)$
5. Analice las trayectorias de tiempo obtenidas en el problema 4.

18.2 Modelo de interacción de multiplicador con acelerador de Samuelson

Como una ilustración del uso de las ecuaciones en diferencias de segundo orden en economía, citemos un trabajo clásico del profesor Paul Samuelson, el primer economista que ganó el premio Nobel. Nos referimos a su modelo clásico de *interacción*, que busca explorar el proceso dinámico de la determinación del ingreso cuando el principio de aceleración está en operación junto con el multiplicador Keynesiano.[3] Entre otras cosas, ese modelo sirve para demostrar que la sola interacción del multiplicador y del acelerador tiene la capacidad de generar fluctuaciones cíclicas en forma endógena.

El marco de referencia

Suponga que el ingreso nacional Y_t está compuesto de tres componentes de corriente de gasto: el consumo C_t, la inversión I_t y el gasto del gobierno G_t. El consumo se concibe como una función no del ingreso presente sino del ingreso del periodo anterior Y_{t-1}; por simplicidad, se supone que C_t es estrictamente proporcional a Y_{t-1}. La inversión, que es de la variedad "inducida", es una función de la tendencia prevaleciente del gasto del consumidor. Por supuesto que es a través de esta inversión inducida como el principio de aceleración entra en el modelo. Específicamente, vamos a suponer que I_t conserva una relación fija con el incremento del consumo $\Delta C_{t-1} \equiv C_t - C_{t-1}$. Por otro lado, el tercer componente G_t se considera como exógeno; de hecho, vamos a suponerlo como constante y vamos a denotarlo simplemente como G_0.

Estas hipótesis pueden traducirse en el siguiente conjunto de ecuaciones:

$$
\begin{aligned}
Y_t &= C_t + I_t + G_0 \\
C_t &= \gamma Y_{t-1} & (0 < \gamma < 1) \\
I_t &= \alpha(C_t - C_{t-1}) & (\alpha > 0)
\end{aligned}
\tag{18.13}
$$

donde γ (la letra griega gamma) representa la propensión marginal al consumo y α representa el acelerador (abreviatura de *coeficiente de aceleración*). Observe que si se cancela la inversión inducida en el modelo, nos queda una ecuación en diferencias de primer orden que abarca al proceso del multiplicador dinámico (vea el ejemplo 2 de la sección 17.2). Sin embargo, si se incluye la inversión inducida, tenemos una ecuación en diferencias de segundo orden que ilustra la interacción del multiplicador y del acelerador.

En virtud de la segunda ecuación, podemos expresar I_t en términos del ingreso como sigue:

$$I_t = \alpha(\gamma Y_{t-1} - \gamma Y_{t-2}) = \alpha\gamma(Y_{t-1} - Y_{t-2})$$

Al sustituir esto y la ecuación de C_t en la primera ecuación en (18.13) y reordenando, el modelo puede condensare en la ecuación individual

$$Y_t - \gamma(1 + \alpha)Y_{t-1} + \alpha\gamma Y_{t-2} = G_0$$

o en forma equivalente (después de desplazar los subíndices hacia delante por dos periodos),

$$Y_{t+2} - \gamma(1 + \alpha)Y_{t+1} + \alpha\gamma Y_t = G_0 \tag{18.14}$$

Dado que ésta es una ecuación en diferencias lineal de segundo orden con coeficientes constantes y término constante, podemos resolverla por el método recién aprendido.

[3] Paul A. Samuelson, "Interactions between the Multiplier Analysis and the Principle of Acceleration", *Review of Economic Statistics*, mayo de 1939, pp. 75-78; reimpreso en American Economic Association, *Readings in Business Cycle Theory*, Richard D. Irwin, Inc., Homewood, Ill., 1944, pp. 261-269.

La solución

Como solución particular tenemos, por (18.2),

$$Y_p = \frac{G_0}{1-\gamma(1+\alpha)+\alpha\gamma} = \frac{G_0}{1-\gamma}$$

Podemos observar que la expresión $1/(1-\gamma)$ es simplemente el multiplicador que prevalecería en ausencia de la inversión inducida. Por lo tanto $G_0/(1-\gamma)$ —el elemento exógeno de gasto por el multiplicador— debe darnos el ingreso de equilibrio Y^* en el sentido de que este nivel de ingreso satisface la condición de equilibrio "ingreso nacional = gasto total" [ver (3.24)]. Sin embargo, siendo la solución particular del modelo, también nos da el ingreso $\overline{Y}$ de equilibrio intertemporal.

Respecto a la función complementaria, hay tres casos posibles. El caso 1 ($a_1^2 > 4a_2$), en el presente contexto, se caracteriza por

$$\gamma^2(1+\alpha)^2 > 4\alpha\gamma \qquad \text{o} \qquad \gamma(1+\alpha)^2 > 4\alpha$$

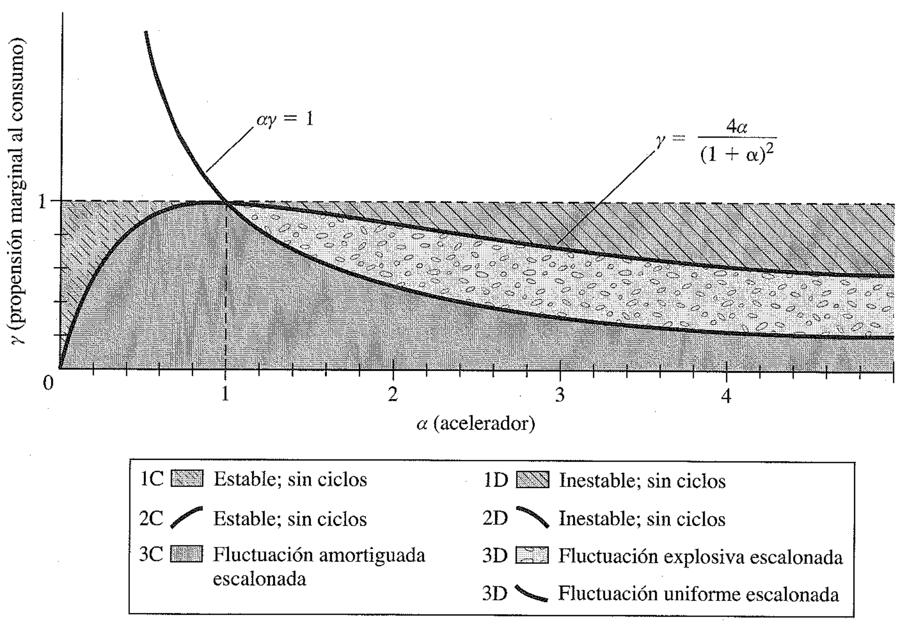
o sea

$$\gamma > \frac{4\alpha}{(1+\alpha)^2}$$

En forma similar, para caracterizar los casos 2 y 3, sólo necesitamos cambiar el signo > de la última desigualdad por = y <, respectivamente. En la figura 18.2 hemos dibujado la gráfica de la ecuación $\gamma = 4\alpha/(1+\alpha)^2$. De acuerdo con la discusión anterior, los pares (α, γ) que se localizan exactamente *sobre* esta curva pertenecen al caso 2. Por otro lado, los pares (α, γ) situados *arriba* de esta curva (que implica valores más altos de γ) tienen que ver con el caso 1 y los situados *debajo* de la curva con el caso 3.

Esta clasificación tripartita, con su representación gráfica en la figura 18.2, es de interés porque revela claramente las condiciones bajo las cuales las fluctuaciones cíclicas pueden sur-

FIGURA 18.2



gir en forma endógena de la interacción del multiplicador con el acelerador. Pero esto no nos dice nada acerca de la convergencia o la divergencia de la trayectoria de tiempo de Y. Por lo tanto, a nosotros nos toca distinguir, para cada caso, entre los subcasos *amortiguado* y *explosivo*. Por supuesto que podríamos tomar el camino fácil ilustrando simplemente estos subcasos al citar ejemplos numéricos específicos. Pero intentemos la tarea más gratificante, aun cuando más ardua, de encontrar las condiciones generales bajo las cuales prevalecen la convergencia y la divergencia.

Convergencia contra divergencia

La ecuación en diferencias (18.14) tiene la ecuación característica

$$b^2 - \gamma(1+\alpha)b + \alpha\gamma = 0$$

que suministra las dos raíces

$$b_1, b_2 = \frac{\gamma(1+\alpha) \pm \sqrt{\gamma^2(1+\alpha)^2 - 4\alpha\gamma}}{2}$$

Puesto que la cuestión de la convergencia contra la divergencia depende de los valores de b_1 y b_2, y como b_1 y b_2 a su vez dependen de los valores de los parámetros α y γ, las condiciones para convergencia y divergencia deben ser expresables en términos de los valores de α y γ. Para hacer esto, podemos usar el hecho de que —por (16.6)— las dos raíces características siempre se relacionan entre sí por las dos siguientes ecuaciones:

$$b_1 + b_2 = \gamma(1+\alpha) \tag{18.15}$$

$$b_1 b_2 = \alpha\gamma \tag{18.15'}$$

Basándonos en estas dos ecuaciones, podemos observar que

$$\begin{aligned}(1-b_1)(1-b_2) &= 1 - (b_1+b_2) + b_1 b_2 \\ &= 1 - \gamma(1+\alpha) + \alpha\gamma = 1 - \gamma\end{aligned} \tag{18.16}$$

En vista de la especificación del modelo de que $0 < \gamma < 1$, se hace necesario imponer sobre las dos raíces la condición

$$0 < (1-b_1)(1-b_2) < 1 \tag{18.17}$$

Examinemos ahora la cuestión de la convergencia para el caso 1, en el cual las raíces son reales y diferentes. Dado que por hipótesis α y γ son ambos positivos, (18.15′) nos dice que $b_1 b_2 > 0$, lo que implica que b_1 y b_2 poseen el mismo signo algebraico. Aún más, ya que $\gamma(1+\alpha) > 0$ (18.15) indica que tanto b_1 como b_2 deben ser positivos. Así, la trayectoria de tiempo Y_t no puede tener oscilaciones en el caso 1.

Aun cuando los signos de b_1 y b_2 se conocen ahora, para el caso 1 existen en realidad cinco combinaciones posibles de valores de (b_1, b_2), cada uno con su propia implicación respecto a los valores correspondientes de α y γ:

$$\begin{array}{lll}(i) & 0 < b_2 < b_1 < 1 & \Rightarrow \quad 0 < \gamma < 1;\ \alpha\gamma < 1 \\ (ii) & 0 < b_2 < b_1 = 1 & \Rightarrow \quad \gamma = 1 \\ (iii) & 0 < b_2 < 1 < b_1 & \Rightarrow \quad \gamma > 1 \\ (iv) & 1 = b_2 < b_1 & \Rightarrow \quad \gamma = 1 \\ (v) & 1 < b_2 < b_1 & \Rightarrow \quad 0 < \gamma < 1;\ \alpha\gamma > 1\end{array}$$

La posibilidad *i*, para la cual b_1 y b_2 son fracciones positivas, satisface debidamente la condición (18.17) y por lo tanto, está de acuerdo con la especificación del modelo $0 < \gamma < 1$. El producto de las dos raíces también debe ser una fracción positiva bajo esta posibilidad y mediante (18.15′), implica que $\alpha\gamma < 1$. En contraste, las tres siguientes posibilidades violan la condición (18.17) y son valores inadmisibles de γ (vea el ejercicio 18.2-3). Por lo tanto, deben descartarse. Pero la posibilidad v puede ser aceptable. Con b_1 y b_2 mayores que uno, (18.17) todavía puede satisfacerse si $(1 - b_1)(1 - b_2) < 1$. Pero esta vez tenemos $\alpha\gamma > 1$ (en vez de < 1) de (18.15′). La conclusión es que existen sólo dos subcasos admisibles para el caso 1. El primero —la posibilidad *i*— incluye las raíces fraccionarias b_1 y b_2, y por lo tanto suministra una trayectoria de tiempo convergente de Y. El otro subcaso —la posibilidad v— se caracteriza por raíces mayores que uno, y por lo tanto produce una trayectoria de tiempo divergente. Sin embargo, por lo que toca a los valores de α y γ, el asunto de la convergencia y la divergencia depende sólo de si $\alpha\gamma < 1$ o $\alpha\gamma > 1$. Esta información se resume en la parte superior de la tabla 18.1, donde el subcaso convergente se rotula como 1C, y el subcaso divergente como 1D.

El análisis del caso 2, con raíces repetidas, es de naturaleza similar. Las raíces son ahora $b = \gamma(1 + \alpha)/2$, con un signo positivo porque α y γ son positivos. Entonces, nuevamente no hay oscilación. Esta vez podemos clasificar el valor de b sólo de acuerdo con tres posibilidades:

$$\begin{array}{lll} (vi) & 0 < b < 1 & \Rightarrow \quad \gamma < 1; \alpha\gamma < 1 \\ (vii) & b = 1 & \Rightarrow \quad \gamma = 1 \\ (viii) & b > 1 & \Rightarrow \quad \gamma < 1; \alpha\gamma > 1 \end{array}$$

Para la posibilidad vi, $b\ (= b_1 = b_2)$ es una fracción positiva; entonces las implicaciones relativas a α y γ son totalmente idénticas con las de la posibilidad *i* para el caso 1. De manera análoga, la posibilidad *viii*, con $b\ (= b_1 = b_2)$ mayor que uno, puede satisfacer a (18.17) sólo si $1 < b < 2$; si es así, suministra los mismos resultados que la posibilidad v. Por otro lado, la posibilidad *vii* viola (18.17) y debe descartarse. Entonces de nuevo hay sólo dos subcasos admisibles. El primero —la posibilidad *vi*— suministra una trayectoria de tiempo convergente, mientras que el otro —la posibilidad *viii*— da una divergente. En términos de α y γ, los subcasos convergente y divergente se asocian de nuevo, respectivamente, con $\alpha\gamma < 1$ y $\alpha\gamma > 1$. Estos resultados se listan en la parte media de la tabla 18.1, donde los dos subcasos se rotulan 2C (convergente) y 2D (divergente).

TABLA 18.1 Casos y subcasos del modelo de Samuelson

Caso	Subcaso	Valores de α y γ	Trayectoria de tiempo Y_t
1. Raíces reales diferentes			
$\gamma > \dfrac{4\alpha}{(1+\alpha)^2}$	1C: $0 < b_2 < b_1 < 1$ 1D: $1 < b_2 < b_1$	$\alpha\gamma < 1$ $\alpha\gamma > 1$	No oscilatoria y no fluctuante
2. Raíces reales repetidas			
$\gamma = \dfrac{4\alpha}{(1+\alpha)^2}$	2C: $0 < b < 1$ 2D: $b > 1$	$\alpha\gamma < 1$ $\alpha\gamma > 1$	No oscilatoria y no fluctuante
3. Raíces complejas			
$\gamma < \dfrac{4\alpha}{(1+\alpha)^2}$	3C: $R < 1$ 3D: $R \geq 1$	$\alpha\gamma < 1$ $\alpha\gamma \geq 1$	Con fluctuación escalonada

Finalmente, en el caso 3, con raíces complejas, tenemos fluctuación escalonada y por lo tanto ciclos de negocios endógenos. En este caso, debemos ver el valor absoluto $R = \sqrt{a_2}$ [vea (18.8)] para la pista de la convergencia y la divergencia, donde a_2 es el coeficiente del término y_t en la ecuación en diferencias (18.1). En el presente modelo, tenemos $R = \sqrt{\alpha\gamma}$, lo que da lugar a las tres siguientes posibilidades:

$$\begin{aligned} &(ix) \quad R < 1 \quad \Rightarrow \quad \alpha\gamma < 1 \\ &(x) \quad R = 1 \quad \Rightarrow \quad \alpha\gamma = 1 \\ &(xi) \quad R > 1 \quad \Rightarrow \quad \alpha\gamma > 1 \end{aligned}$$

Aun cuando todos éstos resultan ser admisibles (vea el ejercicio 18.2-4), sólo la posibilidad $R < 1$ implica una trayectoria de tiempo convergente y califica como el subcaso 3C en la tabla 18.1. Las otras dos se rotulan colectivamente como el subcaso 3D.

En resumen, de la tabla 18.1 podemos concluir que puede presentarse una trayectoria de tiempo convergente si y sólo si $\alpha\gamma < 1$.

Un resumen gráfico

El análisis anterior nos ha conducido a una clasificación algo compleja de casos y subcasos. Nos ayudaría tener una representación visual del esquema de clasificación. Esto se ofrece en la figura 18.2.

El conjunto de todos los pares admisibles (α, γ) en el modelo se muestra en la figura 18.2 por el área rectangular sombreada de varias maneras. Dado que se excluyen los valores de $\gamma = 0$ y $\gamma = 1$, así como el valor $\alpha = 0$, el área sombreada es una especie de rectángulo sin lados. Ya hemos graficado la ecuación $\gamma = 4\alpha/(1+\alpha)^2$ para delimitar los tres casos principales de la tabla 18.1: los puntos sobre esa curva pertenecen al caso 2; los puntos situados al norte de la curva (que representan valores más altos de γ) pertenecen al caso 1; los situados al sur (con valores más bajos de γ) son del caso 3. Para distinguir entre los subcasos convergente y divergente, ahora añadimos la gráfica de $\alpha\gamma = 1$ (una hipérbola rectangular) como otra línea de demarcación. Los puntos situados al norte de esta hipérbola rectangular satisfacen la desigualdad $\alpha\gamma > 1$, mientras que los ubicados por debajo de aquella corresponden a $\alpha\gamma < 1$. Entonces podemos demarcar fácilmente los subcasos. Para el caso 1, la región sombreada con líneas discontinuas, que está por debajo de la hipérbola, corresponde al subcaso 1C, pero la región sombreada con líneas continuas se asocia con el subcaso 1D. Para el caso 2, que se relaciona con los puntos situados sobre la curva $\gamma = 4\alpha/(1+\alpha)^2$, el subcaso 2C cubre la parte con pendiente hacia arriba de esa curva, y el subcaso 2D las partes con pendiente hacia abajo. Finalmente, para el caso 3, la hipérbola rectangular sirve para separar la región sombreada llana (subcaso 3C) de la región sombreada y granulada (subcaso 3D). Debe observar que esta última también incluye los puntos localizados sobre la hipérbola rectangular misma, debido a la *desigualdad débil* en la especificación $\alpha\gamma \geq 1$.

Como la figura 18.2 es la depositaria de todas las conclusiones cualitativas del modelo, dado cualquier par ordenado (α, γ), siempre podemos encontrar gráficamente al subcaso correcto al graficar al par ordenado en el diagrama.

Ejemplo 1

Si el acelerador es 0.8 y la propensión marginal al consumo es 0.7, ¿qué tipo de trayectoria de tiempo de interacción va a resultar? El par ordenado (0.8, 0.7) se ubica en la región de gris liso, subcaso 3C; entonces la trayectoria de tiempo se caracteriza por la fluctuación escalonada amortiguada.

Ejemplo 2

¿Qué tipo de interacción está implicada con $\alpha = 2$ y $\gamma = 0.5$? El par ordenado (2, 0.5) está situado exactamente sobre la hipérbola rectangular, bajo el subcaso 3D. La trayectoria de tiempo de Y nuevamente exhibe fluctuación escalonada, pero ésta no va a ser ni explosiva ni amortiguada.

Por analogía con los casos de oscilación uniforme y fluctuación uniforme, podemos denominar esta situación como "fluctuación escalonada uniforme". Sin embargo, no podemos esperar en general que el aspecto de uniformidad en este último caso sea perfecto porque, en forma similar a lo que se hizo en la figura 18.1, podemos aceptar sólo aquellos puntos sobre una curva seno o coseno que correspondan a valores enteros de t, pero éstos pueden incidir sobre un conjunto de puntos totalmente diferentes sobre la curva para cada periodo de fluctuación.

EJERCICIO 18.2

1. Consulte la figura 18.2 y encuentre los subcasos a los que pertenecen los siguientes conjuntos de valores de α y γ, y describa en forma cualitativa la trayectoria de tiempo de interacción.
   (*a*) $\alpha = 3.5$; $\gamma = 0.8$ (*c*) $\alpha = 0.2$; $\gamma = 0.9$
   (*b*) $\alpha = 2$; $\gamma = 0.7$ (*d*) $\alpha = 1.5$; $\gamma = 0.6$
2. De los valores de α y γ dados en las partes (*a*) y (*c*) del problema 1, encuentre los valores numéricos de las raíces características en cada caso, y analice la naturaleza de la trayectoria de tiempo. ¿Concuerdan sus resultados con los obtenidos anteriormente?
3. Verifique que las posibilidades *ii*, *iii* y *iv* en el caso 1 implican valores inadmisibles de γ.
4. Muestre que en el caso 3 nunca podemos encontrar $\gamma \geq 1$.

18.3 La inflación y el desempleo en tiempo discreto

La interacción de la inflación y del desempleo, estudiada antes en el marco de referencia de tiempo continuo, también puede abordarse en tiempo discreto. Usando esencialmente las mismas hipótesis económicas, ilustraremos en esta sección la forma en que puede reformularse este modelo como un modelo de ecuación en diferencias.

El modelo

La formulación anterior en tiempo continuo (sección 16.5) consistía en tres ecuaciones diferenciales:

$$p = \alpha - T - \beta U + g\pi \qquad \text{[relación de Phillips con expectativas]} \tag{16.33}$$

$$\frac{d\pi}{dt} = j(p - \pi) \qquad \text{[expectativas adaptativas]} \tag{16.34}$$

$$\frac{dU}{dt} = -k(m - p) \qquad \text{[política monetaria]} \tag{16.35}$$

Están presentes tres variables endógenas: p (tasa real de inflación), π (tasa esperada de inflación) y U (tasa de desempleo). Aparecen hasta seis parámetros en el modelo; entre éstos, el parámetro m —la tasa de crecimiento del dinero nominal (o sea la tasa de expansión monetaria)— difiere de los demás en que su magnitud se establece como una decisión de política.

Cuando se vierte en el molde del análisis de periodos, la relación de Phillips (16.33) simplemente se transforma en

$$p_t = \alpha - T - \beta U_t + g\pi_t \qquad (\alpha, \beta > 0;\ 0 < g \leq 1) \tag{18.18}$$

En la ecuación de las expectativas adaptativas, la derivada debe reemplazarse por una expresión en diferencias:

$$\pi_{t+1} - \pi_t = j(p_t - \pi_t) \qquad (0 < j \leq 1) \tag{18.19}$$

Por la misma razón, la ecuación de política monetaria debe cambiarse a[4]

$$U_{t+1} - U_t = -k(m - p_{t+1}) \qquad (k > 0) \tag{18.20}$$

Estas tres ecuaciones constituyen la nueva versión del modelo de inflación-desempleo.

La ecuación en diferencias en p

Como primer paso en el análisis de este nuevo modelo, una vez más tratamos de condensar el modelo en una sola ecuación con una sola variable. Sea esa variable p. De acuerdo con esto, centraremos nuestra atención en (18.18). Sin embargo, ya que (18.18) —a diferencia de las otras dos ecuaciones— no describe por sí misma un patrón de cambio, depende de nosotros la creación de este patrón. Esto se logra al *obtener las diferencias de* p_t, es decir, al tomar la primera diferencia de p_t, de acuerdo con la definición

$$\Delta p_t \equiv p_{t+1} - p_t$$

Intervienen dos pasos en esto. Primero, desplazamos hacia delante los subíndices de tiempo en (18.18) un periodo, para obtener

$$p_{t+1} = \alpha - T - \beta U_{t+1} + g\pi_{t+1} \tag{18.18'}$$

Entonces restamos (18.18) de (18.18′), para obtener la primera diferencia de p_t que da el patrón deseado de cambio:

$$\begin{aligned} p_{t+1} - p_t &= -\beta(U_{t+1} - U_t) + g(\pi_{t+1} - \pi_t) \\ &= \beta k(m - p_{t+1}) + gj(p_t - \pi_t) \qquad \text{[mediante (18.20) y (18.19)]} \end{aligned} \tag{18.21}$$

Observe que en el segundo renglón de (18.21) los patrones de cambio de las otras dos variables tal como están dadas por (18.19) y (18.20) se han incorporado al patrón de cambio de la variable p. Entonces (18.21) incorpora ahora toda la información del presente modelo.

Sin embargo, el término π_t es ajeno al estudio de p y es necesario eliminarlo de (18.21). Con este fin, hacemos uso del hecho de que

$$g\pi_t = p_t - (\alpha - T) + \beta U_t \qquad \text{[mediante (18.18)]} \tag{18.22}$$

Al sustituir esto en (18.21) y reducir términos, obtenemos

$$(1 + \beta k)p_{t+1} - [1 - j(1 - g)]\, p_t + j\beta U_t = \beta km + j(\alpha - T) \tag{18.23}$$

Pero ahora aparece un término U_t que debe eliminarse. Para hacerlo, diferenciamos (18.23) para obtener un término $(U_{t+1} - U_t)$ y luego usar (18.20) para eliminar este último. Sólo después de este proceso más bien engorroso de sustituciones, obtenemos la ecuación en diferencias deseada sólo en la variable p, la cual, cuando se normaliza apropiadamente, adopta la forma

$$p_{t+2} - \underbrace{\frac{1 + gj + (1 - j)(1 + \beta k)}{1 + \beta k}}_{a_1} p_{t+1} + \underbrace{\frac{1 - j(1 - g)}{1 + \beta k}}_{a_2} p_t = \underbrace{\frac{j\beta km}{1 + \beta k}}_{c} \tag{18.24}$$

[4] Hemos supuesto que el cambio en U_t depende de $(m - p_{t+1})$, la tasa de crecimiento del dinero real en el periodo $(t + 1)$. Como alternativa, es posible hacerlo depender de la tasa de crecimiento del dinero real en el periodo t, $(m - p_t)$ (vea el ejercicio 18.3-4).

La trayectoria de tiempo de p

El valor de equilibrio intertemporal de p, dado por la integral particular de (18.24), es

$$\overline{p} = \frac{c}{1 + a_1 + a_2} = \frac{j\beta km}{\beta kj} = m \qquad \text{[mediante (18.2)]}$$

Por lo tanto, al igual que en el modelo de tiempo continuo, la tasa de inflación de equilibrio es exactamente igual a la tasa de expansión monetaria.

En cuanto a la función complementaria, pueden surgir ya sea raíces reales diferentes (caso 1), o raíces reales repetidas (caso 2), o raíces complejas (caso 3), dependiendo de la magnitud relativa de a_1^2 y $4a_2$. En el presente modelo,

$$a_1^2 \gtreqless 4a_2 \text{ si y sólo si } [1 + gj + (1 - j)(1 + \beta k)]^2 \gtreqless 4[1 - j(1 - g)](1 + \beta k) \qquad \textbf{(18.25)}$$

Si $g = \frac{1}{2}$, $j = \frac{1}{3}$ y $\beta k = 5$, por ejemplo, entonces $a_1^2 = (5\frac{1}{6})^2$ mientras que $4a_2 = 20$; por lo tanto resulta el caso 1. Pero si $g = j = 1$, entonces $a_1^2 = 4$ mientras que $4a_2 = 4(1 + \beta k) > 4$, y tenemos el caso 3. Sin embargo, en vista del mayor número de parámetros en el presente modelo, no es factible construir una gráfica de clasificación como la figura 18.2 en el modelo de Samuelson.

Sin embargo, todavía puede proceder el análisis de convergencia a lo largo de la misma línea que en la sección 18.2. Específicamente, recordamos (16.6) que las dos raíces características b_1 y b_2 deben satisfacer las dos siguientes relaciones:

$$b_1 + b_2 = -a_1 = \frac{1 + gj}{1 + \beta k} + 1 - j > 0 \qquad \textbf{(18.26)}$$

[vea (18.24)]

$$b_1 b_2 = a_2 = \frac{1 - j(1 - g)}{1 + \beta k} \in (0, 1) \qquad \textbf{(18.26')}$$

Aún más, en el presente modelo tenemos

$$(1 - b_1)(1 - b_2) = 1 - (b_1 + b_2) + b_1 b_2 = \frac{\beta jk}{1 + \beta k} > 0 \qquad \textbf{(18.27)}$$

Ahora considere el caso 1, donde las dos raíces b_1 y b_2 son reales y diferentes. Como el producto $b_1 b_2$ es positivo, b_1 y b_2 deben tener el mismo signo. Aún más, como la suma es positiva, b_1 y b_2 deben ser ambas positivas, lo que implica que no puede ocurrir oscilación. A partir de (18.27), podemos inferir que ni b_1 ni b_2 pueden ser iguales a uno; ya que de otro modo $(1 - b_1)(1 - b_2)$ sería cero, violando la desigualdad indicada. Esto significa que, en términos de las diferentes posibilidades de las combinaciones de (b_1, b_2) enumeradas en el modelo de Samuelson, las posibilidades *ii* y *iv* no pueden surgir aquí. También es inaceptable tener una raíz mayor que uno y la otra raíz menor que uno; puesto que de otra manera $(1 - b_1)(1 - b_2)$ sería negativo. Entonces también se descarta la posibilidad *iii*. Se sigue que b_1 y b_2 deben ser *ya sea* los dos mayores que uno, *o* ambos menores que uno. Si $b_1 > 1$ y $b_2 > 1$ (posibilidad *v*), sin embargo, se violaría (18.26'). Por esto lo único viable es la posibilidad *i*, con b_1 y b_2 siendo ambas fracciones positivas, de modo que la trayectoria de tiempo de p es convergente.

El análisis del caso 2 no es muy diferente. Mediante un razonamiento prácticamente idéntico, podemos concluir que la raíz repetida b sólo puede ser una fracción positiva en este modelo; es decir, la posibilidad *vi* es factible, pero no las posibilidades *vii* y *viii*. La trayectoria de tiempo de p en el caso 2 es nuevamente no oscilatoria y convergente.

Para el caso 3, la convergencia requiere que R (el valor absoluto de las raíces complejas) sea menor que uno. Por (18.8), $R = \sqrt{a_2}$. Siempre que a_2 sea una fracción positiva [vea (18.26′)], ciertamente tenemos $R < 1$. Entonces, la trayectoria de tiempo de p en el caso 3 también es convergente, aunque esta vez habrá una fluctuación escalonada.

El análisis de U

Si deseamos analizar esta vez la trayectoria de tiempo de la tasa de desempleo, podemos tomar (18.20) como el punto de partida. Para deshacernos del término p en esa ecuación, sustituimos primero (18.18′) para obtener

$$(1+\beta k)U_{t+1} - U_t = k(\alpha - T - m) + kg\pi_{t+1} \tag{18.28}$$

Enseguida, para prepararse para la sustitución de la otra ecuación, (18.19), al tomar la diferencia (18.28) encontramos que

$$(1+\beta k)U_{t+2} - (2+\beta k)U_{t+1} + U_t = kg(\pi_{t+2} - \pi_{t+1}) \tag{18.29}$$

En vista de la presencia de una expresión en diferencias en π a la derecha, podemos sustituirla por una versión desplazada hacia delante de la ecuación de expectativas adaptativas. El resultado de esto,

$$(1+\beta k)U_{t+2} - (2+\beta k)U_{t+1} + U_t = kgj(p_{t+1} - \pi_{t+1}) \tag{18.30}$$

es la incorporación de toda la información del modelo.

Sin embargo, debemos eliminar las variables p y π antes de que surja una ecuación en diferencias propia en U. Para este propósito, observamos de (18.20) que

$$kp_{t+1} = U_{t+1} - U_t + km \tag{18.31}$$

Aún más, al multiplicar (18.22) por $(-kj)$ y al desplazar los subíndices de tiempo, podemos escribir

$$\begin{aligned}
-kjg\pi_{t+1} &= -kjp_{t+1} + kj(\alpha - T) - \beta kjU_{t+1} \\
&= -j(U_{t+1} - U_t + km) + kj(\alpha - T) - \beta kjU_{t+1} \\
&\qquad\qquad\qquad\qquad\qquad\qquad [\text{por } (18.31)] \\
&= -j(1+\beta k)U_{t+1} + jU_t + kj(\alpha - T - m)
\end{aligned} \tag{18.32}$$

Estos dos resultados expresan a p_{t+1} y π_{t+1} en términos de la variable U y por lo tanto nos permite obtener, al sustituir en (18.30) —¡por fin!— la ecuación en diferencias deseada sólo en la variable U:

$$\begin{aligned}
U_{t+2} - \frac{1+gj+(1-j)(1+\beta k)}{1+\beta k}U_{t+1} + \frac{1-j(1-g)}{1+\beta k}U_t \\
= \frac{kj[\alpha - T - (1-g)m]}{1+\beta k}
\end{aligned} \tag{18.33}$$

Vale la pena observar que los dos coeficientes constantes de la izquierda (a_1 y a_2) son idénticos a los de la ecuación en diferencias para p [es decir, (18.24)]. Como resultado, el análisis anterior de la función complementaria de la trayectoria p debe ser igualmente aplicable al contexto presente. Pero el término constante de la derecha de (18.33) en realidad difiere del de (18.24). En consecuencia, las soluciones particulares serán diferentes en las dos situaciones. Esto es como debe ser, ya que, aparte de las coincidencias, no hay una razón inherente para esperar que la tasa de equilibrio intertemporal del desempleo sea la misma que la tasa de equilibrio de la inflación.

La relación de Phillips de largo plazo

Se verifica rápidamente que la tasa de equilibrio intertemporal del desempleo es

$$\overline{U} = \frac{1}{\beta}[\alpha - T - (1-g)m]$$

Pero como se halló que la tasa de inflación de equilibrio es $\overline{p} = m$, podemos enlazar a $\overline{U}$ con $\overline{p}$ mediante la ecuación

$$\overline{U} = \frac{1}{\beta}[\alpha - T - (1-g)\overline{p}] \qquad \textbf{(18.34)}$$

Dado que esta ecuación tiene que ver sólo con las tasas de desempleo y de inflación de *equilibrio*, se dice que ilustra la relación de Phillips *de largo plazo*.

Un caso especial de (18.34) ha recibido mucha atención entre los economistas: el caso de $g = 1$. Si $g = 1$, el término $\overline{p}$ tendrá coeficiente cero y por tanto sale de la escena. En otras palabras, $\overline{U}$ va a transformarse en una función constante de $\overline{p}$. En el diagrama estándar de Phillips, donde la tasa de desempleo se grafica en el eje horizontal, este resultado da lugar a una curva vertical de Phillips de largo plazo. El valor de $\overline{U}$ en este caso, denominado *tasa natural de desempleo*, es consistente con cualquier tasa de inflación de equilibrio, con la notable implicación de política de que, a largo plazo, no hay intercambios entre los demonios gemelos de la inflación y del desempleo tal como sucede a corto plazo.

Pero, ¿qué ocurre si $g < 1$? En ese caso, el coeficiente de $\overline{p}$ de (18.34) será negativo. Entonces, la curva de Phillips de largo plazo tendrá pendiente hacia abajo, suministrando con ello todavía una relación de intercambio entre inflación y desempleo. Ya sea que la curva de Phillips de largo plazo tenga pendiente vertical o negativa, depende en forma crítica por lo tanto del valor del parámetro g, el cual, de acuerdo con la relación de Phillips con expectativas, mide hasta qué punto la tasa de inflación esperada logra influir en la estructura de los salarios y en la tasa de inflación real. Todo esto puede parecerle conocido. Esto se debe a que estudiamos el tema en el ejemplo 1 de la sección 16.5 y también ha trabajado en esto en el ejercicio 16.5-4.

EJERCICIO 18.3

1. Realice los pasos intermedios que llevan de (18.23) a (18.24).
2. Muestre que si el modelo estudiado en esta sección se condensa en una ecuación en diferencias en la variable π, el resultado será el mismo que (18.24) excepto por la sustitución de π por p.
3. Hemos hallado que las trayectorias de tiempo de p y U en el modelo estudiado en esta sección son consistentemente convergentes. ¿Pueden surgir trayectorias de tiempo divergentes si anulamos la hipótesis de que $g \leq 1$? Si así es, ¿qué "posibilidades" divergentes en los casos 1, 2 y 3 se harán factibles ahora?
4. Conserve las ecuaciones (18.18) y (18.19), pero cambie (18.20) a
   $U_{t+1} - U_t = -k(m - p_t)$
   (*a*) Obtenga una nueva ecuación en diferencias en la variable p.
   (*b*) ¿Suministra la nueva ecuación en diferencias un $\overline{p}$ diferente?
   (*c*) Suponga que $j = g = 1$. Encuentre las condiciones para las cuales las raíces características se sitúan en los casos 1, 2 y 3, respectivamente.
   (*d*) Sea $j = g = 1$. Describa la trayectoria de tiempo de p (incluyendo la convergencia o la divergencia) cuando $\beta k = 3$, 4 y 5, respectivamente.

18.4 Generalizaciones a ecuaciones con términos variables y de orden superior

Estamos ahora listos para extender nuestros métodos en dos direcciones, al caso de términos variables y a las ecuaciones en diferencias de orden superior.

El término variable con forma de cm^t

Cuando el término constante c de (18.1) se reemplaza por un término variable —alguna función de t— el único efecto recaerá sobre la solución particular. (¿Por qué?). Para encontrar la nueva solución particular, podemos aplicar de nuevo el método de los coeficientes indeterminados. En el contexto de las ecuaciones diferenciales (sección 16.6), ese método requiere que el término variable y sus derivadas sucesivas juntas tomen sólo un número finito de tipos diferentes de expresiones, aparte de las constantes multiplicativas. Aplicado a las ecuaciones en diferencias, el requerimiento debe enmendarse para que diga: "el término variable y sus *diferencias* sucesivas deben tomar juntas sólo un número finito de tipos diferentes de expresiones, aparte de las constantes multiplicativas". Ilustremos este método con ejemplos concretos, tomando primero un término variable con forma cm^t, donde c y m son constantes.

Ejemplo 1

Encuentre la solución particular de

$$y_{t+2} + y_{t+1} - 3y_t = 7^t$$

Aquí tenemos $c = 1$ y $m = 7$. Evaluemos primero si el término variable 7^t suministra un número finito de tipos de expresiones de diferencias sucesivas. De acuerdo con la regla para tomar diferencias ($\Delta y_t = y_{t+1} - y_t$), la *primera* diferencia del término es

$$\Delta 7^t = 7^{t+1} - 7^t = (7-1)7^t = 6(7)^t$$

En forma similar, la *segunda* diferencia, $\Delta^2(7^t)$, puede expresarse como

$$\Delta(\Delta 7^t) = \Delta 6(7^t) = 6(7)^{t+1} - 6(7)^t = 6(7-1)7^t = 36(7)^t$$

Aún más, como puede verificarse, todas las diferencias sucesivas serán algún múltiplo de 7^t como la primera y la segunda. Como solamente hay un tipo de expresión, podemos intentar una solución $y_t = B(7)^t$ para la solución particular, donde B es un coeficiente indeterminado.

Al sustituir en la solución de prueba y sus versiones correspondientes los periodos $(t+1)$ y $(t+2)$ en la ecuación en diferencias dada, obtenemos

$$B(7)^{t+2} + B(7)^{t+1} - 3B(7)^t = 7^t \qquad \text{o} \qquad B(7^2 + 7 - 3)(7)^t = 7^t$$

Así,

$$B = \frac{1}{49 + 7 - 3} = \frac{1}{53}$$

y podemos escribir la solución particular como

$$y_p = B(7)^t = \frac{1}{53}(7)^t$$

Esto puede tomarse como un equilibrio móvil. Usted puede verificar la corrección de la solución al sustituirla en la ecuación en diferencias y comprobar que va a obtenerse una identidad, $7^t = 7^t$.

El resultado obtenido en el ejemplo 1 puede generalizarse fácilmente del término variable 7^t al de cm^t. De nuestra experiencia, esperamos que todas las diferencias sucesivas de cm^t

adopten la misma forma de expresión, es decir, Bm^t, donde B es una constante multiplicativa. Entonces podemos intentar una solución $y_t = Bm^t$ para la solución particular, cuando se da la ecuación en diferencias

$$y_{t+2} + a_1 y_{t+1} + a_2 y_t = cm^t \qquad \textbf{(18.35)}$$

Usando la solución de prueba $y_t = Bm^t$, que implica que $y_{t+1} = Bm^{t+1}$, etc., podemos expresar la ecuación (18.35) como

$$Bm^{t+2} + a_1 Bm^{t+1} + a_2 Bm^t = cm^t$$

es decir,
$$B(m^2 + a_1 m + a_2)m^t = cm^t$$

Entonces, el coeficiente B en la solución de prueba debe ser

$$B = \frac{c}{m^2 + a_1 m + a_2}$$

y la solución particular deseada de (18.35) puede escribirse como

$$y_p = Bm^t = \frac{c}{m^2 + a_1 m + a_2} m^t \qquad (m^2 + a_1 m + a_2 \neq 0) \qquad \textbf{(18.36)}$$

Observe que no se permite que el denominador de B sea cero. Si lo es,[5] entonces debemos usar en su lugar la solución de prueba $y_t = Btm^t$; o si ésta también falla, $y_t = Bt^2 m^t$.

El término variable con forma de ct^n

Consideremos ahora los términos variables con forma ct^n, donde c es cualquier constante y n es un entero positivo.

Ejemplo 2

Encuentre la solución particular de

$$y_{t+2} + 5y_{t+1} + 2y_t = t^2$$

Las tres primeras diferencias de t^2 (un caso especial de ct^n con $c = 1$ y $n = 2$) se encuentran como sigue:[6]

$$\begin{aligned}
\Delta t^2 &= (t+1)^2 - t^2 = 2t + 1 \\
\Delta^2 t^2 &= \Delta(\Delta t^2) = \Delta(2t+1) = \Delta 2t + \Delta 1 \\
&= 2(t+1) - 2t + 0 = 2 \qquad [\Delta \text{ constante} = 0] \\
\Delta^3 t^2 &= \Delta(\Delta^2 t^2) = \Delta 2 = 0
\end{aligned}$$

Como las diferencias adicionales sólo van a dar cero, hay en conjunto tres tipos diferentes de expresiones: t^2 (del término variable mismo), t y una constante (a partir de las diferencias sucesivas).

Por lo tanto intentemos la solución

$$y_t = B_0 + B_1 t + B_2 t^2$$

[5] Análoga a la situación del ejemplo 3 de la sección 16.6, esta eventualidad se materializa cuando la constante m es igual a una raíz característica de la ecuación en diferencias. Las raíces características de la ecuación en diferencias de (18.35) son los valores de b que satisfacen la ecuación $b^2 + a_1 b + a_2 = 0$. Si una raíz tiene el valor m, entonces debe seguir que $m^2 + a_1 m + a_2 = 0$.

[6] Estos resultados deben compararse con las tres primeras derivadas de t^2:

$$\frac{d}{dt}t^2 = 2t \qquad \frac{d^2}{dt^2}t^2 = 2 \qquad \text{y} \qquad \frac{d^3}{dt^3}t^2 = 0$$

para la solución particular, con coeficientes indeterminados B_0, B_1 y B_2. Observe que esta solución implica

$$\begin{aligned} y_{t+1} &= B_0 + B_1(t+1) + B_2(t+1)^2 \\ &= (B_0 + B_1 + B_2) + (B_1 + 2B_2)t + B_2t^2 \\ y_{t+2} &= B_0 + B_1(t+2) + B_2(t+2)^2 \\ &= (B_0 + 2B_1 + 4B_2) + (B_1 + 4B_2)t + B_2t^2 \end{aligned}$$

Cuando esto se sustituye en la ecuación en diferencias, obtenemos

$$(8B_0 + 7B_1 + 9B_2) + (8B_1 + 14B_2)t + 8B_2t^2 = t^2$$

Al igualar los dos lados término a término, vemos que se requiere que los coeficientes indeterminados satisfagan las siguientes ecuaciones simultáneas:

$$\begin{aligned} 8B_0 + 7B_1 + 9B_2 &= 0 \\ 8B_1 + 14B_2 &= 0 \\ 8B_2 &= 1 \end{aligned}$$

Entonces, los valores deben ser $B_0 = \frac{13}{256}$, $B_1 = -\frac{7}{32}$ y $B_2 = \frac{1}{8}$, dándonos la solución particular

$$y_p = \frac{13}{256} - \frac{7}{32}t + \frac{1}{8}t^2$$

Nuestra experiencia con el término variable t^2 debe permitirnos generalizar el método al caso de ct^n. En la nueva solución de prueba, obviamente debe estar el término B_nt^n, que corresponda al término variable dado. Aún más, ya que la diferenciación sucesiva del término suministra las expresiones diferentes $t^{n-1}, t^{n-2}, \ldots, t$ y B_0 (constante), la nueva solución de prueba para el caso del término variable ct^n debe escribirse como

$$y_t = B_0 + B_1t + B_2t^2 + \cdots + B_nt^n$$

Pero el resto del procedimiento es totalmente el mismo.

Debemos añadir que esta solución de prueba también puede fallar. En ese caso, el truco —ya empleado en incontables ocasiones— nuevamente es multiplicar la solución original de prueba por una potencia suficientemente alta de t, es decir, podemos intentar en su lugar $y_t = t(B_0 + B_1t + B_2t^2 + \cdots + B_nt^n)$, etcétera.

Ecuaciones en diferencias lineales de orden superior

El *orden* de una ecuación en diferencias indica la diferencia de mayor orden presente en la ecuación; pero también indica el número máximo considerado de periodos de desfasamiento de tiempo. Una ecuación en diferencias lineal de orden n (con coeficientes constantes y término constante) puede entonces escribirse en general como

$$y_{t+n} + a_1y_{t+n-1} + \cdots + a_{n-1}y_{t+1} + a_ny_t = c \qquad \textbf{(18.37)}$$

El método para encontrar la solución particular de esto no difiere de una manera sustantiva. Como inicio, podemos todavía intentar $y_t = k$ (el caso del equilibrio intertemporal estacionario). Si esto falla, intentamos entonces $y_t = kt$ o $y_t = kt^2$, etc., en ese orden.

Sin embargo, en la búsqueda de la función complementaria, vamos a confrontarnos con una ecuación característica que es una ecuación polinomial de grado n:

$$b^n + a_1b^{n-1} + \cdots + a_{n-1}b + a_n = 0 \qquad \textbf{(18.38)}$$

Ahora habrá n raíces características b_i $(i = 1, 2, \ldots, n)$, todo lo cual debe entrar en la función complementaria, por lo tanto:

$$y_c = \sum_{i=1}^{n} A_i b_i^t \qquad \textbf{(18.39)}$$

suponiendo que las raíces son todas reales y diferentes. En caso de que haya raíces reales repetidas (digamos, $b_1 = b_2 = b_3$), entonces los tres primeros términos de la suma de (18.39) deben modificarse como

$$A_1 b_1^t + A_2 t b_1^t + A_3 t^2 b_1^t \qquad [\text{ver } (18.6)]$$

Aún más, si existe un par de raíces complejas conjugadas —digamos, b_{n-1}, b_n— entonces, los dos últimos términos de la suma de (18.39) deben combinarse según la expresión

$$R^t (A_{n-1} \cos \theta t + A_n \operatorname{sen} \theta t)$$

También puede asignarse una expresión similar a cualquier otro par de raíces complejas. Sin embargo, en caso de dos pares repetidos, a uno de los dos debe dársele un factor multiplicativo de tR^t en lugar de R^t.

Después de haber encontrado y_p y y_c, la solución general de la ecuación en diferencias completa (81.37) se obtiene nuevamente mediante una suma, es decir

$$y_t = y_p + y_c$$

Pero ya que va a haber un total de n constantes arbitrarias en esta solución, requeriremos no menos de n condiciones iniciales para determinarlas.

Ejemplo 3

Encuentre la solución general de la ecuación en diferencias de tercer orden

$$y_{t+3} - \frac{7}{8} y_{t+2} + \frac{1}{8} y_{t+1} + \frac{1}{32} y_t = 9$$

Al intentar la solución $y_t = k$ se encuentra fácilmente que la solución particular es $y_p = 32$. En cuanto a la función complementaria, como la ecuación característica cúbica

$$b^3 - \frac{7}{8} b^2 + \frac{1}{8} b + \frac{1}{32} = 0$$

puede factorizarse en la forma

$$\left(b - \frac{1}{2}\right)\left(b - \frac{1}{2}\right)\left(b + \frac{1}{8}\right) = 0$$

las raíces son $b_1 = b_2 = \frac{1}{2}$ y $b_3 = -\frac{1}{8}$. Esto nos permite escribir

$$y_c = A_1 \left(\frac{1}{2}\right)^t + A_2 t \left(\frac{1}{2}\right)^t + A_3 \left(-\frac{1}{8}\right)^t$$

Observe que el segundo término contiene un multiplicativo t; esto se debe a la presencia de raíces repetidas. La solución general de la ecuación en diferencias dada es entonces simplemente la suma de y_c y y_p.

En este ejemplo, las tres raíces características resultan ser menores que 1 en valor absoluto. Por lo tanto podemos concluir que la solución obtenida representa una trayectoria de tiempo que converge al equilibrio estacionario en nivel de 32.

La convergencia y el teorema de Schur

Cuando tengamos una ecuación en diferencias de orden superior que no se resuelva fácilmente, podemos determinar la convergencia de la trayectoria de tiempo relevante en forma cualitativa sin tener que batallar con la solución cuantitativa real. El lector recordará que la trayectoria de tiempo puede convergir si y sólo si cada una de las raíces de la ecuación característica es menor que 1 en valor absoluto. En vista de esto, el siguiente teorema —conocido como el *teorema de Schur*—[7] se aplica directamente:

Las raíces de la ecuación polinomial de grado n

$$a_0 b^n + a_1 b^{n-1} + \cdots + a_{n-1} b + a_n = 0$$

serán todas menores que la unidad en valor absoluto si y sólo si los siguientes n determinantes

$$\Delta_1 = \left|\begin{array}{c:c} a_0 & a_n \\ \hdashline a_n & a_0 \end{array}\right| \qquad \Delta_2 = \left|\begin{array}{cc:cc} a_0 & 0 & a_n & a_{n-1} \\ a_1 & a_0 & 0 & a_n \\ \hdashline a_n & 0 & a_0 & a_1 \\ a_{n-1} & a_n & 0 & a_0 \end{array}\right| \qquad \cdots$$

$$\Delta_n = \left|\begin{array}{cccc:cccc} a_0 & 0 & \cdots & 0 & a_n & a_{n-1} & \cdots & a_1 \\ a_1 & a_0 & \cdots & 0 & 0 & a_n & \cdots & a_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n-1} & a_{n-2} & \cdots & a_0 & 0 & 0 & \cdots & a_n \\ \hdashline a_n & 0 & \cdots & 0 & a_0 & a_1 & \cdots & a_{n-1} \\ a_{n-1} & a_n & \cdots & 0 & 0 & a_0 & \cdots & a_{n-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_1 & a_2 & \cdots & a_n & 0 & 0 & \cdots & a_0 \end{array}\right|$$

son todos positivos.

Observe que, como la condición en el teorema está dada sobre la base de "si y sólo si", es una condición necesaria y suficiente. Entonces, el teorema de Schur es una contraparte perfecta de la ecuación en diferencias del teorema de Routh introducido antes en el marco de las ecuaciones diferenciales.

La construcción de estos determinantes se basa en un procedimiento sencillo. Esto se explica de la mejor manera con ayuda de las líneas segmentadas que forman las particiones de cada determinante para formar cuatro *áreas*. Cada área del determinante k-ésimo, Δ_k, siempre consiste de un subdeterminante $k \times k$. El área *superior izquierda* tiene sólo a_0 en la diagonal, ceros por arriba de la diagonal, y subíndices progresivamente mayores para los coeficientes sucesivos en cada columna que está debajo de los elementos de la diagonal. Cuando transponemos los elementos del área superior izquierda, obtenemos al área *inferior derecha*. Regresando al área *superior derecha*, colocamos ahora al coeficiente a_n solo en la diagonal, con ceros debajo de la diagonal, y subíndices progresivamente más pequeños para los coeficientes sucesivos a medida que subimos por cada colina desde la diagonal. Cuando se transponen los elementos de esta área, obtenemos el área *inferior izquierda*.

La aplicación de este teorema es directa. Dado que los coeficientes de la ecuación característica son los mismos que los que aparecen en el lado izquierdo de la ecuación en diferencias original, podemos introducirlos directamente en los determinantes citados. Observe que, en nuestro contexto, siempre tenemos $a_0 = 1$.

Ejemplo 4

¿Converge la trayectoria de tiempo de la ecuación $y_{t+2} + 3y_{t+1} + 2y_t = 12$? Aquí tenemos $n = 2$ y los coeficientes son $a_0 = 1$, $a_1 = 3$ y $a_2 = 2$. Entonces obtenemos

$$\Delta_1 = \begin{vmatrix} a_0 & a_2 \\ a_2 & a_0 \end{vmatrix} = \begin{vmatrix} 1 & 2 \\ 2 & 1 \end{vmatrix} = -3 < 0$$

[7] Para una discusión de este teorema y su historia, vea John S. Chipman, *The Theory of Inter-Sectoral Money Flows and Income Formation*, The John Hopkins Press, Baltimore, 1951, pp. 119-120.

Dado que esto ya viola la condición de convergencia, no hay necesidad de proseguir con Δ_2.

En realidad, las raíces características de la ecuación en diferencias dada se encuentran fácilmente como $b_1, b_2 = -1, -2$, lo que realmente implica una trayectoria de tiempo divergente.

Ejemplo 5 Pruebe la convergencia de la trayectoria de $y_{t+2} + \frac{1}{6}y_{t+1} - \frac{1}{6}y_t = 2$ mediante el teorema de Schur. Aquí los coeficientes son $a_1 = \frac{1}{6}$, $a_2 = -\frac{1}{6}$ (con $n = 2$). Entonces tenemos

$$\Delta_1 = \begin{vmatrix} a_0 & a_2 \\ a_2 & a_0 \end{vmatrix} = \begin{vmatrix} 1 & -\frac{1}{6} \\ -\frac{1}{6} & 1 \end{vmatrix} = \frac{35}{36} > 0$$

$$\Delta_2 = \begin{vmatrix} a_0 & 0 & a_2 & a_1 \\ a_1 & a_0 & 0 & a_2 \\ a_2 & 0 & a_0 & a_1 \\ a_1 & a_2 & 0 & a_0 \end{vmatrix} = \begin{vmatrix} 1 & 0 & -\frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & 1 & 0 & -\frac{1}{6} \\ -\frac{1}{6} & 0 & 1 & \frac{1}{6} \\ \frac{1}{6} & -\frac{1}{6} & 0 & 1 \end{vmatrix} = \frac{1\,176}{1\,296} > 0$$

Éstos realmente satisfacen la condición necesaria y suficiente para la convergencia.

EJERCICIO 18.4

1. Aplique la definición del símbolo de "diferencias" Δ, para encontrar:
   (*a*) Δt (*b*) $\Delta^2 t$ (*c*) Δt^3
   Compare los resultados de las diferencias con los de la diferenciación.
2. Encuentre la solución particular para cada una de las siguientes ecuaciones:
   (*a*) $y_{t+2} + 2y_{t+1} + y_t = 3^t$
   (*b*) $y_{t+2} - 5y_{t+1} - 6y_t = 2(6)^t$
   (*c*) $3y_{t+2} + 9y_t = 3(4)^t$
3. Encuentre las soluciones particulares de:
   (*a*) $y_{t+2} - 2y_{t+1} + 5y_t = t$
   (*b*) $y_{t+2} - 2y_{t+1} + 5y_t = 4 + 2t$
   (*c*) $y_{t+2} + 5y_{t+1} + 2y_t = 18 + 6t + 8t^2$
4. ¿Esperaría que, cuando el término variable adopte la forma $m^t + t^n$, la solución de prueba debe ser $B(m)^t + (B_0 + B_1 t + \cdots + B_n t^n)$? ¿Por qué?
5. Encuentre las raíces características y la función complementaria de:
   (*a*) $y_{t+3} - \frac{1}{2}y_{t+2} - y_{t+1} + \frac{1}{2}y_t = 0$
   (*b*) $y_{t+3} - 2y_{t+2} + \frac{5}{4}y_{t+1} - \frac{1}{4}y_t = 1$
   [*Sugerencia*: Intente factorizar $(b - \frac{1}{4})$ en ambas ecuaciones características.]
6. Pruebe la convergencia de las soluciones de las siguientes ecuaciones en diferencias por el teorema de Schur:
   (*a*) $y_{t+2} + \frac{1}{2}y_{t+1} - \frac{1}{2}y_t = 3$
   (*b*) $y_{t+2} - \frac{1}{9}y_t = 1$
7. En el caso de una ecuación en diferencias de tercer orden
   $y_{t+3} + a_1 y_{t+2} + a_2 y_{t+1} + a_3 y_t = c$
   ¿cuál es la forma exacta de los determinantes requeridos por el teorema de Schur?

Capítulo 19

Ecuaciones diferenciales y ecuaciones en diferencias simultáneas

Hasta aquí, nuestro estudio de la dinámica de la economía se ha limitado al análisis de una sola ecuación dinámica (diferencial o en diferencias). En este capítulo se introducen los métodos para el análisis de un sistema de ecuaciones dinámicas simultáneas. Dado que esto implicaría el manejo de diversas variables al mismo tiempo, puede esperar que surjan muchas complicaciones nuevas. Sin embargo, la verdad es que gran parte de lo que ya hemos aprendido sobre las ecuaciones dinámicas individuales puede ampliarse rápidamente a los sistemas de ecuaciones dinámicas simultáneas. Por ejemplo, la solución de un sistema dinámico consistiría todavía en un conjunto de integrales particulares o de soluciones particulares (los valores de equilibrio intertemporal de las diferentes variables) y de funciones complementarias (desviaciones respecto al equilibrio). Las funciones complementarias todavía se basarían en las ecuaciones reducidas, es decir, en las versiones homogéneas de las ecuaciones del sistema. Y la estabilidad dinámica del sistema todavía dependería de los signos (si es un sistema de ecuaciones diferenciales) o de los valores absolutos (si es un sistema de ecuaciones en diferencias) de las raíces características de las funciones complementarias. Entonces, el problema de un sistema dinámico es sólo un poco más complicado que el de una ecuación dinámica individual.

19.1 Génesis de los sistemas dinámicos

Hay dos formas generales mediante las cuales puede constituirse un sistema dinámico. Éste puede emanar de un *conjunto* dado de patrones interactuantes de cambio; o tal vez obtenerse de un *solo* patrón de cambio dado, siempre que este último consista en una ecuación dinámica de segundo orden (o superior).

Los patrones interactuantes del cambio

El caso más obvio de un conjunto dado de patrones interactuantes del cambio es el de un modelo multisectorial en el cual cada sector, como lo describe una ecuación dinámica, incide cuando menos sobre uno de los sectores. Una versión dinámica del modelo de insumo-producto, por ejemplo, podría incluir n industrias cuyos cambios en el producto ocasionan repercusiones dinámicas en las otras industrias. Entonces constituye un sistema dinámico. En forma similar, un modelo de mercado dinámico de equilibrio general incluiría n artículos que se interrelacionan por el ajuste de sus precios. Entonces hay nuevamente un sistema dinámico.

Sin embargo, los patrones interactuantes de cambio pueden encontrarse aun en un modelo de sector único. Las diferentes variables en un modelo como éste representan, no diferentes sectores ni diferentes artículos, sino diferentes *aspectos* de la economía. Pero pueden afectarse entre sí en su comportamiento dinámico, de modo que formen una red de interacciones.[1] De hecho, en el capítulo 18 hemos encontrado un ejemplo concreto de esto. En el modelo de inflación-desempleo, la tasa esperada de inflación π sigue un patrón de cambio, (18.19), que depende no sólo de π sino también de la tasa de desempleo U (a través de la tasa real de inflación p). En forma recíproca, el patrón de cambio de U, (18.20), depende de π (nuevamente a través de p). Entonces, el comportamiento dinámico de π y de U debe determinarse en forma simultánea. Por lo tanto, de manera retrospectiva, el modelo de inflación-desempleo pudo haberse tratado como un modelo dinámico de ecuaciones simultáneas. Y eso habría obviado la larga secuencia de sustituciones y eliminaciones que se llevaron a cabo para condensar el modelo en una sola ecuación y en una sola variable. Más adelante, en la sección 19.4, reformularemos ese modelo, considerándolo como un sistema dinámico. Mientras tanto, la noción de que el mismo modelo puede analizarse ya sea como una sola ecuación o como un sistema de ecuaciones suministra una pista natural para el estudio de la segunda manera de tener un sistema dinámico.

Transformación de una ecuación dinámica de orden superior

Supongamos que se nos da una ecuación diferencial (o en diferencias) de orden n en *una* variable. Entonces, como se va a mostrar, siempre es posible transformar esa ecuación en un sistema matemáticamente equivalente de n ecuaciones diferenciales (o en diferencias) simultáneas de *primer* orden con n variables. En particular, una ecuación diferencial de segundo orden puede reescribirse como dos ecuaciones diferenciales simultáneas de primer orden con dos variables.[2] Entonces, aun si resulta que comenzamos sólo con una ecuación dinámica (de orden superior), podemos obtener un sistema dinámico a través del mecanismo de la transformación matemática. Incidentalmente, este hecho tiene una implicación importante: en la discusión que sigue de los sistemas dinámicos, sólo tenemos que preocuparnos por los sistemas de ecuaciones de *primer* orden, ya que si está presente una ecuación de orden superior siempre podemos transformarla primero en un conjunto de ecuaciones de primer orden. Esto va a conducir a un número mayor de ecuaciones en el sistema, pero entonces el orden va a reducirse al mínimo.

Para ilustrar el procedimiento de transformación, consideremos la ecuación individual en diferencias

$$y_{t+2} + a_1 y_{t+1} + a_2 y_t = c \tag{19.1}$$

Si proponemos una nueva variable artificial x_t, definida por

$$x_t \equiv y_{t+1} \qquad \text{(lo que implica que } x_{t+1} \equiv y_{t+2}\text{)}$$

entonces podemos expresar la ecuación original de segundo orden mediante *dos* ecuaciones simultáneas de primer orden (con desfasamiento de un periodo) como sigue:

$$\begin{aligned} x_{t+1} \qquad + a_1 x_t + a_2 y_t &= c \\ y_{t+1} - \quad x_t \qquad\qquad &= 0 \end{aligned} \tag{19.1'}$$

[1] Observe que si tenemos dos ecuaciones dinámicas con las dos variables, y_1 y y_2, tales que el patrón de cambio de y_1 depende exclusivamente de y_1 mismo, y en forma similar para y_2, realmente no tenemos un sistema de ecuaciones simultáneas. En vez de esto, tenemos simplemente dos ecuaciones dinámicas separadas, cada una de las cuales puede analizarse por sí misma, sin ningún requerimiento de "simultaneidad".

[2] Inversamente, dos ecuaciones diferenciales (o en diferencias) de primer orden con dos variables pueden consolidarse en una ecuación individual de segundo orden con una variable, como lo hicimos en las secciones 16.5 y 18.3.

Se ve fácilmente que, siempre que se satisfaga la segunda ecuación (que define a la variable x_t), la primera es idéntica a la ecuación original dada. Mediante un procedimiento similar, y usando más variables artificiales, podemos transformar de manera similar una ecuación original de orden superior en un sistema equivalente de ecuaciones simultáneas de primer orden. Por ejemplo, usted puede verificar que la ecuación de tercer orden

$$y_{t+3} + y_{t+2} - 3y_{t+1} + 2y_t = 0 \tag{19.2}$$

puede expresarse como

$$\begin{aligned} w_{t+1} \qquad\qquad + w_t - 3x_t + 2y_t &= 0 \\ x_{t+1} \qquad - w_t \qquad\qquad &= 0 \\ y_{t+1} \qquad - x_t \qquad &= 0 \end{aligned} \tag{19.2'}$$

donde $x_t \equiv y_{t+1}$ (de modo que $x_{t+1} \equiv y_{t+2}$) y $w_t \equiv x_{t+1}$ (de modo que $w_{t+1} \equiv x_{t+2} \equiv y_{t+3}$).

Mediante un procedimiento similar, también podemos transformar una ecuación *diferencial* de orden n en un sistema de n ecuaciones de primer orden. Dada la ecuación diferencial de segundo orden

$$y''(t) + a_1 y'(t) + a_2 y(t) = 0 \tag{19.3}$$

por ejemplo, podemos introducir una nueva variable $x(t)$, definida como

$$x(t) \equiv y'(t) \qquad \left[\begin{matrix}\text{lo que} \\ \text{implica que}\end{matrix} \quad x'(t) \equiv y''(t)\right]$$

Entonces (19.3) puede reescribirse como el siguiente sistema de dos ecuaciones de primer orden:

$$\begin{aligned} x'(t) \qquad + a_1 x(t) + a_2 y(t) &= 0 \\ y'(t) - \quad x(t) \qquad\qquad &= 0 \end{aligned} \tag{19.3'}$$

donde usted puede observar que la segunda ecuación realiza la función de definición de la variable x recién introducida, como lo hizo la segunda ecuación de (19.1′). Esencialmente, también podemos usar el mismo procedimiento para transformar una ecuación diferencial de orden superior. La única modificación es que debemos introducir un número correspondientemente mayor de variables nuevas.

19.2 Solución de ecuaciones dinámicas simultáneas

Los métodos de solución de ecuaciones diferenciales simultáneas y de ecuaciones en diferencias simultáneas son bastante similares, por lo que las estudiaremos juntas en esta sección. Para nuestros propósitos, limitaremos la discusión sólo a ecuaciones lineales con coeficientes constantes.

Ecuaciones en diferencias simultáneas

Suponga que tenemos el siguiente sistema de ecuaciones en diferencias lineales:

$$\begin{aligned} x_{t+1} \qquad + 6x_t + 9y_t &= 4 \\ y_{t+1} - \quad x_t \qquad &= 0 \end{aligned} \tag{19.4}$$

¿Cómo encontramos las trayectorias de tiempo de x y y tales que se satisfagan ambas ecuaciones en este sistema? Nuestra tarea, nuevamente, es buscar las integrales particulares y las funciones complementarias, y sumarlas para obtener las trayectorias de tiempo deseadas de las dos variables.

Dado que las integrales particulares representan los valores de equilibrio intertemporales, vamos a denotarlos por $\overline{x}$ y $\overline{y}$. Como antes, es aconsejable probar primero las soluciones constantes, es decir, $x_{t+1} = x_t = \overline{x}$ y $y_{t+1} = y_t = \overline{y}$. Esto realmente funciona en el presente caso, ya que al sustituir estas soluciones de prueba en (19.4) obtenemos

$$\left.\begin{aligned} 7\overline{x} + 9\overline{y} &= 4 \\ -\overline{x} + \ \ \overline{y} &= 0 \end{aligned}\right\} \quad \Rightarrow \quad \overline{x} = \overline{y} = \frac{1}{4} \tag{19.5}$$

(Sin embargo, en caso de que estas soluciones constantes no funcionen, debemos probar soluciones de la forma $x_t = k_1 t$, $y_t = k_2 t$, etcétera.)

Para las funciones complementarias y apoyándonos en nuestra experiencia previa, debemos adoptar soluciones de prueba de la forma

$$x_t = mb^t \qquad \text{y} \qquad y_t = nb^t \tag{19.6}$$

donde m y n son constantes arbitrarias y la base b representa la raíz característica. Entonces esto implica automáticamente que

$$x_{t+1} = mb^{t+1} \qquad \text{y} \qquad y_{t+1} = nb^{t+1} \tag{19.7}$$

Observe que para simplificar las cosas empleamos la misma base $b \neq 0$ para ambas variables, aunque se permite que sus coeficientes difieran. Nuestro objetivo es encontrar los valores de b, m y n que puedan hacer que las soluciones de prueba (19.6) satisfagan la versión *reducida* (homogénea) de (19.4).

Al sustituir las soluciones de prueba en la versión reducida de (19.4) y cancelar al factor común $b^t \neq 0$, obtenemos las dos ecuaciones

$$\begin{aligned} (b+6)m + 9n &= 0 \\ -m + bn &= 0 \end{aligned} \tag{19.8}$$

Esto puede considerarse como un sistema de ecuaciones lineales homogéneas en las dos variables m y n, si queremos considerar a b como parámetro por el momento. Como el sistema (19.8) es homogéneo, puede arrojar sólo la solución trivial $m = n = 0$ si su matriz de coeficientes es no singular (véase la tabla 5.1 de la sección 5.5). En ese caso, ambas funciones complementarias de (19.6) serán idénticas a cero, lo que significa que x y y nunca se desvían de los valores de equilibrio intertemporal. Puesto que eso sería un caso especial poco interesante, trataremos de descartar esa solución trivial requiriendo que la matriz de coeficientes del sistema sea *singular*. Es decir, vamos a requerir que el determinante de esa matriz se anule:

$$\begin{vmatrix} b+6 & 9 \\ -1 & b \end{vmatrix} = b^2 + 6b + 9 = 0 \tag{19.9}$$

A partir de esta ecuación cuadrática, encontramos que $b(= b_1 = b_2) = -3$ es el único valor que puede evitar que m y n sean ambos iguales a cero en (19.8). Por lo tanto, usaremos sólo este valor de b. La ecuación (19.9) se llama la *ecuación característica*, y sus raíces son las *raíces características* del sistema dado de ecuaciones en diferencias simultáneas.

Una vez que tenemos un valor específico de b, (19.8) nos da los valores correspondientes de la solución para m y n. Sin embargo, al ser homogéneo el sistema, en realidad va a surgir un número infinito de soluciones para (m, n), expresables en la forma de una ecuación $m = kn$, donde k es una constante. De hecho, para cada raíz b_i habrá en general una ecuación diferente $m_i = k_i n_i$. Aun con raíces repetidas, con $b_1 = b_2$, debemos usar todavía dos ecuaciones de este tipo, $m_1 = k_1 n_1$ y $m_2 = k_2 n_2$ en las funciones complementarias. Aún más, con raíces repetidas, recordamos de (18.6) que las funciones complementarias deben escribirse como

$$\begin{aligned} x_t &= m_1(-3)^t + m_2 t(-3)^t \\ y_t &= n_1(-3)^t + n_2 t(-3)^t \end{aligned}$$

Los factores de proporcionalidad entre m_i y n_i deben satisfacer al sistema de ecuaciones dado (19.4), que exige que $y_{t+1} = x_t$, es decir,

$$n_1(-3)^{t+1} + n_2(t+1)(-3)^{t+1} - m_1(-3)^t + m_2 t(-3)^t$$

Al dividir entre $(-3)^t$, obtenemos

$$-3n_1 - 3n_2(t+1) = m_1 + m_2 t$$

o después de reordenar,

$$-3(n_1 + n_2) - 3n_2 t - m_1 + m_2 t$$

Al igualar los términos con t en ambos lados del signo de igualdad, y en forma similar para los términos sin t, encontramos

$$m_1 = -3(n_1 + n_2) \qquad \text{y} \qquad m_2 = -3n_2$$

Si ahora escribimos $n_1 = A_3$, $n_2 = A_4$, entonces se sigue que

$$m_1 = -3(A_3 + A_4) \qquad m_2 = -3A_4$$

Entonces, las funciones complementarias pueden escribirse como

$$\begin{aligned} x_c &= -3(A_3 + A_4)(-3)^t - 3A_4 t(-3)^t \\ &= -3A_3(-3)^t - 3A_4(t+1)(-3)^t \\ y_c &= A_3(-3)^t + A_4 t(-3)^t \end{aligned} \tag{19.10}$$

donde A_3 y A_4 son constantes arbitrarias. La solución general se da fácilmente al combinar las soluciones particulares de (19.5) con las funciones complementarias recién encontradas. Todo lo que resta es determinar las dos constantes arbitrarias A_3 y A_4 con la ayuda de condiciones iniciales o de frontera que sean apropiadas.

Un aspecto importante de la solución anterior es que, como ambas trayectorias de tiempo tienen expresiones idénticas b^t, ambas deben convergir o divergir. Esto tiene sentido, ya que en un modelo con variables dinámicamente interdependientes no puede prevalecer un equilibrio intertemporal general a menos que no esté presente el movimiento dinámico en ninguna parte del sistema. En el presente caso, con raíces repetidas $b = -3$, las trayectorias de tiempo de x y y presentarán oscilación explosiva.

Notación matricial

Con objeto de resaltar el paralelismo básico entre los métodos de solución de una ecuación individual y de un sistema de ecuaciones, desarrollamos la exposición anterior sin el beneficio de la notación matricial. Veamos ahora cómo puede utilizarse aquí esta última. Aun cuando llega a parecer superfluo aplicar la notación matricial a un sistema sencillo de sólo dos ecuaciones, la posibilidad de ampliar esa notación al caso de n ecuaciones debe hacer que este ejercicio valga la pena.

En primer lugar, el sistema dado (19.4) puede expresarse como

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{t+1} \\ y_{t+1} \end{bmatrix} + \begin{bmatrix} 6 & 9 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \end{bmatrix} \tag{19.4'}$$

o en forma más sucinta como

$$Iu + Kv = d \tag{19.4''}$$

donde I es la matriz identidad de 2×2; K es la matriz de 2×2 de los coeficientes de los términos x_t y y_t; y u, v y d son vectores columna que se definen como sigue:[3]

$$u = \begin{bmatrix} x_{t+1} \\ y_{t+1} \end{bmatrix} \qquad v = \begin{bmatrix} x_t \\ y_t \end{bmatrix} \qquad d = \begin{bmatrix} 4 \\ 0 \end{bmatrix}$$

El lector puede juzgar confuso un aspecto: dado que sabemos que $Iu = u$, ¿por qué no eliminar I? La respuesta es que, aun cuando ahora parezca redundante, será necesaria la matriz identidad en las operaciones subsiguientes, y por lo tanto vamos a retenerla como en (19.4″).

Cuando probamos las soluciones constantes $x_{t+1} = x_t = \overline{x}$ y $y_{t+1} = y_t = \overline{y}$ en las soluciones particulares, estamos estableciendo $u = v = \begin{bmatrix} \overline{x} \\ \overline{y} \end{bmatrix}$; esto reducirá (19.4″) a

$$(I + K) \begin{bmatrix} \overline{x} \\ \overline{y} \end{bmatrix} = d$$

Si existe el inverso $(I + K)^{-1}$, podemos expresar las soluciones particulares como

$$\begin{bmatrix} \overline{x} \\ \overline{y} \end{bmatrix} = (I + K)^{-1}\, d \tag{19.5'}$$

Esto es, por supuesto, una fórmula general, ya que es válida para cualquier matriz K y vector d siempre que exista $(I + K)^{-1}$. Aplicado a nuestro ejemplo numérico, tenemos

$$(I + K)^{-1}\, d = \begin{bmatrix} 7 & 9 \\ -1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 4 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{16} & -\frac{9}{16} \\ \frac{1}{16} & \frac{7}{16} \end{bmatrix} \begin{bmatrix} 4 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{4} \\ \frac{1}{4} \end{bmatrix}$$

por lo tanto, $\overline{x} = \overline{y} = \frac{1}{4}$, que concuerda con (19.5).

Haciendo referencia a las funciones complementarias, vemos que las soluciones de prueba (19.6) y (19.7) dan a los vectores u y v las formas específicas

$$u = \begin{bmatrix} mb^{t+1} \\ nb^{t+1} \end{bmatrix} = \begin{bmatrix} m \\ n \end{bmatrix} b^{t+1} \qquad \text{y} \qquad v = \begin{bmatrix} mb^t \\ nb^t \end{bmatrix} = \begin{bmatrix} m \\ n \end{bmatrix} b^t$$

Cuando se sustituyen en la ecuación reducida $Iu + Kv = 0$, estas soluciones de prueba van a transformar esta última en

$$I \begin{bmatrix} m \\ n \end{bmatrix} b^{t+1} + K \begin{bmatrix} m \\ n \end{bmatrix} b^t = 0$$

[3] El símbolo v aquí denota un vector. No lo confunda con la v de la notación de números complejos $h \pm vi$, donde representa un escalar.

o después de multiplicar por b^{-t} (un escalar) y de factorizar,

$$(bI + K)\begin{bmatrix} m \\ n \end{bmatrix} = 0 \qquad \textbf{(19.8')}$$

donde 0 es un vector cero. A partir de este sistema de ecuaciones homogéneas debemos encontrar los valores apropiados de b, m y n que deben usarse en las soluciones de prueba con objeto de hacer que este último sea determinado.

Para evitar las soluciones triviales para m y n, es necesario que

$$|bI + K| = 0 \qquad \textbf{(19.9')}$$

Y ésta es la ecuación característica que nos va a dar las raíces características b_i. Usted puede verificar que si sustituimos

$$bI = \begin{bmatrix} b & 0 \\ 0 & b \end{bmatrix} \qquad \text{y} \qquad K = \begin{bmatrix} 6 & 9 \\ -1 & 0 \end{bmatrix}$$

en esta ecuación, el resultado será precisamente (19.9), arrojando las raíces repetidas $b = -3$.

En general, a partir de (19.8′) cada raíz b_i generará un conjunto particular de un número infinito de valores de solución de m y n que están enlazados entre sí por la ecuación $m_i = k_i n_i$. Por lo tanto, es posible escribir, para cada valor de b_i,

$$n_i = A_i \qquad \text{y} \qquad m_i = k_i A_i$$

donde A_i son constantes arbitrarias que posteriormente van a determinarse. Cuando se sustituyen en las soluciones de prueba, estas expresiones para n_i y m_i junto con los valores de b_i van a conducir a formas específicas de funciones complementarias. Si todas las raíces son números reales diferentes, podemos aplicar (18.5) y escribir

$$\begin{bmatrix} x_c \\ y_c \end{bmatrix} = \begin{bmatrix} \Sigma m_i b_i^t \\ \Sigma n_i b_i^t \end{bmatrix} = \begin{bmatrix} \Sigma k_i A_i b_i^t \\ \Sigma A_i b_i^t \end{bmatrix}$$

Sin embargo, con raíces repetidas debemos aplicar (18.6) en su lugar, y como resultado las funciones complementarias van a contener términos con un multiplicador extra t, tal como $m_1 b^t + m_2 t b^t$ (para x_c) y $n_1 b^t + n_2 t b^t$ (para y_c). Los factores de proporcionalidad entre m_i y n_i debemos determinarlos mediante la relación entre las variables x y y tal como se estipula en el sistema de ecuaciones dado, como se ilustra en (19.10) en nuestro ejemplo numérico. Finalmente, en el caso de raíces complejas las funciones complementarias deben escribirse con (18.10) como su prototipo.

Por último, para obtener la solución general podemos simplemente formar la suma

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} x_c \\ y_c \end{bmatrix} + \begin{bmatrix} \overline{x} \\ \overline{y} \end{bmatrix}$$

Entonces, lo único que resta es determinar las constantes arbitrarias A_i.

La ampliación de este procedimiento a un sistema de n ecuaciones debe ser evidente. Sin embargo, cuando n es grande, puede no ser fácil resolver cuantitativamente la ecuación característica: una ecuación polinomial de grado n. En ese caso, podemos ver nuevamente que el teorema de Schur nos ayuda a realizar ciertas conclusiones cualitativas acerca de las trayectorias de tiempo de las variables en el sistema. Recordemos que a todas estas variables se les asigna la misma base b en las soluciones de prueba, de modo que deben llevarnos a las mismas expresiones b_i^t en las funciones complementarias y deben compartir las mismas propiedades de convergencia. Así, una sola aplicación del teorema de Schur nos permitirá determinar la convergencia o divergencia de la trayectoria de tiempo de cada una de las variables en el sistema.

Ecuaciones diferenciales simultáneas

El método de solución recién descrito también puede aplicarse a un sistema de ecuaciones *diferenciales* lineales de primer orden. Poco más o menos, la única modificación importante necesaria es cambiar las soluciones de prueba a

$$x(t) = me^{rt} \qquad \text{y} \qquad y(t) = ne^{rt} \tag{19.11}$$

lo que implica que

$$x'(t) = rme^{rt} \qquad \text{y} \qquad y'(t) = rne^{rt} \tag{19.12}$$

De acuerdo con nuestra convención sobre la notación, las raíces características se denotan ahora por r en lugar de b.

Suponga que se nos da el siguiente sistema de ecuaciones:

$$\begin{aligned} x'(t) + 2y'(t) + 2x(t) + 5y(t) &= 77 \\ y'(t) + \, x(t) + 4y(t) &= 61 \end{aligned} \tag{19.13}$$

Primero, expresémosla en notación matricial como

$$Ju + Mv = g \tag{19.13'}$$

donde las matrices son

$$J = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \qquad u = \begin{bmatrix} x'(t) \\ y'(t) \end{bmatrix} \qquad M = \begin{bmatrix} 2 & 5 \\ 1 & 4 \end{bmatrix} \qquad v = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} \qquad g = \begin{bmatrix} 77 \\ 61 \end{bmatrix}$$

Observe que en vista de la aparición del término $2y'(t)$ en la primera ecuación de (19.13), tenemos que usar la matriz J en lugar de la matriz identidad I, como en (19.4″). Si J es no singular (de modo que existe J^{-1}), entonces podemos en cierto sentido *normalizar* (19.13′) al premultiplicar cada término por J^{-1}, para obtener

$$J^{-1}Ju + J^{-1}Mv = J^{-1}g \qquad \text{o sea} \qquad Iu + Kv = d$$
$$(K \equiv J^{-1}M;\ d \equiv J^{-1}g) \tag{19.13''}$$

Este nuevo formato es un duplicado exacto de (19.4″), aunque debemos recordar que los vectores u y v tienen significados totalmente diferentes en los dos contextos diferentes. En el desarrollo siguiente vamos a adherirnos a la formulación $Ju + Mv = g$ dada en (19.13′).

Para encontrar las integrales particulares intentemos las soluciones constantes $x(t) = \overline{x}$ y $y(t) = \overline{y}$, lo que implica que $x'(t) = y'(t) = 0$. Si estas soluciones son válidas, los vectores v y u se transformarán en $v = \begin{bmatrix} \overline{x} \\ \overline{y} \end{bmatrix}$ y $u = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, y (19.13′) va a reducirse a $Mv = g$. Entonces, la solución para $\overline{x}$ y $\overline{y}$ puede escribirse como

$$\begin{bmatrix} \overline{x} \\ \overline{y} \end{bmatrix} = \overline{v} = M^{-1}g \tag{19.14}$$

la cual debemos comparar con (19.5′). En términos numéricos, nuestro problema plantea las siguientes integrales particulares:

$$\begin{bmatrix} \overline{x} \\ \overline{y} \end{bmatrix} = \begin{bmatrix} 2 & 5 \\ 1 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 77 \\ 61 \end{bmatrix} = \begin{bmatrix} \frac{4}{3} & -\frac{5}{3} \\ -\frac{1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} 77 \\ 61 \end{bmatrix} = \begin{bmatrix} 1 \\ 15 \end{bmatrix}$$

Busquemos las funciones complementarias. A través de las soluciones de prueba sugeridas en (19.11) y (19.12), los vectores u y v se transforman en

$$u = \begin{bmatrix} m \\ n \end{bmatrix} re^{rt} \qquad \text{y} \qquad v = \begin{bmatrix} m \\ n \end{bmatrix} e^{rt}$$

La sustitución de éstas en la ecuación reducida

$$Ju + Mv = 0$$

arroja el resultado

$$J\begin{bmatrix} m \\ n \end{bmatrix} re^{rt} + M\begin{bmatrix} m \\ n \end{bmatrix} e^{rt} = 0$$

o, después de multiplicar todo por el escalar e^{-rt} y factorizar,

$$(rJ + M)\begin{bmatrix} m \\ n \end{bmatrix} = 0 \tag{19.15}$$

Usted debe comparar esto con (19.8′). Como nuestro objetivo es encontrar las soluciones *no triviales* para m y n (de modo que nuestras soluciones de prueba también serán no triviales), es necesario que

$$|rJ + M| = 0 \tag{19.16}$$

Al ser la analogía de (19.9′), esta última ecuación —la ecuación característica del sistema de ecuaciones dado— nos ofrecerá las raíces r_i que necesitamos. Entonces podemos encontrar los valores correspondientes (no triviales) de m_i y n_i.

En nuestro presente ejemplo, la ecuación característica es

$$|rJ + M| = \begin{vmatrix} r+2 & 2r+5 \\ 1 & r+4 \end{vmatrix} = r^2 + 4r + 3 = 0 \tag{19.16′}$$

con raíces $r_1 = -1, r_2 = -3$. Sustituyéndolas en (19.15), obtenemos

$$\begin{bmatrix} 1 & 3 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} m_1 \\ n_1 \end{bmatrix} = 0 \qquad (\text{para } r_1 = -1)$$

$$\begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} m_2 \\ n_2 \end{bmatrix} = 0 \qquad (\text{para } r_2 = -3)$$

Se sigue que $m_1 = -3n_1$ y $m_2 = -n_2$, que también podemos expresar como

$$\begin{aligned} m_1 &= 3A_1 \qquad & \text{y} \qquad m_2 &= A_2 \\ n_1 &= -A_1 & n_2 &= -A_2 \end{aligned}$$

Ahora que han sido encontradas r_i, m_i y n_i, las funciones complementarias podemos escribirlas como las siguientes combinaciones lineales de expresiones exponenciales:

$$\begin{bmatrix} x_c \\ y_c \end{bmatrix} = \begin{bmatrix} \Sigma m_i e^{r_i t} \\ \Sigma n_i e^{r_i t} \end{bmatrix} \qquad [\text{raíces reales diferentes}]$$

Y la solución general surgirá en la forma

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} x_c \\ y_c \end{bmatrix} + \begin{bmatrix} \overline{x} \\ \overline{y} \end{bmatrix}$$

En este ejemplo, la solución es

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} 3A_1e^{-t} + A_2e^{-3t} + 1 \\ -A_1e^{-t} - A_2e^{-3t} + 15 \end{bmatrix}$$

Aún más, si se nos dan las condiciones iniciales $x(0) = 6$ y $y(0) = 12$, podemos encontrar que las constantes arbitrarias son $A_1 = 1$ y $A_2 = 2$. Éstas van a servir para hacer definitiva la solución anterior.

Una vez más podemos observar que, ya que las expresiones $e^{r_i t}$ son compartidas por ambas trayectorias de tiempo $x(t)$ y $y(t)$, estas últimas deben convergir o divergir. Siendo las raíces -1 y -3 en el presente caso, ambas trayectorias de tiempo convergen en sus equilibrios respectivos, es decir, $\overline{x} = 1$ y $\overline{y} = 15$.

Aun cuando nuestro ejemplo consiste sólo en un sistema con dos ecuaciones, el método se amplía al sistema general de n ecuaciones. Cuando n es grande, las soluciones cuantitativas pueden ser difíciles, pero una vez que se encuentra la solución característica, siempre será posible un análisis cualitativo empleando el teorema de Routh.

Comentarios adicionales sobre la ecuación característica

El término "ecuación característica" ya lo hemos encontrado en *tres* contextos separados: en la sección 11.3 hablamos de la ecuación característica de una matriz; en las secciones 16.1 y 18.1, este término se aplicó a una ecuación diferencial lineal individual y a una ecuación en diferencias; ahora, en esta sección, hemos introducido la ecuación característica de un sistema de ecuaciones lineales diferenciales o en diferencias. ¿Hay alguna conexión entre los tres?

Realmente la hay, y la conexión es muy cercana. En primer lugar, dada una ecuación individual y un sistema equivalente de ecuaciones —como lo ejemplifican la ecuación (19.1) y el sistema (19.1′), o la ecuación (19.3) y el sistema (19.3′)— las ecuaciones características deben ser idénticas. Como ilustración, considere la ecuación en diferencias (19.1), $y_{t+2} + a_1y_{t+1} + a_2y_t = c$. Antes aprendimos a escribir la ecuación característica al trasladar directamente los coeficientes constantes a una ecuación cuadrática:

$$b^2 + a_1b + a_2 = 0$$

¿Qué pasa con el sistema equivalente (19.1′)? Considerando que este sistema está en la forma $Iu + Kv = d$, como en (19.4″), tenemos la matriz $K = \begin{bmatrix} a_1 & a_2 \\ -1 & 0 \end{bmatrix}$. De modo que la ecuación característica es

$$|bI + K| = \begin{vmatrix} b + a_1 & a_2 \\ -1 & b \end{vmatrix} = b^2 + a_1b + a_2 = 0 \qquad \text{[por (19.9′)]} \qquad \textbf{(19.17)}$$

que es precisamente la misma que la obtenida a partir de la ecuación individual, como se aseveró. Naturalmente que el mismo tipo de resultado es válido también en el marco de referencia de las ecuaciones diferenciales, siendo la única diferencia que nosotros reemplazaríamos el símbolo b por el símbolo r de acuerdo con nuestra convención en este último marco de referencia.

También es posible enlazar la ecuación característica de un sistema de ecuaciones en diferencias (o diferenciales) con la de una matriz cuadrada específica, a la cual vamos a llamar D. Haciendo referencia a la definición de (11.14), pero usando el símbolo b (en lugar de r) para el marco de referencia de las ecuaciones en diferencias, podemos escribir la ecuación característica de la matriz D como sigue:

$$|D - bI| = 0 \tag{19.18}$$

En general, si multiplicamos cada uno de los elementos del determinante $|D - bI|$ por -1, el valor del determinante permanecerá sin cambios si la matriz D contiene un número *par* de filas (o de columnas), y cambiará de signo si D contiene un número *impar* de filas. Sin embargo en el presente caso, como $|D - bI|$ se igualará a cero, la multiplicación de cada uno de los elementos por -1 no tendrá ninguna consecuencia, independientemente de la dimensión de la matriz D. Pero la multiplicación de cada uno de los elementos del determinante $|D - bI|$ por -1 es equivalente a la multiplicación de la matriz $(D - bI)$ por -1 (vea el ejemplo 6 de la sección 5.3) antes de calcular su determinante. Entonces, (19.18) podemos expresarla como

$$|bI - D| = 0 \tag{19.18'}$$

Cuando igualamos esto con (19.17), se hace evidente que si escogemos la matriz $D = -K$, entonces su ecuación característica será idéntica a la del sistema (19.1′). Esta matriz, $-K$, tiene un significado especial: si consideramos la versión *reducida* del sistema, $Iu + Kv = 0$, y la expresamos en la forma de $Iu = -Kv$, o simplemente $u = -Kv$, vemos que $-K$ es la matriz que puede transformar al vector $v = \begin{bmatrix} x_t \\ y_t \end{bmatrix}$ en el vector $u = \begin{bmatrix} x_{t+1} \\ y_{t+1} \end{bmatrix}$ en esa ecuación específica.

Nuevamente, podemos adaptar el mismo razonamiento al sistema de ecuaciones diferenciales (19.3′). Sin embargo, en el caso de un sistema tal como (19.13′), $Ju + Mv = g$, donde —a diferencia del sistema (19.3′)— el primer término es Ju en vez de Iu, la ecuación característica está en la forma

$$|rJ + M| = 0 \qquad [\text{vea } (19.16')]$$

Para este caso, si deseamos encontrar la expresión para la matriz D, primero debemos normalizar la ecuación $Ju + Mv = g$ a la forma de (19.13″), y luego tomar $D = -K = -J^{-1}M$.

En resumen, dada (1) una ecuación individual diferencial o en diferencias, y (2) un sistema de ecuaciones equivalente, a partir del cual también podemos obtener (3) una matriz apropiada D, si tratamos de encontrar las ecuaciones características de estos tres, el resultado debe ser uno y el mismo.

EJERCICIO 19.2

1. Verifique que el sistema de ecuaciones en diferencias (19.4) es equivalente a la ecuación individual $y_{t+2} + 6y_{t+1} + 9y_t = 4$, que se resolvió anteriormente como ejemplo 4 en la sección 18.1. ¿Cómo se comparan las soluciones obtenidas por los dos métodos diferentes?
2. Muestre que la ecuación característica de la ecuación en diferencias (19.2) es idéntica a la del sistema equivalente (19.2′).
3. Resuelva los dos siguientes sistemas de ecuaciones en diferencias:

(*a*)
$$\begin{aligned} x_{t+1} \qquad + x_t + 2y_t &= 24 \\ y_{t+1} + 2x_t - 2y_t &= 9 \qquad (\text{con } x_0 = 10 \text{ y } y_0 = 9) \end{aligned}$$

(*b*)
$$\begin{aligned} x_{t+1} \qquad - x_t - \tfrac{1}{3}y_t &= -1 \\ x_{t+1} + y_{t+1} \qquad - \tfrac{1}{6}y_t &= 8\tfrac{1}{2} \qquad (\text{con } x_0 = 5 \text{ y } y_0 = 4) \end{aligned}$$

4. Resuelva los dos siguientes sistemas de ecuaciones diferenciales:

   (*a*) $x'(t) \quad - x(t) - 12y(t) = -60$

   $y'(t) + x(t) + 6y(t) = 36 \quad$ [con $x(0) = 13$ y $y(0) = 4$]

   (*b*) $x'(t) \quad - 2x(t) + 3y(t) = 10$

   $y'(t) - x(t) + 2y(t) = 9 \quad$ [con $x(0) = 8$ y $y(0) = 5$]

5. Basándose en el sistema de ecuaciones diferenciales (19.13), encuentre la matriz D cuya ecuación característica sea idéntica a la del sistema. Verifique que las ecuaciones características de los dos sean realmente la misma.

19.3 Modelos dinámicos de insumo-producto

Nuestro primer encuentro con el análisis de insumo-producto encaró la pregunta: ¿qué cantidad debe producirse en cada industria de modo que los requerimientos de insumos de todas las industrias, así como la demanda final (sistema abierto), se satisfagan con exactitud? El contexto era estático, y el problema consistía en resolver un sistema de ecuaciones simultáneas para los niveles de *equilibrio* de producto de todas las industrias. Cuando incorporamos al modelo ciertas consideraciones económicas adicionales, el sistema de insumo-producto puede asumir un carácter dinámico, por lo que ofrece un sistema de ecuaciones diferenciales o en diferencias del tipo estudiado en la sección 19.2.

Aquí vamos a considerar tres aspectos de dinámica económica. Sin embargo, para mantener sencilla la exposición, ilustraremos sólo con dos industrias en un sistema abierto. No obstante, ya que vamos a emplear notación matricial, no debe resultar difícil la generalización al caso de n industrias, ya que podemos lograrlo simplemente mediante el cambio apropiado de las dimensiones de las matrices involucradas. Para los propósitos de esta generalización, es aconsejable que denotemos las variables no como x_t y y_t, sino como $x_{1,t}$ y $x_{2,t}$, de modo que podamos hacer extensiva la notación a $x_{n,t}$ cuando sea necesario. Recordará que en el contexto de insumo-producto x_i representa al producto (medido en dólares) de la industria i-ésima; el nuevo subíndice t va a añadir ahora la dimensión del tiempo. El símbolo a_{ij} del coeficiente de insumos todavía representa el valor en dólares del articulo i-ésimo requerido en la producción de un monto en dólares del artículo j-ésimo, y d_i nuevamente indica la demanda final del artículo i-ésimo.

El desfasamiento de tiempo en la producción

En los sistemas abiertos estáticos de dos industrias, el producto de la industria I debe establecerse al nivel de la demanda como sigue:

$$x_1 = a_{11}x_1 + a_{12}x_2 + d_1$$

Supongamos ahora un desfasamiento de un periodo en la producción, de modo que la cantidad demandada para el periodo t determina no al producto presente sino al producto del periodo $(t+1)$. Para describir esta nueva situación, debemos modificar la ecuación anterior a la forma

$$x_{1,t+1} = a_{11}x_{1,t} + a_{12}x_{2,t} + d_{1,t} \tag{19.19}$$

En forma similar, podemos escribir para la industria II:

$$x_{2,t+1} = a_{21}x_{1,t} + a_{22}x_{2,t} + d_{2,t} \tag{19.19'}$$

Entonces tenemos ahora un sistema de ecuaciones simultáneas en diferencias; esto constituye una versión dinámica del modelo de insumo-producto.

En notación matricial, el sistema consiste en la ecuación

$$x_{t+1} - Ax_t = d_t \tag{19.20}$$

donde $x_{t+1} = \begin{bmatrix} x_{1,t+1} \\ x_{2,t+1} \end{bmatrix} \quad x_t = \begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} \quad A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad d_t = \begin{bmatrix} d_{1,t} \\ d_{2,t} \end{bmatrix}$

Es claro que (19.20) está en la forma de (19.4″), con sólo dos excepciones. Primero, a diferencia del vector u, el vector x_{t+1} no tiene una matriz identidad I como su "coeficiente". Sin embargo, como explicamos antes, esto realmente no constituye ninguna diferencia analítica. El segundo punto que es más sustantivo es que el vector d_t, con un subíndice de tiempo, implica que el vector de demanda final se considera como una función del tiempo. Si esta función es no constante, se requerirá una modificación del método para encontrar las soluciones particulares, aunque las funciones complementarias van a permanecer sin ser afectadas. El siguiente ejemplo ilustrará el procedimiento modificado.

Ejemplo 1 Dado el vector exponencial de demanda final

$$d_t = \begin{bmatrix} \delta^t \\ \delta^t \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \delta^t \qquad (\delta = \text{un escalar positivo})$$

encuentre las soluciones particulares del modelo dinámico de insumo-producto (19.20). De acuerdo con el método de coeficientes indeterminados introducido en la sección 18.4, debemos intentar soluciones de la forma $x_{1,t} = \beta_1 \delta^t$ y $x_{2,t} = \beta_2 \delta^t$, donde β_1 y β_2 son coeficientes indeterminados. Es decir, debemos pues, intentar

$$x_t = \begin{bmatrix} \beta_1 \delta^t \\ \beta_2 \delta^t \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \delta^t \tag{19.21}$$

lo que implica[4]

$$x_{t+1} = \begin{bmatrix} \beta_1 \delta^{t+1} \\ \beta_2 \delta^{t+1} \end{bmatrix} = \begin{bmatrix} \beta_1 \delta \\ \beta_2 \delta \end{bmatrix} \delta^t = \begin{bmatrix} \delta & 0 \\ 0 & \delta \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \delta^t$$

Si las soluciones de prueba indicadas son válidas, entonces el sistema (19.20) se transforma en

$$\begin{bmatrix} \delta & 0 \\ 0 & \delta \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \delta^t - \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \delta^t = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \delta^t$$

o, al cancelar el multiplicador escalar común $\delta^t \neq 0$,

$$\begin{bmatrix} \delta - a_{11} & -a_{12} \\ -a_{21} & \delta - a_{22} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{19.22}$$

[4] Observe que el vector $\begin{bmatrix} \beta_1 \delta \\ \beta_2 \delta \end{bmatrix}$ puede reescribirse en varias formas equivalentes:

$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \delta \quad \text{o} \quad \delta \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \quad \text{o} \quad \delta \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \delta & 0 \\ 0 & \delta \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

Escogemos la tercera alternativa porque en un paso subsiguiente vamos a sumar $\begin{bmatrix} \delta & 0 \\ 0 & \delta \end{bmatrix}$ a otra matriz de 2×2. Las dos primeras formas alternativas presentan problemas de conformabilidad de dimensiones.

Suponiendo que la matriz de coeficientes del extremo izquierdo es no singular, podemos encontrar rápidamente que los valores de β_1 y β_2 (mediante la regla de Cramer) son

$$\beta_1 = \frac{\delta - a_{22} + a_{12}}{\Delta} \qquad \text{y} \qquad \beta_2 = \frac{\delta - a_{11} + a_{21}}{\Delta} \tag{19.22'}$$

donde $\Delta \equiv (\delta - a_{11})(\delta - a_{22}) - a_{12}a_{21}$. Ya que ahora β_1 y β_2 se expresan completamente en términos de los valores conocidos de los parámetros, sólo necesitamos insertarlos en la solución de prueba (19.21) para obtener las expresiones definidas para las soluciones particulares.

El ejercicio 19.3-1 proporciona una versión más general del tipo de vector de demanda final estudiado aquí.

El procedimiento para encontrar las funciones complementarias de (19.20) no es diferente del presentado en la sección 19.2. Dado que la versión homogénea del sistema de ecuaciones es $x_{t+1} - Ax_t = 0$, la ecuación característica debe ser

$$|bI - A| = \begin{vmatrix} b - a_{11} & -a_{12} \\ -a_{21} & b - a_{22} \end{vmatrix} = 0 \qquad \text{[vea (19.9')]}$$

A partir de esto podemos encontrar las raíces características b_1 y b_2 y de ahí proseguir a los pasos restantes del proceso de solución.

La demanda excedente y el ajuste de la producción

La formulación del modelo en (19.20) también puede surgir a partir de una hipótesis económica diferente. Considere la situación en la cual la demanda excedente de cada producto siempre tiende a inducir un incremento del producto igual a la demanda excedente. Como la demanda excedente para el primer producto en el periodo t es de

$$\underbrace{a_{11}x_{1,t} + a_{12}x_{2,t} + d_{1,t}}_{\text{demandada}} - \underbrace{x_{1,t}}_{\text{ofertada}}$$

el ajuste del producto (incremento) $\Delta x_{1,t}$ debe hacerse exactamente igual a ese nivel:

$$\Delta x_{1,t}(\equiv x_{1,t+1} - x_{1,t}) = a_{11}x_{1,t} + a_{12}x_{2,t} + d_{1,t} - x_{1,t}$$

Sin embargo, si sumamos $x_{1,t}$ a ambos lados de esta ecuación, el resultado será idéntico a (19.19). En forma similar, nuestra hipótesis de ajuste del producto nos dará una ecuación igual a (19.19′) para la segunda industria. En resumen, podemos llegar al mismo modelo matemático a partir de hipótesis económicas totalmente diferentes.

Hasta ahora, el sistema de insumo-producto lo hemos considerado sólo en el marco de referencia de tiempo discreto. Para propósitos de comparación, vaciemos ahora al proceso de ajuste de la producción en el molde de tiempo continuo.

En general, esto requeriría el uso del símbolo $x_i(t)$ en lugar de $x_{i,t}$ y de la derivada $x_i'(t)$ en lugar de la diferencia $\Delta x_{i,t}$. Con estos cambios, nuestra hipótesis de ajuste de la producción se manifestará en el siguiente par de ecuaciones diferenciales:

$$\begin{aligned} x_1'(t) &= a_{11}x_1(t) + a_{12}x_2(t) + d_1(t) - x_1(t) \\ x_2'(t) &= a_{21}x_1(t) + a_{22}x_2(t) + d_2(t) - x_2(t) \end{aligned}$$

Para cualquier instante de tiempo $t = t_0$, el símbolo $x_i(t_0)$ nos da la tasa de flujo de producto por unidad de tiempo (digamos, por mes) que prevalece en dicho instante, y $d_i(t_0)$ indica la demanda final por mes que prevalece en ese instante. Entonces, la suma en el miembro derecho de cada ecuación indica la tasa de demanda excedente por mes, medida para $t = t_0$.

Por otro lado, la derivada $x_i'(t_0)$ a la izquierda, representa la tasa de ajuste de producto por mes determinada por la demanda excedente para $t = t_0$. Este ajuste va a erradicar la demanda excedente (y ocasionar el equilibrio) en el plazo de un mes, pero sólo si tanto la demanda excedente como el ajuste del producto permanecen sin cambios para las tasas presentes. En realidad, la demanda excedente variará con el tiempo, al igual que el ajuste inducido del producto, lo que conduce al juego del gato y del ratón. La solución del sistema, que consiste en las trayectorias de tiempo del producto x_i, es una crónica de este juego. Si la solución es convergente, el gato (ajuste del producto) finalmente podrá atrapar al ratón (demanda excedente) asintóticamente (cuando $t \to \infty$).

Después del ordenamiento apropiado, este sistema de ecuaciones diferenciales puede escribirse en el formato de (19.13′) como sigue:

$$Ix' + (I - A)x = d \qquad \textbf{(19.23)}$$

donde $x' = \begin{bmatrix} x_1'(t) \\ x_2'(t) \end{bmatrix} \quad x = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \quad A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad d = \begin{bmatrix} d_1(t) \\ d_2(t) \end{bmatrix}$

(la prima representa la derivada, no la transpuesta). Las funciones complementarias pueden encontrarse por el método estudiado antes. En particular, las raíces características deben encontrarse a partir de la ecuación

$$|rI + (I - A)| = \begin{vmatrix} r + 1 - a_{11} & -a_{12} \\ -a_{21} & r + 1 - a_{22} \end{vmatrix} = 0 \qquad \text{[vea (19.16)]}$$

En cuanto a las integrales particulares, si el vector de demanda final contiene funciones del tiempo no constantes $d_1(t)$ y $d_2(t)$ como sus elementos, será necesaria una modificación del método de solución. Ilustremos esto con un ejemplo sencillo.

Ejemplo 2

Dado el vector de demanda final

$$d = \begin{bmatrix} \lambda_1 e^{\rho t} \\ \lambda_2 e^{\rho t} \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} e^{\rho t}$$

donde λ_i y ρ son constantes, encuentre las integrales particulares del modelo dinámico (19.23). Usando el método de los coeficientes indeterminados, podemos intentar soluciones de la forma $x_i(t) = \beta_i e^{\rho t}$, que, por supuesto, implican que $x_i'(t) = \rho\beta_i e^{\rho t}$. En notación matricial, pueden escribirse como

$$x = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} e^{\rho t} \qquad \textbf{(19.24)}$$

y $$x' = \rho \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} e^{\rho t} = \begin{bmatrix} \rho & 0 \\ 0 & \rho \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} e^{\rho t} \qquad \text{[vea el pie de página del ejemplo 1]}$$

Al sustituir en (19.23) y cancelar el multiplicador escalar común $e^{\rho t}$ (diferente de cero), obtenemos

$$\begin{bmatrix} \rho & 0 \\ 0 & \rho \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} 1 - a_{11} & -a_{12} \\ -a_{21} & 1 - a_{22} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}$$

o sea

$$\begin{bmatrix} \rho + 1 - a_{11} & -a_{12} \\ -a_{21} & \rho + 1 - a_{22} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} \qquad \textbf{(19.25)}$$

Si la matriz de la extrema izquierda es no singular, podemos aplicar la regla de Cramer y determinar que los valores de los coeficientes β_i son

$$\begin{aligned} \beta_1 &= \frac{\lambda_1(\rho+1-a_{22})+\lambda_2 a_{12}}{\Delta} \\ \beta_2 &= \frac{\lambda_2(\rho+1-a_{11})+\lambda_1 a_{21}}{\Delta} \end{aligned} \tag{19.25'}$$

donde $\Delta \equiv (\rho+1-a_{11})(\rho+1-a_{22}) - a_{12}a_{21}$. Entonces habiendo encontrado los *coeficientes indeterminados*, podemos introducir estos valores en la solución de prueba (19.24) para obtener las integrales particulares deseadas.

La formación de capital

Otra consideración económica que puede dar lugar a un sistema dinámico de insumo-producto es la formación de capital, que incluye la acumulación del inventario.

En la discusión estática, consideramos sólo el nivel de producción de cada producto necesario para satisfacer la demanda presente. Las necesidades de acumulación de inventario o de formación de capital se ignoraron, o se incluyeron bajo el vector de demanda final. Para hacer aparecer la formación de capital, consideremos ahora —junto con la matriz de coeficientes de insumos $A = [a_{ij}]$— una matriz de coeficientes de capital

$$C = [c_{ij}] = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

donde c_{ij} denota el valor en dólares del artículo i-ésimo que necesita la industria j-ésima como capital nuevo (ya sea equipo o inventario, dependiendo de la naturaleza del artículo i-ésimo) como resultado de un incremento de \$1 en la producción de la industria j-ésima. Por ejemplo, si un incremento de \$1 en la producción de la industria (j-ésima) de los refrescos la induce a añadir \$2 en equipo de embotellado (artículo i-ésimo), entonces $c_{ij} = 2$. Este coeficiente de capital revela una especie de relación de capital marginal-producto, limitándose esta relación sólo a un tipo de capital (el artículo i-ésimo). Al igual que los coeficientes de insumo a_{ij}, se supone que los coeficientes de capital son fijos. La idea es que la economía produzca cada artículo en una cantidad tal que satisfaga no sólo la demanda de requerimiento de insumos más la demanda final, sino también la demanda de requerimiento de capital.

Si el tiempo es *continuo*, el incremento de producto se indica por las derivadas $x_i'(t)$; entonces, la producción de cada industria debe establecerse como

$$\begin{aligned} x_1(t) &= a_{11}x_1(t) + a_{12}x_2(t) + c_{11}x_1'(t) + c_{12}x_2'(t) + d_1(t) \\ x_2(t) &= \underbrace{a_{21}x_1(t) + a_{22}x_2(t)}_{\text{requerimiento de insumos}} + \underbrace{c_{21}x_1'(t) + c_{22}x_2'(t)}_{\text{requerimiento de capital}} + \underbrace{d_2(t)}_{\text{demanda final}} \end{aligned}$$

En notación matricial, esto se expresa mediante la ecuación

$$Ix = Ax + Cx' + d$$

o sea

$$Cx' + (A - I)x = -d \tag{19.26}$$

Si el tiempo es *discreto*, el requerimiento de capital en el periodo t se basará en el incremento de la producción $x_{i,t} - x_{i,t-1}$ $(\equiv \Delta x_{i,t-1})$; entonces, los niveles de producción deben establecerse como

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} = \underbrace{\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix}}_{\text{requerimiento de insumos}} + \underbrace{\begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} x_{1,t} - x_{1,t-1} \\ x_{2,t} - x_{2,t-1} \end{bmatrix}}_{\text{requerimiento de capital}} + \underbrace{\begin{bmatrix} d_{1,t} \\ d_{2,t} \end{bmatrix}}_{\text{demanda final}}$$

o sea
$$Ix_t = Ax_t + C(x_t - x_{t-1}) + d_t$$

Sin embargo, al desplazar el subíndice de tiempo hacia delante un periodo y al reducir términos podemos escribir la ecuación en la forma

$$(I - A - C)x_{t+1} + Cx_t = d_{t+1} \qquad \textbf{(19.27)}$$

El sistema de ecuaciones diferenciales (19.26) y el sistema de ecuaciones en diferencias (19.27) pueden resolverse nuevamente mediante el método de la sección 19.2. Esto también vale sin mencionar que estas dos ecuaciones matriciales son extensibles al caso n-ésimo de la industria simplemente mediante una redefinición apropiada de las matrices y un cambio correspondiente de sus dimensiones.

En lo anterior, hemos estudiado la forma en que un modelo dinámico de insumo-producto puede surgir de consideraciones tales como desfasamientos de tiempo y mecanismos de ajuste. Cuando se aplican consideraciones similares a los modelos de mercado de equilibrio general, éstos tienden a hacerse dinámicos de una manera muy parecida. Pero dado que la formulación de estos modelos tiene un espíritu análogo a los modelos de insumo-producto, evitaremos una discusión formal de esto y simplemente lo referiremos a los casos ilustrativos de los ejercicios 19.3-6 y 19.3-7.

EJERCICIO 19.3

1. En el ejemplo 1, si el vector de la demanda final se cambia a $d_t = \begin{bmatrix} \lambda_1 \delta^t \\ \lambda_2 \delta^t \end{bmatrix}$, ¿cuáles serán las soluciones particulares? Después de encontrar sus respuestas, muestre que para el ejemplo 1 son sólo un caso especial de éstas, con $\lambda_1 = \lambda_2 = 1$.
2. (*a*) Muestre que (19.22) puede escribirse en forma más concisa como $(\delta I - A)\beta = u$.
   (*b*) De los cinco símbolos que se usan, ¿cuáles son escalares?, ¿cuáles, vectores?, ¿cuáles, matrices?
   (*c*) Escriba la solución para β en forma matricial, suponiendo que $(\delta I - A)$ es no singular.
3. (*a*) Muestre que (19.25) puede escribirse en forma más concisa como $(\rho I + I - A)\beta = \lambda$.
   (*b*) ¿Cuáles de los cinco símbolos representan escalares, vectores y matrices, respectivamente?
   (*c*) Escriba la solución para β en forma matricial, suponiendo que $(\rho I + I - A)$ es no singular.
4. Dado $A = \begin{bmatrix} \frac{3}{10} & \frac{4}{10} \\ \frac{3}{10} & \frac{2}{10} \end{bmatrix}$ y $d_t = \begin{bmatrix} \left(\frac{12}{10}\right)^t \\ \left(\frac{12}{10}\right)^t \end{bmatrix}$ para el modelo de insumo-producto desfasado en la producción de tiempo discreto descrito en (19.20), encuentre (*a*) las soluciones particulares; (*b*) las funciones complementarias y (*c*) las trayectorias de tiempo definidas, suponiendo productos iniciales $x_{1,0} = \frac{187}{39}$ y $x_{2,0} = \frac{72}{13}$ (use fracciones en vez de decimales en todos los cálculos).

5. Dados $A = \begin{bmatrix} \frac{3}{10} & \frac{4}{10} \\ \frac{3}{10} & \frac{2}{10} \end{bmatrix}$ y $d = \begin{bmatrix} e^{t/10} \\ 2e^{t/10} \end{bmatrix}$ para el modelo de insumo-producto con ajuste de producto y tiempo continuo descrito en (19.23), encuentre (*a*) las integrales particulares; (*b*) las funciones complementarias; y (*c*) las trayectorias de tiempo definidas, suponiendo condiciones iniciales $x_1(0) = \frac{53}{6}$ y $x_2(0) = \frac{25}{6}$ (use fracciones en vez de decimales en todos los cálculos).
6. En un mercado con n artículos, todos los Q_{di} y Q_{si} (con $i = 1, 2, \ldots, n$) pueden considerarse como funciones de los n precios $P_1, \ldots, P_n$, así como la demanda excedente para cada artículo $E_i \equiv Q_{di} - Q_{si}$. Suponiendo linealidad, podemos escribir

$$\begin{aligned} E_1 &= a_{10} + a_{11}P_1 + a_{12}P_2 + \cdots + a_{1n}P_n \\ E_2 &= a_{20} + a_{21}P_1 + a_{22}P_2 + \cdots + a_{2n}P_n \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ E_n &= a_{n0} + a_{n1}P_1 + a_{n2}P_2 + \cdots + a_{nn}P_n \end{aligned}$$

o con notación matricial

$$E = a + AP$$

(*a*) ¿Qué representan estos últimos cuatro símbolos: escalares, vectores o matrices? ¿Cuáles son sus respectivas dimensiones?

(*b*) Considere que todos los precios son funciones del tiempo, y suponga que $dP_i/dt = \alpha_i E_i$ $(i = 1, 2, \ldots, n)$. ¿Cuál es la interpretación económica de este último conjunto de ecuaciones?

(*c*) Escriba las ecuaciones diferenciales que muestran que cada dP_i/dt es una función lineal de los n precios.

(*d*) Muestre que, si hacemos que P' denote al vector columna $n \times 1$ de las derivadas dP_i/dt, y si hacemos que α denote una matriz diagonal $n \times n$, con $\alpha_1, \alpha_2, \ldots, \alpha_n$ (en ese orden) en la diagonal principal y ceros en las demás posiciones, podemos escribir el anterior sistema de ecuaciones diferenciales con notación matricial como $P' - \alpha AP = \alpha a$.

7. Para el mercado de n artículos del problema 6, la versión en tiempo discreto consistiría en un conjunto de ecuaciones en diferencias $\Delta P_{i,t} = \alpha_i E_{i,t}$ $(i = 1, 2, \ldots, n)$, donde $E_{i,t} = a_{i0} + a_{i1}P_{1,t} + a_{i2}P_{2,t} + \cdots + a_{in}P_{n,t}$.

(*a*) Escriba el sistema de ecuaciones de demanda excedente y muestre que puede expresarse con notación matricial como $E_t = a + AP_t$.

(*b*) Muestre que las ecuaciones de ajuste de precios pueden escribirse como $P_{t+1} - P_t = \alpha E_t$, donde α es la matriz diagonal $n \times n$ definida en el problema 6.

(*c*) Muestre que el sistema de ecuaciones en diferencias del presente modelo en tiempo discreto puede expresarse en la forma $P_{t+1} - (I + \alpha A)P_t = \alpha a$.

19.4 Modelo de inflación-desempleo, una vez más

Habiendo ilustrado los sistemas dinámicos de tipo multisectorial con modelos de insumo-producto, suministraremos ahora un ejemplo en economía de las ecuaciones dinámicas simultáneas en el escenario de un sector individual. Para este propósito, podemos invocar una vez más al modelo de inflación-desempleo, el cual ya lo encontramos antes dos veces en dos modalidades diferentes.

Ecuaciones diferenciales simultáneas

En la sección 16.5 se presentó el modelo de inflación-desempleo en el marco de referencia de tiempo continuo vía las tres siguientes ecuaciones:

$$p = \alpha - T - \beta U + g\pi \qquad (\alpha, \beta > 0; 0 < g \leq 1) \tag{16.33}$$

$$\frac{d\pi}{dt} = j(p - \pi) \qquad (0 < j \leq 1) \tag{16.34}$$

$$\frac{dU}{dt} = -k(\mu - p) \qquad (k > 0) \tag{16.35}$$

excepto que hemos adoptado la letra griega μ (mi) aquí para reemplazar a m en (16.35) con objeto de evitar confusiones con el uso anterior del símbolo m en la discusión metodológica de la sección 19.2. En el tratamiento de este modelo en la sección 16.5, pues todavía no estábamos preparados para manejar ecuaciones dinámicas simultáneas, enfocamos el problema al condensar el modelo en una sola ecuación con una variable. Esto requirió un proceso muy laborioso de sustituciones y eliminaciones. Ahora, en vista de la coexistencia de dos patrones dados de cambio en el modelo para π y U, trataremos al modelo considerando dos ecuaciones diferenciales simultáneas.

Cuando (16.33) se sustituye en las otras dos ecuaciones, y las derivadas $d\pi/dt \equiv \pi'(t)$ y $dU/dt \equiv U'(t)$ se escriben de forma más sencilla como π' y U', el modelo adopta la forma

$$\begin{aligned} \pi' \quad + j(1-g)\pi + j\beta U &= j(\alpha - T) \\ U' - \qquad kg\pi + k\beta U &= k(\alpha - T - \mu) \end{aligned} \tag{19.28}$$

o con notación matricial,

$$\underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{J} \begin{bmatrix} \pi' \\ U' \end{bmatrix} + \underbrace{\begin{bmatrix} j(1-g) & j\beta \\ -kg & k\beta \end{bmatrix}}_{M} \begin{bmatrix} \pi \\ U \end{bmatrix} = \begin{bmatrix} j(\alpha - T) \\ k(\alpha - T - \mu) \end{bmatrix} \tag{19.28'}$$

A partir de este sistema, las trayectorias de tiempo de π y U pueden encontrarse en forma simultánea. Entonces, si se desea, podemos obtener la trayectoria p usando (16.33).

Trayectorias de solución

Para encontrar las integrales particulares, podemos tomar simplemente $\pi' = U' = 0$ (para hacer a π y a U estacionarias respecto al tiempo) en (19.28') y despejar π y a U. En nuestra discusión anterior, en (19.14), estas soluciones se obtuvieron a través de la inversión de matrices, pero también puede usarse la regla de Cramer. De cualquiera de las dos maneras, podemos encontrar que

$$\overline{\pi} = \mu \qquad \text{y} \qquad \overline{U} = \frac{1}{\beta}[\alpha - T - (1-g)\mu] \tag{19.29}$$

El resultado de que $\overline{\pi} = \mu$ (la tasa esperada de inflación en equilibrio es igual a la tasa de expansión monetaria) coincide con el alcanzado en la sección 16.5. Por lo que toca a la tasa de desempleo U, no intentamos encontrar su nivel de equilibrio en esa sección. Sin embargo, si lo hiciéramos (basándonos en la ecuación diferencial en U dada en el ejercicio 16.5-2), la respuesta no sería diferente de la solución $\overline{U}$ en (19.29).

Refiriéndonos a las funciones complementarias, que se basan en las soluciones de prueba me^{rt} y ne^{rt}, podemos determinar m, n y r a partir de la ecuación de la matriz reducida

$$(rJ + M)\begin{bmatrix} m \\ n \end{bmatrix} = 0 \qquad [\text{de } (19.15)]$$

la cual, en el presente contexto, adopta la forma

$$\begin{bmatrix} r+j(1-g) & j\beta \\ -kg & r+k\beta \end{bmatrix}\begin{bmatrix} m \\ n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{19.30}$$

Para evitar las soluciones triviales para m y n a partir de este sistema homogéneo, debemos hacer que se anule el determinante de la matriz de coeficientes; es decir, requerimos

$$|rJ + M| = r^2 + [k\beta + j(1-g)]r + k\beta j = 0 \tag{19.31}$$

Esta ecuación cuadrática, una versión específica de la ecuación característica $r^2 + a_1 r + a_2 = 0$, tiene coeficientes

$$a_1 = k\beta + j(1-g) \qquad \text{y} \qquad a_2 = k\beta j$$

Y éstos son precisamente los valores a_1 y a_2 en (16.37″), como es de esperarse: una versión del presente modelo en la variable π de una sola ecuación. Como resultado, el análisis previo de los tres casos de raíces características debemos aplicarlo aquí con igual validez. Entre otras conclusiones, podemos recordar que, independientemente de si las raíces son reales o complejas, la parte real de cada raíz en este modelo siempre es negativa. Por lo tanto, las trayectorias de solución siempre son convergentes.

Ejemplo 1 Encuentre las trayectorias de tiempo de π y U, dados los valores de los parámetros

$$\alpha - T = \frac{1}{6} \qquad \beta = 3 \qquad g = 1 \qquad j = \frac{3}{4} \qquad \text{y} \qquad k = \frac{1}{2}$$

Ya que los valores de estos parámetros son el doble de los del ejemplo 1 en la sección 16.5, los resultados del presente análisis pueden verificarse rápidamente respecto a los de dicha sección.

Primero, es fácil determinar que las integrales particulares son

$$\overline{\pi} = \mu \qquad \text{y} \qquad \overline{U} = \frac{1}{3}\left(\frac{1}{6}\right) = \frac{1}{18} \qquad [\text{por } (19.29)] \tag{19.32}$$

Siendo la ecuación característica

$$r^2 + \frac{3}{2}r + \frac{9}{8} = 0 \qquad [\text{por } (19.31)]$$

las dos raíces resultan ser complejas:

$$r_1, r_2 = \frac{1}{2}\left(-\frac{3}{2} \pm \sqrt{\frac{9}{4} - \frac{9}{2}}\right) = -\frac{3}{4} \pm \frac{3}{4}i \qquad \left(\text{con } h = -\frac{3}{4} \text{ y } v = \frac{3}{4}\right) \tag{19.33}$$

La sustitución de las dos raíces (junto con los valores de los parámetros) de (19.30) arroja, respectivamente, las ecuaciones matriciales

$$\begin{bmatrix} -\frac{3}{4}(1-i) & \frac{9}{4} \\ -\frac{1}{2} & \frac{3}{4}(1+i) \end{bmatrix}\begin{bmatrix} m_1 \\ n_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \left[\text{de } r_1 = -\frac{3}{4} + \frac{3}{4}i\right] \tag{19.34}$$

$$\begin{bmatrix} -\frac{3}{4}(1+i) & \frac{9}{4} \\ -\frac{1}{2} & \frac{3}{4}(1-i) \end{bmatrix}\begin{bmatrix} m_2 \\ n_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \left[\text{de } r_2 = -\frac{3}{4} - \frac{3}{4}i\right] \tag{19.34'}$$

Ya que r_1 y r_2 están calculados —vía (19.31)— para hacer que la matriz de coeficientes sea singular, cada una de las dos ecuaciones matriciales anteriores contiene en realidad sólo una ecuación independiente, lo que puede determinar una relación de proporcionalidad entre las constantes arbitrarias m_i y n_i. Específicamente, tenemos

$$\frac{1}{3}(1-i)m_1 = n_1 \qquad \text{y} \qquad \frac{1}{3}(1+i)m_2 = n_2$$

De acuerdo con esto, las funciones complementarias pueden expresarse como

$$\begin{aligned}\begin{bmatrix}\pi_c \\ U_c\end{bmatrix} &= \begin{bmatrix} m_1e^{r_1t} + m_2e^{r_2t} \\ n_1e^{r_1t} + n_2e^{r_2t}\end{bmatrix} \\ &= e^{ht}\begin{bmatrix} m_1e^{vit} + m_2e^{-vit} \\ n_1e^{vit} + n_2e^{-vit}\end{bmatrix} \qquad \text{[por (16.11)]} \\ &= e^{ht}\begin{bmatrix}(m_1+m_2)\cos vt + (m_1-m_2)i\,\text{sen}\,vt \\ (n_1+n_2)\cos vt + (n_1-n_2)i\,\text{sen}\,vt\end{bmatrix} \qquad \text{[por (16.24)]}\end{aligned}$$

Si, por sencillez en la notación, definimos constantes arbitrarias nuevas

$$A_5 \equiv m_1 + m_2 \qquad \text{y} \qquad A_6 \equiv (m_1 - m_2)i$$

entonces se sigue que[5]

$$n_1 + n_2 = \frac{1}{3}(A_5 - A_6) \qquad (n_1 - n_2)i = \frac{1}{3}(A_5 + A_6)$$

Así que, usando esto, e incorporando los valores de h y v de (19.33) en las funciones complementarias, terminamos con

$$\begin{bmatrix}\pi_c \\ \\ U_c\end{bmatrix} = e^{-3t/4}\begin{bmatrix} A_5\cos\frac{3}{4}t + A_6\,\text{sen}\,\frac{3}{4}t \\ \\ \frac{1}{3}(A_5 - A_6)\cos\frac{3}{4}t + \frac{1}{3}(A_5 + A_6)\,\text{sen}\,\frac{3}{4}t\end{bmatrix} \tag{19.35}$$

Finalmente, al combinar las integrales particulares en (19.32) con las funciones complementarias anteriores, podemos obtener las trayectorias de solución de π y de U. Como es de esperarse, estas trayectorias son exactamente las mismas que las de (16.43) y (16.45) de la sección 16.5.

Ecuaciones en diferencias simultáneas

El tratamiento de las ecuaciones simultáneas del modelo de inflación-desempleo en tiempo discreto es similar en su concepción a la discusión anterior en tiempo continuo. Por lo tanto sólo vamos a dar los lineamientos generales.

[5] Esto puede verse a partir de lo siguiente:

$$\begin{aligned} n_1 + n_2 &= \frac{1}{3}(1-i)m_1 + \frac{1}{3}(1+i)m_2 = \frac{1}{3}[(m_1+m_2) - (m_1-m_2)i] \\ &= \frac{1}{3}(A_5 - A_6) \\ (n_1 - n_2)i &= \left[\frac{1}{3}(1-i)m_1 - \frac{1}{3}(1+i)m_2\right]i = \frac{1}{3}[(m_1-m_2) - (m_1+m_2)i]i \\ &= \frac{1}{3}(A_6 + A_5) \qquad [i^2 \equiv -1]\end{aligned}$$

El modelo en cuestión, tal como está dado en la sección 18.3, consiste en tres ecuaciones, dos de las cuales describen a los patrones de cambio de π y U, respectivamente:

$$p_t = \alpha - T - \beta U_t + g\pi_t \tag{18.18}$$

$$\pi_{t+1} - \pi_t = j(p_t - \pi_t) \tag{18.19}$$

$$U_{t+1} - U_t = -k(\mu - p_{t+1}) \tag{18.20}$$

Eliminando p y agrupando términos, podemos reescribir el modelo como el sistema de ecuaciones en diferencias

$$\underbrace{\begin{bmatrix} 1 & 0 \\ -kg & 1+\beta k \end{bmatrix}}_{J} \begin{bmatrix} \pi_{t+1} \\ U_{t+1} \end{bmatrix} + \underbrace{\begin{bmatrix} -(1-j+jg) & j\beta \\ 0 & -1 \end{bmatrix}}_{K} \begin{bmatrix} \pi_t \\ U_t \end{bmatrix} = \begin{bmatrix} j(\alpha - T) \\ k(\alpha - T - \mu) \end{bmatrix} \tag{19.36}$$

Trayectorias de solución

Si hay equilibrio estacionario, las soluciones particulares de (19.36) podemos expresarlas como $\overline{\pi} = \pi_t = \pi_{t+1}$ y $\overline{U} = U_t = U_{t+1}$. Al sustituir $\overline{\pi}$ y $\overline{U}$ en (19.36), y al resolver el sistema (mediante la inversión de matrices o la regla de Cramer), obtenemos

$$\overline{\pi} = \mu \qquad \text{y} \qquad \overline{U} = \frac{1}{\beta}[\alpha - T - (1-g)\mu] \tag{19.37}$$

El valor de $\overline{U}$ es el mismo que el encontrado en la sección 18.3. Aunque no encontramos $\overline{\pi}$ en la última sección, la información en el ejercicio 18.3-2 indica que $\overline{\pi} = \mu$, lo que concuerda con (19.37). De hecho, puede observar que los resultados de (19.37) también son idénticos a los valores de equilibrio intertemporal obtenidos en el marco de referencia de tiempo continuo de (19.29).

La búsqueda de las funciones complementarias, basada esta vez en las soluciones de prueba mb^t y nb^t, incluye la ecuación matricial reducida

$$(bJ + K)\begin{bmatrix} m \\ n \end{bmatrix} = 0$$

o, en vista de (19.36),

$$\begin{bmatrix} b-(1-j+jg) & j\beta \\ -bkg & b(1+\beta k)-1 \end{bmatrix} \begin{bmatrix} m \\ n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{19.38}$$

Con objeto de evitar las soluciones triviales para este sistema homogéneo, requerimos

$$\begin{aligned} |bJ + K| &= (1+\beta k)b^2 - [1 + gj + (1-j)(1+\beta k)]b \\ &\quad + (1 - j + jg) = 0 \end{aligned} \tag{19.39}$$

La versión normalizada de esta ecuación cuadrática es la ecuación característica $b^2 + a_1 b + a_2 = 0$, con los mismos coeficientes a_1 y a_2 que en (18.24) y (18.33) de la sección 18.3. En consecuencia, el análisis de los tres casos de raíces características considerado en esa sección debe ser igualmente aplicable aquí.

Para cada raíz, b_i, (19.38) es una relación de proporcionalidad específica entre las constantes arbitrarias m_i y n_i, y esto nos permite enlazar las constantes arbitrarias de la función complementaria de U con las de la función complementaria para π. Entonces, al combinar las funciones complementarias con las soluciones particulares, podemos obtener las trayectorias de tiempo de π y de U.

EJERCICIO 19.4

1. Verifique (19.29) usando la regla de Cramer.
2. Verifique que surge la misma relación de proporcionalidad entre m_1 y n_1 independientemente de si usamos la primera o la segunda ecuación del sistema (19.34).
3. Encuentre las trayectorias de tiempo (soluciones generales) de π y de U, dados:

$$p = \frac{1}{6} - 2U + \frac{1}{3}\pi$$
$$\pi' = \frac{1}{4}(p - \pi)$$
$$U' = -\frac{1}{2}(\mu - p)$$

4. Encuentre las trayectorias de tiempo (soluciones generales) de π y U, dados

(*a*)
$$p_t = \frac{1}{2} - 3U_t + \frac{1}{2}\pi_t$$
$$\pi_{t+1} - \pi_t = \frac{1}{4}(p_t - \pi_t)$$
$$U_{t+1} - U_t = -(\mu - p_{t+1})$$

(*b*)
$$p_t = \frac{1}{4} - 4U_t + \pi_t$$
$$\pi_{t+1} - \pi_t = \frac{1}{4}(p_t - \pi_t)$$
$$U_{t+1} - U_t = -(\mu - p_{t+1})$$

19.5 Diagramas de fase de dos variables

En las secciones anteriores hemos expuesto las soluciones *cuantitativas* para los sistemas dinámicos *lineales*. En esta sección estudiaremos el análisis *gráfico-cuantitativo* (diagrama de fase) de un sistema de ecuaciones diferenciales *no lineales*. Más específicamente, nuestra atención se centrará en un sistema de ecuaciones diferenciales de primer orden con dos variables, con la forma general de

$$x'(t) = f(x, y)$$
$$y'(t) = g(x, y)$$

Observe que las derivadas respecto al tiempo $x'(t)$ y $y'(t)$ dependen sólo de x y y, mientras que la variable t no interviene en las funciones f y g como un argumento separado. Esta característica, que convierte al sistema un *sistema autónomo*, es un prerrequisito para la aplicación de la técnica del diagrama de fase.[6]

El diagrama de fase de dos variables, al igual que la versión de una variable de la sección 15.6, es limitado porque puede responder sólo a preguntas cualitativas: las relacionadas con la ubicación y la estabilidad dinámica del(os) equilibrio(s) intertemporal(es). Pero nuevamente, al igual que la versión de una variable, tiene las ventajas compensatorias de poder manejar los sistemas no lineales con tanta comodidad como los lineales y de enfrentar los problemas expresados en términos de funciones generales con tanta facilidad como en términos de las funciones específicas.

[6] En el diagrama de fase con una variable introducido en la sección 15.6, la ecuación $dy/dt = f(y)$ también se restringe a ser *autónoma*, prohibiéndose el tener la variable t como un argumento explícito en la función f.

Espacio de fase

Cuando se construye un diagrama de fase de una variable (figura 15.3) para la ecuación diferencial (autónoma) $dy/dt = f(y)$, simplemente se grafica dy/dt contra y en los dos ejes en un espacio de fase bidimensional. Sin embargo, ahora que el número de variables se ha *duplicado*, ¿cómo podemos lograr satisfacer la necesidad aparente de más ejes? Afortunadamente, la respuesta es que todo lo que necesitamos es el espacio bidimensional.

Para ver por qué es factible esto, observe que la tarea más crucial en la construcción de un diagrama de fase es determinar la dirección del movimiento de la(s) variable(s) respecto al tiempo. Esta información es, tal como lo engloban las cabezas de flecha de la figura 15.3, lo que nos permite obtener las inferencias cualitativas finales. Para dibujar dichas cabezas de flecha sólo requerimos dos cosas: (1) una línea de demarcación —podemos llamarla la línea "$dy/dt = 0$"— que suministre la ubicación de cualquier prospecto de equilibrio(s) y, lo que es más importante, que separe al espacio de fase en dos regiones, una caracterizada por $dy/dt > 0$ y la otra por $dy/dt < 0$ y (2) una línea real sobre la cual pueda indicarse el incremento y el decremento de y implicados por cualquier valor de dy/dt diferente de cero. En la figura 15.3, la línea de demarcación citada en el inciso 1 se encuentra sobre el eje horizontal. Pero ese eje en realidad también sirve como la línea real citada en el inciso 2. Esto significa que el eje vertical, para dy/dt, puede abandonarse sin pérdida, siempre que tengamos cuidado de diferenciar entre la región $dy/dt > 0$ y la región $dy/dt < 0$, digamos al rotular a la primera con un signo más, y a la segunda con un signo menos. La calidad de prescindible de uno de los ejes es lo que hace factible la colocación de un diagrama de fase de dos variables en el espacio bidimensional. Ahora necesitamos *dos* líneas reales en lugar de una. Pero esto es considerado automáticamente por los ejes estándar x y y de un diagrama bidimensional. Necesitamos también dos líneas de demarcación (o curvas), una para $dx/dt = 0$ y la otra para $dy/dt = 0$. Pero ambas son graficables en un espacio de fase bidimensional. Una vez que se dibujan, no sería difícil decidir qué lados de estas líneas o curvas deben marcarse con signos de más y de menos, respectivamente.

Curvas de demarcación

Dado el siguiente sistema autónomo de ecuaciones diferenciales

$$\begin{aligned} x' &= f(x, y) \\ y' &= g(x, y) \end{aligned} \qquad \textbf{(19.40)}$$

donde x' y y' son abreviaturas de las derivadas respecto al tiempo $x'(t)$ y $y'(t)$, las dos curvas de demarcación —que deben denotarse por $x' = 0$ y $y' = 0$— representan las gráficas de las dos ecuaciones

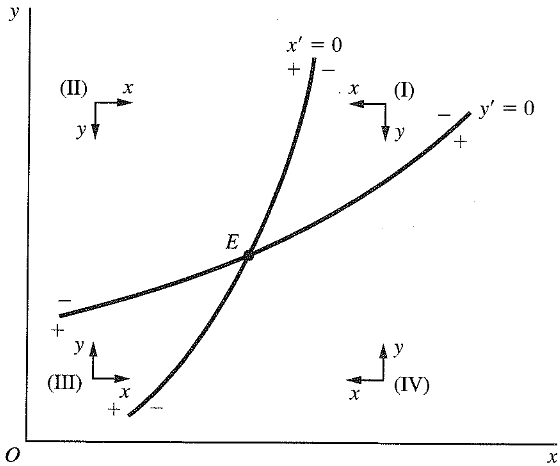
$$f(x, y) = 0 \qquad [\text{curva } x' = 0] \qquad \textbf{(19.41)}$$

$$g(x, y) = 0 \qquad [\text{curva } y' = 0] \qquad \textbf{(19.42)}$$

Si conocemos la forma específica de la función f, podemos despejar y de (19.41) en términos de x y la solución podemos graficarla en el plano xy como la curva $x' = 0$. Sin embargo, aun si no fuera así, podemos, sin embargo, hacer uso de la regla de la función implícita y evaluar la pendiente de la curva $x' = 0$ como

$$\left.\frac{dy}{dx}\right|_{x'=0} = -\frac{\partial f/\partial x}{\partial f/\partial y} = -\frac{f_x}{f_y} \qquad (f_y \neq 0) \qquad \textbf{(19.43)}$$

FIGURA 19.1



Siempre que conozcamos el signo de las derivadas parciales f_x y f_y ($\neq 0$), disponemos de (19.43) de una pista cualitativa de la pendiente de la curva $x' = 0$. Por la misma razón, la pendiente de la curva $y' = 0$ puede inferirse a partir de la derivada

$$\left.\frac{dy}{dx}\right|_{y'=0} = -\frac{g_x}{g_y} \qquad (g_y \neq 0) \tag{19.44}$$

Como un ejemplo más concreto, supongamos que

$$f_x < 0 \qquad f_y > 0 \qquad g_x > 0 \qquad \text{y} \qquad g_y < 0 \tag{19.45}$$

Entonces, ambas curvas $x' = 0$ y $y' = 0$ tendrán pendiente positiva. Si suponemos además que

$$-\frac{f_x}{f_y} > -\frac{g_x}{g_y} \qquad [\text{la curva } x' = 0 \text{ es más empinada que la curva } y' = 0]$$

entonces podemos encontrar una situación tal como la que se muestra en la figura 19.1. Observe que ahora las líneas de demarcación posiblemente son curvas. Observe también que ya no se requiere que coincidan con los ejes.

Las dos curvas de demarcación, que se cortan en el punto E, dividen al espacio de fase en cuatro regiones diferentes, rotuladas de I a IV. El punto E, donde x y y son estacionarios ($x' = y' = 0$), representa el equilibrio intertemporal del sistema. Sin embargo, para cualquier otro punto, ya sea x o y (o ambos) estarían cambiando respecto al tiempo, en direcciones determinadas por el signo de las derivadas respecto al tiempo x' y y' para ese punto. En este ejemplo, tenemos $x' > 0$ ($x' < 0$) a la izquierda (derecha) de la curva $x' = 0$; de ahí el signo más (menos) a la izquierda (derecha) de esa curva. Estos signos se basan en el hecho de que

$$\frac{\partial x'}{\partial x} = f_x < 0 \qquad [\text{por (19.40) y (19.45)}] \tag{19.46}$$

lo que implica que, a medida que nos movemos continuamente de oeste a este en el espacio de fase (a medida que x se incrementa), x' experimenta un decremento uniforme, de modo que el signo de x' debe pasar por tres etapas, en el orden $+$, 0, $-$. En forma análoga, la derivada

$$\frac{\partial y'}{\partial y} = g_y < 0 \qquad [\text{por (19.40) y (19.45)}] \tag{19.47}$$

implica que, a medida que nos movemos continuamente de sur a norte (a medida que y se incrementa), y' disminuye uniformemente, de modo que el signo de y' debe pasar por tres etapas, en el orden +, 0, −. Esto nos lleva a asignar el signo más debajo de la curva $y' = 0$ y el signo menos arriba de ésta en la figura 19.1.

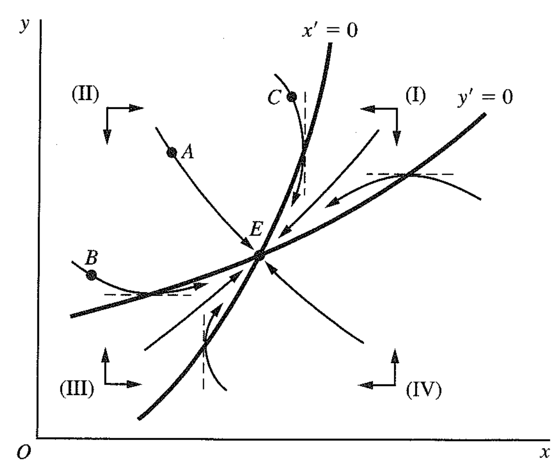
Basándose en estos signos más y menos, ahora puede dibujarse un conjunto de flechas direccionales para indicar el movimiento intertemporal de x y de y. Para cualquier punto de la región I, x' y y' son ambos negativos. Entonces x y y deben disminuir respecto al tiempo, produciendo un movimiento *hacia el oeste* para x, y un movimiento *hacia el sur* para y. Como lo indican las dos flechas de la región I, dado un punto inicial ubicado en la región I, el movimiento intertemporal debe ir en la dirección general hacia el suroeste. El opuesto exacto es verdad en la región III, donde x' y y' son positivos, de modo que ambas variables x y y deben incrementarse respecto al tiempo. En contraste, x' y y' tienen signos diferentes en la región II. Con x' positivo y y' negativo, x debe moverse *hacia el este* y y *hacia el sur*. Y la región IV exhibe una tendencia exactamente opuesta a la región II.

Líneas de corriente

Para una mejor comprensión de las implicaciones de las flechas direccionales, podemos esbozar una serie de *líneas de corriente* en el diagrama de fase. También llamadas *trayectorias de fase* (o simplemente *trayectorias*) o *rutas de fase*, estas líneas de corriente sirven para mapear el movimiento dinámico del sistema desde cualquier punto inicial que pueda concebirse. Algunas de éstas se ilustran en la figura 19.2, la cual reproduce las curvas $x' = 0$ y $y' = 0$ de la figura 19.1. Dado que cada punto en el espacio de fase debe ubicarse en una u otra línea de corriente, ha de existir un número infinito de líneas de corriente, todas las cuales se conforman con los requerimientos direccionales impuestos por las flechas xy de cada región. Sin embargo, para describir el carácter cualitativo general del diagrama de fase, normalmente deben ser suficientes unas cuantas líneas de corriente representativas.

En la figura 19.2 pueden observarse varios aspectos acerca de las líneas de corriente. Primero, todas ellas conducen hasta el punto E. Esto hace de E un equilibrio intertemporal estable (aquí, globalmente estable). Posteriormente vamos a encontrar otros tipos de configuraciones de líneas de corriente. Segundo, mientras que algunas líneas de corriente nunca se aventuran más allá de una sola región (tal como la que pasa por el punto A), otras pueden cruzar de una región a otra (tales como las que pasan por B y C). Tercero, en los puntos de cruce de una línea de corriente, ésta debe tener ya sea una pendiente infinita (al cruzar a la curva $x' = 0$) o una pendiente cero (al cruzar a la curva $y' = 0$). Esto se debe al hecho de que, a lo largo de la curva $x' = 0$ ($y' = 0$), $x(y)$ es estacionaria respecto al tiempo, de modo que la línea de corriente no

FIGURA 19.2



debe tener ningún movimiento horizontal (vertical) al cruzar aquella curva. Para asegurar que estos requerimientos de pendiente se cumplen consistentemente, sería aconsejable, tan pronto como las curvas de demarcación se coloquen en su lugar, añadir unas cuantas barras cortas *verticales* que crucen la curva $x' = 0$ y unas cuantas *horizontales* que crucen la curva $y' = 0$, como lineamientos para dibujar las líneas de corriente.[7] Cuarto, y por último, aunque las líneas de corriente señalan en forma explícita la dirección de movimiento de x y y respecto al tiempo, no suministran información específica respecto a la velocidad y la aceleración, ya que el diagrama de fase no contempla un eje para t (tiempo). Por supuesto que es por esta razón por la que las líneas de corriente llevan el nombre alternativo de trayectorias de *fase*, en contraposición a trayectorias de *tiempo*. La única observación que podemos hacer acerca de la velocidad es de naturaleza cualitativa: a medida que nos movemos a lo largo de la línea de corriente cada vez más cerca de la curva $x' = 0$ ($y' = 0$), la velocidad de aproximación en la dirección horizontal (vertical) debe disminuir en forma progresiva. Esto se debe al decremento uniforme del valor absoluto de la derivada $x' \equiv dx/dt$ ($y' \equiv dy/dt$) que se presenta a medida que nos movemos hacia la línea de demarcación sobre la cual $x'(y')$ adopta un valor cero.

Tipos de equilibrio

Dependiendo de la configuración de las líneas de corriente que circundan a un equilibrio intertemporal específico, ese equilibrio debe situarse en una de cuatro categorías: (1) nodos, (2) puntos de silla, (3) focos y (4) vórtices.

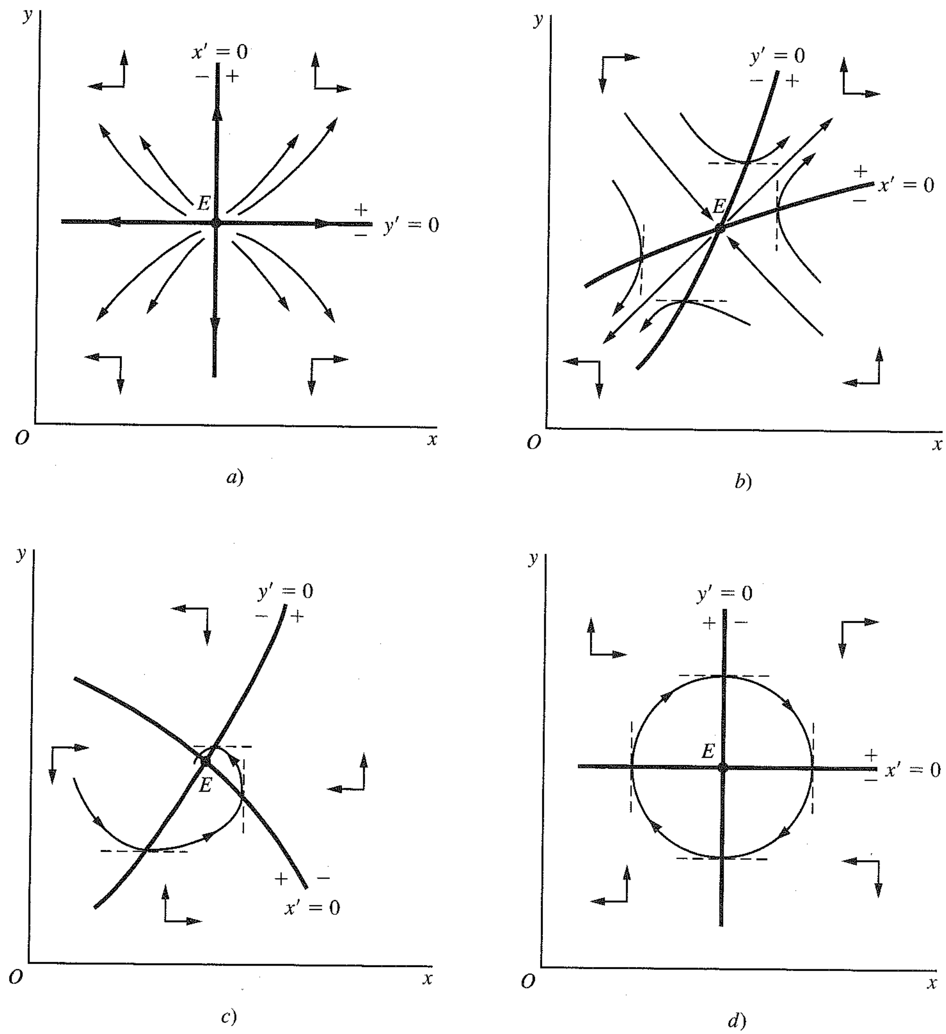
Un *nodo* es un equilibrio tal que todas las líneas de corriente asociadas con éste fluyen en forma no cíclica hacia él (*nodo estable*) o fluyen en forma no cíclica alejándose de él (*nodo inestable*). Ya hemos encontrado un nodo estable en la figura 19.2. En la figura 19.3a se muestra un nodo inestable. Observe que en esta ilustración específica ocurre que las líneas de corriente nunca cruzan de una región a otra. También, las curvas $x' = 0$ y $y' = 0$ son lineales y, de hecho, ellas mismas sirven como líneas de corriente.

Un *punto silla* es un equilibrio con doble personalidad: es estable en algunas direcciones pero inestable en otras. Con más exactitud, en referencia a la ilustración de la figura 19.3b, un punto de silla tiene exactamente un par de líneas de corriente, llamadas las *ramas estables* del punto de silla, que fluyen de manera directa y consistente hacia el equilibrio y exactamente un par de líneas de corriente, las *ramas inestables*, que fluyen de modo directo y consistente alejándose de él. Todas las otras trayectorias apuntan inicialmente hacia el punto de silla pero tarde o temprano se desvían de él. Por supuesto que esta doble personalidad es lo que inspiró el nombre de "punto de silla". Como la estabilidad se observa sólo en las ramas estables, que de hecho no son alcanzables, un punto de silla se clasifica genéricamente como un equilibrio *inestable*.

El tercer tipo de equilibrio, el *foco*, se caracteriza por trayectorias giratorias, las cuales fluyen en forma cíclica hacia él (*foco estable*) o fluyen en forma cíclica alejándose de él (*foco inestable*). La figura 19.3c ilustra un foco estable, con sólo una línea de corriente que se dibuja explícitamente con objeto de evitar amontonamientos. ¿Qué causa la ocurrencia del movimiento giratorio? La respuesta radica en la manera en que se posicionan las curvas $x' = 0$ y $y' = 0$. En la figura 19.3c, las dos curvas de demarcación presentan su pendiente de una manera tal que se turnan para bloquear la línea de corriente que fluye en una dirección prescrita por un conjunto particular de flechas xy. Como resultado, con frecuencia la línea de corriente se ve empujada a cruzar de una región a otra, describiendo una espiral. El que obtengamos un

[7] Para ayudar a su memoria, observe que las barras que atraviesan la curva $x' = 0$ deben ser *perpendiculares* al eje x. En forma similar, las barras que atraviesan la curva $y' = 0$ deben ser *perpendiculares* al eje y.

FIGURA 19.3



foco estable (como es el caso aquí) o uno inestable depende de la colocación relativa de las dos curvas de demarcación. Pero en cualquiera de los dos casos, la pendiente de la línea de corriente en los puntos de cruce debe ser ya sea infinita (cuando cruza a $x' = 0$) o cero (cuando cruza a $y' = 0$).

Por último, podemos tener un *vórtice* (o *centro*). Esto es nuevamente un equilibrio con las líneas de corriente giratorias, pero ahora estas líneas de corriente forman una familia de bucles (círculos concéntricos u óvalos) que orbitan alrededor del equilibrio en un movimiento perpetuo. Un ejemplo de esto está dado en la figura 19.3*d*, donde nuevamente, sólo se muestra una línea de corriente individual. Puesto que este tipo de equilibrio es inalcanzable a partir de cualquier posición inicial alejada del punto *E*, un vórtice se clasifica automáticamente como un equilibrio inestable.

Todas las ilustraciones de la figura 19.3 exhiben un equilibrio único. Sin embargo, cuando hay suficiente no linealidad, las dos curvas de demarcación pueden intersecarse más de una vez, produciendo con ello equilibrios múltiples. En ese caso, puede haber en el mismo diagrama de fase una combinación de los tipos de equilibrio intertemporal previamente citados. Aunque entonces habrá más de cuatro regiones con las cuales confrontarse, el principio subyacente del análisis del diagrama de fase va a permanecer prácticamente igual.

La inflación y la regla monetaria según Obst

Como una ilustración en economía del diagrama de fase de dos variables, presentaremos un modelo del profesor Obst,[8] que intenta mostrar la ineficacia del tipo convencional de regla de política monetaria contracíclica (de ahí la necesidad de una nueva), cuando está funcionando un "mecanismo de ajuste de la inflación". Este modelo contrasta con nuestra discusión anterior de la inflación en que, en lugar de estudiar las implicaciones de una tasa *dada* de expansión monetaria, busca más a fondo en la eficacia de dos *reglas* monetarias diferentes, cada una que prescribe un conjunto diferente de acciones monetarias que deben seguirse en vista de las diferentes condiciones inflacionarias.

Una hipótesis crucial del modelo es el mecanismo de ajuste de la inflación

$$\frac{dp}{dt} = h\left(\frac{M_s - M_d}{M_s}\right) = h\left(1 - \frac{M_d}{M_s}\right) \qquad (h > 0) \tag{19.48}$$

que muestra que el efecto de una oferta excesiva de dinero ($M_s > M_d$) es elevar la tasa de inflación p, en vez del nivel de precios P. La puesta en ceros del mercado de dinero implicaría entonces no la estabilidad de los precios, sino sólo una tasa de inflación estable. Para facilitar el análisis, la segunda igualdad de (19.48) sirve para desplazar el enfoque de la oferta monetaria excesiva a la relación de la demanda-oferta de dinero, M_d/M_s, que denotaremos por μ. Con la hipótesis de que M_d es directamente proporcional al producto nacional nominal PQ, podemos escribir

$$\mu \equiv \frac{M_d}{M_s} = \frac{aPQ}{M_s} \qquad (a > 0)$$

Entonces, las tasas de crecimiento de las diversas variables están relacionadas mediante

$$\begin{aligned} \frac{d\mu/dt}{\mu} &= \frac{da/dt}{a} + \frac{dP/dt}{P} + \frac{dQ/dt}{Q} - \frac{dM_s/dt}{M_s} \\ & \qquad\qquad\qquad\qquad [\text{por } (10.24) \text{ y } (10.25)] \\ &\equiv p + q - m \qquad\qquad [a = \text{una constante}] \end{aligned} \tag{19.49}$$

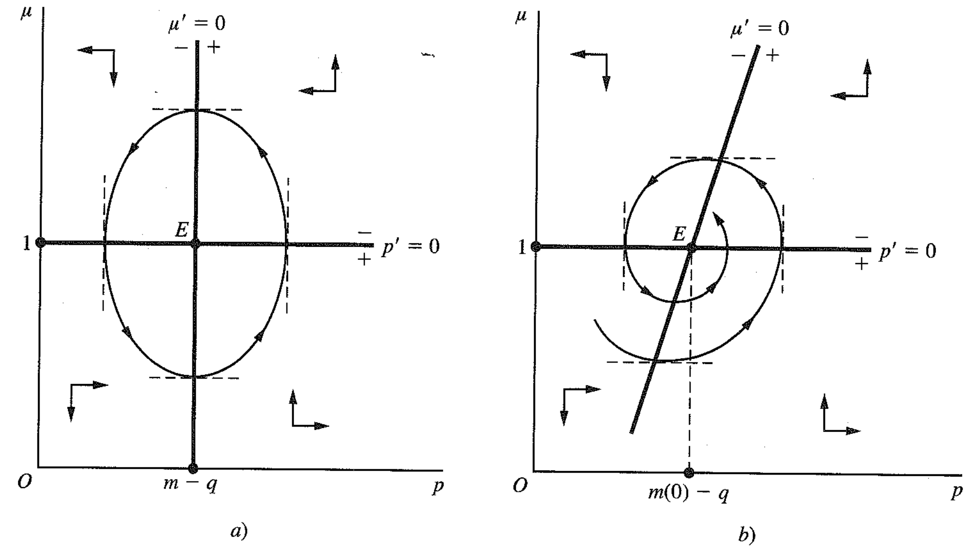
donde las letras minúsculas p, q y m denotan, respectivamente, la tasa de inflación, la tasa de crecimiento (exógena) del producto nacional real y la tasa de expansión monetaria.

Las ecuaciones (19.48) y (19.49), un conjunto de dos ecuaciones diferenciales, pueden determinar conjuntamente las trayectorias de tiempo de p y μ, si por el momento, m se considera como exógena. Usando los símbolos p' y μ' para representar las derivadas de tiempo $p'(t)$ y $\mu'(t)$, podemos expresar este sistema en forma más concisa como

$$\begin{aligned} p' &= h(1 - \mu) \\ \mu' &= (p + q - m)\mu \end{aligned} \tag{19.50}$$

[8]Norman P. Obst, "Stabilization Policy with an Inflation Adjustment Mechanism", *Quarterly Journal of Economics*, mayo de 1978, pp. 355-359. No se dan diagramas de fase en el artículo de Obst, pero pueden construirse rápidamente a partir del modelo.

FIGURA 19.4



Dado que h es positivo, podemos tener $p' = 0$ si y sólo si $1 - \mu = 0$. En forma similar, ya que μ siempre es positivo, $\mu' = 0$ si y sólo si $p + q - m = 0$. Entonces, las curvas de demarcación $p' = 0$ y $\mu' = 0$ se asocian con las ecuaciones

$$\mu = 1 \qquad [\text{curva } p' = 0] \tag{19.51}$$

$$p = m - q \qquad [\text{curva } \mu' = 0] \tag{19.52}$$

Como se muestra en la figura 19.4*a*, éstas se grafican como una línea horizontal y una línea vertical, respectivamente, y arrojan un equilibrio único en E. El valor de equilibrio $\overline{\mu} = 1$ significa que en el *equilibrio* M_d y M_s son iguales, poniendo en ceros al mercado de dinero. El hecho de que se muestra que la tasa de inflación de equilibrio es positiva refleja la hipótesis implícita de que $m > q$.

Como la curva $p' = 0$ corresponde a la curva $x' = 0$ en nuestra discusión anterior, debe tener barras verticales. Y la otra curva deberá tenerlas horizontales. De (19.50) encontramos que

$$\frac{\partial p'}{\partial \mu} = -h < 0 \qquad \text{y} \qquad \frac{\partial \mu'}{\partial p} = \mu > 0 \tag{19.53}$$

con la implicación de que un movimiento hacia el norte a través de la curva $p' = 0$ pasa por la secuencia de signos $(+, 0, -)$ para p', y para un movimiento hacia el este a través de la curva $\mu' = 0$, por la secuencia de signos $(-, 0, +)$ para μ'. De este modo obtenemos los cuatro conjuntos de flechas direccionales tal como están dibujadas, con lo que se generan líneas de corriente (de las cuales sólo se muestra una) que orbitan en sentido contrario a las manecillas del reloj alrededor del punto E. Esto hace que E sea un vórtice. A menos que resulte que la economía esté inicialmente en E, es imposible alcanzar el equilibrio. En lugar de eso, siempre va a haber una fluctuación que nunca cesa.

Sin embargo, la conclusión anterior es la consecuencia de una tasa *exógena* de expansión monetaria. ¿Qué pasa si ahora hacemos endógena a m adoptando una regla monetaria antiinflacionaria? La regla monetaria "convencional" requeriría el engranado de la tasa de expansión monetaria en sentido negativo con la tasa de inflación:

$$m = m(p) \qquad m'(p) < 0 \qquad [\text{regla monetaria convencional}] \tag{19.54}$$

Esta regla modificaría a la segunda ecuación en (19.50) como

$$\mu' = [p + q - m(p)]\mu \tag{19.55}$$

y alteraría a (19.52) como

$$p = m(p) - q \qquad [\text{curva } \mu' = 0 \text{ bajo la regla monetaria convencional}] \tag{19.56}$$

Dado que $m(p)$ es monótona, existe sólo un valor de p —digamos p_1— que satisface esta ecuación. Entonces la nueva curva $\mu' = 0$ va a aparecer todavía como una línea recta vertical, aun cuando con una intercepción horizontal diferente $p_1 = m(p_1) - q$. Aún más, de (19.55) encontramos que

$$\frac{\partial \mu'}{\partial p} = [1 - m'(p)]\mu > 0 \qquad [\text{por (19.54)}]$$

que *cualitativamente* no es diferente de la derivada de (19.53). Se sigue que las flechas direccionales también deben permanecer como están en la figura 19.4*a*. En resumen, terminaríamos con un vórtice como antes.

La regla monetaria alternativa propuesta por Obst es acoplar m con la *tasa de cambio* de la tasa de inflación (en vez del *nivel* de inflación):

$$m = m(p') \qquad m'(p') < 0 \qquad [\text{regla monetaria alternativa}] \tag{19.57}$$

Bajo esta regla, (19.55) y (19.56) van a transformarse, respectivamente, en,

$$\mu' = [p + q - m(p')]\mu \tag{19.58}$$

$$p = m(p') - q \qquad [\text{curva } \mu' = 0 \text{ bajo la regla monetaria alternativa}] \tag{19.59}$$

Esta vez la curva $\mu' = 0$ presentaría la pendiente hacia arriba. Al diferenciar (19.59) respecto a μ por la regla de la cadena, tenemos

$$\frac{dp}{d\mu} = m'(p')\frac{dp'}{d\mu} = m'(p')(-h) > 0 \qquad [\text{por (19.50)}]$$

de modo que por la regla de la función inversa, $d\mu/dp$—la pendiente de la curva $\mu' = 0$— también es positiva. Esta nueva situación se ilustra en la figura 19.4*b*, donde por simplicidad, la curva $\mu' = 0$ se traza como una línea recta, con una pendiente arbitrariamente asignada.[9] A pesar del cambio de pendiente, la derivada parcial

$$\frac{\partial \mu'}{\partial p} = \mu > 0 \qquad [\text{de (19.58)}]$$

permanece igual desde (19.53), de modo que las flechas μ deben conservar su orientación original en la figura 19.4*a*. Las líneas de corriente (de las cuales sólo se muestra una) van a girar ahora hacia dentro y hacia el equilibrio para $\overline{\mu} = 1$ y $\overline{p} = m(0) - q$, donde $m(0)$ denota a $m(p')$ evaluada para $p' = 0$. Entonces se ve que la regla monetaria alternativa tiene la capacidad de convertir un vórtice en un foco estable, haciendo posible con ello la eliminación asintótica de la fluctuación perpetua de la tasa de inflación. En verdad que con una curva $\mu' = 0$ suficientemente plana, es posible aun transformar el vórtice en un nodo estable.

[9] La pendiente es inversamente proporcional al valor absoluto de $m'(p')$. Cuanto mayor sea la sensibilidad de la tasa de expansión monetaria m con que se haga responder a la tasa de cambio de la tasa de inflación p', tanto más plana es la curva $\mu' = 0$ de la figura 19.4*b*.

EJERCICIO 19.5

1. Muestre que el diagrama de fase de dos variables también puede usarse si el modelo consta de una ecuación diferencial individual de segundo orden $y''(t) = f(y', y)$ en lugar de un sistema de dos ecuaciones de primer orden.
2. Los signos más y menos añadidos a ambos lados de las curvas $x' = 0$ y $y' = 0$ en la figura 19.1 se basan en las derivadas parciales $\partial x'/\partial x$ y $\partial y'/\partial y$, respectivamente. ¿Pueden obtenerse las mismas conclusiones a partir de las derivadas $\partial x'/\partial y$ y $\partial y'/\partial x$?
3. Usando la figura 19.2, verifique que si una línea de corriente no tiene una pendiente infinita (o bien cero) cuando cruza a la curva $x' = 0$ (o bien $y' = 0$), será necesario violar las restricciones direccionales impuestas por las flechas *xy*.
4. Como casos especiales del sistema de ecuaciones diferenciales (19.40), suponga que
   (*a*) $f_x = 0 \quad f_y > 0 \quad g_x > 0 \quad$ y $\quad g_y = 0$
   (*b*) $f_x = 0 \quad f_y < 0 \quad g_x < 0 \quad$ y $\quad g_y = 0$
   Para cada caso, construya un diagrama de fase apropiado, trace las líneas de corriente y determine la naturaleza del equilibrio.
5. (*a*) Muestre que es posible producir *ya sea* un nodo estable o un foco estable a partir del sistema de ecuaciones diferenciales (19.40) si
   $f_x < 0 \quad f_y > 0 \quad g_x < 0 \quad$ y $\quad g_y < 0$
   (*b*) ¿Qué característica(s) especial(es) de la construcción del diagrama de fase son responsables de la diferencia en los resultados (nodo contra foco)?
6. Respecto al modelo de Obst, verifique que si la curva $\mu' = 0$ con pendiente positiva en la figura 19.4*b* se hace suficientemente plana, aunque todavía está caracterizada por cruces, va a convergir al equilibrio a la manera de un nodo en vez de un foco.

19.6 Linealización de un sistema de ecuaciones diferenciales no lineales

Otra técnica cualitativa para analizar un sistema de ecuaciones diferenciales *no lineales* es realizar una *aproximación lineal* de ese sistema, mediante una expansión de Taylor del sistema dado alrededor de su equilibrio.[10] En la sección 9.5 aprendimos que una aproximación lineal (o incluso un polinomio de orden superior) de una función arbitraria $\phi(x)$ puede darnos el valor exacto de $\phi(x)$ en el punto de la expansión, pero va a generar errores de aproximación progresivamente mayores a medida que nos alejamos del punto de la expansión. Lo mismo es verdad de la aproximación lineal de un sistema no lineal. En el punto de la expansión —aquí el punto de equilibrio *E*— la aproximación lineal puede localizar exactamente el mismo equilibrio que el sistema original no lineal. Y en una vecindad lo suficientemente pequeña de *E*, la aproximación lineal debe tener la misma configuración general de las líneas de corriente que el sistema original. Por lo tanto, siempre que tengamos el deseo de confinar nuestras inferencias de estabilidad a la vecindad inmediata del equilibrio, la aproximación lineal puede servir como una fuente adecuada de información. Este análisis, denominado *análisis local de*

[10] En el caso de los equilibrios múltiples, cada equilibrio requiere una aproximación lineal por separado.

estabilidad, puede usarse ya sea por sí mismo o como un complemento del análisis de diagrama de fase. Lo estudiaremos sólo para el caso de dos variables.

Expansión de Taylor y linealización

Dada una función $\phi(x)$ arbitraria de una variable (sucesivamente diferenciable), la expansión de Taylor alrededor de un punto x_0 da la serie

$$\begin{aligned}\phi(x) = \phi(x_0) + \phi'(x_0)(x - x_0) + \frac{\phi''(x_0)}{2!}(x - x_0)^2 + \cdots \\ + \frac{\phi^{(n)}(x_0)}{n!}(x - x_0)^n + R_n\end{aligned}$$

donde aparece a la derecha un polinomio que incluye diferentes potencias de $(x - x_0)$. Una estructura similar caracteriza a la expansión de Taylor de una función de dos variables $f(x, y)$ alrededor de cualquier punto (x_0, y_0). Sin embargo, con dos variables en escena, el polinomio resultante comprendería diferentes potencias de $(y - y_0)$, así como de $(x - x_0)$, de hecho, también los productos de estas dos expresiones:

$$\begin{aligned}f(x, y) = f(x_0, y_0) &+ f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) \\ &+ \frac{1}{2!}[f_{xx}(x_0, y_0)(x - x_0)^2 + 2f_{xy}(x_0, y_0)(x - x_0)(y - y_0) \\ &+ f_{yy}(x_0, y_0)(y - y_0)^2] + \cdots + R_n\end{aligned} \tag{19.60}$$

Observe que los coeficientes de las expresiones $(x - x_0)$ y $(y - y_0)$ son ahora las derivadas parciales de f, todas evaluadas en el punto de la expansión (x_0, y_0).

A partir de la serie de Taylor de una función, la aproximación lineal —o brevemente, *linealización*— se obtiene anulando todos los términos de orden mayor que uno. Así, para el caso de una variable, la linealización es la siguiente función lineal de x:

$$\phi(x_0) + \phi'(x_0)(x - x_0)$$

En forma similar, la linealización de (19.60) es la siguiente función lineal de x y y:

$$f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$$

Además, al sustituir el símbolo de función g en vez de f en este resultado, también podemos obtener la correspondiente linealización de $g(x, y)$. Se sigue que, dado el sistema no lineal

$$\begin{aligned}x' &= f(x, y) \\ y' &= g(x, y)\end{aligned} \tag{19.61}$$

su linealización alrededor del punto de expansión (x_0, y_0) puede escribirse como

$$\begin{aligned}x' &= f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) \\ y' &= g(x_0, y_0) + g_x(x_0, y_0)(x - x_0) + g_y(x_0, y_0)(y - y_0)\end{aligned} \tag{19.62}$$

Si se conocen las formas específicas de las funciones f y g, entonces a $f(x_0, y_0)$, $f_x(x_0, y_0)$, $f_y(x_0, y_0)$ y a sus contrapartes para la función g pueden asignárseles valores específicos y el sistema lineal (19.62) puede resolverse cuantitativamente. Sin embargo, aun si las funciones f y g están dadas en formas *generales*, el análisis cualitativo todavía es posible, considerando que los signos de f_x, f_y, g_x y g_y se pueden conocer.

Linealización reducida

Para propósitos del análisis local de estabilidad, la linealización (19.62) puede exponerse en una forma más sencilla. Primero, ya que nuestro punto de expansión debe ser el punto de equilibrio $(\bar{x}, \bar{y})$, debemos reemplazar (x_0, y_0) por $(\bar{x}, \bar{y})$. En forma más sustantiva, ya que para el punto de equilibrio tenemos $x' = y' = 0$ por definición, se sigue que

$$f(\bar{x}, \bar{y}) = g(\bar{x}, \bar{y}) = 0 \qquad \text{[por (19.61)]}$$

de modo que pueda anularse el primer término en el miembro derecho de cada ecuación de (19.62). Haciendo estos cambios, luego efectuando las multiplicaciones de los términos restantes a la derecha de (19.62) y reordenando, obtenemos otra versión de la linealización:

$$\begin{aligned} x' - f_x(\bar{x}, \bar{y})x - f_y(\bar{x}, \bar{y})y &= -f_x(\bar{x}, \bar{y})\bar{x} - f_y(\bar{x}, \bar{y})\bar{y} \\ y' - g_x(\bar{x}, \bar{y})x - g_y(\bar{x}, \bar{y})y &= -g_x(\bar{x}, \bar{y})\bar{x} - g_y(\bar{x}, \bar{y})\bar{y} \end{aligned} \qquad \textbf{(19.63)}$$

Observe que, en (19.63), cada término situado a la derecha del signo de igualdad representa una constante. Nos tomamos la molestia de apartar estos términos constantes de modo que ahora podamos anularlos, para llegar a las ecuaciones reducidas de la linealización. El resultado, que puede escribirse en notación matricial como

$$\begin{bmatrix} x' \\ y' \end{bmatrix} - \begin{bmatrix} f_x & f_y \\ g_x & g_y \end{bmatrix}_{(\bar{x}, \bar{y})} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \textbf{(19.64)}$$

constituye la *linealización reducida* de (19.61). Dado que el análisis cualitativo depende exclusivamente del conocimiento de las raíces características, las cuales a su vez dependen sólo de las ecuaciones reducidas de un sistema, (19.64) es todo lo que necesitamos para el análisis local de estabilidad deseado.

Avanzando un paso más, podemos observar que la única propiedad distintiva de la linealización reducida radica en la matriz de las derivadas parciales —la matriz jacobiana del sistema no lineal (19.61)— evaluada para el equilibrio $(\bar{x}, \bar{y})$. Así en el análisis final, la estabilidad o inestabilidad local del equilibrio se basa sólo en la composición de dicho jacobiano. Por conveniencia en la notación en el análisis que viene, denotaremos al jacobiano evaluado mediante J_E en el equilibrio y sus elementos por a, b, c y d:

$$J_E \equiv \begin{bmatrix} f_x & f_y \\ g_x & g_y \end{bmatrix}_{(\bar{x}, \bar{y})} \equiv \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad \textbf{(19.65)}$$

Vamos a suponer que las dos ecuaciones diferenciales son funcionalmente independientes. Entonces siempre vamos a tener $|J_E| \neq 0$. (Para algunos casos en los cuales $|J_E| = 0$, vea el ejercicio 19.6-4.)

Análisis local de estabilidad

De acuerdo con (19.16), y usando (19.65), la ecuación característica de la linealización reducida debe ser

$$\begin{vmatrix} r - a & -b \\ -c & r - d \end{vmatrix} = r^2 - (a + d)r + (ad - bc) = 0$$

Es claro que las raíces características dependen en forma crítica de las expresiones $(a + d)$ y $(ad - bc)$. Esta última es simplemente el determinante del jacobiano de (19.65):

$$ad - bc = |J_E|$$

Y la primera, que representa la *suma* de los elementos de la diagonal principal de ese jacobiano, se llama la *traza* de J_E, simbolizada por tr J_E:

$$a + d = \text{tr } J_E$$

De acuerdo con esto, las raíces características pueden expresarse como

$$r_1, r_2 = \frac{\text{tr } J_E \pm \sqrt{(\text{tr } J_E)^2 - 4|J_E|}}{2}$$

Las magnitudes relativas de $(\text{tr } J_E)^2$ y $4|J_E|$ van a determinar si las dos raíces son reales o complejas, es decir, si las trayectorias de tiempo de x y y son uniformes o fluctuantes. Por otro lado, para verificar la estabilidad dinámica de equilibrio, necesitamos evaluar los signos algebraicos de las dos raíces. Para ese propósito, las dos relaciones siguientes nos son muy útiles:

$$r_1 + r_2 = \text{tr } J_E \qquad \textbf{(19.66)}$$

$$r_1 r_2 = |J_E| \qquad \textbf{(19.67)}$$

[vea (16.5) y (16.6)]

Caso 1 $(\text{tr } J_E)^2 > 4|J_E|$ En este caso, las raíces son reales y diferentes, y no es posible ninguna fluctuación. Entonces, el equilibrio puede ser un nodo o un punto silla, pero nunca un foco o un vórtice. En vista de que $r_1 \neq r_2$, existen tres posibilidades diferentes de combinación de signos: ambas raíces negativas, ambas raíces positivas, y dos raíces con signos opuestos.[11] Considerando la información de (19.66) y (19.67), estas tres posibilidades se caracterizan por:

$$\begin{array}{lll} (i) \;\; r_1 < 0, r_2 < 0 & \Rightarrow & |J_E| > 0; \text{tr } J_E < 0 \\ (ii) \;\; r_1 > 0, r_2 > 0 & \Rightarrow & |J_E| > 0; \text{tr } J_E > 0 \\ (iii) \;\; r_1 > 0, r_2 < 0 & \Rightarrow & |J_E| < 0; \text{tr } J_E \gtreqless 0 \end{array}$$

Para la posibilidad *i*, con ambas raíces negativas, ambas funciones complementarias x_c y y_c tienden a cero cuando t se hace infinito. El equilibrio es entonces un nodo estable. Lo opuesto es verdad para la posibilidad *ii*, la cual describe un nodo inestable. En contraste, con dos raíces de signos opuestos, la posibilidad *iii* arroja un punto silla.

Para ver este último caso con mayor claridad, recuerde que las funciones complementarias de las dos variables para el caso 1 adoptan la forma general

$$x_c = A_1 e^{r_1 t} + A_2 e^{r_2 t}$$
$$y_c = k_1 A_1 e^{r_1 t} + k_2 A_2 e^{r_2 t}$$

donde las constantes arbitrarias A_1 y A_2 deben determinarse a partir de las condiciones iniciales. Si las condiciones iniciales son tales que $A_1 = 0$, la raíz positiva r_1 va a desaparecer de la escena, dejando que la raíz negativa r_2 alcance el equilibrio estable. Estas condiciones iniciales pertenecen a los puntos ubicados sobre las ramas estables del punto silla. Por otro lado, si las condiciones iniciales son tales que $A_2 = 0$, la raíz negativa r_2 desaparecerá de la escena, dejando que la raíz positiva r_1 alcance el equilibrio estable. Estas condiciones iniciales se relacionan con los puntos situados sobre las ramas inestables. Como todas las condiciones iniciales restantes también implican $A_1 \neq 0$, todas deben dar lugar también a funciones complementarias divergentes. Entonces la posibilidad *iii* arroja un punto silla.

[11] Dado que hemos descartado a $|J_E| = 0$, ninguna raíz puede adoptar un valor cero.

Caso 2 $(\text{tr } J_E)^2 = 4|J_E|$ Como en este caso las raíces se repiten, pueden surgir sólo dos posibilidades de combinación de signos:

$$(iv) \quad r_1 < 0, r_2 < 0 \quad \Rightarrow \quad |J_E| > 0; \text{tr } J_E < 0$$
$$(v) \quad r_1 > 0, r_2 > 0 \quad \Rightarrow \quad |J_E| > 0; \text{tr } J_E > 0$$

Estas dos posibilidades son simples duplicados de las posibilidades *i* y *ii*. Entonces apuntan a un nodo estable y a un nodo inestable, respectivamente.

Caso 3 $(\text{tr } J_E)^2 < 4|J_E|$ Esta vez, con raíces complejas $h \pm vi$, está presente la fluctuación cíclica y debemos encontrar ya sea un foco o un vórtice. Basándose en (19.66) y (19.67), tenemos en el presente caso

$$\text{tr } J_E = r_1 + r_2 = (h + vi) + (h - vi) = 2h$$
$$|J_E| = r_1 r_2 = (h + vi)(h - vi) = h^2 + v^2$$

Entonces tr J_E debe adoptar el mismo signo que h, mientras que $|J_E|$ es invariablemente positivo. En consecuencia, hay tres resultados posibles:

$$(vi) \quad h < 0 \quad \Rightarrow \quad |J_E| > 0; \text{tr } J_E < 0$$
$$(vii) \quad h > 0 \quad \Rightarrow \quad |J_E| > 0; \text{tr } J_E > 0$$
$$(viii) \quad h = 0 \quad \Rightarrow \quad |J_E| > 0; \text{tr } J_E = 0$$

Éstos se asocian, respectivamente, con la fluctuación amortiguada, la fluctuación explosiva y la fluctuación uniforme. En otras palabras, la posibilidad *vi* implica un foco estable; la posibilidad *vii*, un foco inestable; y la posibilidad *viii*, un vórtice.

Las conclusiones de la revisión anterior se resumen en la tabla 19.1 para facilitar las inferencias cualitativas a partir de los signos de $|J_E|$ y tr J_E. Tres características de la tabla merecen atención especial. Primero, se enlaza en forma exclusiva un $|J_E|$ negativo al tipo de equilibrio de punto silla. Esto sugiere que $|J_E| < 0$ es una condición necesaria y suficiente para un punto silla. Segundo, un valor cero de tr J_E se da sólo bajo dos circunstancias: cuando hay un punto silla o un vórtice. Sin embargo estas dos circunstancias son discernibles entre sí por el signo de $|J_E|$. De acuerdo con esto, una tr J_E acoplada con un $|J_E|$ positivo es necesario y suficiente para un vórtice. Tercero, aun cuando es necesario un signo negativo en tr J_E para la estabilidad dinámica, *no* es suficiente, considerando la posibilidad de un punto silla. Sin embargo, cuando

TABLA 19.1
Análisis local de estabilidad de un sistema de ecuaciones diferenciales no lineales con dos variables

Caso	Signo de $\lvert J_E\rvert$	Signo de tr J_E	Tipo de equilibrio
1. $(\text{tr } J_E)^2 > 4\lvert J_E\rvert$	+	−	Nodo estable
	+	+	Nodo inestable
	−	+, 0, −	Punto silla
2. $(\text{tr } J_E)^2 = 4\lvert J_E\rvert$	+	−	Nodo estable
	+	+	Nodo inestable
3. $(\text{tr } J_E)^2 < 4\lvert J_E\rvert$	+	−	Foco estable
	+	+	Foco inestable
	+	0	Vórtice

una tr J_E negativa está acompañada por un $|J_E|$ positivo, tenemos una condición necesaria y suficiente para la estabilidad dinámica.

El análisis que conduce al resumen de la tabla 19.1 se ha realizado en el contexto de una aproximación lineal a un sistema *no lineal*. Sin embargo, el contenido de esa tabla obviamente es aplicable también al análisis cualitativo de un sistema que, para comenzar, es *lineal*. En este último caso, los elementos de la matriz jacobiana van a ser un conjunto de constantes dadas, de modo que no hay necesidad de evaluarlos para el equilibrio. Dado que no interviene ningún proceso de aproximación, las inferencias de la estabilidad ya no van a ser de naturaleza "local" sino que tendrán una validez global.

Ejemplo 1

Analice la estabilidad local del sistema no lineal

$$\begin{aligned} x' &= f(x, y) = xy - 2 \\ y' &= g(x, y) = 2x - y \end{aligned} \qquad (x, y \geq 0)$$

Primero, haciendo $x' = y' = 0$, y observando la no negatividad de x y y, encontramos un solo equilibrio E para $(\overline{x}, \overline{y}) = (1, 2)$. Entonces, al calcular las derivadas parciales de x' y y', y al evaluarlas para E, obtenemos

$$J_E = \begin{bmatrix} f_x & f_y \\ g_x & g_y \end{bmatrix}_{(\overline{x},\overline{y})} = \begin{bmatrix} y & x \\ 2 & -1 \end{bmatrix}_{(1,2)} = \begin{bmatrix} 2 & 1 \\ 2 & -1 \end{bmatrix}$$

Dado que $|J_E| = -4$ es negativo, podemos concluir de inmediato que el equilibrio es localmente un punto silla.

Observe que mientras el primer renglón de la matriz jacobiana contiene originalmente las variables y y x, el segundo renglón no las contiene. La razón de esta diferencia es que la segunda ecuación en el sistema dado es originalmente lineal, y no requiere linealización.

Ejemplo 2

Dado el sistema no lineal

$$\begin{aligned} x' &= x^2 - y \\ y' &= 1 - y \end{aligned}$$

al hacer $x' = y' = 0$, podemos encontrar dos puntos de equilibrio: $E_1 = (1, 1)$ y $E_2 = (-1, 1)$. Entones necesitamos dos linealizaciones separadas. Al evaluar el jacobiano $\begin{bmatrix} 2x & -1 \\ 0 & -1 \end{bmatrix}$ para los dos equilibrios en turno, obtenemos

$$J_{E1} = \begin{bmatrix} 2 & -1 \\ 0 & -1 \end{bmatrix} \qquad \text{y} \qquad J_{E2} = \begin{bmatrix} -2 & -1 \\ 0 & -1 \end{bmatrix}$$

El primero de éstos tiene un determinante negativo; entonces $E_1 = (1, 1)$ es localmente un punto silla. A partir del segundo encontramos que $|J_{E2}| = 2$ y tr $J_{E2} = -3$. Por lo tanto, de la tabla 19.1, $E_2 = (-1, 1)$ es localmente un nodo estable para el caso 1.

Ejemplo 3

¿Posee el sistema lineal

$$\begin{aligned} x' &= x - y + 2 \\ y' &= x + y + 4 \end{aligned}$$

un equilibrio estable? Para responder esta pregunta cualitativa, simplemente nos concentramos en las ecuaciones reducidas e ignoramos por completo las constantes 2 y 4. Como puede esperarse de un sistema lineal, el jacobiano $\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ tiene cuatro constantes como sus elementos.

Como su determinante y la traza son ambos iguales a 2, el equilibrio se sitúa en el caso 3 y es un foco inestable. Observe que esta conclusión se alcanza sin tener que resolver para el equilibrio. Observe también que la conclusión en este caso es globalmente válida.

Ejemplo 4 Analice la estabilidad local del modelo de Obst (19.50),

$$p' = h\,(1-\mu)$$
$$\mu' = (p+q-m)\mu$$

suponiendo que la tasa de expansión monetaria m es exógena (no se sigue ninguna regla monetaria). De acuerdo con la figura 19.4*a*, el equilibrio de este modelo se presenta en $E = (\overline{p}, \overline{\mu}) = (m-q, 1)$. La matriz jacobianna evaluada en E es

$$J_E = \begin{bmatrix} \dfrac{\partial p'}{\partial p} & \dfrac{\partial p'}{\partial \mu} \\ \dfrac{\partial \mu'}{\partial p} & \dfrac{\partial \mu'}{\partial \mu} \end{bmatrix}_E = \begin{bmatrix} 0 & -h \\ \mu & p+q-m \end{bmatrix}_{(m-q,1)} = \begin{bmatrix} 0 & -h \\ 1 & 0 \end{bmatrix}$$

Como $|J_E| = h > 0$ y tr $J_E = 0$, la tabla 19.1 indica que el equilibrio es localmente un vórtice. Esta conclusión es consistente con la del análisis del diagrama de fase de la sección 19.5.

Ejemplo 5 Analice la estabilidad local del modelo de Obst, suponiendo que la regla monetaria alternativa es:

$$p' = h\,(1-\mu) \qquad \text{[de (19.50)]}$$
$$\mu' = [p+q-m\,(p')]\mu \qquad \text{[de (19.58)]}$$

Observe que dado que p' es una función de μ, la función $m\,(p')$ es también en el presente modelo una función de μ. Entonces tenemos que aplicar la regla del producto para encontrar $\partial\mu'/\partial\mu$. Para el equilibrio E, donde $p' = \mu' = 0$, tenemos $\overline{\mu} = 1$ y $\overline{p} = m\,(0) - q$. Por lo tanto el jacobiano evaluado en E es,

$$J_E = \begin{bmatrix} 0 & -h \\ \mu & p+q-m\,(p') - m'(p')(-h)\mu \end{bmatrix}_E = \begin{bmatrix} 0 & -h \\ 1 & m'(0)h \end{bmatrix}$$

donde $m'(0)$ es negativo mediante (19.57). De acuerdo con la tabla 19.1, con $|J_E| = h > 0$ y tr $J_E = m'(0)h < 0$, podemos tener ya sea un foco estable o un nodo estable, dependiendo de las magnitudes relativas de (tr $J_E)^2$ y $4|J_E|$. Para ser específicos, cuanto mayor sea el valor absoluto de la derivada $m'(0)$, tanto mayor es el valor absoluto de tr J_E y es más probable que (tr $J_E)^2$ sobrepase a $4|J_E|$ para producir un nodo estable en lugar de un foco estable. Esta conclusión es nuevamente consistente con lo que aprendimos del análisis de diagrama de fase.

EJERCICIO 19.6

1. Analice la estabilidad local de cada uno de los siguientes sistemas no lineales:

(*a*) $x' = e^x - 1$, $y' = ye^x$

(*b*) $x' = x + 2y$, $y' = x^2 + y$

(*c*) $x' = 1 - e^y$, $y' = 5x - y$

(*d*) $x' = x^3 + 3x^2y + y$, $y' = x(1 + y^2)$

2. Use la tabla 19.1 para determinar el tipo de equilibrio que un sistema no lineal tendría localmente, dado que:

   (*a*) $f_x = 0 \qquad f_y > 0 \qquad g_x > 0 \qquad$ y $\qquad g_y = 0$

   (*b*) $f_x = 0 \qquad f_y < 0 \qquad g_x < 0 \qquad$ y $\qquad g_y = 0$

   (*c*) $f_x < 0 \qquad f_y > 0 \qquad g_x < 0 \qquad$ y $\qquad g_y < 0$

   ¿Son sus resultados consistentes con sus respuestas a los ejercicios 19.5-4 y 19.5-5?

3. Analice la estabilidad local del modelo de Obst, suponiendo que se sigue la regla monetaria convencional.

4. Los dos siguientes sistemas poseen jacobianos con valor cero. Construya un diagrama de fase para cada uno y deduzca las ubicaciones de todos los equilibrios que existan:

   (*a*) $x' = x + y$ $\qquad$ (*b*) $x' = 0$

   $\quad\;\; y' = -x - y$ $\qquad\qquad\quad y' = 0$

Capítulo 20

Teoría de control óptimo

Al final del capítulo 13 nos referimos a la optimización dinámica como un tipo de problema que no estábamos en condiciones de enfrentar, porque todavía no teníamos las herramientas del análisis dinámico tales como las ecuaciones diferenciales. Ahora que hemos adquirido estas herramientas, finalmente podemos hacer la prueba con la optimización dinámica.

El enfoque clásico de la optimización dinámica se llama *cálculo de variaciones*. Sin embargo, en el desarrollo más reciente de esta metodología, un enfoque más fuerte conocido como *teoría de control óptimo* ha sustituido en su mayor parte al cálculo de variaciones. Por esta razón, en este capítulo vamos a limitar nuestra atención a la teoría de control óptimo, explicando su naturaleza básica, introduciendo la principal herramienta de solución, llamada *principio del máximo*, ejemplificando su uso en algunos modelos económicos elementales.[1]

20.1 Naturaleza del control óptimo

En la optimización estática, la tarea es encontrar un valor individual para cada variable de elección, con el fin de maximizar o minimizar una función objetivo propuesta, cualquiera que sea el caso. El problema de optimización estática no contempla la dimensión del tiempo. En contraste, el tiempo interviene en forma explícita y prominente en el problema de optimización dinámica. En este problema, debemos recordar siempre un periodo de planeación, digamos desde un tiempo inicial $t = 0$ hasta un tiempo terminal $t = T$, y trataremos de encontrar el mejor curso de acción a seguir durante el periodo completo. Entonces la solución para cualquier variable adoptará la forma no de un solo valor, sino de una trayectoria de tiempo completa.

Supongamos que el problema tiene que ver con la maximización de ganancia para un periodo. Para cualquier punto de tiempo t, tendremos que escoger el valor de alguna *variable de control*, $u(t)$, que entonces afectará el valor de alguna *variable de estado*, $y(t)$, vía la así llamada *ecuación de movimiento*. A su vez, $y(t)$ determinará la ganancia $\pi(t)$. Como nuestro objetivo es maximizar la ganancia durante el periodo completo, la función objetivo debe adoptar la forma de una integral definida de π de $t = 0$ a $t = T$. Para ser completo, el problema

[1] Para un tratamiento más completo de la teoría de control óptimo (así como del "cálculo de variaciones"), se refiere al estudiante a *Elements of Dynamic Optimization* de Alpha C. Chiang, McGraw-Hill, Nueva York, 1992, publicado actualmente por Waveland Press, Inc., Prospect Heights, Illinois. Este capítulo se basa en gran parte en el material de este libro citado.

también especifica el valor inicial de la variable de estado y, $y(0)$, y el valor terminal de y, $y(T)$, o en forma alterna, el intervalo de valores que $y(T)$ puede asumir.

Considerando lo anterior, podemos enunciar el problema más sencillo de control óptimo como:

$$\begin{aligned} &\text{Maximizar} && \int_0^T F(t, y, u)\, dt \\ &\text{sujeto a} && \frac{dy}{dt} \equiv y' = f(t, y, u) \\ & && y(0) = A \qquad y(T) \ \text{libre} \\ &\text{y} && u(t) \in U \ \text{ para todo } t \in [0, T] \end{aligned} \tag{20.1}$$

El primer renglón de (20.1), la función objetivo, es una integral cuyo integrando $F(t, y, u)$ estipula la forma en que la elección de la variable de control u para el tiempo t, junto con la y resultante para el tiempo t, determina nuestro objeto de maximización para t. El segundo renglón es la ecuación de movimiento para la variable de estado y. Esta ecuación suministra el mecanismo mediante el cual nuestra elección de la variable de control u puede traducirse a un patrón específico de movimiento de la variable de estado y. Normalmente, el enlace entre u y y puede describirse adecuadamente mediante una ecuación diferencial de primer orden $y' = f(t, y, u)$. Sin embargo, si el patrón de cambio de la variable de estado requiere una ecuación diferencial de segundo orden, entonces debemos transformar esta ecuación en un par de ecuaciones diferenciales de primer orden. En ese caso hay que introducir una variable de estado adicional. Tanto el integrando F como la ecuación de movimiento se suponen continuos para todos sus argumentos y poseen derivadas parciales continuas de primer orden respecto a la variable de estado y y la variable de tiempo t, pero no necesariamente la variable de control u. En el tercer renglón, indicamos que el estado inicial, el valor de y para $t = 0$, es una constante A, pero el estado terminal $y(T)$ se deja sin restricciones. Finalmente, el cuarto renglón indica que las elecciones permisibles de u se limitan a una región de control U. Por supuesto que puede suceder que $u(t)$ no tenga restricciones.

Ejemplo: un modelo macroeconómico simple

Considere una economía que elabora una producción Y con el uso del capital K y una cantidad fija de mano de obra L, de acuerdo con la función de producción

$$Y = Y(K, L)$$

Adicionalmente, la producción se usa tanto para el consumo C como para la inversión I. Si ignoramos el problema de la depreciación, entonces

$$I \equiv \frac{dK}{dt}$$

En otras palabras, la inversión es el cambio de las existencias de capital respecto al tiempo. Entonces también podemos expresar la inversión como

$$I = Y - C = Y(K, L) - C = \frac{dK}{dt}$$

que nos da una ecuación diferencial de primer orden en la variable K.

Si nuestro objetivo fuera maximizar alguna forma de utilidad social para un periodo fijo de planeación, entonces el problema se transformaría en

$$\begin{aligned}
&\text{Maximizar} && \int_0^T U(C)\,dt \\
&\text{sujeto a} && \frac{dK}{dt} = Y(K, L) - C \qquad\qquad (20.2)\\
&\text{y} && K(0) = K_0 \qquad K(T) = K_T
\end{aligned}$$

donde K_0 y K_T son el valor inicial y el valor terminal (objetivo) de K. Observe que en (20.2), el estado terminal es un valor fijo, que no se deja libre como en (20.1). Aquí C sirve como la variable de control y K es la variable de estado. El problema es escoger la trayectoria de control óptimo $C(t)$ de tal manera que su efecto sobre la producción Y y el capital K, y las repercusiones de eso sobre C mismo, van a maximizar juntos la utilidad agregada para el periodo de planeación.

El principio del máximo de Pontryagin

La clave para la teoría de control óptimo es una condición necesaria de primer orden conocida como el *principio del máximo*.[2] El enunciado del principio del máximo implica un enfoque que es afín a la función lagrangiana y a la variable multiplicadora de Lagrange. Para los problemas de control óptimo, éstas se conocen como la *función hamiltoniana* y la *variable de coestado*, conceptos que ahora vamos a desarrollar.

El hamiltoniano

En (20.1) hay tres variables: el tiempo t, la variable de estado y y la variable de control u. Ahora introducimos una nueva variable, conocida como la variable de coestado, y la denotamos como $\lambda(t)$. Al igual que el multiplicador de Lagrange, la variable de coestado mide el precio sombra de la variable de estado.

La variable de coestado se introduce en el problema de control óptimo vía una *función hamiltoniana* (abreviada como hamiltoniano). El hamiltoniano se define como

$$H(t, y, u, \lambda) \equiv F(t, y, u) + \lambda(t) f(t, y, u) \qquad (20.3)$$

donde H denota al hamiltoniano y es una función de cuatro variables: t, y, u y λ.

El principio máximo

El principio del máximo —la herramienta principal para la solución de problemas de control óptimo— debe su nombre a que una condición necesaria de primer orden requiere que escojamos a u de modo que se maximice al hamiltoniano H *para todos los instantes de tiempo.*

Además de la variable de control u, como H implica a la variable de estado y y a la variable de coestado λ, el enunciado del principio máximo también estipula como la forma en que y y λ deben cambiar respecto al tiempo, por medio de una *ecuación de movimiento para la va-*

[2] El término "principio del máximo" se atribuye a L. S. Pontryagin y asociados, y frecuentemente se denomina principio del máximo de Pontryagin. Vea *The Mathematical Theory of Optimal Control Processes* de L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze y E. F. Mishchenko, Interscience, Nueva York, 1962 (traducido por K. N. Trirogoff).

riable de estado y (abreviada como *ecuación de estado*), así como una *ecuación de movimiento para la variable de coestado* λ (abreviada como *ecuación de coestado*). La ecuación de estado siempre viene como parte del enunciado mismo del problema, como en la segunda ecuación de (20.1). Pero en vista de que (20.39) implica $\partial H/\partial\lambda = f(t, y, u)$, el principio del máximo describe la ecuación de estado

$$y' = f(t, y, u) \text{ como } y' = \frac{\partial H}{\partial \lambda} \tag{20.4}$$

En contraste, λ no aparece en el enunciado del problema (20.1) y su ecuación de movimiento entra en escena sólo como una condición de optimización. La ecuación de coestado es

$$\lambda' \left(\equiv \frac{d\lambda}{dt}\right) = -\frac{\partial H}{\partial y} \tag{20.5}$$

Observe que ambas ecuaciones de movimiento se enuncian en términos de las derivadas parciales de *H*, sugiriendo alguna simetría, pero hay un signo negativo añadido a $\partial H/\partial y$ en (20.5).

Las ecuaciones (20.4) y (20.5) constituyen un sistema de dos ecuaciones diferenciales. Así, necesitamos dos condiciones de frontera para determinar las dos constantes arbitrarias que van a surgir en el proceso de solución. Si tanto el estado inicial $y(0)$ como el estado terminal $y(T)$ son fijos, entonces podemos usar estas especificaciones para determinar las constantes. Pero si, como en el problema (20.1), el estado terminal no está fijo, entonces debemos incluir algo llamado *condición de transversalidad* como parte del principio del máximo, para cubrir la brecha dejada por la condición de frontera faltante.

Resumiendo, podemos identificar los diferentes componentes del principio del máximo para el problema (20.1) como sigue:

$$\begin{aligned}
&(i) \quad H(t, y, u^*, \lambda) \geq H(t, y, u, \lambda) \quad \text{para todo } t \in [0, T] \\
&(ii) \quad y' = \frac{\partial H}{\partial \lambda} \qquad \text{(ecuación de estado)} \\
&(iii) \quad \lambda' = -\frac{\partial H}{\partial y} \qquad \text{(ecuación de coestado)} \\
&(iv) \quad \lambda(T) = 0 \qquad \text{(condición de transversalidad)}
\end{aligned} \tag{20.6}$$

La condición *i* de (20.6) establece que para todo instante *t* el valor de $u(t)$, el control óptimo, debe escogerse de modo que se maximice el valor del hamiltoniano para todos los valores admisibles de $u(t)$. En el caso en el cual el hamiltoniano es diferenciable respecto a *u* y ofrece una solución interior, la condición *i* puede reemplazarse con

$$\frac{\partial H}{\partial u} = 0$$

Sin embargo, si la región de control es un conjunto cerrado, entonces son posibles las soluciones de frontera y $\partial H/\partial u = 0$ puede no aplicar. De hecho, el principio del máximo ni siquiera requiere que el hamiltoniano sea diferenciable respecto a *u*.

Las condiciones *ii* y *iii* del principio del máximo, $y' = \partial H/\partial\lambda$ y $\lambda' = -\partial H/\partial y$, nos dan dos ecuaciones de movimiento, denominadas como sistema hamiltoniano para el problema dado. La condición *iv*, $\lambda(T) = 0$, es la condición de transversalidad apropiada sólo para el problema de estado terminal libre.

Ejemplo 1

Para ilustrar el uso del principio del máximo, consideremos primero un ejemplo simple fuera de la economía: el de encontrar la trayectoria más corta desde un punto dado A hasta una línea recta dada. En la figura 20.1 hemos graficado el punto A sobre el eje vertical en el plano ty, y hemos dibujado la línea recta como una vertical para $t = T$. Se muestran tres trayectorias admisibles (del número infinito de ellas), cada una con una longitud diferente. La longitud de cualquier trayectoria es el agregado de pequeños segmentos de trayectoria, cada uno de los cuales puede considerarse como la hipotenusa (que no se dibuja) de un triángulo formado por pequeños movimientos dt y dy. Si denotamos la hipotenusa como dh, por el teorema de Pitágoras tenemos,

$$dh^2 = dt^2 + dy^2$$

La división de ambos lados entre dt^2 y la extracción de la raíz cuadrada arrojan

$$\frac{dh}{dt} = \left[1 + \left(\frac{dy}{dt}\right)^2\right]^{1/2} = [1 + (y')^2]^{1/2} \tag{20.7}$$

La longitud total de la trayectoria puede encontrarse entonces por integración de (20.7) respecto a t, de $t = 0$ a $t = T$. Si hacemos que $y' = u$ sea la variable de control, (20.7) puede expresarse como

$$\frac{dh}{dt} = (1 + u^2)^{1/2} \tag{20.7'}$$

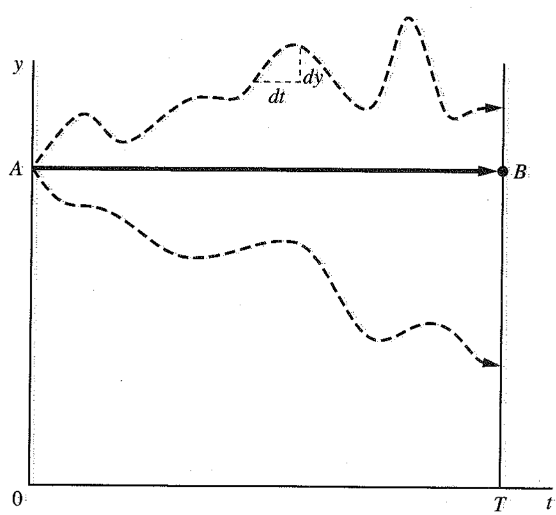
La minimización de la integral de (20.7′) es equivalente a maximizar al *negativo* de (20.7′). Así, el problema de trayectoria más corta es:

$$\begin{array}{lll} \text{Maximizar} & \displaystyle\int_0^T -(1 + u^2)^{1/2} dt & \\ \text{sujeto a} & y' = u & \\ \text{y} & y(0) = A & y(T) \text{ libre} \end{array}$$

El hamiltoniano para el problema, mediante (20.3), es

$$H = -(1 + u^2)^{1/2} + \lambda u$$

FIGURA 20.1



Como H es diferenciable en u, y u no está restringida, podemos usar la siguiente condición de primer orden para maximizar H:

$$\frac{\partial H}{\partial u} = -\frac{1}{2}(1+u^2)^{-1/2}(2u) + \lambda = 0$$

o sea
$$u(t) = \lambda(1-\lambda^2)^{-1/2}$$

Al revisar la condición de segúndo orden, encontramos que

$$\frac{\partial^2 H}{\partial u^2} = -(1+u^2)^{-3/2} < 0$$

lo que verifica que la solución de $u(t)$ maximiza el hamiltoniano. Dado que $u(t)$ es una función de λ, necesitamos una solución para la variable de coestado. A partir de las condiciones de primer orden, la ecuación de movimiento para la variable de coestado es

$$\lambda' = -\frac{\partial H}{\partial y} = 0$$

ya que H es independiente de y. Entonces λ es una constante. Para hacer definitiva esta constante, podemos hacer uso de la condición de transversalidad $\lambda(T) = 0$. Dado que λ puede adoptar un solo valor, que ahora se sabe que es cero, en realidad tenemos $\lambda(t) = 0$ para todo t. Entonces podemos escribir

$$\lambda^*(t) = 0 \text{ para todo } t \in [0, T]$$

De ahí se sigue que el control óptimo es

$$u^*(t) = \lambda^*[1-(\lambda^*)^2]^{-1/2} = 0$$

Finalmente, usando la ecuación de movimiento para la variable de estado, vemos que

$$y' = u = 0$$

o sea
$$y^*(t) = c_0 \qquad \text{(una constante)}$$

Incorporando la condición inicial

$$y(0) = A$$

podemos concluir que $c_0 = A$, y escribir

$$y^*(t) = A \qquad \text{para todo } t$$

En la figura 20.1 esta trayectoria es la línea AB. Se ve que la trayectoria más corta es una línea recta con pendiente cero.

Ejemplo 2 Encuentre la trayectoria de control óptimo que resuelva el siguiente problema

$$\begin{aligned} &\text{Maximizar} && \int_0^1 (y-u^2)\,dt \\ &\text{sujeto a} && y' = u \\ &\text{y} && y(0) = 5 \qquad y(1) \text{ libre} \end{aligned}$$

Este problema está en el formato de (20.1), excepto que u es no restringida.

El hamiltoniano para este problema,

$$H = y - u^2 + \lambda u$$

es cóncavo en u, y u no está restringido, de modo que podemos maximizar H aplicando la condición de primer orden (también suficiente debido a la concavidad de H):

$$\frac{\partial H}{\partial u} = -2u + \lambda = 0$$

lo que nos da

$$u(t) = \frac{\lambda}{2} \quad \text{o sea } y' = \frac{\lambda}{2} \tag{20.8}$$

La ecuación de movimiento para λ es

$$\lambda' = -\frac{\partial H}{\partial y} = -1 \tag{20.8'}$$

Las dos últimas ecuaciones constituyen el sistema de ecuaciones diferenciales para este problema.

Primero podemos despejar λ mediante integración directa de (20.8′) para obtener

$$\lambda(t) = c_1 - t \qquad (c_1 \text{ arbitrario})$$

Aún más, por la condición de transversalidad en (20.6), debemos tener $\lambda(1) = 0$. Haciendo $t = 1$ en la última ecuación arroja $c_1 = 1$. Así, la trayectoria óptima de coestado es

$$\lambda^*(t) = 1 - t$$

Se sigue que $y' = \frac{1}{2}(1 - t)$, por (20.8), y por integración,

$$y(t) = \frac{1}{2}t - \frac{1}{4}t^2 + c_2 \qquad (c_2 \text{ arbitrario})$$

La constante arbitraria puede determinarse mediante el uso de la condición inicial $y(0) = 5$. Haciendo $t = 0$ en la ecuación anterior, obtenemos $5 = y(0) = c_2$. Entonces, la trayectoria óptima para la variable de estado es

$$y^*(t) = \frac{1}{2}t - \frac{1}{4}t^2 + 5$$

y la trayectoria de control óptimo correspondiente es

$$u^*(t) = \frac{1}{2}(1 - t)$$

Ejemplo 3 Encuentre la trayectoria de control óptimo que resuelva el siguiente problema

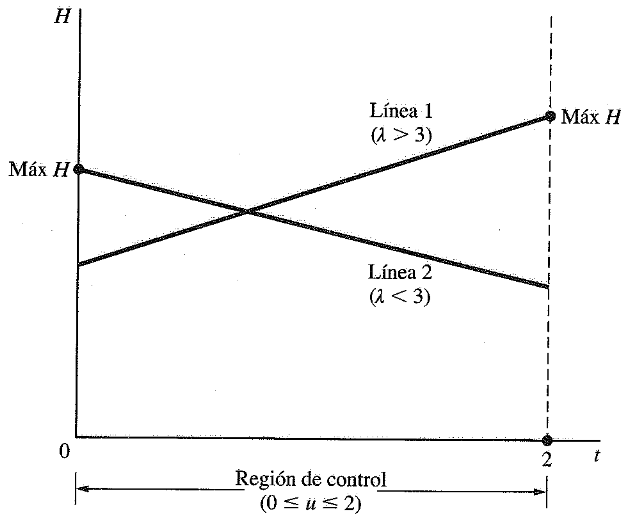
$$\begin{array}{lll} \text{Maximizar} & \displaystyle\int_0^2 (2y - 3u)\, dt & \\ \text{sujeto a} & y' = y + u & \\ & y(0) = 4 & y(2) \text{ libre} \\ \text{y} & u(t) \in [0, 2] & \end{array}$$

El hecho de que la variable de control se restrinja al conjunto cerrado [0, 2] da lugar a la posibilidad de las soluciones de frontera.

La función hamiltoniana

$$H = 2y - 3u + \lambda(y + u) = (2 + \lambda)y + (\lambda - 3)u$$

FIGURA 20.2



es lineal en u. Si graficamos H contra u en el plano uH, obtenemos una línea recta con pendiente $\partial H/\partial u = \lambda - 3$, que es positiva si $\lambda > 3$ (línea 1), pero negativa si $\lambda < 3$ (línea 2), como se ilustra en la figura 20.2. Si en cualquier momento λ sobrepasa a 3, entonces H máximo se presenta en la frontera superior de la región de control y debemos escoger $u = 2$. Si por otro lado, λ se sitúa por debajo de 3, entonces, con objeto de maximizar H, debemos escoger $u = 0$. En resumen, $u^*(t)$ depende de $\lambda(t)$ como sigue:

$$u^*(t) = \begin{Bmatrix} 2 \\ 0 \end{Bmatrix} \quad \text{si} \quad \lambda(t) \begin{Bmatrix} > \\ < \end{Bmatrix} 3 \tag{20.9}$$

Entonces es crítico encontrar $\lambda(t)$. Para hacer esto, comenzamos desde la ecuación de coestado

$$\lambda' = -\frac{\partial H}{\partial y} = -2 - \lambda \quad \text{o sea} \quad \lambda' + \lambda = -2$$

La solución general de esta ecuación es

$$\lambda(t) = Ae^{-t} - 2 \qquad [\text{por } (15.5)]$$

donde A es una constante arbitraria. Usando la condición de transversalidad $\lambda(T) = \lambda(2) = 0$, encontramos que $A = 2e^2$. Entonces la solución definida para λ es

$$\lambda^*(t) = 2e^{2-t} - 2 \tag{20.10}$$

que es una función decreciente de t, que decrece uniformemente a partir del valor inicial $\lambda^*(0) = 2e^2 - 2 = 12.778$ hasta un valor terminal $\lambda^*(2) = 2e^0 - 2 = 0$. Esto significa que λ^* debe pasar por el punto $\lambda = 3$ para un tiempo crítico τ, cuando u óptimo tiene que cambiar de $u^* = 2$ a $u^* = 0$.

Para encontrar este tiempo crítico τ, hacemos $\lambda^*(\tau) = 3$ en (20.10):

$$3 = \lambda^*(\tau) = 2e^{2-\tau} - 2 \quad \text{o sea} \quad e^{2-\tau} = \frac{5}{2} = 2.5$$

Tomando el logaritmo natural de ambos lados, obtenemos

$$\ln e^{2-\tau} = \ln 2.5 \quad \text{o sea} \quad 2 - \tau = \ln 2.5$$

Entonces,

$$\tau = 2 - \ln 2.5 = 1.084 \qquad \text{(aproximadamente)}$$

y resulta que el control óptimo consta de dos fases en el intervalo [0, 2]:

$$\text{Fase 1: } u^*[0, \tau) = 2 \qquad \text{Fase 2: } u^*[\tau, 2] = 0$$

20.2 Condiciones terminales alternativas

¿Qué ocurre con el principio del máximo cuando la condición terminal es diferente de la de (20.1)? En (20.1) enfrentamos una línea terminal vertical, con un tiempo terminal fijo, pero un estado terminal no restringido, como se ilustra en la figura 20.1. El principio del máximo para el problema de maximización requiere que

$$(i) \quad H(t, y, u^*, \lambda) \geq H(t, y, u, \lambda) \quad \text{para todo } t \in [0, T]$$

$$(ii) \quad y' = \frac{\partial H}{\partial \lambda}$$

$$(iii) \quad \lambda' = -\frac{\partial H}{\partial y}$$

con la condición de transversalidad

$$(iv) \quad \lambda(T) = 0$$

Con las condiciones terminales alternativas, las condiciones *i*, *ii* y *iii* van a permanecer iguales, pero la condición *iv* (la condición de transversalidad) debe modificarse oportunamente.

Punto terminal fijo

Si el punto terminal es fijo de modo que la condición terminal es $y(T) = y_T$ con T y y_T dados, entonces la condición terminal misma debe dar la información para determinar una constante. En este caso, no se necesita ninguna condición de transversalidad.

Línea terminal horizontal

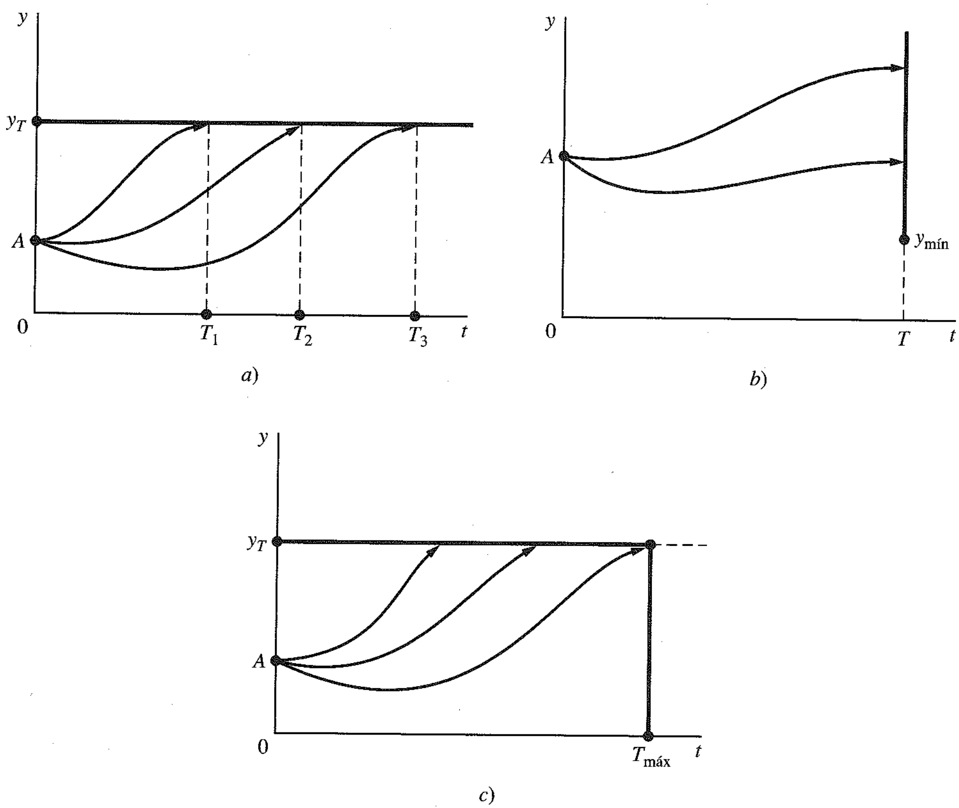
Suponga que el estado terminal es fijo para un nivel objetivo dado y_T, pero el tiempo terminal T es libre, de modo que tenemos la flexibilidad de alcanzar el objetivo apresuradamente o a un paso lento. Entonces tenemos una línea terminal horizontal como se ilustra en la figura 20.3*a*, lo que nos permite escoger entre T_1, T_2, T_3, u otros tiempos terminales para alcanzar el nivel objetivo de y. Para este caso, la condición de transversalidad es una restricción del hamiltoniano (en vez de la variable de coestado) para $t = T$:

$$H_{t=T} = 0 \qquad \textbf{(20.11)}$$

Línea terminal vertical truncada

Si tenemos un tiempo terminal fijo T, y el estado terminal es libre pero está sujeto a la disposición de que $y_T \geq y_{\text{mín}}$, donde $y_{\text{mín}}$ denota un nivel permisible de y mínimo dado, enfrentamos una línea terminal vertical truncada, como se ilustra en la figura 20.3*b*.

FIGURA 20.3



La condición de transversalidad para este caso puede enunciarse como la condición de holgura complementaria encontrada en las condiciones de Kuhn-Tucker:

$$\lambda(T) \geq 0 \qquad y_T \geq y_{\text{mín}} \qquad (y_T - y_{\text{mín}})\,\lambda(T) = 0 \tag{20.12}$$

El enfoque práctico para resolver este tipo de problema es probar primero $\lambda(T) = 0$ como la condición de transversalidad y probar si la y_T^* resultante satisface la restricción $y_T^* \geq y_{\text{mín}}$. Si es así, el problema está resuelto. Si no, entonces se debe tratar el problema como un problema de punto terminal dado con $y_{\text{mín}}$ como el estado terminal.

Línea terminal horizontal truncada

Cuando el estado terminal está fijo en y_T y el tiempo terminal es libre pero está sujeto a la restricción $T \leq T_{\text{máx}}$, donde $T_{\text{máx}}$ denota el tiempo permisible más reciente (una fecha límite) para alcanzar el y_T dado, enfrentamos una línea terminal horizontal truncada, como se ilustra en la figura 20.3*c*. La condición de transversalidad se transforma en

$$H_{t=T_{\text{máx}}} \geq 0 \qquad T \leq T_{\text{máx}} \qquad (T - T_{\text{máx}})H_{t=T_{\text{máx}}} = 0 \tag{20.13}$$

Nuevamente esto aparece en el formato de la condición de holgura complementaria.

El enfoque práctico para resolver ese tipo de problema es probar primero $H_{t=T_{\text{máx}}} = 0$. Si el valor de solución resultante es $T^* \leq T_{\text{máx}}$, entonces el problema está resuelto. Si no, debemos entonces tomar a $T_{\text{máx}}$ como un tiempo terminal fijo, el cual, junto con el y_T dado, define un punto final fijo, y resuelve el problema como un problema de punto final fijo.

Ejemplo 1 En el problema

$$\begin{aligned} &\text{Maximizar} && \int_0^1 (y - u^2)\, dt \\ &\text{sujeto a} && y' = u \\ &\text{y} && y(0) = 2 \qquad y(1) = a \end{aligned}$$

el punto terminal está fijo, aun cuando a $y(1)$ se le asigna aquí un valor paramétrico en vez de numérico.

La función hamiltoniana

$$H = y - u^2 + \lambda u$$

es cóncava en u, de modo que podemos hacer $\partial H/\partial u = 0$ para maximizar H:

$$\frac{\partial H}{\partial u} = -2u + \lambda = 0$$

Entonces,

$$u = \frac{\lambda}{2}$$

lo que muestra que con objeto de despejar $u(t)$, necesitamos despejar primero $\lambda(t)$.

Las dos ecuaciones de movimiento son

$$y'(=u) = \frac{\lambda}{2}$$

$$\lambda'\left(= -\frac{\partial H}{\partial y}\right) = -1$$

La integración directa de la última ecuación arroja

$$\lambda(t) = c_1 - t \qquad (c_1 \text{ arbitrario})$$

lo que implica que

$$y' = \frac{1}{2}c_1 - \frac{1}{2}t$$

Nuevamente, por integración directa, encontramos que

$$y(t) = \frac{c_1}{2}t - \frac{1}{4}t^2 + c_2 \qquad (c_2 \text{ arbitrario})$$

Para hacer definitivas las dos constantes arbitrarias, usamos la condición inicial $y(0) = 2$ y la condición terminal $y(1) = a$. Haciendo $t = 0$ y $t = 1$, sucesivamente, en la ecuación anterior, obtenemos

$$2 = y(0) = c_2 \qquad a = y(1) = \frac{c_1}{2} - \frac{1}{4} + c_2$$

Entonces, $c_2 = 2$ y $c_1 = 2a - \frac{7}{2}$.

Por lo tanto, podemos expresar las trayectorias óptimas de este problema como:

$$y^*(t) = (a - \frac{7}{4})t - \frac{1}{4}t^2 + 2$$

$$\lambda^*(t) = 2a - \frac{7}{2} - t$$

$$u^*(t) = a - \frac{7}{4} - \frac{1}{2}t$$

Ejemplo 2 El problema

$$\begin{array}{lll} \text{Maximizar} & \displaystyle\int_0^T -(t^2 + u^2)\, dt & \\ \text{sujeto a} & y' = u & \\ \text{y} & y(0) 4 \qquad y(T) = 5 & T \text{ libre} \end{array}$$

ejemplifica el caso de la línea terminal horizontal donde el estado terminal está fijo, pero el tiempo de llegada al nivel objetivo de y no tiene restricciones. De hecho, una de nuestras tareas es despejar el valor óptimo de T.

Ya que el hamiltoniano

$$H = -t^2 - u^2 + \lambda u$$

es cóncavo en u, nuevamente podemos maximizar H usando la condición de primer orden

$$\frac{\partial H}{\partial u} = -2u + \lambda = 0$$

que nos da

$$u = \frac{\lambda}{2} \tag{20.14}$$

La concavidad de H hace innecesario verificar la condición de segundo orden; pero, si lo deseamos, es fácil verificar que $\partial^2 H/\partial u^2 = -2 < 0$, suficiente para un máximo de H.

La ecuación de movimiento para λ es

$$\lambda' = -\frac{\partial H}{\partial y} = 0$$

Lo que implica que λ es constante. Pero todavía no podemos determinar su valor exacto en este punto.

Refiriéndonos a la ecuación de movimiento para y,

$$y' = u = \frac{\lambda}{2} \qquad \text{[por (20.14)]}$$

por integración directa, podemos obtener,

$$y(t) = \frac{\lambda}{2}t + c \tag{20.15}$$

Dado que $y(0) = 4$, vemos que $c = 4$. Aún más, la condición de transversalidad (20.11) requiere que

$$H_{t=T} = -T^2 - \frac{\lambda^2}{4} + \frac{\lambda^2}{2} = -T^2 + \frac{\lambda^2}{4} = 0 \qquad \text{[por (20.14)]}$$

Despejando el valor de T y extrayendo la raíz cuadrada positiva, obtenemos

$$T = \frac{\lambda}{2} \tag{20.16}$$

Como λ es constante, también lo es T. Ahora tratamos de encontrar su valor exacto.

Aplicando la especificación de estado terminal $y(T) = 5$ a (20.15), y recordando que $c = 4$, obtenemos

$$y(T) = \frac{\lambda}{2}T + 4 = 5$$

En vista de (20.16), la última ecuación puede reescribirse como $T^2 = 1$. Así, al extraer la raíz cuadrada, podemos determinar al tiempo óptimo de llegada como

$$T^* = 1 \qquad \text{(raíz negativa inaceptable)}$$

De esto, podemos deducir rápidamente que

$$\begin{aligned} \lambda^*(t) &= 2T^* = 2 && \text{[por (20.16)]} \\ u^*(t) &= \frac{\lambda}{2} = 1 && \text{[por (20.14)]} \\ y^*(t) &= t + 4 && \text{[por (20.15)]} \end{aligned}$$

El último resultado muestra que, en este ejemplo, la trayectoria óptima y es una línea recta que va desde el punto inicial dado hasta la línea terminal horizontal.

EJERCICIO 20.2

Encuentre las trayectorias óptimas de las variables de control, de estado y de coestado que

1. Maximizar $\displaystyle\int_0^1 (y - u^2)\, dt$
   sujeto a $y' = u$
   y $y(0) = 2 \qquad y(1)$ libre
2. Maximizar $\displaystyle\int_0^8 6y\, dt$
   sujeto a $y' = y + u$
   $y(0) = 10 \qquad y(8)$ libre
   y $u(t) \in [0, 2]$
3. Maximizar $\displaystyle\int_0^T -(au + bu^2)\, dt$
   sujeto a $y' = y - u$
   y $y(0) = y_0 \qquad y(t)$ libre
4. Maximizar $\displaystyle\int_0^T (yu - u^2 - y^2)\, dt$
   sujeto a $y' = u$
   y $y(0) = y_0 \qquad y(t)$ libre

5. Maximizar $\int_0^{20} -\frac{1}{2}u^2\,dt$

   sujeto a $y' = u$

   y $y(0) = 10 \qquad y(20) = 0$

6. Maximizar $\int_0^{4} 3y\,dt$

   sujeto a $y' = y + u$

   $y(0) = 5 \qquad y(4) \geq 300$

   y $0 \leq u(t) \leq 2$

7. Maximizar $\int_0^{1} -u^2\,dt$

   sujeto a $y' = y + u$

   y $y(0) = 1 \qquad y(1) = 0$

8. Maximizar $\int_1^{2} (y + ut - u^2)\,dt$

   sujeto a $y' = u$

   y $y(1) = 3 \qquad y(2) = 4$

9. Maximizar $\int_0^{2} (2y - 3u - au^2)\,dt$

   sujeto a $y' = u + y$

   y $y(0) = 5 \qquad y(2)$ libre

20.3 Problemas autónomos

En el marco de referencia del problema general de control, la variable t puede intervenir en la función objetivo y se puede enunciar la ecuación directamente. La especificación general

$$\begin{array}{ll} \text{Maximizar} & \displaystyle\int_0^T F(t, y, u)\,dt \\ \text{sujeto a} & y' = f(t, y, u) \\ \text{y} & \text{condiciones de frontera} \end{array}$$

donde t interviene explícitamente en F y f significa que la fecha tiene importancia. Es decir, el valor generado por la actividad $u(t)$ depende no sólo del nivel, sino también de donde tiene lugar exactamente esta actividad.

Los problemas en los cuales t está ausente de la función objetivo y la ecuación se enuncia tal como

$$\begin{array}{ll} \text{Maximizar} & \displaystyle\int_0^T F(y, u)\,dt \\ \text{sujeto a} & y' = f(y, u) \\ \text{y} & \text{condiciones de frontera} \end{array}$$

se llaman *problemas autónomos*. En estos problemas, como el hamiltoniano

$$H = F(y, u) + \lambda f(y, u)$$

no contiene a t como un argumento, las ecuaciones de movimiento son más fáciles de resolver; aún más, son asequibles al uso del análisis del diagrama de fase.

Todavía en otros casos, en un problema que por lo demás es autónomo, el tiempo t interviene en la escena como parte del factor de descuento e^{-rt}, pero no en otro lado, de modo que la función objetivo adopta la forma de

$$\int_0^T G(y, u)e^{-rt}\,dt$$

Hablando estrictamente, este problema es no autónomo. Sin embargo, es fácil convertir el problema en autónomo al emplear el así llamado *hamiltoniano a valor presente*, definido como:

$$H_c \equiv He^{rt} = G(y, u) + \mu f(y, u) \qquad \textbf{(20.17)}$$

donde

$$\mu \equiv \lambda e^{rt} \qquad \textbf{(20.18)}$$

es el *multiplicador de Lagrange a valor presente*. Al centrarse en el valor presente (sin descuento), podemos eliminar t del hamiltoniano original.

Usando H_c en lugar de H, debemos revisar el principio del máximo hasta que:

$$\begin{aligned}
&(i)\ H_c(y, u^*, \mu) \geq H_c(y, u, \mu) \qquad \text{para todo } t \in [0, T] \\
&(ii)\ y' = \frac{\partial H_c}{\partial \mu} \\
&(iii)\ \mu' = -\frac{\partial H_c}{\partial y} + r\mu \qquad \textbf{(20.19)} \\
&(iv)\ \mu(T) = 0 \qquad \text{(para la línea terminal vertical)} \\
&\quad\ \text{o sea } [H_c]_{t=T} = 0 \quad \text{(para la línea terminal horizontal)}
\end{aligned}$$

20.4 Aplicaciones económicas

Maximización de utilidad a lo largo de todo el tiempo de vida

Supongamos que un consumidor tiene la función utilidad $U(C(t))$, donde $C(t)$ es el consumo para el tiempo t. La función utilidad del consumidor es cóncava y tiene las siguientes propiedades:

$$U' > 0 \qquad U'' < 0$$

El consumidor también está dotado con una dotación inicial de riqueza, o capital K_0, con una corriente de ingreso derivada de la dotación capital de acuerdo con lo siguiente:

$$Y = rK$$

donde r es la tasa de interés del mercado. El consumidor usa el ingreso para comprar C. Además, el consumidor puede consumir la inicial de capital. Cualquier ingreso que no se consuma se añade a la existencia de capital como una inversión. Entonces,

$$K' \equiv I = Y - C = rK - C$$

El problema de la maximización de utilidad de tiempo de vida es

$$\begin{aligned} &\text{Maximizar} \quad && \int_0^T U(C(t))e^{-\delta t}\,dt \\ &\text{sujeto a} && K' = rK(t) - C(t) \\ &\text{y} && K(0) = K_0 \qquad K(T) \geq 0 \end{aligned}$$

donde δ es la tasa personal de preferencia de tiempo del consumidor ($\delta \geq 0$). Se supone que $C(t) > 0$ y $K(t) > 0$ para todo t.

El hamiltoniano es

$$H = U(C(t))e^{-\delta t} + \lambda(t)\,[rK(t) - C(t)]$$

donde C es la variable de control, y K es la variable de estado. Ya que $U(C)$ es cóncava, y la restricción es lineal en C, sabemos que el hamiltoniano es cóncavo y la maximización de H puede alcanzarse haciendo simplemente $\partial H/\partial C = 0$. Entonces tenemos

$$\frac{\partial H}{\partial C} = U'(C)e^{-\delta t} - \lambda = 0 \tag{20.20}$$

$$K' = rK(t) - C(t) \tag{20.20'}$$

$$\lambda' = -\frac{\partial H}{\partial K} = -r\lambda \tag{20.20''}$$

La ecuación (20.20) establece que la utilidad marginal de descuento debe igualarse al precio sombra presente de una unidad adicional de capital. Al diferenciar (20.20) respecto a t, obtenemos

$$U''(C)C'e^{-\delta t} - \delta U'(C)e^{-\delta t} = \lambda' \tag{20.21}$$

En vista de (20.20) y (20.20″) tenemos

$$\lambda' = -r\lambda = -rU'(C)e^{-\delta t}$$

que puede sustituirse en (20.21) para dar

$$U''(C)C'(t)e^{-\delta t} - \delta U'(C)e^{-\delta t} = -rU'(C)\,e^{-\delta t}$$

o después de cancelar el factor común $e^{-\delta t}$ y reordenar,

$$\frac{-U''(C(t))}{U'(C(t))}C'(t) = r - \delta$$

Dado que $U' > 0$ y $U'' < 0$, el signo de la derivada $C'(t)$ tiene que ser el mismo que $(r - \delta)$. Por lo tanto, si $r > \delta$, el consumo óptimo va a aumentar con el tiempo; si $r < \delta$, el consumo óptimo va a disminuir con el tiempo.

La solución directa de (20.20″) nos da

$$\lambda(t) = \lambda_0 e^{-rt}$$

donde $\lambda_0 > 0$ es la constante de integración. La combinación de esto con (20.20) nos da

$$U'(C(t)) = \lambda e^{\delta t} = \lambda_0 e^{(\delta - r)t}$$

lo que muestra que la utilidad marginal del consumo va a disminuir en forma óptima con el tiempo si $r > \delta$, pero va a aumentar con el tiempo si $r < \delta$.

Como la condición terminal $K(T) \geq 0$ identifica al problema presente como uno que tiene una línea terminal vertical truncada, la condición de transversalidad apropiada es, por (20.12),

$$\lambda(T) \geq 0 \quad K(T) \geq 0 \quad K(T)\lambda(T) = 0$$

La condición clave es la estipulación de holgura complementaria, que significa ya sea que la dotación de capital K debe agotarse en la fecha terminal, o que el precio sombra del capital λ debe caer hasta cero en la fecha terminal. Por hipótesis, $U'(C) > 0$, la utilidad marginal nunca puede ser cero. Por lo tanto, el valor marginal del capital no puede ser cero. Esto implica que el inventario de capital debe agotarse en forma óptima para la fecha terminal T en este modelo.

Recurso no renovable

Sea $s(t)$ un inventario de un recurso no renovable y $q(t)$ la tasa de extracción para un instante t tal que

$$s' = -q$$

El recurso extraído produce un bien de consumidor final c tal que

$$c = c(q) \qquad \text{donde} \qquad c' > 0, c'' < 0 \tag{20.22}$$

El bien de consumo es el único argumento en la función utilidad de un consumidor representativo con las siguientes propiedades:

$$U = U(c) \qquad \text{donde} \qquad U' > 0, U'' < 0 \tag{20.22'}$$

El consumidor desea maximizar la función utilidad para un intervalo dado $[0, T]$. Como c es una función de q, la tasa de extracción, q va a servir como la variable de control. Por simplicidad, ignoramos el aspecto del descuento respecto al tiempo. Entonces, el problema dinámico es escoger la tasa óptima de extracción que maximice a la función utilidad sujeta sólo a una restricción no negativa sobre la variable de estado $s(t)$, el inventario del recurso no renovable. La formulación es

$$\begin{aligned} &\text{Maximizar} && \int_0^T U(c(q))\, dt \\ &\text{sujeto a} && s' = -q \\ &\text{y} && s(0) = s_0 \qquad s(T) \geq 0 \end{aligned} \tag{20.23}$$

donde s_0 y T están dados.

El hamiltoniano del problema es

$$H = U(c(q)) - \lambda q$$

Como H es cóncava en q por especificaciones del modelo para la función $U(c(q))$, podemos maximizar H haciendo $\partial H/\partial q = 0$:

$$\frac{\partial H}{\partial q} = U'(c(q))c'(q) - \lambda = 0 \tag{20.24}$$

La concavidad de H nos asegura que (20.24) maximiza a H, pero podemos verificar fácilmente la condición de segundo orden y confirmar que $\partial^2 H/\partial q^2$ es negativo.

El principio del máximo estipula que

$$\lambda' = -\frac{\partial H}{\partial s} = 0$$

lo que implica que

$$\lambda(t) = c_0 \qquad \text{una constante} \tag{20.25}$$

Para determinar c_0, hacemos referencia a las condiciones de transversalidad. Dado que el modelo especifica $K(T) \geq 0$, tiene una línea terminal vertical truncada, de modo que aplica (20.12):

$$\lambda(T) \geq 0 \qquad s(T) \geq 0 \qquad s(T)\lambda(T) = 0$$

En las aplicaciones prácticas, el paso inicial es probar $\lambda(T) = 0$, despejar q, y ver si la solución va a funcionar. Puesto que $\lambda(T)$ es una constante, el probar $\lambda(T) = 0$ implica que $\lambda(t) = 0$ para todo t, y $\partial H/\partial q$ en (20.24) se reduce a

$$U'(c)c'(q) = 0$$

de donde (en principio) podemos despejar q. Como t no es un argumento explícito de U o de c, la trayectoria de solución para q es constante respecto al tiempo:

$$q^*(t) = q^*$$

Ahora revisamos si q^* satisface la restricción $s(T) \geq 0$. Si q^* es una constante, entonces la ecuación de movimiento

$$s' = -q$$

puede integrarse rápidamente, dando

$$s(t) = -qt + c_1 \qquad [c_1 = \text{constante de integración}]$$

Usando la condición inicial $s(0) = s_0$ da una solución para la constante de integración

$$c_1 = s_0$$

y la trayectoria de estado óptima es

$$s(t) = s_0 - q^*t \tag{20.26}$$

Sin especificar las formas funcionales de U y c, no podemos encontrar ninguna solución numérica para q^*. Sin embargo, a partir de las condiciones de transversalidad, podemos concluir que si $s(T) \geq 0$, entonces q^* es aceptable tal como se obtiene en la solución. Pero si $s(T) < 0$ para el q^* dado, entonces la tasa de extracción es demasiado alta y necesitamos encontrar una solución diferente. Dado que falló la solución de prueba $\lambda(T) = 0$, ahora tomamos la alternativa de $\lambda(T) > 0$. Aun en este caso, sin embargo, λ es todavía una constante mediante (20.25). Y (20.24) puede suministrar aún (en principio) un valor solución q_2^* constante pero diferente. Se sigue que (20.26) permanece válido. Pero esta vez, con $\lambda(T) > 0$, la condición de transversalidad (20.12) exige que $s(T) = 0$, o en vista de (20.26),

$$s_0 - q_2^*T = 0$$

Entonces podemos escribir la tasa de extracción óptima revisada (constante) como

$$q_2^* = \frac{s_0}{T}$$

Este nuevo valor solución debe representar una tasa de extracción más baja que no viole la condición de frontera $s(T) \geq 0$.

EJERCICIO 20.4

1. Maximizar $\int_0^T (K - \alpha K^2 - I^2)dt \qquad (\alpha > 0)$

 sujeto a $K' = I - \delta K \qquad (\delta > 0)$

 y $K(0) = K_0 \qquad K(T)$ libre

2. Resuelva el siguiente problema de recurso no renovable para la trayectoria de extracción óptima:

 Maximizar $\int_0^T \ln(q)e^{-\delta t}dt$

 sujeto a $s' = -q$

 y $s(0) = s_o \qquad s(t) \geq 0$

20.5 Horizonte de tiempo infinito

En esta sección introducimos el problema de la optimización dinámica para un periodo infinito de planeación. Los modelos de horizonte de tiempo infinito tienden a introducir complejidades respecto a las condiciones de transversalidad y las trayectorias de tiempo óptimo que difieren de las desarrolladas antes. En vez de enfrentar estos aspectos aquí, vamos a ilustrar la metodología de estos modelos con una versión del *modelo neoclásico de crecimiento óptimo*.

Modelo neoclásico de crecimiento óptimo

La función de producción neoclásica estándar expresa el producto Y como una función de dos insumos: mano de obra L y capital K. Su forma general es

$$Y = Y(K, L)$$

donde $Y(K, L)$ es una función linealmente homogénea con las propiedades

$$Y_L > 0 \qquad Y_K > 0 \qquad Y_{LL} < 0 \qquad Y_{KK} < 0$$

Reescribiendo la función de producción en términos per cápita da

$$y = \phi(k) \qquad \text{con} \quad \phi'(k) > 0 \qquad \text{y} \qquad \phi''(k) < 0$$

donde $y = Y/L$ y $k = K/L$. El producto total Y se asigna al consumo C o a la inversión bruta I. Sea δ la tasa de depreciación del inventario de capital K. Entonces, la inversión neta o los cambios del inventario de capital pueden escribirse como

$$K' = I - \delta K = Y - C - \delta K$$

Denotando el consumo per cápita como $c \equiv C/L$, podemos escribir como

$$\frac{1}{L}K' = y - c - \delta k \qquad \textbf{(20.27)}$$

El miembro derecho de (20.27) está en términos per cápita, pero el miembro izquierdo no lo está. Para unificar observamos que

$$K' = \frac{dk}{dt} = \frac{d}{dt}(kL) = k\frac{dL}{dt} + L\frac{dk}{dt} \tag{20.28}$$

Si la tasa de crecimiento de la población es[3]

$$\frac{dL/dt}{L} = n \text{ de modo que } \frac{dL}{dt} = nL$$

entonces (20.28) se transforma en

$$K' = knL + Lk' \quad \text{o sea} \quad \frac{1}{L}K' = kn + k'$$

La sustitución de esto en (20.27) la transforma en una ecuación totalmente en términos per cápita:

$$k' = y - c - (n+\delta)k = \phi(k) - c - (n+\delta)k \tag{20.27$'$}$$

Sea $U(c)$ la función de bienestar social (expresada en términos per cápita), donde

$$U'(c) > 0 \qquad \text{y} \qquad U''(c) < 0$$

y para eliminar las soluciones de esquina, también suponemos

$$U'(c) \to \infty \text{ cuando } c \to 0$$

$$\text{y} \qquad U'(c) \to 0 \text{ cuando } c \to \infty$$

Si ρ denota la tasa de descuento social y la población inicial se normaliza a uno, la función objetivo puede expresarse como

$$V = \int_0^\infty U(c)e^{-\rho t}L_0e^{nt}\,dt = \int_0^\infty U(c)e^{-(\rho-n)t}\,dt$$
$$= \int_0^\infty U(c)e^{-rt}\,dt \qquad \text{donde } r = \rho - n$$

En esta versión del modelo neoclásico de crecimiento óptimo, la utilidad se sopesa por una población que crece continuamente a una tasa n. Sin embargo, si $r = \rho - n > 0$, entonces el modelo no es diferente matemáticamente de uno sin pesos de población pero con una tasa de descuento positiva r.

Ahora el problema óptimo de crecimiento podemos enunciarlo como

$$\begin{aligned} &\text{Maximizar} && \int_0^\infty U(c)e^{-rt}\,dt \\ &\text{sujeto a} && k' = \phi(k) - c - (n+\delta)k \\ & && k(0) = k_0 \\ &\text{y} && 0 \le c(t) \le \phi(k) \end{aligned} \tag{20.29}$$

donde k es la variable de estado y c es la variable de control.

[3] En este modelo suponemos que la fuerza laboral y la población son una y la misma.

El hamiltoniano para el problema es

$$H = U(c)e^{-rt} + \lambda[\phi(k) - c - (n+\delta)k]$$

Dado que H es cóncavo en c, el máximo de H corresponde a una solución interior en la región de control $[0 < c < f(k)]$, y por lo tanto podemos encontrar el máximo de H a partir de

$$\frac{\partial H}{\partial c} = U'(c)e^{-rt} - \lambda = 0$$

o sea

$$U'(c) = \lambda e^{rt} \tag{20.30}$$

La interpretación económica de (20.30) es que, a lo largo de la trayectoria óptima, la utilidad marginal del consumo per cápita debe ser igual al precio sombra del capital (λ) sopesado por e^{rt}. Al verificar las condiciones de segundo orden, encontramos

$$\frac{\partial^2 H}{\partial c^2} = U''(c)e^{-rt} < 0$$

Por lo tanto se maximiza el hamiltoniano.

A partir del principio del máximo, tenemos dos ecuaciones de movimiento

$$k' = \frac{\partial H}{\partial \lambda} = \phi(k) - c - (n+\delta)k$$

y

$$\lambda' = -\frac{\partial H}{\partial k} = -\lambda[\phi'(k) - (n+\delta)]$$

Las dos ecuaciones de movimiento combinadas con $U'(c) = \lambda e^{rt}$ deben en principio definir una solución para c, k, λ. Sin embargo, para este nivel de generalidad no podemos hacer más que acometer el análisis cualitativo del modelo. Cualquier otro aspecto adicional requeriría formas específicas de ambas funciones de utilidad y de producción.

El hamiltoniano a valor presente

Como el modelo anterior es un ejemplo de un problema autónomo (t no es un argumento separado en la función utilidad o en la ecuación de estado, sino que aparece sólo en el factor de descuento), podemos usar el hamiltoniano a valor presente escrito como

$$H_c = He^{rt} = U(c) + \mu\,[\phi(k) - c - (n+\delta)k] \qquad \text{[vea (20.17)]}$$

donde $\mu = \lambda e^{rt}$.

El principio del máximo requiere que

$$\frac{\partial H_c}{\partial c} = U'(c) - \mu = 0 \quad \text{o sea} \quad \mu = U'(c) \tag{20.31}$$

$$k' = \frac{\partial H_c}{\partial \mu} = \phi(k) - c - (n+\delta)k \tag{20.31'}$$

$$\begin{aligned}\mu' &= -\frac{\partial H_c}{\partial k} + r\mu = -\mu[\phi'(k) - (n+\delta)] + r\mu \\ &= -\mu[\phi'(k) - (n+\delta+r)]\end{aligned} \tag{20.31''}$$

Las ecuaciones (20.31′) y (20.31″) constituyen un sistema de ecuaciones diferenciales autónomas. Esto hace posible un análisis cualitativo mediante el diagrama de fase.

Construcción de un diagrama de fase

Las variables de las ecuaciones diferenciales (20.31′) y (20.31″) son k y μ. Como (20.31) implica una función de c, a saber $U'(c)$, en vez de c sola, sería más sencillo construir un diagrama de fase en el espacio kc en vez del espacio $k\mu$. Para hacer esto, trataremos de eliminar μ. Como $\mu = U'(c)$, mediante (20.31), la diferenciación respecto a t nos da

$$\mu' = U''(c)c'$$

La sustitución de estas expresiones de μ y μ' en (20.31″) da

$$c' = -\frac{U'(c)}{U''(c)}[\phi'(k) - (n + \delta + r)]$$

que es una ecuación diferencial en c. Ahora tenemos el sistema de ecuaciones diferenciales autónomas

$$k' = \phi(k) - c - (n + \delta)k \qquad \textbf{(20.31′)}$$

y

$$c' = -\frac{U'(c)}{U''(c)}[\phi'(k) - (n + \delta + r)] \qquad \textbf{(20.32)}$$

Para construir el diagrama de fase en el espacio kc, dibujamos primero las curvas $k' = 0$ y $c' = 0$ que se definen por

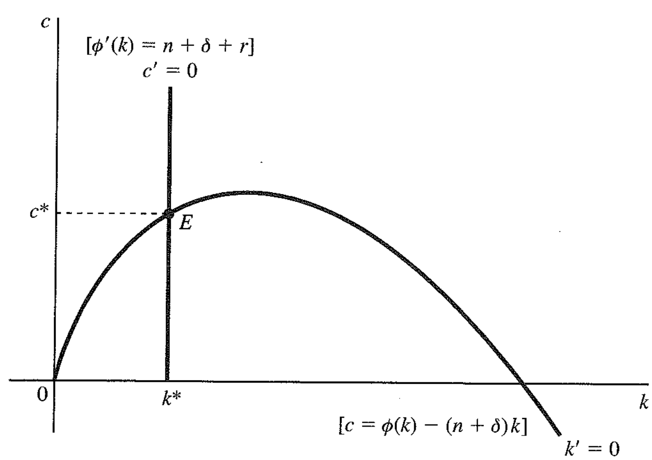
$$c = \phi(k) - (n + \delta)k \qquad (k' = 0) \qquad \textbf{(20.33)}$$

y

$$\phi'(k) = n + \delta + r \qquad (c' = 0) \qquad \textbf{(20.34)}$$

Estas dos curvas se ilustran en la figura 20.4. La ecuación para la curva $k' = 0$, (20.33), tiene la misma estructura que la ecuación fundamental del modelo de crecimiento de Solow, (15.30). Entonces, la curva $k' = 0$ tiene la misma forma general que la de la figura 15.5*b*. Por otro lado, la curva $c' = 0$, se grafica como una línea vertical, ya que dadas las especificaciones del modelo $\phi'(k) > 0$ y $\phi''(k) < 0$, $\phi(k)$ se asocia con una curva cóncava con pendiente hacia arriba, con una pendiente diferente para cada uno de los puntos en la curva, de modo que sólo un valor único de k puede satisfacer (20.34). La intersección de las dos curvas en el punto E determina los valores de equilibrio intertemporal de k y c, ya que para el punto E, ni k ni c van

FIGURA 20.4



a cambiar de valor respecto al tiempo, lo que conduce a un *estado permanente*. Podemos rotular estos valores como $\bar{k}$ y $\bar{c}$ para los valores de equilibrio intertemporal, pero vamos a rotularlos como k^* y c^*, ya que también representan a los valores de equilibrio para el crecimiento óptimo.

Análisis del diagrama de fase

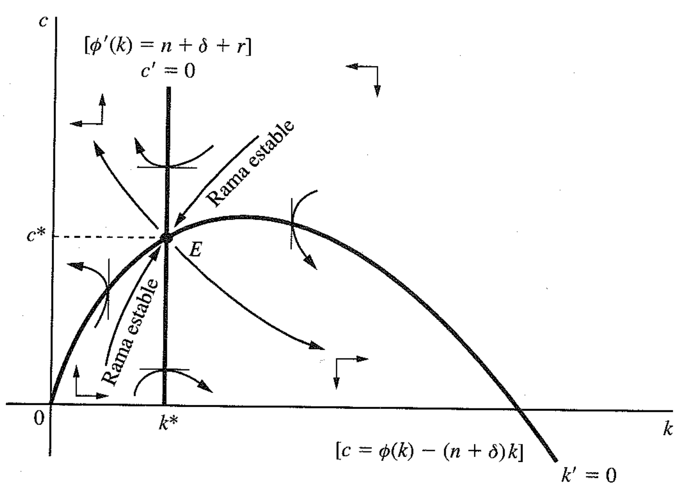
El punto de intersección E de la figura 20.4 nos da un estado permanente único. ¿Pero qué ocurre si inicialmente nos encontramos en algún punto que no sea E? Regresando a nuestro sistema de ecuaciones diferenciales de primer orden (20.31′) y (20.32), podemos deducir que

$$\frac{\partial k'}{\partial c} = -1 < 0 \qquad \text{y} \qquad \frac{\partial c'}{\partial k} = -\frac{U'(c)}{U''(c)}\phi''(k) < 0$$

Como $\partial k'/\partial c < 0$, todos los puntos debajo de la curva $k' = 0$ se caracterizan por $k' > 0$ y todos los puntos arriba de la curva por $k' < 0$. En forma similar, dado que $\partial c'/\partial k < 0$, todos los puntos a la izquierda de la línea $c' = 0$ se caracterizan por $c' > 0$ y todos los puntos a la derecha de la línea por $c' < 0$. Entonces, la curva $k' = 0$ y la línea $c' = 0$ dividen el espacio de fase en cuatro regiones, cada una con sus propios pares de signos de c' y k'. Éstos se reflejan en la figura 20.5 mediante las flechas direccionales en ángulo recto en cada región.

Las líneas de corriente que siguen las flechas direccionales en cada región nos dicen que el estado permanente en el punto E es un punto silla. Si tenemos un punto inicial que está situado en una de las dos ramas estables del punto silla, la dinámica del sistema nos va a conducir al punto E. Pero cualquier punto inicial que no esté situado sobre una rama estable nos va a obligar, ya sea a bordear alrededor del punto E, sin alcanzarlo nunca, o a alejarse consistentemente de él. Si seguimos las líneas de corriente de los últimos casos, finalmente terminaremos (cuando $t \to \infty$) ya sea con $k = 0$ (agotamiento del capital) o $c = 0$ (consumo per cápita que disminuye hasta cero), siendo los dos económicamente inaceptables. Así, la única alternativa viable es escoger al par (k, c) a fin de ubicar nuestra economía sobre una rama estable —por así decirlo, un "camino de ladrillos amarillos"— que nos lleve al estado permanente en E. No hemos hablado explícitamente acerca de la condición de transversalidad, pero si lo hubiéramos hecho, nos hubiera guiado al estado permanente en E, donde el consumo per cápita puede conservarse a un nivel constante a partir de ese momento.

FIGURA 20.5



20.6 Limitaciones del análisis dinámico

El análisis estático presentado en la parte 2 de este volumen trató sólo el asunto de dónde estará la posición de equilibrio para ciertas condiciones de un modelo. La principal interrogante era: ¿Qué valores de las variables, si se *alcanzaran*, van a tender a perpetuarse por sí solas? Pero la *posibilidad de alcanzar* la posición de equilibrio se dio como un hecho. Cuando avanzamos al campo de la estática comparativa, en la parte 3, la interrogante central se desplazó hacia un problema más interesante: ¿cómo va a desplazarse la posición de equilibrio en respuesta a cierto cambio en un parámetro? Pero el aspecto de asequibilidad lo dejamos de lado otra vez. Fue ahora que llegamos al análisis dinámico en la parte 5 cuando abordamos frente a frente el aspecto de la asequibilidad. Aquí preguntamos específicamente: si inicialmente estamos lejos de una posición de equilibrio —digamos, debido a un cambio en un parámetro que nos lleva fuera del equilibrio— ¿tenderán las diferentes fuerzas del modelo a dirigirnos hacia la nueva posición de equilibrio? Aún más, en un análisis dinámico, también aprendemos el carácter específico de la trayectoria (si es permanente, fluctuante, u oscilatoria) que la variable va a seguir camino del equilibrio (si es el caso). Por lo tanto, la importancia del análisis dinámico debe ser evidente.

Sin embargo, al terminar esta discusión, también debemos reconocer las limitaciones del análisis dinámico. Simplemente, para hacer que el análisis sea manejable, con frecuencia los modelos dinámicos se formulan en términos de ecuaciones lineales. Aun cuando podemos ganar simplicidad de esta manera, la hipótesis de linealidad acarrea en muchos casos un sacrificio considerable de realismo. Dado que una trayectoria de tiempo que sea aplicable a un modelo lineal no siempre va a aproximarse a la de la contraparte no lineal, como lo hemos visto en el ejemplo de precios máximos en la sección 17.6, debemos tener cuidado en la interpretación y aplicación de los resultados de los modelos dinámicos lineales. Sin embargo, a este respecto, el enfoque cualitativo gráfico puede prestar un servicio muy valioso, ya que en condiciones muy generales nos permite incorporar la no linealidad a un modelo sin añadir una complejidad excesiva al análisis.

Otra desventaja que generalmente se encuentra en los modelos de la economía dinámica es el uso de coeficientes constantes en las ecuaciones diferenciales o de diferencias. Dado que el papel principal de los coeficientes es especificar los parámetros del modelo, el carácter constante de los coeficientes —nuevamente supuestos por la conveniencia de la manejabilidad matemática— sirve esencialmente para "congelar" el ambiente económico del problema investigado. En otras palabras, esto significa que el ajuste endógeno del modelo se está estudiando en una especie de vacío económico, de modo que no se permite la intrusión de factores exógenos. Por supuesto que en ciertos casos este problema puede no ser muy grave, ya que ciertamente muchos parámetros económicos tienden a permanecer relativamente constantes para periodos largos. Y en algunos otros casos, podremos encarar un tipo dinámico comparativo de análisis, para ver cómo la trayectoria de tiempo de una variable va a ser afectada por un cambio en ciertos parámetros. Sin embargo, cuando estamos interpretando una trayectoria que se proyecta hacia el futuro distante, siempre debemos tener cuidado de no confiarnos demasiado acerca de la validez de la trayectoria en sus tramos más alejados, si se han hecho hipótesis del carácter de constante por simplicidad.

Por supuesto que usted se da cuenta de que el señalamiento de las limitaciones como lo hemos hecho aquí no conlleva la intención de desprestigiar al análisis dinámico. Debemos recordar que se ha mostrado que cada tipo de análisis presentado hasta este momento tiene su propia clase de limitaciones. Por lo tanto, siempre que se interprete debidamente y se aplique con propiedad, el análisis dinámico —como cualquier otro tipo de análisis— puede desempeñar una parte importante en el estudio de los fenómenos económicos. En particular, las técnicas del análisis dinámico nos han permitido extender el estudio de la optimización hasta el campo de la optimización dinámica en este capítulo, en el cual la solución que buscamos ya no es un estado óptimo estático, sino una trayectoria completa óptima de tiempo.

El alfabeto griego

A	α	alfa
B	β	beta
Γ	γ	gamma
Δ	δ	delta
E	ε	épsilon
Z	ζ	zeta
H	η	eta
Θ	θ	theta
I	ι	iota
K	κ	kappa
Λ	λ	lambda
M	μ	mi
N	ν	ni
Ξ	ξ	xi
O	o	ómicron
Π	π	pi
P	ρ	rho
Σ	σ	sigma
T	τ	tau
Υ	υ	ípsilon
Φ	ϕ (o φ)	fi
X	χ	ji
Ψ	ψ	psi
Ω	ω	omega

Símbolos matemáticos

1. **Conjuntos**

$a \in S$	a es un elemento del (pertenece a) conjunto S
$b \notin S$	b no es un elemento del conjunto S
$S \subset T$	el conjunto S es un subconjunto del (está contenido en) conjunto S
$T \supset S$	el conjunto T contiene al conjunto S
$A \cup B$	unión del conjunto A y el conjunto B
$A \cap B$	intersección del conjunto A y del conjunto B
$\tilde{S}$	complemento del conjunto S
$\{\ \}$ o $\varnothing$	conjunto nulo (conjunto vacío)
$\{a, b, c\}$	conjunto con los elementos a, b y c
$\{x \mid x \text{ tiene la propiedad } P\}$	conjunto de todos los objetos con la propiedad P
$\min\{a, b, c\}$	el elemento más pequeño del conjunto especificado
R	el conjunto de todos los números reales
R^2	el espacio real bidimensional
R^n	el espacio real de n dimensiones
(x, y)	par ordenado
(x, y, z)	terna ordenada
(a, b)	intervalo abierto de a a b
$[a, b]$	intervalo cerrado de a a b

2. **Matrices y determinantes**

A' o A^T	transpuesta de la matriz A
A^{-1}	inversa de la matriz A
$\lvert A \rvert$	determinante de la matriz A
$\lvert J \rvert$	determinante jacobiano
$\lvert H \rvert$	determinante hessiano
$\lvert \overline{H} \rvert$	determinante hessiano orlado
$r(A)$	rango de la matriz A
$\mathrm{tr}A$	traza de A

0	matriz nula (matriz cero)
$u \cdot v$	producto interno (producto punto) de los vectores u y v
$u'v$	producto escalar de dos vectores

3. **Cálculo**

Dada $y = f(x)$, una función de una sola variable x:

$\lim\limits_{x\to\infty} f(x)$	límite de $f(x)$ cuando x se aproxima a infinito
dy	primera diferencial de y
d^2y	segunda diferencial de y
$\dfrac{dy}{dx}$ o $f'(x)$	primera derivada de la función $y = f(x)$
$\left.\dfrac{dy}{dx}\right\vert_{x=x_0}$ o $f'(x_0)$	primera derivada evaluada en $x = x_0$
$\dfrac{d^2y}{dx^2}$ o $f''(x)$	segunda derivada de $y = f(x)$
$\dfrac{d^ny}{dx^n}$ o $f^{(n)}(x)$	derivada enésima de $y = f(x)$
$\int f(x)\,dx$	integral indefinida de $f(x)$
$\int_a^b f(x)\,dx$	integral definida de $f(x)$ de $x = a$ a $x = b$

Dada la función $y = f(x_1, x_2, \ldots, x_n)$:

$\dfrac{\partial y}{\partial x_i}$ o f_i	derivada parcial de f respecto a x_i
$\nabla f \equiv \text{grad } f$	gradiente de f
$\dfrac{dy}{dx_i}$	derivada total de f respecto a x_i
$\dfrac{\S y}{\S x_i}$	derivada total parcial de f respecto a x_i

4. **Ecuaciones diferenciales y en diferencias**

$\dot{y} \equiv \dfrac{dy}{dt}$	derivada de y respecto al tiempo
Δy_t	primera diferencia de y_t
$\Delta^2 y_t$	segunda diferencia de y_t
y_p	integral particular
y_c	función complementaria

5. **Otros**

$\sum\limits_{i=1}^{n} x_i$	la suma de x_i cuando i varía de 1 a n

$p \Rightarrow q$	p solamente si q (p implica a q)
$p \Leftarrow q$	p si q (p está implicado por q)
$p \Leftrightarrow q$	p si y sólo si q
iff	si y sólo si
$\lvert m \rvert$	valor absoluto del número m
$n!$	n factorial $\equiv n(n-1)(n-2)\cdots(3)(2)(1)$
$\log_b x$	logaritmo de x con base b
$\log_e x$ o $\ln x$	logaritmo natural de x (con base e)
e	base de los logaritmos naturales y de las funciones exponenciales naturales
$\operatorname{sen} \theta$	función seno de θ
$\cos \theta$	función coseno de θ
R_n	término residual cuando la serie de Taylor implica un polinomio de grado n

Breve lista de lecturas

Abadie, J. (ed.): *Nonlinear Programming,* North-Holland Publishing Company, Amsterdam, 1967. (Una antología de artículos sobre ciertos aspectos teóricos y computacionales de la programación no lineal; el capítulo 2, de Abadie, trata del teorema de Kuhn-Tucker en relación con la clasificación de la restricción.)

Allen, R. G. D.: *Mathematical Analysis for Economists,* Macmillan & Co., Ltd., Londres, 1938. (Una clara exposición del cálculo diferencial e integral; se estudian los determinantes, pero no las matrices; no se incluye la teoría de conjuntos ni la programación matemática.)

______: *Mathematical Economics,* 2ª ed., St. Martin's Press, Inc., Nueva York, 1959. (Estudia una gran cantidad de modelos de economía matemática; explica las ecuaciones lineales diferenciales y en diferencias y el álgebra matricial.)

Almon, C.: *Matrix Methods in Economics,* Addison-Wesley Publishing Company, Inc., Reading, Mass., 1967. (Se estudian los métodos matriciales en relación con los sistemas de ecuaciones lineales, los modelos de insumo-producto, la programación lineal y la programación no lineal. También se cubren las raíces características y los vectores característicos.)

Baldani, J., J. Bradfield, y R. Turner: *Mathematical Economics,* The Dryden Press, Orlando, 1996.

Baumol, W. J.: *Economic Dynamics: An Introduction,* 3ª ed., The Macmillan Company, Nueva York, 1970 (la parte IV ofrece una explicación lúcida de ecuaciones diferenciales simples; la parte V trata sobre ecuaciones en diferencias simultáneas; las ecuaciones diferenciales se estudian sólo brevemente).

Braun, M.: *Differential Equations and Their Applications: An Introduction to Applied Mathematics,* 4ª ed., Springer-Verlag, Inc., Nueva York, 1993. (Contiene aplicaciones interesantes de las ecuaciones diferenciales, tales como la detección de las falsificaciones de las obras de arte, la diseminación de las epidemias, la carrera armamentista y la disposición de los residuos nucleares.)

Burmeister, E., y A. R. Dobell: *Mathematical Theories of Economic Growth,* The Macmillan Company, Nueva York, 1970. (Una exposición completa de los modelos de crecimiento con diferentes grados de complejidad.)

Chiang, Alpha C.: *Elements of Dynamic Optimization,* McGraw-Hill Book Company, 1992, ahora publicado por Waveland Press, Inc., Prospect Heights, Ill.

Clark, Colin W.: *Mathematical Bioeconomics: The Optimal Management of Renewable Resources,* 2ª ed., John Wiley & Sons, Inc., Toronto, 1990. (Una explicación completa de la teoría de control óptimo y su uso tanto en recursos renovables como no renovables.)

Coddington, E. A., y N. Levinson: *Theory of Ordinary Differential Equations,* McGraw-Hill Book Company, Nueva York, 1955. (Un texto matemático básico sobre ecuaciones diferenciales.)

Courant, R.: *Differential and Integral Calculus* (trad. E. J. McShane), Interscience Publishers, Inc., Nueva York, vol. I, 2ª ed., 1937, vol. II, 1936. (Un tratado clásico del cálculo.)

______, y F. John: *Introduction to Calculus and Analysis,* Interscience Publishers, Inc., Nueva York, vol. I, 1965, vol. II, 1974. (Una versión actualizada del título anterior.)

Dorfman, R., P. A. Samuelson, y R. M. Solow: *Linear Programming and Economic Analysis,* McGraw-Hill Book Company, Nueva York, 1958. (Un tratamiento detallado de la programación lineal, la teoría de juegos y el análisis de insumo-producto.)

Franklin, J.: *Methods of Mathematical Economics: Linear and Nonlinear Programming, Fixed-Point Theorems,* Springer-Verlag, Inc., Nueva York, 1980. (Una presentación fascinante de la programación matemática.)

Frisch, R.: *Maxima and Minima: Theory and Economic Applications* (en colaboración con A. Nataf), Rand McNally & Company, Chicago, Ill., 1966. (Un tratamiento completo de los problemas de extremos, hecho principalmente según la tradición clásica.)

Goldberg, S.: *Introduction to Difference Equations,* John Wiley & Sons, Inc., Nueva York, 1958. (Con aplicaciones en la economía.)

Hadley, G.: *Linear Algebra,* Addison-Wesley Publishing Company, Inc., Reading, Mass., 1961. (Cubre matrices, determinantes, conjuntos convexos, etcétera).

______: *Linear Programming,* Addison-Wesley Publishing Company, Inc., Reading, Mass., 1962. (Una exposición escrita con claridad y orientada hacia las matemáticas.)

______: *Nonlinear and Dynamic Programming,* Addison-Wesley Publishing Company, Inc., Reading, Mass., 1964. (Cubre la programación no lineal, la programación estocástica, la programación entera y la programación dinámica; se enfatizan los aspectos computacionales.)

Halmos, P. R.: *Naive Set Theory,* D. Van Nostrand Company, Inc., Princeton, N.J., 1960. (Una introducción informal y por tanto legible de las bases de la teoría de conjuntos.)

Hands, D. Wade: *Introductory Mathematical Economics,* 2ª ed., Oxford University Press, Nueva York, 2004.

Henderson, J. M., y R. E. Quandt: *Microeconomic Theory: A Mathematical Approach,* 3ª ed., McGraw-Hill Book Company, Nueva York, 1980. (Un tratamiento matemático comprensivo de temas de microeconomía.)

Hoy, M., J. Livernois, C. McKenna, R. Rees, y T. Stengos: *Mathematics for Economics,* 2ª ed., The MIT Press, Cambridge, Mass., 2001.

Intriligator, M. D.: *Mathematical Optimization and Economic Theory,* Prentice Hall, Inc., Englewood Cliffs, N.J., 1971. (Un estudio completo de los métodos de optimización, incluyendo las técnicas clásicas, la programación lineal y no lineal, y la optimización dinámica; también aplicaciones a las teorías del consumidor y de la firma, el equilibrio general y el bienestar en la economía y las teorías del crecimiento.)

Kemeny, J. G., J. L. Snell, y G. L. Thompson: *Introduction to Finite Mathematics,* 3ª ed., Prentice Hall, Inc., Englewood Cliffs, N.J., 1974. (Cubre temas tales como conjuntos, matrices, probabilidad y la programación lineal.)

Klein, Michael W.: *Mathematical Methods for Economics,* 2ª ed., Addison-Wesley Publishing Company, Inc., Reading, Mass., 2002.

Koo, D.: *Elements of Optimization: With Applications in Economics and Business,* Springer-Verlag, Inc., Nueva York, 1977. (Un claro estudio de los métodos clásicos de optimización, la programación matemática así como la teoría de control óptimo.)

Koopmans, T. C. (ed.): *Activity Analysis of Production and Allocation,* John Wiley & Sons, Inc., Nueva York, 1951, reimpreso por Yale University Press, 1972. (Contiene varios artículos importantes sobre programación lineal y análisis de la actividad.)

______: *Three Essays on the State of Economic Science,* McGraw-Hill Book Company, Nueva York, 1957. (El primer ensayo contiene una buena exposición de los conjuntos convexos; el tercer ensayo estudia la interacción de las herramientas y de los problemas en la economía.)

Lambert, Peter J., *Advanced Mathematics for Economists: Static and Dynamic Optimization,* Blackwell Publishers, Nueva York, 1985.

Leontief, W. W.: *The Structure of American Economy, 1919-1939,* 2ª ed., Oxford University Press, Fair Lawn, N.J., 1951. (El trabajo pionero en el análisis de insumo-producto.)

Samuelson, P. A.: *Foundations of Economic Analysis,* Harvard University Press, Cambridge, Mass., 1947. (Un clásico de la economía matemática, pero muy difícil de leer.)

Silberberg, Eugene, y Wing Suen: *The Structure of Economics: A Mathematical Analysis,* 3ª ed., McGraw-Hill Book Company, Nueva York, 2001. (Principalmente con un enfoque macroeconómico, este libro contiene un estudio muy profundo del teorema de la envolvente, y una amplia variedad de aplicaciones.)

Sydsæter, Knut, y Peter Hammond: *Essential Mathematics for Economic Analysis,* Prentice Hall, Inc., Londres, 2002.

Takayama, A.: *Mathematical Economics,* 2ª ed., The Dryden Press, Hinsdale, Ill., 1985. (Ofrece un tratamiento comprensivo de la teoría económica en términos matemáticos, con énfasis en dos temas específicos: el equilibrio competitivo y el crecimiento económico.)

Thomas, G. B., y R. L. Finney: *Calculus and Analytic Geometry,* 9ª ed., Addison-Wesley Publishing Company, Inc., Reading, Mass., 1996. (Una introducción al cálculo muy bien escrita.)

Respuestas a ejercicios seleccionados

Ejercicio 2.3

1. (*a*) $\{x \mid x > 34\}$

3. (*a*) $\{2, 4, 6, 7\}$ (*c*) $\{2, 6\}$ (*e*) $\{2\}$

8. Hay 16 subconjuntos.

9. *Sugerencia*: distinga entre los símbolos $\notin$ y $\not\subset$.

Ejercicio 2.4

1. (*a*) $\{(3, a), (3, b), (6, a), (6, b), (9, a), (9, b)\}$

3. Ninguna.

5. Imagen= $\{y \mid 8 \leq y \leq 32\}$

Ejercicio 2.5

2. (*a*) y (*b*) difieren en el signo de la medida de la pendiente; (*a*) y (*c*) difieren en la intersección vertical.

4. Cuando se permiten valores negativos, se tiene que usar también el cuadrante III.

5. (*a*) x^{19}

6. (*a*) x^6

Ejercicio 3.2

1. $P^* = 2\frac{3}{11}$, y $Q^* = 14\frac{2}{11}$

3. *Nota*: en 2(*a*), $c = 10$ (no 6).

5. *Sugerencia*: $b + d = 0$ implica $d = -b$.

Ejercicio 3.3

1. (*a*) $x_1^* = 5$, y $x_2^* = 3$

3. (*a*) $(x - 6)(x + 1)(x - 3) = 0$, o bien, $x^3 - 8x^2 + 9x + 18 = 0$

5. (*a*) $-1, 2,$ y 3 (*c*) $-1, \frac{1}{2}$ y $-\frac{1}{4}$

Ejercicio 3.4

3. $P_1^* = 3\frac{6}{17}$ $\quad P_2^* = 3\frac{8}{17}$ $\quad Q_1^* = 11\frac{7}{17}$ $\quad Q_2^* = 8\frac{7}{17}$

Ejercicio 3.5

1. (*b*) $Y^* = (a - bd + I_0 + G_0)/[1 - b(1-t)]$

$T^* = [d(1-b) + t(a + I_0 + G_0)]/[1 - b(1-t)]$

$C^* = [a - bd + b(1-t)(I_0 + G_0)]/[1 - b(1-t)]$

3. *Sugerencia*: después de sustituir las dos últimas ecuaciones en la primera, considere la ecuación resultante como una ecuación cuadrática en la variable $w \equiv Y^{1/2}$. Sólo una raíz es aceptable, $w_1^* = 11$, dando $Y^* = 121$ y $C^* = 91$. La otra raíz conduce a una C^* negativa.

Ejercicio 4.1

1. Los componentes del vector (columna) de constantes son: 0, a, $-c$.

Ejercicio 4.2

1. (*a*) $\begin{bmatrix} 7 & 3 \\ 9 & 7 \end{bmatrix}$ $\quad$ (*c*) $\begin{bmatrix} 21 & -3 \\ 18 & 27 \end{bmatrix}$

3. En este caso especial, sucede que AB es igual a $BA = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

4. (*b*) $\underset{(2 \times 2)}{\begin{bmatrix} 49 & 3 \\ 4 & 3 \end{bmatrix}}$ $\quad$ (*c*) $\underset{(2 \times 1)}{\begin{bmatrix} 3x + 5y \\ 4x + 2y - 7z \end{bmatrix}}$

6. (*a*) $x_2 + x_3 + x_4 + x_5$ $\quad$ (*c*) $b(x_1 + x_2 + x_3 + x_4)$

7. (*b*) $\sum_{i=2}^{4} a_i(x_{i+1} + i)$ $\quad$ (*d*) *Sugerencia*: $x^0 = 1$ para $x \neq 0$

Ejercicio 4.3

1. (*a*) $uv' = \begin{bmatrix} 15 & 5 & -5 \\ 3 & 1 & -1 \\ 9 & 3 & -3 \end{bmatrix}$ $\quad$ (*c*) $xx' = \begin{bmatrix} x_1^2 & x_1x_2 & x_1x_3 \\ x_2x_1 & x_2^2 & x_2x_3 \\ x_3x_1 & x_3x_2 & x_3^2 \end{bmatrix}$

(*e*) $u'v = 13$ $\quad$ (*g*) $u'u = 35$

3. (*a*) $\sum_{i=1}^{n} P_i Q_i$ $\quad$ (*b*) $P \cdot Q$ o $P'Q$ o $Q'P$

5. (*a*) $2v = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$ $\quad$ (*c*) $u - v = \begin{bmatrix} 5 \\ -2 \end{bmatrix}$

7. (*a*) $d = \sqrt{27}$

9. (*c*) $d(v, 0) = (v \cdot v)^{1/2}$

Ejercicio 4.4

1. (*a*) $\begin{bmatrix} 5 & 17 \\ 11 & 17 \end{bmatrix}$

2. No; debe ser $A - B = -B + A$.

4. (*a*) $k(A + B) = k[a_{ij} + b_{ij}] = [ka_{ij} + kb_{ij}] = [ka_{ij}] + [kb_{ij}] = k[a_{ij}] + k[b_{ij}] = kA + kB$ (¿Puede justificar cada paso?)

Ejercicio 4.5

1. (*a*) $AI_3 = \begin{bmatrix} -1 & 5 & 7 \\ 0 & -2 & 4 \end{bmatrix}$ (*c*) $I_2x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

3. (*a*) 5×3 (*c*) 2×1

4. *Sugerencia*: multiplique por sí misma la matriz diagonal dada y examine la matriz producto resultante en relación con las condiciones de idempotencia.

Ejercicio 4.6

1. $A' = \begin{bmatrix} 0 & -1 \\ 4 & 3 \end{bmatrix}$ y $B' = \begin{bmatrix} 3 & 0 \\ -8 & 1 \end{bmatrix}$

3. *Sugerencia*: defina $D \equiv AB$ y aplique (4.11).

5. *Sugerencia*: defina $D \equiv AB$ y aplique (4.14).

Ejercicio 5.1

1. (*a*) (5.2) (*c*) (5.3) (*e*) (5.3)

3. (*a*) Sí. (*d*) No.

5. (*a*) $r(A) = 3$; A es no singular. (*b*) $r(B) = 2$; B es singular.

Ejercicio 5.2

1. (*a*) -6 (*c*) 0 (*e*) $3abc - a^3 - b^3 - c^3$

3. $|M_b| = \begin{vmatrix} d & f \\ g & i \end{vmatrix}$ $|C_b| = -\begin{vmatrix} d & f \\ g & i \end{vmatrix}$

4. (*a*) *Sugerencia*: desarrolle mediante la tercera columna.

5. 20 (no -20)

Ejercicio 5.3

3. (*a*) Propiedad IV. (*b*) Propiedad III (aplicada a ambos renglones).

4. (*a*) Singular. (*c*) Singular.

5. (*a*) Rango < 3 (*c*) Rango < 3

7. A es no singular porque $|A| = 1 - b \neq 0$.

Ejercicio 5.4

1. $\sum_{i=1}^{4} a_{i3}\,|C_{i2}|$ $\sum_{j=1}^{4} a_{2j}\,|C_{4j}|$

3. (*a*) Intercambie los dos elementos de la diagonal de A; multiplique los dos elementos fuera de la diagonal de A por -1.

(*b*) Divida entre $|A|$.

4. (*a*) $E^{-1} = \frac{1}{20}\begin{bmatrix} 3 & 2 & -3 \\ -7 & 2 & 7 \\ -6 & -4 & 26 \end{bmatrix}$ (*c*) $G^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$

Ejercicio 5.5

1. (*a*) $x_1^* = 4$, y $x_2^* = 3$ (*c*) $x_1^* = 2$, y $x_2^* = 1$

2. (*a*) $A^{-1} = \frac{1}{7}\begin{bmatrix} 1 & 2 \\ -2 & 3 \end{bmatrix}; x^* = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$ (*c*) $A^{-1} = \frac{1}{15}\begin{bmatrix} 1 & 7 \\ -1 & 8 \end{bmatrix}; x^* = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

3. (*a*) $x_1^* = 2, x_2^* = 0, x_3^* = 1$ (*c*) $x^* = 0, y^* = 3, z^* = 4$

4. *Sugerencia*: aplique (5.8) y (5.13).

Ejercicio 5.6

1. (*a*) $A^{-1} = \frac{1}{1-b+bt}\begin{bmatrix} 1 & 1 & -b \\ b(1-t) & 1 & -b \\ t & t & 1-b \end{bmatrix}$

$$\begin{bmatrix} Y^* \\ C^* \\ T^* \end{bmatrix} = \frac{1}{1-b+bt}\begin{bmatrix} I_0 + G_0 + a - bd \\ b(1-t)(I_0+G_0) + a - bd \\ t(I_0+G_0) + at + d(1-b) \end{bmatrix}$$

(*b*) $|A| = 1 - b + bt$ $|A_1| = I_0 + G_0 - bd + a$

$|A_2| = a - bd + b(1-t)(I_0 + G_0)$ $|A_3| = d(1-b) + t(a + I_0 + G_0)$

Ejercicio 5.7

1. $x_1^* = 69.53, x_2^* = 57.03$, y $x_3^* = 42.58$

3. (*a*) $A = \begin{bmatrix} 0.10 & 0.50 \\ 0.60 & 0 \end{bmatrix}$; la ecuación matricial es $\begin{bmatrix} 0.90 & -0.50 \\ -0.60 & 1.00 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1\,000 \\ 2\,000 \end{bmatrix}$.

(*c*) $x_1^* = 3\,333\frac{1}{3}$, y $x_2^* = 4\,000$

4. Elemento 0.33: 33¢ del artículo II son necesarios como insumo para producir \$1 del artículo 1.

Ejercicio 6.2

1. (*a*) $\Delta y/\Delta x = 8x + 4\Delta x$ (*b*) $dy/dx = 8x$ (*c*) $f'(3) = 24, f'(4) = 32$

3. (*a*) $\Delta y/\Delta x = 5$; una función constante.

Ejercicio 6.4

1. Límite izquierdo = límite derecho = 15; el límite es 15.

3. (*a*) 5 (*b*) 5

Ejercicio 6.5

1. (*a*) $-3/4 < x$ (*c*) $x < 1/2$

3. (*a*) $-7 < x < 5$ (*c*) $-4 \leq x \leq 1$

Ejercicio 6.6

1. (*a*) 7 (*c*) 17

3. (*a*) $2\frac{1}{2}$ (*c*) 2

Ejercicio 6.7

2. (*a*) $N^2 - 5N - 2$ (*b*) Sí. (*c*) Sí.
3. (*a*) $(N+2)/(N^2+2)$ (*b*) Sí. (*c*) Continua en el dominio.
6. Sí; cada función es continua y uniforme.

Ejercicio 7.1

1. (*a*) $dy/dx = 12x^{11}$ (*c*) $dy/dx = 35x^4$ (*e*) $dw/du = -2u^{-1/2}$
3. (*a*) $f'(x) = 18$; $f'(1) = f'(2) = 18$
(*c*) $f'(x) = 10x^{-3}$; $f'(1) = 10$, $f'(2) = 1\frac{1}{4}$

Ejercicio 7.2

1. $\text{VC} = Q^3 - 5Q^2 + 12Q$; $\dfrac{d\text{VC}}{dQ} = 3Q^2 - 10Q + 12$ es la función MC.
3. (*a*) $3(27x^2 + 6x - 2)$ (*c*) $12x(x+1)$ (*e*) $-x(9x+14)$
4. (*b*) $\text{MR} = 60 - 6Q$
7. (*a*) $(x^2-3)/x^2$ (*c*) $30/(x+5)^2$
8. (*a*) a (*c*) $-a/(ax+b)^2$

Ejercicio 7.3

1. $-2x[3(5-x^2)^2 + 2]$
3. (*a*) $18x(3x^2-13)^2$ (*c*) $5a(ax+b)^4$
5. $x = \frac{1}{7}y - 3$, $dx/dy = \frac{1}{7}$

Ejercicio 7.4

1. (*a*) $\partial y/\partial x_1 = 6x_1^2 - 22x_1x_2$ y $\partial y/\partial x_2 = -11x_1^2 + 6x_2$
(*c*) $\partial y/\partial x_1 = 2(x_2 - 2)$ y $\partial y/\partial x_2 = 2x_1 + 3$
3. (*a*) 12 (*c*) 10/9
5. (*a*) $U_1 = 2(x_1+2)(x_2+3)^3$ y $U_2 = 3(x_1+2)^2(x_2+3)^2$

Ejercicio 7.5

1. $\partial Q^*/\partial a = d/(b+d) > 0$ $\partial Q^*/\partial b = -d(a+c)/(b+d)^2 < 0$
$\partial Q^*/\partial c = -b/(b+d) < 0$ $\partial Q^*/\partial d = b(a+c)/(b+d)^2 > 0$
2. $\partial Y^*/\partial I_0 = \partial Y^*/\partial \alpha = 1/(1-\beta+\beta\delta) > 0$

Ejercicio 7.6

1. (*a*) $|J| = 0$; las funciones son dependientes.
(*b*) $|J| = -20x_2$; las funciones son independientes.

Ejercicio 8.1

1. (*a*) $dy = -3(x^2+1)\,dx$ (*c*) $dy = [(1-x^2)/(x^2+1)^2]\,dx$
3. (*a*) $dC/dY = b$, y $C/Y = (a+bY)/Y$

Ejercicio 8.2

2. (*a*) $dz = (6x+y)\,dx + (x-6y^2)\,dy$

3. (*a*) $dy = [x_2/(x_1+x_2)^2]\,dx_1 - [x_1/(x_1+x_2)^2]\,dx_2$

4. $\varepsilon_{QP} = 2bP^2/(a+bP^2+R^{1/2})$

6. $\varepsilon_{XP} = -2/(Y_f^{1/2}P^2+1)$

Ejercicio 8.3

3. (*a*) $dy = 3[(2x_2-1)(x_3+5)\,dx_1 + 2x_1(x_3+5)\,dx_2 + x_1(2x_2-1)\,dx_3]$

4. *Sugerencia*: aplique las definiciones de diferencial y diferencial total.

Ejercicio 8.4

1. (*a*) $dz/dy = x + 10y + 6y^2 = 28y + 9y^2$

(*c*) $dz/dy = -15x + 3y = 108y - 30$

3. $dQ/dt = [a\alpha A/K + b\beta A/L + A'(t)]K^{\alpha}L^{\beta}$

4. (*b*) $\S W/\S u = 10uf_1 + f_2 \quad \S W/\S v = 3f_1 - 12v^2 f_2$

Ejercicio 8.5

5. (*a*) definida; $dy/dx = -(3x^2-4xy+3y^2)/(-2x^2+6xy) = -9/8$

(*b*) definida; $dy/dx = -(4x+4y)/(4x-4y^3) = 2/13$

7. La condición $F_y \neq 0$ se viola en (0, 0).

8. El producto de las derivadas parciales es igual a -1.

Ejercicio 8.6

1. (*c*) $(dY^*/dG_0) = 1/(S'+T'-I') > 0$

3. $(\partial P^*/\partial Y_0) = D_{Y_0}/(S_{P^*}-D_{P^*}) > 0 \quad (\partial Q^*/\partial Y_0) = D_{Y_0}S_{P^*}/(S_{P^*}-D_{P^*}) > 0$

$(\partial P^*/\partial T_0) = -S_{T_0}/(S_{P^*}-D_{P^*}) > 0 \quad (\partial Q^*/\partial T_0) = -S_{T_0}D_{P^*}/(S_{P^*}-D_{P^*}) < 0$

Ejercicio 9.2

1. (*a*) Cuando $x = 2$, $y = 15$ (un máximo relativo).

(*c*) Cuando $x = 0$, $y = 3$ (un mínimo relativo).

2. (*a*) El valor crítico $x = -1$ queda fuera del dominio; el valor crítico $x = 1$ conduce a $y = 3$ (un máximo relativo).

4. (*d*) La elasticidad es uno.

Ejercicio 9.3

1. (*a*) $f''(x) = 2a$, $f'''(x) = 0$ (*c*) $f''(x) = 6(1-x)^{-3}$, $f'''(x) = 18(1-x)^{-4}$

3. (*b*) Una línea recta.

5. Cada punto de $f(x)$ es un punto estacionario, pero el único punto estacionario $g(x)$ del que se sabe está en $x = 3$.

Ejercicio 9.4

1. (*a*) $f(2) = 33$ es un máximo.

(*c*) $f(1) = 5\frac{1}{3}$ es un máximo; $f(5) = -5\frac{1}{3}$ es un mínimo.

2. *Sugerencia*: primero se escribe una función de área A en términos de una variable (ya sea L o W) solamente.

3. (*d*) $Q^* = 11$ (*e*) Ganancia máxima $= 111\frac{1}{3}$

5. (*a*) $k < 0$ (*b*) $h < 0$ (*c*) $j > 0$

7. (*b*) S se maximiza en el nivel de producción 20.37 (aproximadamente).

Ejercicio 9.5

1. (*a*) 120 (*c*) 4 (*e*) $(n+2)(n+1)$

2. (*a*) $1 + x + x^2 + x^3 + x^4$

3. (*b*) $-63 - 98x - 62x^2 - 18x^3 - 2x^4 + R_4$

Ejercicio 9.6

1. (*a*) $f(0) = 0$ es un punto de inflexión. (*c*) $f(0) = 5$ es un mínimo relativo.

2. (*b*) $f(2) = 0$ es un mínimo relativo.

Ejercicio 10.1

1. (*a*) Sí. (*b*) Sí.

3. (*a*) $5e^{5t}$ (*c*) $-12e^{-2t}$

5. (*a*) La curva con $a = -1$ es la imagen especular de la curva con $a = 1$ con referencia al eje horizontal.

Ejercicio 10.2

1. (*a*) 7.388 (*b*) 1.649

2. (*c*) $1 + 2x + \frac{1}{2!}(2x)^2 + \frac{1}{3!}(2x)^3 + \cdots$

3. (*a*) $\$70e^{0.12}$ (*b*) $\$690e^{0.10}$

Ejercicio 10.3

1. (*a*) 4 (*c*) 4

2. (*a*) 7 (*c*) -3 (*e*) 6

3. (*a*) 26 (*c*) $\ln 3 - \ln B$ (*f*) 3

Ejercicio 10.4

1. El requerimiento evita que la función degenere en una función constante.

3. *Sugerencia*: tome el log de base b.

4. (*a*) $y = e^{(3\ln 8)t}$ o $y = e^{6.2385t}$ (*c*) $y = 5e^{(\ln 5)t}$ o $y = 5e^{1.6095t}$

5. (*a*) $t = (\ln y)/(\ln 7)$ o $t = 0.5139 \ln y$

(*c*) $t = 3\ln(9y)/(\ln 15)$ o $t = 1.1078 \ln(9y)$

6. (*a*) $r = \ln 1.05$ (*c*) $r = 2\ln 1.03$

Ejercicio 10.5

1. (*a*) $2e^{2t+4}$ (*c*) $2te^{t^2+1}$ (*e*) $(2ax+b)e^{ax^2+bx+c}$

3. (*a*) $5/t$ (*c*) $1/(t+19)$ (*e*) $1/[x(1+x)]$

5. *Sugerencia*: use (10.21) y aplique la regla de la cadena.

7. (*a*) $3(8-x^2)/[(x+2)^2(x+4)^2]$

Ejercicio 10.6

1. $t^* = 1/r^2$

2. $d^2A/dt^2 = -A(\ln 2)/4\sqrt{t^3} < 0$

Ejercicio 10.7

1. (*a*) $2/t$ (*c*) $\ln b$ (*e*) $1/t - \ln 3$

3. $r_y = kr_x$

7. $|\varepsilon_d| = n$

11. $r_Q = \varepsilon_{QK} r_K + \varepsilon_{QL} r_L$

Ejercicio 11.2

1. $z^* = 3$ es un mínimo.

3. $z^* = c$, que es un mínimo en el caso (*a*), un máximo en el caso (*b*) y un punto silla en el caso (*c*).

5. (*a*) Cualquier par (x, y) distinto de $(2, 3)$ produce un valor de z positivo.
(*b*) Sí. (*c*) No. (*d*) Sí ($d^2z = 0$).

Ejercicio 11.3

1. (*a*) $q = 4u^2 + 4uv + 3v^2$ (*c*) $q = 5x^2 + 6xy$

3. (*a*) Definida positiva. (*c*) Ninguna.

5. (*a*) Definida positiva. (*c*) Definida negativa. (*e*) Definida positiva.

6. (*a*) $r_1, r_2 = \frac{1}{2}(7 \pm \sqrt{17})$; $u'Du$ es definida positiva.
(*c*) $r_1, r_2 = \frac{1}{2}(5 \pm \sqrt{61})$; $u'Fu$ es indefinida.

7. $v_1 = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}, v_2 = \begin{bmatrix} -1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$

Ejercicio 11.4

1. $z^* = 0$ (mínimo)

3. $z^* = -11/40$ (mínimo)

4. $z^* = 2 - e$ (mínimo), alcanzado en $(x^*, y^*, w^*) = (0, 0, 1)$

5. (*b*) *Sugerencia*: véase (11.16).

6. (*a*) $r_1 = 2$ $r_2 = 4 + \sqrt{6}$ $r_3 = 4 - \sqrt{6}$

Ejercicio 11.5

1. (*a*) Estrictamente convexa. (*c*) Estrictamente convexa.

2. (*a*) Estrictamente cóncava. (*c*) Ninguna.

3. No.

5. (*a*) Disco. (*b*) Sí.

7. (*a*) Combinación convexa, con $\theta = 0.5$. (*b*) Combinación convexa, con $\theta = 0.2$.

Ejercicio 11.6

1. (*a*) No. (*b*) $Q_1^* = P_{10}/4$ y $Q_2^* = P_{20}/4$

3. $|\varepsilon_{d1}| = 1\frac{5}{8}$ $|\varepsilon_{d2}| = 1\frac{1}{3}$ $|\varepsilon_{d3}| = 1\frac{1}{2}$

5. (*a*) $\pi = P_0 Q(a, b)(1 + \frac{1}{2}i_0)^{-2} - P_{a0}a - P_{b0}b$

Ejercicio 11.7

1. $(\partial a^*/\partial P_{a0}) = P_0 Q_{bb} e^{-rt}/|J| < 0$ $(\partial b^*/\partial P_{a0}) = -P_0 Q_{ab} e^{-rt}/|J| < 0$

2. (*a*) Cuatro. (*b*) $(\partial a^*/\partial P_0) = (Q_b Q_{ab} - Q_a Q_{bb})P_0(1 + i_0)^{-2}/|J| > 0$

(*c*) $(\partial a^*/\partial i_0) = (Q_a Q_{bb} - Q_b Q_{ab})P_0^2(1 + i_0)^{-3}/|J| < 0$

Ejercicio 12.2

1. (*a*) $z^* = 1/2$, alcanzado cuando $\lambda^* = 1/2, x^* = 1$, y $y^* = 1/2$

(*c*) $z^* = -19$, alcanzado cuando $\lambda^* = -4, x^* = 1$, y $y^* = 5$

4. $Z_\lambda = -G(x, y) = 0$ $Z_x = f_x - \lambda G_x = 0$ $Z_y = f_y - \lambda G_y = 0$

5. *Sugerencia*: distinguir entre la igualdad idéntica y la igualdad condicional.

Ejercicio 12.3

1. (*a*) $|\overline{H}| = 4$; z^* es un máximo. (*c*) $|\overline{H}| = -2$; z^* es un mínimo.

Ejercicio 12.4

2. (*a*) Cuasicóncava, pero no estrictamente. (*c*) Estrictamente cuasicóncava.

4. (*a*) Ninguno. (*c*) Cuasiconvexa, pero no cuasicóncava.

5. *Sugerencia*: revise la seccion 9.4.

7. *Sugerencia*: use ya sea (12.21) o (12.25′).

Ejercicio 12.5

1. (*b*) $\lambda^* = 3, x^* = 16, y^* = 11$ (*c*) $|\overline{H}| = 48$; la condición se satisface.

3. $(\partial x^*/\partial B) = 1/2P_x > 0$ $(\partial x^*/\partial P_x) = -(B + P_y)/2P_x^2 < 0$

$(\partial x^*/\partial P_y) = 1/2P_x > 0$ etc.

5. No es válido.

7. No tanto a (*a*) como a (*b*) —vea (12.32) y (12.33′).

Ejercicio 12.6

1. (*a*) Homogénea de grado uno. (*c*) No homogénea.

(*e*) Homogénea de grado dos.

4. Son verdaderas.

7. (*a*) Homogénea de grado $a + b + c$.

8. (*a*) $j^2 Q = g(jK, jL)$ (*b*) *Sugerencia*: sea $j = 1/L$.

(*d*) Homogénea de grado uno en K y L.

Ejercicio 12.7

1. (*a*) $1:2:3$ (*b*) $1:4:9$

2. *Sugerencia*: revise las figuras 8.2 y 8.3.

4. *Sugerencia*: ésta es una derivada total.

6. (*a*) Líneas rectas con pendiente hacia abajo. (*b*) $\sigma \to \infty$ cuando $\rho \to -1$

8. (*a*) 7 (*c*) $\ln 5 - 1$

Ejercicio 13.1

3. Las condiciones $x_j(\partial Z/\partial x_j) = 0$ y las condicioness $\lambda_i(\partial Z/\partial \lambda_i) = 0$ pueden condensarse.

5. Consistente.

Ejercicio 13.2

1. No puede encontrarse ningún arco calificador para un vector de prueba tal como $(dx_1, dx_2) = (1, 0)$.

3. $(x_1^*, x_2^*) = (0, 0)$ es una cúspide. Se satisface la calificación de restricción (todos los vectores de prueba son horizontales y apuntan hacia el este); también se satisfacen las condiciones de Kuhn-Tucker.

4. Todas las condiciones pueden satisfacerse escogiendo $y_0^* = 0$ y $y_1^* \geq 0$.

Ejercicio 13.4

2. (*a*) Sí. (*b*) Sí. (*c*) No.

4. (*a*) Sí. (*b*) Sí.

Ejercicio 14.2

1. (*a*) $-8x^{-2} + c, (x \neq 0)$ (*c*) $\frac{1}{6}x^6 - \frac{3}{2}x^2 + c$

2. (*a*) $13e^x + c$ (*c*) $5e^x - 3x^{-1} + c, (x \neq 0)$

3. (*a*) $3 \ln|x| + c, (x \neq 0)$ (*c*) $\ln(x^2 + 3) + c$

4. (*a*) $\frac{2}{3}(x + 1)^{3/2}(x + 3) - \frac{4}{15}(x + 1)^{5/2} + c$

Ejercicio 14.3

1. (*a*) $4\frac{1}{3}$ (*b*) $3\frac{1}{4}$ (*e*) $2\left(\dfrac{a}{3} + c\right)$

2. (*a*) $\frac{1}{2}(e^{-2} - e^{-4})$ (*c*) $e^2\left(\frac{1}{2}e^4 - \frac{1}{2}e^2 + e - 1\right)$

3. (*b*) Subestimar. (*e*) $f(x)$ es Riemann integrable.

Ejercicio 14.4

1. Ninguno.

2. (*a*), (*c*), (*d*) y (*e*).

3. (*a*), (*c*) y (*d*) convergente; (*e*) divergente.

Ejercicio 14.5

1. (*a*) $R(Q) = 14Q^2 - \frac{10}{3}e^{0.3Q} + \frac{10}{3}$ (*b*) $R(Q) = 10Q/(1 + Q)$

3. (*a*) $K(t) = 9t^{4/3} + 25$

5. (*a*) 29 000

Ejercicio 14.6

1. Se considera sólo el capital. Como normalmente también es necesaria la mano de obra para la producción, la hipótesis subyacente es que K y L siempre se usan en una proporción fija.

3. *Sugerencia*: use (6.8)

4. *Sugerencia*: $\ln u - \ln v = \ln \dfrac{u}{v}$

Ejercicio 15.1

1. (*a*) $y(t) = -e^{-4t} + 3$ (*c*) $y(t) = \frac{3}{2}(1 - e^{-10t})$

3. (*a*) $y(t) = 4(1 - e^{-t})$ (*c*) $y(t) = 6e^{5t}$ (*e*) $y(t) = 8e^{7t} - 1$

Ejercicio 15.2

1. La curva D debe ser más empinada.

3. El mecanismo de ajuste de precios genera una ecuación diferencial.

5. (*a*) $P(t) = A \exp\left(-\dfrac{\beta + \delta}{\eta} t\right) + \dfrac{\alpha}{\beta + \delta}$ (*b*) Sí.

Ejercicio 15.3

1. $y(t) = Ae^{-5t} + 3$

3. $y(t) = e^{-t^2} + \frac{1}{2}$

5. $y(t) = e^{-6t} - \frac{1}{7}e^{t}$

6. *Sugerencia*: revise la sección 14.2, ejemplo 17.

Ejercicio 15.4

1. (*a*) $y(t) = (c/t^3)^{1/2}$ (*c*) $yt + y^2 t = c$

Ejercicio 15.5

1. (*a*) Separable; lineal cuando se escribe como $\dfrac{dy}{dt} + \dfrac{1}{t} y = 0$

(*c*) Separable; reducible a una ecuación de Bernoulli.

3. $y(t) = (A - t^2)^{1/2}$

Ejercicio 15.6

1. (*a*) Línea de fase con pendiente hacia arriba; equilibrio dinámicamente inestable.

(*c*) Línea de fase con pendiente hacia abajo; equilibrio dinámicamente estable.

3. El signo de la derivada mide la pendiente de la línea de fase.

Ejercicio 15.7

1. $r_k = r_K - r_L$ [vea (10.25)]

4. (*a*) Grafique $(3 - y)$ y $\ln y$ como dos curvas separadas, y luego reste. Existe un solo equilibrio (para un valor de y entre 1 y 3) y es dinámicamente estable.

Ejercicio 16.1

1. (*a*) $y_p = 2/5$ (*c*) $y_p = 3$ (*e*) $y_p = 6t^2$

3. (*a*) $y(t) = 6e^t + e^{-4t} - 3$ (*c*) $y(t) = e^t + te^t + 3$

6. *Sugerencia*: aplique la regla de L'Hôpital.

Ejercicio 16.2

1. (*a*) $\frac{3}{2} \pm \frac{3}{2}\sqrt{3}i$ (*c*) $-\frac{1}{4} \pm \frac{3}{4}\sqrt{7}i$

3. (*b*) *Sugerencia*: cuando $\theta = \pi/4$, la línea *OP* es una línea a 45°.

5. (*a*) $\dfrac{d}{d\theta}\operatorname{sen} f(\theta) = f'(\theta)\cos f(\theta)$ (*b*) $\dfrac{d}{d\theta}\cos\theta^3 = -3\theta^2 \operatorname{sen}\theta^3$

7. (*a*) $\sqrt{3} + i$ (c) $1 - i$

Ejercicio 16.3

1. $y(t) = e^{2t}(3\cos 2t + \frac{1}{2}\operatorname{sen} 2t)$

3. $y(t) = e^{-3t/2}\left(-\cos\dfrac{\sqrt{7}}{2}t + \dfrac{\sqrt{7}}{7}\operatorname{sen}\dfrac{\sqrt{7}}{2}t\right) + 3$

5. $y(t) = \frac{2}{3}\cos 3t + \operatorname{sen} 3t + \frac{1}{3}$

Ejercicio 16.4

1. (*a*) $P'' + \dfrac{m-u}{n-w}P' - \dfrac{\beta+\delta}{n-w} = -\dfrac{\alpha+\gamma}{n-w}$ $(n \neq w)$ (*b*) $P_p = \dfrac{\alpha+\gamma}{\beta+\delta}$

3. (*a*) $P(t) = e^{t/2}\left(2\cos\frac{3}{2}t + 2\operatorname{sen}\frac{3}{2}t\right) + 2$

Ejercicio 16.5

1. (*a*) $\dfrac{d\pi}{dt} + j(1-g) = j(\alpha - T - \beta U)$

(*b*) No hay raíces complejas; no hay fluctuación.

3. (*c*) Ambas son ecuaciones diferenciales de primer orden. (*d*) $g \neq 1$

4. (*a*) $\pi(t) = e^{-t}\left(A_5\cos\dfrac{\sqrt{2}}{4}t + A_6\operatorname{sen}\dfrac{\sqrt{2}}{4}t\right) + m$ (*c*) $\overline{P} = m; \overline{U} = \dfrac{1}{18} - \dfrac{2}{9}m$

Ejercicio 16.6

2. (*a*) $y_p = t - 2$ (*c*) $y_p = \frac{1}{4}e^t$

Ejercicio 16.7

1. (*a*) $y_p = 4$ (*c*) $y_p = \frac{1}{18}t^2$

3. (*a*) Divergente. (*c*) Convergente.

Ejercicio 17.2

1. (*a*) $y_{t+1} = y_t + 7$ (*c*) $y_{t+1} = 3y_t - 9$

3. (*a*) $y_t = 10 + t$ (*c*) $y_t = y_0\alpha^t - \beta(1 + \alpha + \alpha^2 + \cdots + \alpha^{t-1})$

Ejercicio 17.3

1. (*a*) No oscilatorio; divergente. (*c*) Oscilatorio; convergente.

3. (*a*) $y_t = -8(1/3)^t + 9$ (*c*) $y_t = -2(-1/4)^t + 4$

Ejercicio 17.4

1. $Q_t = \alpha - \beta(P_0 - \overline{P})(-\delta/\beta)^t - \beta\overline{P}$

3. (*a*) $\overline{P} = 3$; oscilación explosiva. (*c*) $\overline{P} = 2$; oscilación uniforme.

5. El retraso en la función de suministro.

Ejercicio 17.5

1. $a = -1$

3. $P_t = (P_0 - 3)(-1.4)^t + 3$, con oscilación explosiva.

Ejercicio 17.6

1. No.

2. (*b*) Movimiento explosivo hacia abajo, no oscilatorio.

(*d*) Movimiento uniforme hacia abajo y hacia R, amortiguado.

4. (*a*) Primero con pendiente hacia abajo, luego se hace horizontal.

Ejercicio 18.1

1. (*a*) $\frac{1}{2} \pm \frac{1}{2}i$ (*c*) $\frac{1}{2}, -1$

3. (*a*) 4 (estacionario) (*c*) 5 (estacionario)

4. (*b*) $y_t = \sqrt{2}^{\,t}\left(2\cos\frac{\pi}{4}t + \operatorname{sen}\frac{\pi}{4}t\right) + 1$

Ejercicio 18.2

1. (*a*) Subcaso 1D. (*c*) Subcaso 1C.

3. *Sugerencia*: use: (18.16).

Ejercicio 18.3

3. Las posibilidades *v*, *viii*, *x* y *xi* se hacen factibles.

4. (*a*) $p_{t+2} - [2 - j(1-g) - \beta k]p_{t+1} + [1 - j(1-g) - \beta k(1-j)]p_t = j\beta km$

(*c*) $\beta k \gtreqless 4$

Ejercicio 18.4

1. (*a*) 1 (*c*) $3t^2 + 3t + 1$

3. (*a*) $y_p = \frac{1}{4}t$ (*c*) $y_p = 2 - t + t^2$

5. (*a*) $1/2, -1$ y 1

Ejercicio 19.2

2. $b^3 + b^2 - 3b + 2 = 0$

3. (*a*) $x_t = -(3)^t + 4(-2)^t + 7 \qquad y_t = 2(3)^t + 2(-2)^t + 5$

4. (*a*) $x(t) = 4e^{-2t} - 3e^{-3t} + 12 \qquad y(t) = -e^{-2t} + e^{-3t} + 4$

Ejercicio 19.3

2. (*c*) $\beta = (\delta I - A)^{-1}u$

3. (*c*) $\beta = (\rho I + I - A)^{-1}\lambda$

5. (*c*) $x_1(t) = 4e^{-4t/10} + 2e^{-11t/10} + \frac{17}{6}e^{t/10}; \quad x_2(t) = 3e^{-4t/10} - 2e^{-11t/10} + \frac{19}{6}e^{t/10}$

Ejercicio 19.4

4. (*a*)

$$\begin{bmatrix} \pi_c \\ U_c \end{bmatrix} = \begin{bmatrix} A_1 \\ \dfrac{23 - \sqrt{193}}{48} A_1 \end{bmatrix} \left(\frac{33 + \sqrt{193}}{64} \right)^t + \begin{bmatrix} A_2 \\ \dfrac{23 + \sqrt{193}}{48} A_2 \end{bmatrix} \left(\frac{33 - \sqrt{193}}{64} \right)^t + \begin{bmatrix} \mu \\ \dfrac{1}{6}(1 - \mu) \end{bmatrix}$$

Ejercicio 19.5

1. La ecuación individual puede reescribirse como dos ecuaciones de primer orden.

2. Sí.

4. (*a*) Punto silla.

Ejercicio 19.6

1. (*a*) $|J_E| = 1$ y tr $J_E = 2$; nodo localmente inestable

(*c*) $|J_E| = 5$ y tr $J_E = -1$; foco localmente estable

2. (*a*) Localmente un punto silla. (*c*) Nodo localmente estable o foco estable.

4. (*a*) Las curvas $x' = 0$ y $y' = 0$ coinciden, y suministran una amplia gama de líneas que contienen a los puntos de equilibrio.

Ejercicio 20.2

1. $\lambda^* = 1 - t \qquad u^* = \dfrac{1-t}{2} \qquad y^* = \dfrac{t}{2} - \dfrac{t^2}{4} + 2$

6. $\lambda^*(t) = 3e^{4-t} - 3 \qquad u^*(t) = 2 \qquad y^*(t) = 7e^t - 2$

Ejercicio 20.4

1. $\lambda^* = \delta/(\delta^2 + \alpha) \qquad K^* = 1/2(\delta^2 + \alpha)$

Índice

D

G

H

I

N

O