Universidad Autónoma del Estado de Hidalgo
Instituto de Ciencias Básicas e Ingeniería
Centro de Investigación en Matemáticas

La asignación del tiempo de un niño a educación y trabajo: un modelo microeconómico II

Tesis que para obtener el título de

Licenciado en Matemáticas Aplicadas

presenta

Yadira Elizabeth Peralta Torres

bajo la dirección de

Dr. Fernando Barrera Mora

Dr. Héctor V. Robles Vásquez

Pachuca, Hidalgo. Enero de 2008.

Resumen

Se analiza un modelo microeconómico de los usos del tiempo de un niño a educación, ocio y trabajo, con el cual es posible estudiar simultáneamente los determinantes microeconómicos de la participación laboral y de la asistencia escolar de los niños. Uno de los principales determinantes en la asignación óptima del tiempo de un niño es el nivel de riqueza familiar, razón por la cual las distintas soluciones están en función de dicho nivel de riqueza.

Abstract

A dynamic microeconomic model of child's allocation of time to education, leisure and work is shown. This model allows to study the microeconomics determinants for schooling and labor supply participation simultaneously. The level of parental wealth plays an important role as one of these determinants. This is the factor that mostly influences in the child's distribution of time.

Para mis padres, por su apoyo en todo momento.

Agradecimientos

A mis asesores de tesis, los Dres. Fernando Barrera Mora y Héctor Robles Vásquez, mi estimación y agradecimiento por su tiempo y apoyo brindados.

Índice general

Introducción

En términos económicos, el capital son los recursos utilizados para producir bienes y servicios. Lo más común es que por él se entienda el capital físico; pero el capital humano es igualmente importante para la producción y es muy valioso para la persona que cuenta con él. El capital humano son los conocimientos, las aptitudes y la experiencia de los seres humanos, que los hacen económicamente productivos, y se puede incrementar invirtiendo en educación, atención de la salud y capacitación laboral, principalmente. El análisis del capital humano asume que la educación incrementa ganancias y productividad, dando conocimiento, habilidades y una forma diferente de analizar problemas. Probablemente la mayor evidencia de esto es que las personas que poseen un mayor nivel educativo cuentan con mayores ingresos que otras (Becker, 1993).

La riqueza del capital humano y su ritmo de aumento son cruciales para el desarrollo de un país, principalmente porque el capital humano es el factor más importante que determina la capacidad de un país para producir y para implementar nuevas tecnologías. Esto se ve reflejado en los países que han tenido un crecimiento económico significativo gracias a grandes inversiones en la educación y capacitación de su fuerza laboral (Becker, 1993). Por esta razón, los gobiernos invierten en educación, ya que es el eje fundamental para alcanzar la realización personal y para contribuir a un desarrollo más rápido del país.

Sin embargo, a pesar de las enormes ventajas que ofrece la educación muchas personas ni siquiera completan su educación básica, y éste es un problema que se presenta en muchos países subdesarrollados o en vías de desarrollo, como es el caso de México. Por tal razón, uno de los principales problemas a los que se enfrenta el área de economía de la educación es tratar de identificar los factores más importantes que influyen en la inasistencia escolar; uno de éstos, y de gran importancia, es el trabajo infantil.

Cabe destacar que el trabajo infantil parece ser perjudicial para el desarrollo integral de los niños, ya que afecta no sólo su presente sino también sus posibilidades de desarrollo, limitando así sus oportunidades en el futuro. En el largo plazo, conduce al retraso escolar o al abandono del sistema educativo, a menores ingresos en la vida adulta, al acceso a trabajos no calificados y a la reproducción de las condiciones de pobreza que originaron su prematura inserción laboral. La pérdida de años de educación se traduce en una calidad inferior del capital humano disponible en una sociedad. El trabajo infantil contribuye a conspirar contra el desarrollo personal de los niños, ya que entra en conflicto con la educación y sus logros en el aprendizaje, así como con el juego y el esparcimiento. Esto sucede porque el tiempo que los niños utilizan para trabajar es tiempo robado a esas actividades educativas y recreativas.

Durante las etapas iniciales del desarrollo o del proceso de industrialización de países ahora altamente industrializados, los niños fueron activos económicos de las familias; desafortunadamente, todavía lo son en países de ingresos bajos y medios, como es el caso de México. Los niños pueden ser empleados por sus propias familias si el hogar funciona como una unidad económica, o pueden compartir sus ingresos laborales con su núcleo familiar, o bien, el trabajo infantil puede servir como amortiguador en las reducciones repentinas del ingreso paterno (Robles, 2004). En la gran mayoría de los casos, el ingreso temprano en el mercado laboral guarda estrecha vinculación con la problemática de la pobreza y con las estrategias de supervivencia a que deben recurrir los grupos familiares de los sectores de la sociedad que no disponen de los medios necesarios para asegurar la satisfacción de sus necesidades básicas.

Es importante notar que a medida que la sociedad adquiere una mayor preocupación por el bienestar de los niños y valora cada vez más la educación, los padres disminuyen el tiempo que los niños dedican a auxiliar a la familia y aumentan los recursos y el tiempo asignados al desarrollo de su capital humano. La teoría del capital humano señala un mecanismo económico que refuerza este proceso. Cuando los padres obtienen mayores ingresos, deciden tener menos hijos e incrementar la inversión en educación y salud de los mismos; de esta manera el trabajo infantil tiende a disminuir (Becker, 1993).

En el presente trabajo se realiza un análisis detallado del modelo desarrollado por Robles (2000) en su tesis doctoral, titulada *A microeconomic Analysis of Child Labor Participation and Education: The Case of Mexico, 1984-1996*. Dicho modelo estudia la asignación del tiempo de los niños en edad escolar en educación y trabajo, permitiendo identificar los principales factores que determinan la asistencia escolar de los mismos. Este trabajo está basado en el libro *El Trabajo Infantil en México 1984-2000* escrito por Robles (2004), el cual es una extensión de la tesis doctoral del mismo autor. A lo largo de los primeros tres capítulos se introduce la teoría matemática y microeconómica necesaria para abordar el modelo. En el cuarto capítulo se desarrolla detalladamente el modelo y en el quinto se da solución a una parte del mismo mediante la técnica de Kuhn-Tucker. Finalmente, en el último capítulo se presentan las conclusiones.

Capítulo 1

Continuidad y diferenciabilidad de funciones de $\mathbb{R}^n$ a $\mathbb{R}^m$

En este capítulo se encuentran algunos resultados básicos de funciones continuas y funciones diferenciables que son importantes para el desarrollo del presente trabajo. Para ello se requiere introducir algunos conceptos topológicos de $\mathbb{R}^n$.

Notación. Los vectores se designarán con tipo de letra negrita y los escalares con letra más clara. Un vector $\mathbf{x} \in \mathbb{R}^n$ se denotará por un arreglo de una fila y n columnas, es decir,

$$\mathbf{x} = (x_1, x_2, \ldots, x_n).$$

Si f está definida en un punto $\mathbf{x} \in \mathbb{R}^n$, la notación $f(\mathbf{x})$ y $f(x_1, \ldots, x_n)$ se usan indistintamente para designar el valor de f en dicho punto.

1.1

El espacio vectorial $\mathbb{R}^n$

En esta sección suponemos conocidas las propiedades básicas de $\mathbb{R}^n$ como espacio vectorial y sólo recordamos algunos conceptos que nos serán de utilidad.

El *producto escalar* de dos vectores $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ y $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ en $\mathbb{R}^n$, denotado por $\langle \mathbf{x}, \mathbf{y} \rangle$, es definido por la siguiente fórmula:

$$\langle \mathbf{x}, \mathbf{y} \rangle := \sum_{i=1}^{n} x_i y_i.$$

El producto escalar de dos vectores también puede ser denotado por $\mathbf{x} \cdot \mathbf{y}$.

En $\mathbb{R}^n$ se tiene una norma inducida por el producto escalar, en la forma: dado un vector $\mathbf{x} \in \mathbb{R}^n$ se define la norma de $\mathbf{x}$ por

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

A esta norma se le llama *norma euclidiana*.

También se introduce la noción de *distancia* entre dos elementos $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, denotada y definida por

$$d(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}.$$

1.2

Conceptos topológicos básicos de $\mathbb{R}^n$

Utilizaremos la notación y lenguaje estándar de Teoría de Conjuntos. A lo largo de esta sección se introducirán algunos conceptos topológicos de $\mathbb{R}^n$, tales como conjuntos abiertos, cerrados, compactos, convexos y conexos; tomando como referencia Spivak (1965). Empezamos con las siguientes definiciones básicas.

Definición 1.1. Sea $\mathbf{x}_0 \in \mathbb{R}^n$ y $r \in \mathbb{R}$ positivo fijo. Se define la *bola abierta* de centro $\mathbf{x}_0$ y radio r como el conjunto

$$B(\mathbf{x}_0, r) = \{\mathbf{x} \mid \|\mathbf{x} - \mathbf{x}_0\| < r\}.$$

$B(\mathbf{x}_0, r)$ está constituida por todos los puntos cuya distancia a $\mathbf{x}_0$ es menor que r.

Si en la definición anterior tuviéramos el signo $\leq$ en vez de $<$, obtenemos la definición de *bola cerrada*, que es el conjunto que consta de todos los puntos cuya distancia a $\mathbf{x}_0$ es menor o igual que r y se denota por $B[\mathbf{x}_0, r]$.

La Figura 1.1 ilustra una bola abierta en $\mathbb{R}^2$, y si el conjunto incluyera la circunferencia punteada obtendríamos una bola cerrada.

Definición 1.2. Sea $A \subseteq \mathbb{R}^n$ un subconjunto dado.

i) Se dice que $\mathbf{x}_0 \in A$ es un *punto interior* de A si existe $r > 0$ tal que

$$B(\mathbf{x}_0, r) \subseteq A.$$

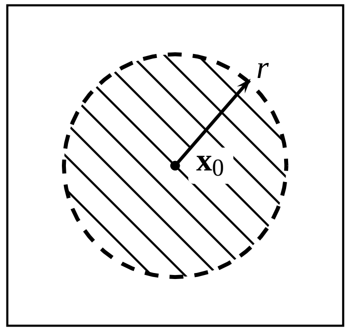

Figura 1.1: Bola abierta.

ii) Se dice que $\mathbf{x}_0 \in \mathbb{R}^n$ es un *punto frontera* de A si para todo $r > 0$ se cumple

$$B(\mathbf{x}_0, r) \cap A \neq \emptyset \text{ y } B(\mathbf{x}_0, r) \cap A^c \neq \emptyset.$$

Donde A^c denota el complemento de A.

iii) Se dice que $\mathbf{x}_0 \in \mathbb{R}^n$ es un *punto adherente* de A si para todo $r > 0$ se cumple

$$B(\mathbf{x}_0, r) \cap A \neq \emptyset.$$

iv) Se dice que $\mathbf{x}_0 \in \mathbb{R}^n$ es un *punto límite* de A si para todo $r > 0$ se tiene

$$(B(\mathbf{x}_0, r) \backslash \{\mathbf{x}_0\}) \cap A \neq \emptyset.$$

El conjunto de todos los puntos interiores de A se llama el *interior de A* y se designa con $int(A)$; y al conjunto de todos los puntos frontera de A se le llama *frontera de A* y se denota por ∂A. La colección de todos los puntos adherentes de A se llama la *adherencia de A* y se denota por $\overline{A}$.

Definición 1.3. Un conjunto $A \subseteq \mathbb{R}^n$ se llama *conjunto abierto* si todos sus puntos son interiores; es decir, A es abierto si $A = int(A)$.

Definición 1.4. Un conjunto $A \subseteq \mathbb{R}^n$ se llama *conjunto cerrado* si su complemento es abierto.

Notemos que $\mathbb{R}^n$ y $\emptyset$ son conjuntos abiertos y cerrados a la vez; y que $B(\mathbf{x}_0, r)$ es un conjunto abierto y $B[\mathbf{x}_0, r]$ es un conjunto cerrado. El conjunto $B(\mathbf{x}_0, r)$ es abierto ya que si tomamos $r_1 = r - \|\mathbf{x}_0 - \mathbf{x}_1\| > 0$, con $\mathbf{x}_1 \in B(\mathbf{x}_0, r)$, se cumple que $B(\mathbf{x}_1, r_1) \subset B(\mathbf{x}_0, r)$; y el conjunto $B[\mathbf{x}_0, r]$ es cerrado porque si tomamos $\mathbf{x}_1 \in (B[\mathbf{x}_0, r])^c$ y $r_1 = \|\mathbf{x}_0 - \mathbf{x}_1\| - r > 0$, entonces se cumple que $B(\mathbf{x}_1, r_1) \subset (B[\mathbf{x}_0, r])^c$.

Algunas propiedades importantes sobre los conjuntos abiertos y cerrados son las siguientes:

i) La unión arbitraria de conjuntos abiertos es un conjunto abierto.

ii) La intersección de cualquier colección finita de conjuntos abiertos es un conjunto abierto.

iii) La intersección arbitraria de conjuntos cerrados es un conjunto cerrado.

iv) La unión de cualquier colección finita de conjuntos cerrados es un conjunto cerrado.

Proposición 1.5. *Un conjunto A es cerrado si y sólo si contiene a todos sus puntos límites.*

Ya que tenemos los conceptos de conjuntos abiertos y cerrados pasamos al análisis de unos conjuntos muy importantes, los conjuntos compactos; para ello empezamos con lo siguiente.

Definición 1.6. Un *recubrimiento abierto* de un conjunto $A \subseteq \mathbb{R}^n$ es una colección $\mathscr{F}$ de conjuntos abiertos tales que

$$A \subseteq \bigcup_{U \in \mathscr{F}} U.$$

Definición 1.7. Un conjunto $A \subseteq \mathbb{R}^n$ se dice *conjunto compacto* si toda cubierta abierta de A posee una subcubierta finita.

Definición 1.8. Un conjunto $A \subseteq \mathbb{R}^n$ se dice *conjunto acotado* si existe $R > 0$ para el cual $A \subseteq B(\mathbf{0}, R)$.

En la Figura 1.2 ilustramos un conjunto acotado para el caso $n = 2$.

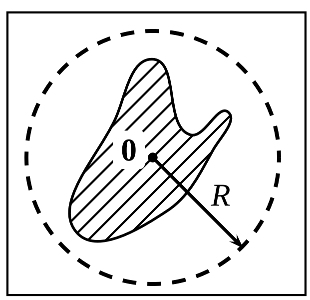

Figura 1.2: Conjunto acotado.

Teorema 1.9. ***Teorema de Heine-Borel****. Un conjunto $A \subseteq \mathbb{R}^n$ es compacto si y sólo si es cerrado y acotado.*

Demostración.
($\Rightarrow$ Supongamos que $A \subseteq \mathbb{R}^n$ es un conjunto compacto, queremos probar que es cerrado y

acotado. Consideremos la colección

$$\mathscr{F} = \{B(\mathbf{0}, k) \mid k \in \mathbb{N}\}.$$

Dicha colección es una cubierta abierta infinita de $\mathbb{R}^n$ y por lo tanto de A. Como A es compacto existe una subcubierta finita $B(\mathbf{0}, k_1), B(\mathbf{0}, k_2), \ldots, B(\mathbf{0}, k_j)$ que lo cubre. Sea

$$R = \text{ máx } \{k_1, k_2, \ldots, k_j\},$$

entonces $A \subseteq B(\mathbf{0}, R)$; es decir, A es acotado.

Ahora para demostrar que A es cerrado probaremos que A^c es abierto. Sea $\mathbf{x}_0 \in A^c$. La colección $\mathscr{F} = \{B[\mathbf{x}_0, \frac{1}{n}]^c \mid n \in \mathbb{N}\}$ es una cubierta abierta de $\mathbb{R}^n \backslash \{\mathbf{x}_0\}$ y por lo tanto de A.

Como A es compacto existe una subcubierta finita $B[\mathbf{x}_0, \frac{1}{n_1}]^c, B[\mathbf{x}_0, \frac{1}{n_2}]^c, \ldots, B[\mathbf{x}_0, \frac{1}{n_k}]^c$ que lo cubre, es decir, $A \subseteq B[\mathbf{x}_0, \frac{1}{n_i}]^c$ para todo $i = 1, 2, \ldots, k$.

Si $n_0 = \text{ máx } \{n_1, n_2, \ldots, n_k\}$, entonces $B[\mathbf{x}_0, \frac{1}{n_0}] \subseteq B[\mathbf{x}_0, \frac{1}{n_i}]$ para todo $i = 1, 2, \ldots, k$.

Se sigue que $B[\mathbf{x}_0, \frac{1}{n_0}]^c \supseteq A$, lo cual implica que $B[\mathbf{x}_0, \frac{1}{n_0}] \subseteq A^c$ y de aquí obtenemos que $B(\mathbf{x}_0, \frac{1}{n_0}) \subseteq A^c$; es decir, A^c es abiero.

$\Leftarrow$) Supongamos que A es cerrado y acotado, queremos demostrar que A es compacto. Sólo daremos un bosquejo de esta prueba, los detalles pueden consultarse en Spivak (1965). Como A es acotado, existe $m > 0$ tal que $A \subseteq B$, donde

$$B = [-m, m] \times \ldots \times [-m, m].$$

Notemos que B es compacto, pues es producto de compactos; más adelante probaremos este hecho. Por otra parte, si K es un subconjunto cerrado de un conjunto compacto $D \subseteq \mathbb{R}^n$, entonces K es compacto. Como A es cerrado y B es compacto, entonces A es compacto.

Con lo anterior finalizamos la demostración del Teorema 1.9, sólo falta probar que B es compacto; para ello vamos a demostrar lo siguiente: Si A es un conjunto compacto en $\mathbb{R}^n$ y B es un conjunto compacto en $\mathbb{R}^m$, entonces $A \times B$ es un conjunto compacto en $\mathbb{R}^n \times \mathbb{R}^m$.

Para la demostración de esta parte vamos a necesitar la siguiente proposición:

Proposición 1.10. *Sean $\mathbf{x} \in \mathbb{R}^n$ y B un subconjunto compacto de $\mathbb{R}^m$, entonces $\{\mathbf{x}\} \times B$ es un conjunto compacto en el espacio $\mathbb{R}^n \times \mathbb{R}^m$.*

Sea $\mathscr{F}$ una cubierta abierta de $A \times B$. Por la Proposición 1.10, para cada $\mathbf{x} \in A$ existe un abierto $U_{\mathbf{x}}$ y una colección finita $\{W_1^{\mathbf{x}}, \ldots, W_{k\mathbf{x}}^{\mathbf{x}}\}$ tal que

$$U_{\mathbf{x}} \times B \subseteq W_1^{\mathbf{x}} \cup \ldots \cup W_{k\mathbf{x}}^{\mathbf{x}}.$$

Claramente, la colección $\{U_{\mathbf{x}}|\mathbf{x} \in A\}$ es una cubierta abierta de A; pero como A es compacto, existe una subcubierta finita $\{U_{\mathbf{x}_1}, \ldots, U_{\mathbf{x}_m}\}$ que lo cubre.

Como para cada $i = 1, \ldots, m$, el conjunto $U_{\mathbf{x}_i} \times B$ esta cubierto por una colección finita $\{W_1^{\mathbf{x}_i}, \ldots, W_{k\mathbf{x}_i}^{\mathbf{x}_i}\}$ de elementos de $\mathscr{F}$, entonces la colección

$$\{W_1^{\mathbf{x}_1}, \ldots, W_{k\mathbf{x}_1}^{\mathbf{x}_1}, \ldots, W_1^{\mathbf{x}_m}, \ldots, W_{k\mathbf{x}_m}^{\mathbf{x}_m}\}$$

es una subcubierta finita de $\mathscr{F}$ para $A \times B$. Luego $A \times B$ es compacto. ∎

Ahora pasamos a definir dos tipos de conjuntos que serán de utilidad para este trabajo, los conjuntos convexos y conexos; así como algunas de sus propiedades más importantes.

Definición 1.11. Un conjunto no vacío $A \subset \mathbb{R}^n$ se dice *conjunto convexo* si el segmento lineal que une a cualesquiera dos puntos del conjunto, está totalmente contenido en el mismo. Es decir, si $\mathbf{x}_1, \mathbf{x}_2 \in A$, entonces

$$\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2 \in A \text{ para cada } \lambda \in [0,1].$$

La Figura 1.3 ilustra un conjunto convexo y otro que no lo es, ya que el segmento que une a $\mathbf{x}_1$ y $\mathbf{x}_2$ no está completamente contenido en el conjunto.

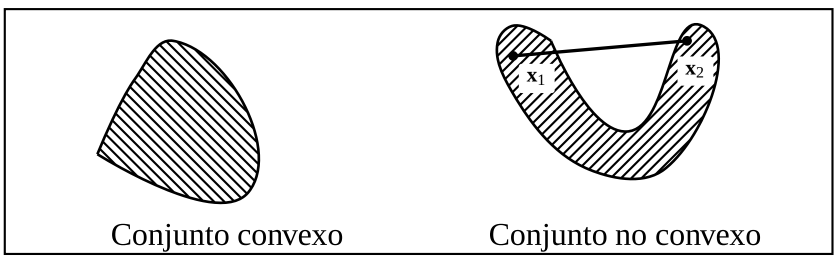

Figura 1.3: A la izquierda se ilustra un conjunto convexo y a la derecha uno que no lo es.

Las propiedades más importantes de conjuntos convexos son las siguientes:

i) $\mathbb{R}^n$ y $\emptyset$ son conjuntos convexos.

ii) Los conjuntos que constan de un único elemento son convexos.

iii) La intersección arbitraria de conjuntos convexos es un conjunto convexo.

iv) La unión de conjuntos convexos no siempre es un conjunto convexo.

Definición 1.12. Un conjunto $A \subseteq \mathbb{R}^n$ se dice *conjunto disconexo* si existen conjuntos abiertos $U, V \subset \mathbb{R}^n$ no vacíos tales que:

i) $A \subseteq U \cup V$

ii) $A \cap U \neq \emptyset$ y $A \cap V \neq \emptyset$

iii) $(A \cap U) \cap (A \cap V) = \emptyset$

En este caso decimos que U y V forman una separación o disconexión del conjunto A.

La Figura 1.4 ilustra un conjunto disconexo.

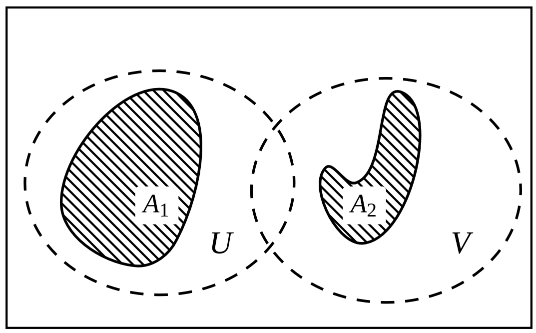

Figura 1.4: El conjunto $A = A_1 \cup A_2$ es disconexo.

Definición 1.13. Un conjunto $A \subseteq \mathbb{R}^n$ se llama *conjunto conexo* si no es disconexo, es decir, si no existen abiertos U y V que formen una separación de A.

La Figura 1.5 ilustra un conjunto conexo, podemos observar que no hay forma de separar tal conjunto.

Definición 1.14. Un ***intervalo***, I, es un subconjunto de $\mathbb{R}$ que verifican la propiedad siguiente: si x e y pertenecen a I, con $x \leq y$, entonces para todo z tal que $x \leq z \leq y$, z pertenece a I.

Lema 1.15. *Un subconjunto $A \subseteq \mathbb{R}$ es conexo si y sólo si A es un intervalo (acotado o no).*

Demostración.
($\Rightarrow$ Supongamos que A es conexo. Si A es el conjunto de un sólo elemento, entonces A es el intervalo $[x_0, x_0]$. Si A contiene dos elementos distintos, supongamos $a < b$:

Afirmamos que el conjunto $(a,b) = \{x \mid a < x < b\}$ está contenido en A. Demostraremos la afirmación anterior por contradicción. Supongamos que $(a,b) \nsubseteq A$, entonces existe al menos un elemento x_0 tal que $x_0 \in (a,b)$ pero $x_0 \notin A$. Sean $U = (-\infty, x_0)$ y $V = (x_0, +\infty)$ dos conjuntos abiertos, notemos que:

i) $A \subseteq (U \cup V)$

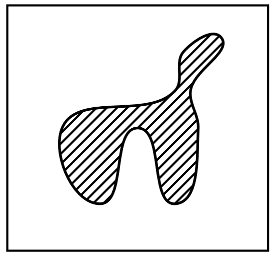

Figura 1.5: Conjunto conexo.

ii) $A \cap U \neq \emptyset$ y $A \cap V \neq \emptyset$

iii) $(A \cap U) \cap (A \cap V) = \emptyset$

Por lo que U y V forman una separación de A, lo cual es una contradicción al hecho de que A es conexo, entonces A es necesariamente un intervalo.

Ahora, sean p y q la cota inferior máxima y la cota superior mínima de A en la adherencia de $\mathbb{R}$, $\overline{\mathbb{R}}$, respectivamente; y sea $c \in A$ tal que $p < c < q$. Por la Afirmación **??** tenemos lo siguiente:

Supongamos que p y q son finitos, entonces se tienen los casos que se enumeran a continuación.

i) Si p, $q \in A$, entonces $A = [p,q]$.

ii) Si $p \in A$ y $q \notin A$, entonces $A = [p,q)$.

iii) Si $p \notin A$ y $q \in A$, entonces $A = (p,q]$.

iv) Si p, $q \notin A$, entonces $A = (p,q)$.

v) Si $p = -\infty$, para todo $x < c$ existe $y \in A$ tal que $y < x$, y ya que $x \in A$, entonces $(-\infty, c]$ está contenido en A.

vi) Análogamente al caso anterior, si $q = +\infty$, entonces $[c, +\infty)$ está contenido en A.

$\Leftarrow$) Con respecto a la prueba de esta implicación, sólo la bosquejaremos y daremos las ideas centrales, los detalles de la prueba se pueden consultar en (Dieudonné, 1969). Primero, procedemos por contradicción suponiendo que existen abiertos U y V que forman una separación de A. Suponemos por ejemplo que $x \in U$, $y \in V$ y $x < y$. Luego sea z la cota superior mínima del conjunto acotado $U \cap [x,y]$; si $z \in U$ se tiene una contradicción con respecto a la definición de z, y similarmente si $z \in V$; por lo que z no puede pertenecer ni a U ni a V, lo cual es absurdo ya que el conjunto cerrado $[x,y]$ está contenido en A. ∎

Del Lema 1.15 surge un importante hecho plasmado en el siguiente corolario.

Corolario 1.16. *El conjunto de los números reales*, $\mathbb{R}$, *es conexo.*

1.3

Continuidad de funciones

En esta sección introduciremos el concepto de función continua, así como algunas de sus propiedades más importantes.

Definición 1.17. Sea $f : A \to \mathbb{R}^m$, $A \subseteq \mathbb{R}^n$, una función. Decimos que f *es continua* en $\mathbf{x}_0 \in A$, si para cada $\varepsilon > 0$ existe $\delta > 0$ tal que

$$\|\mathbf{x}-\mathbf{x}_0\| < \delta \text{ implica que } \|f(\mathbf{x}) - f(\mathbf{x}_0)\| < \varepsilon.$$

Equivalentemente,

$$\mathbf{x} \in B(\mathbf{x_0}, \delta) \cap A \text{ implica que } f(\mathbf{x}) \in B(f(\mathbf{x_0}), \varepsilon).$$

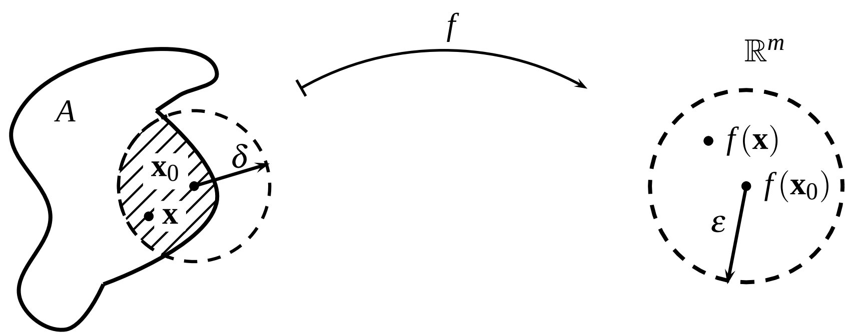

Figura 1.6: Continuidad.

En la Figura 1.6 se representa el concepto de continuidad. También podemos definir dicho concepto utilizando sucesiones, así que las Definiciones 1.17 y 1.18 son equivalentes.

Definición 1.18. Sean $A \subseteq \mathbb{R}^n$, $B \subseteq \mathbb{R}^m$ y $f : A \to B$ una función. Decimos que f es continua en $\mathbf{x}_0 \in A$ si para toda sucesión $\{\mathbf{x}_n\}_{n\in\mathbb{N}} \subset A$ tal que $\mathbf{x}_n \to \mathbf{x}_0$, la sucesión $\{f(\mathbf{x}_n)\}_{n\in\mathbb{N}} \subset B$ converge a $f(\mathbf{x}_0)$.

Definición 1.19. Si $f : A \to \mathbb{R}^m$,donde $A \subseteq \mathbb{R}^n$, entonces

$$f(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x}))\, \forall \mathbf{x} \in A$$

en donde $f_i : A \to \mathbb{R}$ para $i = 1, 2, \ldots, m$. Las funciones f_i se denominan *funciones coordenadas* de f.

Diremos que f es continua en A si f es continua en todo punto de A. Además, $f : A \to \mathbb{R}^m$, $A \subseteq \mathbb{R}^n$, es continua si cada una de sus funciones coordenadas, f_i con $i = 1, 2, \ldots, m$, es continua.

Sean $A \subseteq \mathbb{R}^n$ un conjunto dado, $f, g : A \to \mathbb{R}^m$ funciones continuas en $\mathbf{x}_0 \in A$, entonces se cumple lo siguiente:

i) La función $f + g$ es continua en $\mathbf{x}_0$.

ii) La función λf es continua en $\mathbf{x}_0$, $\lambda \in \mathbb{R}$.

iii) La función $f \cdot g$ es continua en $\mathbf{x}_0$.

iv) La función $\|f\|$ es continua en $\mathbf{x}_0$.

Teorema 1.20. *Si $L : \mathbb{R}^n \to \mathbb{R}^m$ es una función lineal, entonces es continua en todo $\mathbb{R}^n$.*

Demostración.
Para la prueba de este teorema necesitamos el siguiente resultado: Si $L : \mathbb{R}^n \to \mathbb{R}^m$ es una función lineal, entonces existe $M > 0$ tal que

$$\|L(\mathbf{x})\| \leq M\|\mathbf{x}\| \;\; \forall \mathbf{x} \in \mathbb{R}^n.$$

Sea $\varepsilon > 0$ dado. Si $\mathbf{x}_0 \in \mathbb{R}^n$, entonces, por el resultado antes mencionado, tenemos que

$$\|L(\mathbf{x}) - L(\mathbf{x}_0)\| = \|L(\mathbf{x} - \mathbf{x}_0)\| \leq M\|\mathbf{x} - \mathbf{x}_0\|.$$

Elegimos $\delta \leq \frac{\varepsilon}{M}$ y se tiene el resultado deseado. ■

Teorema 1.21. *Sea $f : D \to \mathbb{R}^m$ una función dada, con $D \subseteq \mathbb{R}^n$ un conjunto abierto.*

i) La función f es continua en D si y sólo si $f^{-1}(V)$ es un conjunto abierto en D para todo conjunto abierto $V \subseteq \mathbb{R}^m$.

ii) La función f es continua en D si y sólo si $f^{-1}(F)$ es un conjunto cerrado en D para todo conjunto cerrado $F \subseteq \mathbb{R}^m$.

Demostración.
Probaremos el primer inciso.

($\Rightarrow$ Supongamos que f es continua. Sean $V \subseteq \mathbb{R}^m$ un conjunto abierto y $\mathbf{x}_0 \in f^{-1}(V)$, entonces $f(\mathbf{x}_0) \in V$. Como V es abierto existe $\varepsilon > 0$ tal que $B(f(\mathbf{x}_0), \varepsilon) \subseteq V$ y como f es continua en $\mathbf{x}_0$ existe $\delta_{\mathbf{x}_0} > 0$ tal que

$$f(\mathbf{x}) \in B(f(\mathbf{x}_0), \varepsilon) \;\text{ siempre que }\; \mathbf{x} \in B(\mathbf{x}_0, \delta_{\mathbf{x}_0}) \cap D.$$

Si

$$U = \bigcup_{\mathbf{x} \in f^{-1}(V)} B(\mathbf{x}, \delta_{\mathbf{x}}),$$

entonces U es abierto y $f^{-1}(V) = U \cap D$ es abierto en D.

$\Leftarrow$) Supongamos que $f^{-1}(V)$ es un conjunto abierto en $\mathbb{R}^n$ para todo conjunto abierto $V \subseteq \mathbb{R}^m$. Sean $\varepsilon > 0$ y $\mathbf{x}_0 \in D$ arbitrario, entonces el conjunto $V = B(f(\mathbf{x}_0), \varepsilon) \subseteq \mathbb{R}^m$ es abierto. Luego existe un abierto U en $\mathbb{R}^n$ tal que $f^{-1}(V) = U \cap D$ es abierto en D.

Como $\mathbf{x}_0 \in U$ y U es abierto, existe $\delta > 0$ tal que $B(\mathbf{x}_0, \delta) \subseteq U$. Por lo tanto si

$$\mathbf{x} \in B(\mathbf{x}_0, \delta) \cap D \subseteq U \cap D = f^{-1}(V),$$

entonces $f(\mathbf{x}) \in V = B(f(\mathbf{x}_0), \varepsilon)$. Por lo tanto f es continua.

La prueba para el segundo inciso es análoga a la anterior, sólo hace falta notar que el conjunto $f^{-1}(F)^c = f^{-1}(F^c)$ es abierto. ■

Teorema 1.22. *Sea $f : A \to \mathbb{R}^m$ una función continua, con $A \subseteq \mathbb{R}^n$ abierto.*

i) Si $C \subseteq A$ es conexo entonces $f(C)$ es conexo.

ii) Si $K \subseteq A$ es compacto entonces $f(K)$ es compacto.

Demostración.
Primero probaremos el primer inciso. Supongamos, por contradicción, que U, V son abiertos en $\mathbb{R}^m$ tales que

i) $f(C) \subseteq U \cup V$

ii) $f(C) \cap U \neq \emptyset$ y $f(C) \cap V \neq \emptyset$

iii) $(f(C) \cap U) \cap (f(C) \cap V) = \emptyset$

Como f es continua, entonces $f^{-1}(U)$ y $f^{-1}(V)$ son abiertos en $\mathbb{R}^n$ que satisfacen

i) $C \subseteq f^{-1}(U) \cup f^{-1}(V)$

ii) $C \cap f^{-1}(U) \neq \emptyset$ y $C \cap f^{-1}(V) \neq \emptyset$

iii) $(C \cap f^{-1}(U)) \cap (C \cap f^{-1}(V)) = \emptyset$

Contradicción, pues C es un conjunto conexo. Por lo tanto $f(C)$ es conexo.

Ahora demostraremos el segundo inciso. Sea $\mathscr{F}$ una cubierta abierta de $f(K)$. Por la continuidad de f, la colección

$$\mathscr{C} = \{ f^{-1}(U) \mid U \in \mathscr{F} \}$$

es una cubierta abierta de K. Como K es compacto, existe una colección finita $U_1, \ldots, U_n$ de elementos de $\mathscr{F}$ tales que $K \subseteq f^{-1}(U_1) \cup \ldots \cup f^{-1}(U_n)$ y por lo tanto

$$f(K) \subseteq U_1 \cup \ldots \cup U_n.$$

Es decir, $f(K)$ posee una subcubierta finita; por lo tanto es compacto. ■

Note que si K es un subconjunto compacto de $\mathbb{R}$, $k_* = \text{ínf} K$ y $k^* = \sup K$, entonces k_*, $k^* \in K$. Esto es porque k_*, $k^* \in \overline{K} = K$. Ahora, del segunco inciso del Teorema 1.22 surge el siguiente corolario.

Corolario 1.23. *Si $f : A \to \mathbb{R}$ es continua, con $A \subseteq \mathbb{R}^n$ abierto, y $K \subseteq \mathbb{R}^n$ es compacto, entonces existen $\mathbf{x}_*$ y $\mathbf{x}^*$ en K tales que*

$$f(\mathbf{x}_*) \leq f(\mathbf{x}) \leq f(\mathbf{x}^*) \quad \forall \mathbf{x} \in K.$$

Demostración.

Sabemos que $f(K)$ es un subconjunto compacto de $\mathbb{R}$. Luego $k_* = \text{ínf} f(K)$ y $k^* = \sup f(K)$ son elementos de $f(K)$. Por lo tanto existen $\mathbf{x}_*$ y $\mathbf{x}^*$ en K tales que $k_* = f(\mathbf{x}_*)$ y $k^* = f(\mathbf{x}^*)$. De esta forma, $k_* = f(\mathbf{x}_*) \leq f(\mathbf{x}) \leq f(\mathbf{x}^*) = k^* \ \ \forall \mathbf{x} \in K$. ■

Teorema 1.24. *Si $f : A \to \mathbb{R}^m$, $A \subseteq \mathbb{R}^n$, es una función continua en $\mathbf{x}_0 \in A$ y $g : \mathbb{R}^m \to \mathbb{R}^k$ es continua en $f(\mathbf{x}_0)$, entonces $g \circ f$ es continua en $\mathbf{x}_0$.*

Demostración.

Dado $\varepsilon > 0$, por la continuidad de g en $f(\mathbf{x}_0)$, existe $\delta_1 > 0$ tal que

$$g(\mathbf{y}) \in B(g(f(\mathbf{x_0})), \varepsilon) \text{ siempre que } \mathbf{y} \in B(f(\mathbf{x_0}), \delta_1).$$

Por la continuidad de f en $\mathbf{x}_0$, existe $\delta > 0$ tal que

$$f(\mathbf{x}) \in B(f(\mathbf{x_0}), \delta_1) \text{ siempre que } \mathbf{x} \in B(\mathbf{x_0}, \delta).$$

Por lo tanto, $g(f(\mathbf{x})) \in B(g(f(\mathbf{x_0})), \varepsilon)$ siempre que $\mathbf{x} \in B(\mathbf{x_0}, \delta)$. ■

1.4

Diferenciabilidad de funciones

Dada una función $f(x)$, siempre es útil saber cómo cambia f cuando se da un cambio en x. Por ejemplo, supongamos que conocemos la función de demanda, D, de un bien y, y que dicha función depende del precio p de ese bien. Estamos interesados en analizar el cambio en la función D cuando varía el precio p del bien; para ello es necesario introducir el concepto de derivada.

A lo largo de esta sección Ω denotará un subconjunto de $\mathbb{R}^n$.

Definición 1.25. Sea $f : \Omega \to \mathbb{R}^m$ una función, si $\mathbf{a}$ es un punto interior de Ω y $\mathbf{y}$ un vector cualquiera de $\mathbb{R}^n$ definimos la *derivada* de f en $\mathbf{a}$ con respecto a $\mathbf{y}$, $f'(\mathbf{a};\mathbf{y})$, mediante la fórmula

$$f'(\mathbf{a};\mathbf{y}) = \lim_{h\to 0} \frac{f(\mathbf{a}+h\mathbf{y}) - f(\mathbf{a})}{h}, \tag{1.1}$$

siempre que tal límite exista.

Designemos con f_k la k-ésima componente de f, pues recordemos que

$$f(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x}))\,\forall \mathbf{x} \in \Omega.$$

Notemos que la derivada $f'(\mathbf{a};\mathbf{y})$ existe si y sólo si $f_k'(\mathbf{a};\mathbf{y})$ existe para cada $k = 1, 2, \ldots, m$ en cuyo caso tenemos

$$f'(\mathbf{a};\mathbf{y}) = (f_1'(\mathbf{a};\mathbf{y}), f_2'(\mathbf{a};\mathbf{y}), \ldots, f_m'(\mathbf{a};\mathbf{y})) = \sum_{k=1}^{m} f_k'(\mathbf{a};\mathbf{y})\mathbf{e}_k,$$

donde $\mathbf{e}_k$ es el k-ésimo vector canónico.

Si $\mathbf{y}$ es un vector unitario, es decir, cuando $\|\mathbf{y}\| = 1$, la derivada $f'(\mathbf{a};\mathbf{y})$ se llama la *derivada direccional* de f en $\mathbf{a}$ en la dirección de $\mathbf{y}$. En particular, si $\mathbf{y} = \mathbf{e}_k$ la derivada direccional $f'(\mathbf{a};\mathbf{e}_k)$ se denomina *derivada parcial* de f respecto a $\mathbf{e}_k$ y se representa también mediante el símbolo $D_k f(\mathbf{a})$. De esta manera tenemos que

$$D_k f(\mathbf{a}) = (f_1'(\mathbf{a};\mathbf{e}_k), f_2'(\mathbf{a};\mathbf{e}_k), \ldots, f_m'(\mathbf{a};\mathbf{e}_k)),$$

o también se puede escribir como

$$D_k f(\mathbf{a}) = (D_k f_1(\mathbf{a}), D_k f_2(\mathbf{a}), \ldots, D_k f_m(\mathbf{a})).$$

Por otra parte, las derivadas parciales de $D_1 f, D_2 f, \ldots, D_n f$ se llaman *derivadas parciales de segundo orden* de f y las derivadas parciales de éstas se llaman *derivadas parciales de tercer orden*, y así sucesivamente. En general, las derivadas parciales de este tipo reciben el nombre de *derivadas parciales de orden superior*.

Definición 1.26. Una función $f:\Omega\to\mathbb{R}^m$ se dice *diferenciable* en un punto interior $\mathbf{a}\in\Omega$ si existe una transformación lineal $T_{\mathbf{a}}:\mathbb{R}^n\to\mathbb{R}^m$ tal que

$$f(\mathbf{a}+\mathbf{v})=f(\mathbf{a})+T_{\mathbf{a}}(\mathbf{v})+\|\mathbf{v}\|E(\mathbf{a},\mathbf{v}), \tag{1.2}$$

donde $E(\mathbf{a},\mathbf{v})\to\mathbf{0}$ cuando $\mathbf{v}\to\mathbf{0}$.

Definición 1.27. Si $f:\Omega\to\mathbb{R}^m$ es diferenciable en $\mathbf{a}\in\Omega$, entonces a la transformación lineal T que satisface (1.2) la denominamos la *diferencial total* o simplemente *diferencial* de f en $\mathbf{a}$.

Teorema 1.28. *Supongamos que f es diferenciable en $\mathbf{a}$ con diferencial $T_{\mathbf{a}}$, entonces existe la derivada $f'(\mathbf{a};\mathbf{y})$ $\forall\mathbf{y}\in\mathbb{R}^n$, y tenemos*

$$T_{\mathbf{a}}(\mathbf{y})=f'(\mathbf{a};\mathbf{y}).$$

Además si $f=(f_1,f_2,\ldots,f_m)$ y $\mathbf{y}=(y_1,y_2,\ldots,y_n)$ tenemos

$$T_{\mathbf{a}}(\mathbf{y})=\sum_{k=1}^{m}(\nabla f_k(\mathbf{a})\cdot\mathbf{y})\mathbf{e}_k=(\nabla f_1(\mathbf{a})\cdot\mathbf{y},\nabla f_2(\mathbf{a})\cdot\mathbf{y},\ldots,\nabla f_m(\mathbf{a})\cdot\mathbf{y}). \tag{1.3}$$

La ecuación (1.3) también puede representarse como un producto matricial

$$T_{\mathbf{a}}(\mathbf{y})=Df(\mathbf{a})\mathbf{y},$$

donde $Df(\mathbf{a})$ es la matriz $m\times n$ cuya fila k-ésima es $\nabla f_k(\mathbf{a})$ y su elemento kj es la derivada parcial $D_j f_k(\mathbf{a})$; y $\mathbf{y}\in\mathbb{R}^n$ es un vector columna.

$$Df(\mathbf{a}) \;=\; \begin{pmatrix} D_1f_1(\mathbf{a}) & D_2f_1(\mathbf{a}) & \cdots & D_nf_1(\mathbf{a}) \\ D_1f_2(\mathbf{a}) & D_2f_2(\mathbf{a}) & \cdots & D_nf_2(\mathbf{a}) \\ \vdots & & & \vdots \\ D_1f_m(\mathbf{a}) & D_2f_m(\mathbf{a}) & \cdots & D_nf_m(\mathbf{a}) \end{pmatrix}$$

La matriz $Df(\mathbf{a})$ se llama *matriz jacobiana* de f en $\mathbf{a}$ y está definida en cada punto en el que existan las *mn* derivadas parciales $D_j f_k(\mathbf{a})$.

La diferencial $T_{\mathbf{a}}$ se expresa también poniendo $f'(\mathbf{a})$. La derivada $f'(\mathbf{a})$ es una transformación lineal, la matriz jacobiana $Df(\mathbf{a})$ es la matriz que representa a $Df(\mathbf{a})$ respecto de la base canónica de esa transformación. La fórmula de Taylor de primer orden toma la forma

$$f(\mathbf{a}+\mathbf{v})=f(\mathbf{a})+f'(\mathbf{a})(\mathbf{v})+\|\mathbf{v}\|E(\mathbf{a},\mathbf{v}), \tag{1.4}$$

donde $E(\mathbf{a},\mathbf{v})\to\mathbf{0}$ cuando $\mathbf{v}\to\mathbf{0}$.

La desigualdad enunciada a continuación será de utilidad para la demostración del siguiente teorema. La desigualdad de *Cauchy-Schwarz* establece que para todo $\mathbf{x},\mathbf{y}\in\mathbb{R}^n$

$$|\langle\mathbf{x},\mathbf{y}\rangle|\le\|\mathbf{x}\|\cdot\|\mathbf{y}\|,$$

con igualdad si y sólo si $\mathbf{y}=c\mathbf{x}$ para algún escalar c.

Teorema 1.29. *Si $f : \Omega \to \mathbb{R}^m$ es diferenciable en $\mathbf{a}$ entonces f es continua en $\mathbf{a}$.*

Demostración.

Tomando en cuenta la ecuación (1.4), tenemos que

$$\|f(\mathbf{a}+\mathbf{v}) - f(\mathbf{a})\| = \|f'(\mathbf{a})(\mathbf{v}) + \|\mathbf{v}\|E(\mathbf{a},\mathbf{v})\|.$$

Aplicando la desigualdad de Cauchy-Schwarz obtenemos

$$0 \leq \|f(\mathbf{a}+\mathbf{v}) - f(\mathbf{a})\| \leq \|f'(\mathbf{a})(\mathbf{v})\| + \|\|\mathbf{v}\|E(\mathbf{a},\mathbf{v})\|.$$

Si hacemos que $\mathbf{v} \to \mathbf{0}$, el error $\|\mathbf{v}\|E(\mathbf{a},\mathbf{v}) \to \mathbf{0}$. La parte lineal $f'(\mathbf{a})(\mathbf{v})$ tiende también a $\mathbf{0}$, debido a que las transformaciones lineales son continuas en $\mathbf{0}$. Completando así la prueba. ■

Teorema 1.30. ***Regla de la Cadena***. *Sean $g : \Omega \to \mathbb{R}^m$ y $f : \mathbb{R}^m \to \mathbb{R}^k$ tales que $h = f \circ g$ esté definida en un entorno del punto $\mathbf{a}$. Supongamos que g es diferenciable en $\mathbf{a}$, con diferencial $g'(\mathbf{a})$. Pongamos $\mathbf{b} = g(\mathbf{a})$ y supongamos que f es diferenciable en $\mathbf{b}$, con diferencial $f'(\mathbf{b})$. Entonces h es diferenciable en $\mathbf{a}$, y la diferencial $h'(\mathbf{a})$ viene dada por*

$$h'(\mathbf{a}) = f'(\mathbf{b}) \circ g'(\mathbf{a}),$$

que es la composición de las transformaciones lineales $f'(\mathbf{b})$ y $g'(\mathbf{a})$.

Demostración.

Consideremos la diferencia $h(\mathbf{a}+\mathbf{y}) - h(\mathbf{a})$ para valores pequeños de $\|\mathbf{y}\|$, y demostremos que se obtiene una fórmula de Taylor de primer orden. De la definición de h resulta

$$h(\mathbf{a}+\mathbf{y}) - h(\mathbf{a}) = f[g(\mathbf{a}+\mathbf{y})] - f[g(\mathbf{a})] = f(\mathbf{b}+\mathbf{v}) - f(\mathbf{b}), \tag{1.5}$$

donde $\mathbf{v} = g(\mathbf{a}+\mathbf{y}) - g(\mathbf{a})$. Por la diferenciabilidad de g tenemos:

$$\mathbf{v} = g'(\mathbf{a})(\mathbf{y}) + \|\mathbf{y}\|E_g(\mathbf{a},\mathbf{y}) \quad \text{donde} \quad E_g(\mathbf{a},\mathbf{y}) \to \mathbf{0} \quad \text{cuando} \quad \mathbf{y} \to \mathbf{0}. \tag{1.6}$$

Además, por la diferenciabilidad de f, se cumple:

$$f(\mathbf{b}+\mathbf{v}) - f(\mathbf{b}) = f'(\mathbf{b})(\mathbf{v}) + \|\mathbf{v}\|E_f(\mathbf{b},\mathbf{v}), \tag{1.7}$$

donde $E_f(\mathbf{b},\mathbf{v}) \to \mathbf{0}$ cuando $\mathbf{v} \to \mathbf{0}$. Aplicando (1.6) en (1.7) obtenemos

$$\begin{aligned} f(\mathbf{b}+\mathbf{v}) - f(\mathbf{b}) &= f'(\mathbf{b})g'(\mathbf{a})(\mathbf{y}) + f'(\mathbf{b})(\|\mathbf{y}\|E_g(\mathbf{a},\mathbf{y})) + \|\mathbf{v}\|E_f(\mathbf{b},\mathbf{v}) && (1.8)\\ &= f'(\mathbf{b})g'(\mathbf{a})(\mathbf{y}) + \|\mathbf{y}\|E(\mathbf{a},\mathbf{y}), && (1.9)\end{aligned}$$

donde $E(\mathbf{a},\mathbf{0}) = \mathbf{0}$ y

$$E(\mathbf{a},\mathbf{y}) = f'(\mathbf{b})(E_g(\mathbf{a},\mathbf{y})) + \frac{\|\mathbf{v}\|}{\|\mathbf{y}\|}E_f(\mathbf{b},\mathbf{v}) \quad \text{si} \quad \mathbf{y} \neq \mathbf{0}. \tag{1.10}$$

Para completar la demostración necesitamos probar que $E(\mathbf{a},\mathbf{y}) \to \mathbf{0}$ cuando $\mathbf{y} \to \mathbf{0}$. El primer término del segundo miembro de (1.10) tiende a $\mathbf{0}$ cuando $\mathbf{y} \to \mathbf{0}$ porque $E_g(\mathbf{a},\mathbf{y}) \to \mathbf{0}$ cuando $\mathbf{y} \to \mathbf{0}$ y las transformaciones lineales son continuas en $\mathbf{0}$. En el segundo término del segundo miembro de (1.10) el factor $E_f(\mathbf{b},\mathbf{v}) \to \mathbf{0}$ porque $\mathbf{v} \to \mathbf{0}$ cuando $\mathbf{y} \to \mathbf{0}$. El cociente $\frac{\|\mathbf{v}\|}{\|\mathbf{y}\|}$ permanece acotado porque, según (1.6) y debido a que las transformaciones lineales son acotadas, tenemos

$$\|\mathbf{v}\| \leq M\|\mathbf{y}\| + \|\mathbf{y}\|\|E_g(\mathbf{a},\mathbf{y})\|.$$

Por consiguiente los dos términos del segundo miembro de (1.10) tienden a $\mathbf{0}$ cuando $\mathbf{y} \to \mathbf{0}$, por lo tanto $E(\mathbf{a},\mathbf{y}) \to \mathbf{0}$.

Entonces, de (1.8) y (1.5) obtenemos la fórmula de Taylor

$$h(\mathbf{a}+\mathbf{y}) - h(\mathbf{a}) = f'(\mathbf{b})g'(\mathbf{a})(\mathbf{y}) + \|\mathbf{y}\|E(\mathbf{a},\mathbf{y}),$$

donde $E(\mathbf{a},\mathbf{y}) \to \mathbf{0}$ cuando $\mathbf{y} \to \mathbf{0}$. Esto prueba que h es diferenciable en $\mathbf{a}$ y que la diferencial $h'(\mathbf{a})$ es igual a la composición $f'(\mathbf{b}) \circ g'(\mathbf{a})$. ∎

1.5

Funciones cóncavas y convexas

Las funciones cóncavas y convexas son muy peculiares y tienen muchas propiedades importantes. A lo largo de esta sección analizaremos dichas propiedades, las cuales nos permiten obtener condiciones adecuadas para alcanzar un óptimo.

Definición 1.31. Sea $f : S \to \mathbb{R}$, donde $S \subseteq \mathbb{R}^n$ es un conjunto no vacío. Se dice que la función f es *convexa* en S si

$$f(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) \leq \lambda f(\mathbf{x}_1) + (1-\lambda)f(\mathbf{x}_2)$$

para todo $\mathbf{x}_1,\mathbf{x}_2 \in S$ y para todo $\lambda \in (0,1)$.

Se dice que la función f es *estrictamente convexa* en S si la desigualdad anterior se cumple con desigualdad estricta para todo $\mathbf{x}_1 \neq \mathbf{x}_2 \in S$ y para todo $\lambda \in (0,1)$.

La función $f : S \to \mathbb{R}$ es *cóncava* (*estrictamente cóncava*) en S si $-f$ es convexa (estrictamente convexa) en S.

Ahora hagamos una interpretación geométrica de estos conceptos. Sean $\mathbf{x}_1$ y $\mathbf{x}_2$ dos puntos distintos en el dominio de f y consideremos el punto $\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$ con $\lambda \in (0,1)$. Note que

$\lambda f(\mathbf{x}_1)+(1-\lambda)f(\mathbf{x}_2)$ es un punto que está sobre el segmento que une a $f(\mathbf{x}_1)$ y $f(\mathbf{x}_2)$, mientras que $f(\lambda\mathbf{x}_1+(1-\lambda)\mathbf{x}_2)$ es el valor de f en el punto $\lambda\mathbf{x}_1+(1-\lambda)\mathbf{x}_2$. Entonces para una función convexa f, el valor de f en los puntos de la forma $\lambda\mathbf{x}_1+(1-\lambda)\mathbf{x}_2$ con $\lambda \in (0,1)$ es menor o igual que la altura del segmento que une los puntos $(\mathbf{x}_1, f(\mathbf{x}_1))$ y $(\mathbf{x}_2, f(\mathbf{x}_2))$. Es decir, f es convexa si el segmento que une $(\mathbf{x}_1, f(\mathbf{x}_1))$ con $(\mathbf{x}_2, f(\mathbf{x}_2))$ queda por encima de la gráfica de f.

Para una función cóncava el segmento que une $(\mathbf{x}_1, f(\mathbf{x}_1))$ con $(\mathbf{x}_2, f(\mathbf{x}_2))$ queda por debajo de la gráfica de f. La Figura 1.7 ilustra una función cóncava y una convexa. De aquí en adelante nos concentraremos únicamente en analizar funciones convexas, ya que los resultados para las funciones cóncavas se pueden obtener fácilmente tomando en cuenta que f es cóncava si y sólo si $-f$ es convexa.

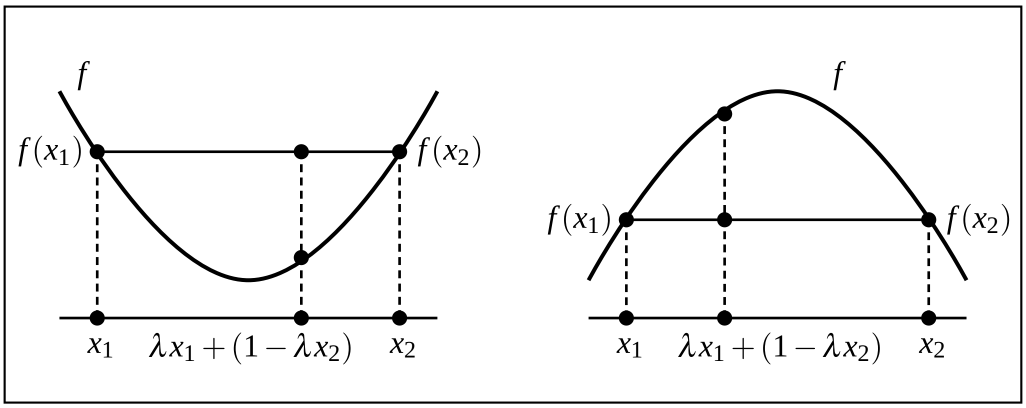

Figura 1.7: A la izquierda tenemos una función convexa y a la derecha una función cóncava.

1.5.1

Continuidad de funciones convexas

Una propiedad importante de las funciones cóncavas y convexas es que son continuas en el interior de su dominio, esto se establece en el siguiente teorema.

Teorema 1.32. *Sean $S \subseteq \mathbb{R}^n$ un conjunto convexo no vacío y $f : S \to \mathbb{R}$ una función convexa, entonces f es continua en el interior de S.*

Demostración.

Sea $\tilde{\mathbf{x}} \in int(S)$. Para demostrar la continuidad de f en $\tilde{\mathbf{x}}$, necesitamos probar que dado $\varepsilon > 0$, existe $\delta > 0$ tal que $\|\mathbf{x}-\tilde{\mathbf{x}}\| \leq \delta$ implica que $|f(\mathbf{x}) - f(\tilde{\mathbf{x}})| \leq \varepsilon$. Como $\tilde{\mathbf{x}} \in int(S)$, existe $\delta' > 0$

tal que $\|\mathbf{x}-\tilde{\mathbf{x}}\| \leq \delta'$ implica que $\mathbf{x} \in S$. Definamos θ de la siguiente forma:

$$\theta = \max_{1\leq i\leq n}\{\max\{f(\tilde{\mathbf{x}}+\delta'\mathbf{e}_i)-f(\tilde{\mathbf{x}}), f(\tilde{\mathbf{x}}-\delta'\mathbf{e}_i)-f(\tilde{\mathbf{x}})\}\} \tag{1.11}$$

donde $\mathbf{e}_i$ es el i-ésimo vector coordenado unidad. Note que $0 \leq \theta < \infty$. Sea

$$\delta = \min\left\{\frac{\delta'}{n}, \frac{\varepsilon\delta'}{n\theta}\right\}. \tag{1.12}$$

Tomemos un $\mathbf{x}$ tal que $\|\mathbf{x}-\tilde{\mathbf{x}}\| \leq \delta$. Sea

$$\mathbf{z}_i = \begin{cases} \delta'\mathbf{e}_i & \text{si } x_i - \tilde{x}_i \geq 0, \\ -\delta'\mathbf{e}_i & \text{de otra forma.} \end{cases}$$

Entonces $\mathbf{x}-\tilde{\mathbf{x}} = \sum_{i=1}^{n}\alpha_i\mathbf{z}_i$, donde $\alpha_i \geq 0$ para $i = 1,2\ldots,n$. Más aún,

$$\|\mathbf{x}-\tilde{\mathbf{x}}\| = \delta'\left(\sum_{i=1}^{n}\alpha_i^2\right)^{\frac{1}{2}}. \tag{1.13}$$

Tomando en cuenta (1.12) y como $\|\mathbf{x}-\tilde{\mathbf{x}}\| \leq \delta$, se sigue que $\alpha_i \leq \frac{1}{n}$ para $i = 1,2\ldots,n$. Por la convexidad de f y como $0 \leq n\alpha_i \leq 1$, obtenemos

$$\begin{aligned} f(\mathbf{x}) = f\left(\tilde{\mathbf{x}}+\sum_{i=1}^{n}\alpha_i\mathbf{z}_i\right) &= f\left[\frac{1}{n}\sum_{i=1}^{n}(\tilde{\mathbf{x}}+n\alpha_i\mathbf{z}_i)\right] \\ &\leq \frac{1}{n}\sum_{i=1}^{n} f(\tilde{\mathbf{x}}+n\alpha_i\mathbf{z}_i) \\ &= \frac{1}{n}\sum_{i=1}^{n} f[(1-n\alpha_i)\tilde{\mathbf{x}}+n\alpha_i(\tilde{\mathbf{x}}+\mathbf{z}_i)] \\ &\leq \frac{1}{n}\sum_{i=1}^{n}[(1-n\alpha_i)f(\tilde{\mathbf{x}})+n\alpha_i f(\tilde{\mathbf{x}}+\mathbf{z}_i)]. \end{aligned}$$

Entonces $f(\mathbf{x})-f(\tilde{\mathbf{x}}) \leq \sum_{i=1}^{n}\alpha_i[f(\tilde{\mathbf{x}}+\mathbf{z}_i)-f(\tilde{\mathbf{x}})]$. Por otra parte, por (1.11) se tiene que $f(\tilde{\mathbf{x}}+\mathbf{z}_i)-f(\tilde{\mathbf{x}}) \leq \theta$ para todo i, y como $\alpha_i \geq 0$, se sigue que

$$f(\mathbf{x})-f(\tilde{\mathbf{x}}) \leq \theta\sum_{i=1}^{n}\alpha_i. \tag{1.14}$$

Tomando en cuenta las ecuaciones (1.12) y (1.13), se sigue que $\alpha_i \leq \frac{\varepsilon}{n\theta}$, y (1.14) implica que $f(\mathbf{x})-f(\tilde{\mathbf{x}}) \leq \varepsilon$. Hemos probado que $\|\mathbf{x}-\tilde{\mathbf{x}}\| \leq \delta$ implica que $f(\mathbf{x})-f(\tilde{\mathbf{x}}) \leq \varepsilon$. Para terminar la prueba necesitamos demostrar que $f(\tilde{\mathbf{x}})-f(\mathbf{x}) \leq \varepsilon$. Sea $\mathbf{y} = 2\tilde{\mathbf{x}}-\mathbf{x}$, notemos que $\|\mathbf{y}-\tilde{\mathbf{x}}\| \leq \delta$. Por lo tanto

$$f(\mathbf{y})-f(\tilde{\mathbf{x}}) \leq \varepsilon. \tag{1.15}$$

Pero $\tilde{\mathbf{x}} = \frac{1}{2}\mathbf{y} + \frac{1}{2}\mathbf{x}$, por la convexidad de f, tenemos

$$f(\tilde{\mathbf{x}}) \leq \frac{1}{2}f(\mathbf{y}) + \frac{1}{2}f(\mathbf{x}). \tag{1.16}$$

Combinando (1.15) y (1.16), se sigue que $f(\tilde{\mathbf{x}}) - f(\mathbf{x}) \leq \varepsilon$, completando la prueba. ■

Note que las funciones cóncavas y convexas, como cualquier función, pueden no ser continuas en todas partes. Sin embargo, por el teorema anterior, los puntos de discontinuidad para estas funciones son permitidos únicamente en la frontera de S.

1.5.2

Optimización de funciones convexas

En este apartado consideraremos el problema de minimizar una función convexa sobre un conjunto convexo y analizaremos las condiciones necesarias para obtener un óptimo. El problema de maximizar una función cóncava es equivalente al de minimizar una función convexa.

Considerando la Definición 1.26 y el Teorema 1.28 de la sección 1.4, establecemos la definición de una función convexa diferenciable; lo hacemos de esta manera ya que será de utilidad en las secciones posteriores.

Definición 1.33. Sean $S \subseteq \mathbb{R}^n$ un conjunto no vacío y $f : S \to \mathbb{R}$ una función. Se dice que f es *diferenciable* en $\mathbf{x}_0 \in int(S)$ si existe un vector $\nabla f(\mathbf{x}_0)$, llamado el *vector gradiente*, y una función $E : \mathbb{R}^n \to \mathbb{R}$ tal que

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0) + \|\mathbf{x} - \mathbf{x}_0\| E(\mathbf{x}_0; \mathbf{x} - \mathbf{x}_0)$$

para todo $\mathbf{x} \in S$, donde $E(\mathbf{x}_0; \mathbf{x} - \mathbf{x}_0) \to \mathbf{0}$ cuando $\mathbf{x} \to \mathbf{x}_0$.

La función f se dice diferenciable en el conjunto abierto $S' \subset S$ si es diferenciable en cada punto de S'.

Definición 1.34. Sea $f : \mathbb{R}^n \to \mathbb{R}$ una función, consideremos el siguiente problema

Problema 1.35.

$$\begin{array}{ll} \text{Minimizar} & f(\mathbf{x}) \\ \text{sujeto a} & \mathbf{x} \in S \end{array}$$

Un punto $\mathbf{x} \in S$ es llamado *solución factible* del problema anterior. Si $\mathbf{x}_0 \in S$ y

$$f(\mathbf{x}_0) \leq f(\mathbf{x}) \text{ para todo } \mathbf{x} \in S,$$

entonces $\mathbf{x}_0$ es una *solución óptima*, solución *óptima global*, o simplemente *solución* del Problema 1.35. Si $\mathbf{x}_0 \in S$ y existe $\varepsilon > 0$ tal que

$$f(\mathbf{x}_0) \leq f(\mathbf{x}) \text{ para cada } \mathbf{x} \in S \cap B(\mathbf{x}_0, \varepsilon),$$

entonces $\mathbf{x}_0$ es una *solución óptima local*.

En el siguiente teorema vamos a probar que cada mínimo local del Problema 1.35 es también un mínimo global. Este resultado es de mucha utilidad en el proceso de optimización.

Teorema 1.36. *Sean $S \subseteq \mathbb{R}^n$ un conjunto convexo no vacío y $f : S \to \mathbb{R}$ una función. Consideremos el Problema 1.35 y supongamos que $\mathbf{x}_0 \in S$ es una solución óptima local del problema.*

i) Si f es convexa, entonces $\mathbf{x}_0$ es una solución óptima global.

ii) Si f es estrictamente convexa, entonces $\mathbf{x}_0$ es la única solución óptima global.

Demostración.
Vamos a probar la parte (i). Supongamos que f es convexa, como $\mathbf{x}_0$ es un óptimo local, existe $\varepsilon > 0$ tal que

$$f(\mathbf{x}_0) \leq f(\mathbf{x}) \text{ para cada } \mathbf{x} \in S \cap B(\mathbf{x}_0, \varepsilon). \tag{1.17}$$

Supongamos, por contradicción, que $\mathbf{x}_0$ no es un óptimo global, entonces se debería cumplir que $f(\widehat{\mathbf{x}}) < f(\mathbf{x}_0)$ para algún $\widehat{\mathbf{x}} \in S$. Ya que f es convexa se tiene que

$$f(\lambda \widehat{\mathbf{x}} + (1-\lambda)\mathbf{x}_0) \leq \lambda f(\widehat{\mathbf{x}}) + (1-\lambda) f(\mathbf{x}_0) < \lambda f(\mathbf{x}_0) + (1-\lambda) f(\mathbf{x}_0) = f(\mathbf{x}_0)$$

para todo $\lambda \in (0,1)$. Pero para $\lambda > 0$ suficientemente pequeño

$$\lambda \widehat{\mathbf{x}} + (1-\lambda)\mathbf{x}_0 \in S \cap B(\mathbf{x}_0, \varepsilon).$$

Entonces la desigualdad anterior contradice (1.17), por lo tanto $\mathbf{x}_0$ es un óptimo global.

Ahora probaremos la segunda parte. Supongamos que f es estrictamente convexa. Como convexidad estricta implica convexidad, entonces por la parte (1) tenemos que $\mathbf{x}_0$ es un óptimo global. Supongamos, por contradicción, que $\mathbf{x}_0$ no es el único óptimo global; entonces existe $\mathbf{x} \in S$, $\mathbf{x} \neq \mathbf{x}_0$, tal que $f(\mathbf{x}) = f(\mathbf{x}_0)$. Como f es estrictamente convexa se cumple lo siguiente

$$f\left(\frac{1}{2}\mathbf{x} + \frac{1}{2}\mathbf{x}_0\right) < \frac{1}{2} f(\mathbf{x}) + \frac{1}{2} f(\mathbf{x}_0) = f(\mathbf{x}_0).$$

Como S es un conjunto convexo, $\frac{1}{2}\mathbf{x} + \frac{1}{2}\mathbf{x}_0 \in S$, y la desigualdad anterior contradice el hecho de que $\mathbf{x}_0$ sea un óptimo global. Probando así que $\mathbf{x}_0$ es el único óptimo global. ∎

1.5.3

Generalizaciones del concepto de convexidad

A lo largo de esta subsección presentaremos algunos tipos de funciones que son similares a las funciones cóncavas y convexas. Varios de los resultados que se presentarán después no requieren el supuesto restrictivo de convexidad, más bien trabajaremos bajo la suposición más general de cuasiconvexidad y pseudoconvexidad. En Bazaara (1993) se encuentran mayores detalles sobre estos conceptos.

Funciones cuasiconvexas

Con la siguiente definición introducimos el concepto de función cuasiconvexa.

Definición 1.37. Sea $f : S \to \mathbb{R}$, donde $S \subset \mathbb{R}^n$ es un conjunto convexo no vacío. La función f se dice *cuasiconvexa* si, para todos $\mathbf{x}_1, \mathbf{x}_2 \in S$, se cumple la siguiente desigualdad

$$f(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) \leq \text{máx}\{f(\mathbf{x}_1), f(\mathbf{x}_2)\} \text{ para todo } \lambda \in (0,1).$$

La función f se dice *cuasicóncava* si $-f$ es cuasiconvexa, es decir, si se cumple que

$$f(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) \geq \text{mín}\{f(\mathbf{x}_1), f(\mathbf{x}_2)\} \text{ para todo } \lambda \in (0,1).$$

Note que de la definición anterior se tiene que una función f es cuasiconvexa si cuando $f(\mathbf{x}_1) \leq f(\mathbf{x}_2)$, entonces $f(\mathbf{x}_2)$ es mayor o igual que el valor de f en todas las combinaciones convexas de $\mathbf{x}_1$ y $\mathbf{x}_2$. Una función f es cuasicóncava si cuando $f(\mathbf{x}_1) \leq f(\mathbf{x}_2)$, entonces el valor de f en todas las combinaciones convexas de $\mathbf{x}_1$ y $\mathbf{x}_2$ es mayor o igual que $f(\mathbf{x}_1)$. La Figura 1.8 ilustra ejemplos de funciones cuasicóncavas y cuasiconvexas.

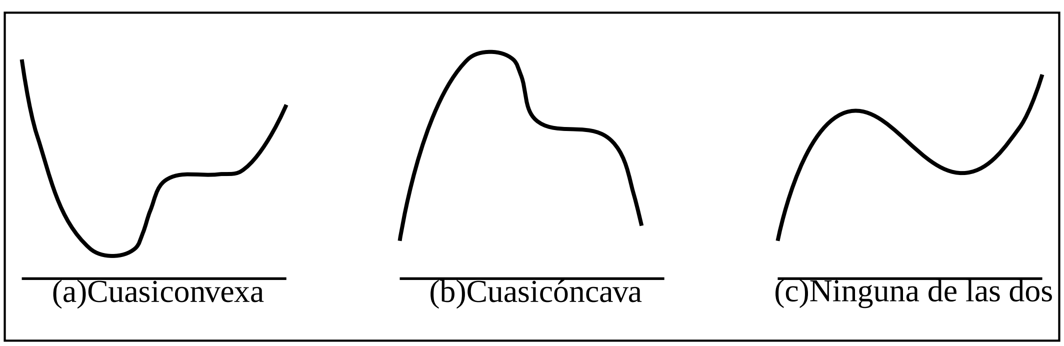

Figura 1.8: Funciones cuasiconvexas y cuasicóncavas.

Note que de la Definición 1.37 se sigue que toda función convexa es también cuasiconvexa, pero al revés no necesariamente. Por ejemplo, $f(x) = x^3$ es una función que es a la vez cuasicóncava y cuasiconvexa, pero que no es ni cóncava ni convexa (Figura1.9).

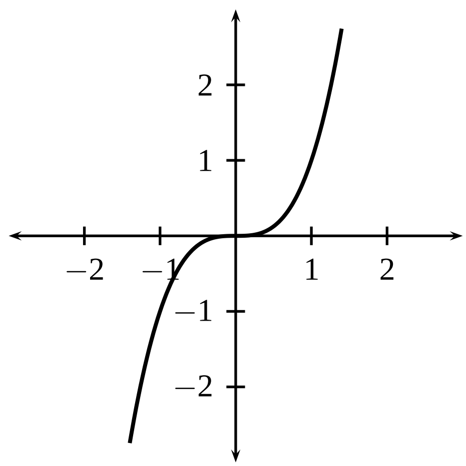

Figura 1.9: Ejemplo de una función cuasiconvexa y cuasicóncava.

Funciones fuertemente cuasiconvexas

A continuación definiremos otra versión de cuasiconvexidad, llamada cuasiconvexidad fuerte, la cual asegura unicidad del mínimo global.

Definición 1.38. Sea $f : S \to \mathbb{R}$, donde $S \subset \mathbb{R}^n$ es un conjunto convexo no vacío. La función f se dice *fuertemente cuasiconvexa* si, para todos $\mathbf{x}_1, \mathbf{x}_2 \in S$, con $\mathbf{x}_1 \neq \mathbf{x}_2$, se tiene que

$$f(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) < \max\{f(\mathbf{x}_1), f(\mathbf{x}_2)\} \text{ para todo } \lambda \in (0,1).$$

La función f se dice *fuertemente cuasicóncava* si $-f$ es fuertemente cuasiconvexa.

La Figura 1.10 ilustra un ejemplo de una función fuertemente cuasiconvexa y un ejemplo de una que no lo es.

Teorema 1.39. *Sea $f : \mathbb{R}^n \to \mathbb{R}$ una función fuertemente cuasiconvexa. Consideremos el Problema 1.35, donde S es un conjunto convexo no vacío en $\mathbb{R}^n$. Si $\mathbf{x}_0$ es un óptimo local, entonces $\mathbf{x}_0$ es un óptimo global y además es único.*

Demostración.
Como $\mathbf{x}_0$ es un óptimo local, entonces existe $\varepsilon > 0$ tal que

$$f(\mathbf{x}_0) \leq f(\mathbf{x}) \text{ para todo } \mathbf{x} \in S \cap B(\mathbf{x}_0, \varepsilon).$$

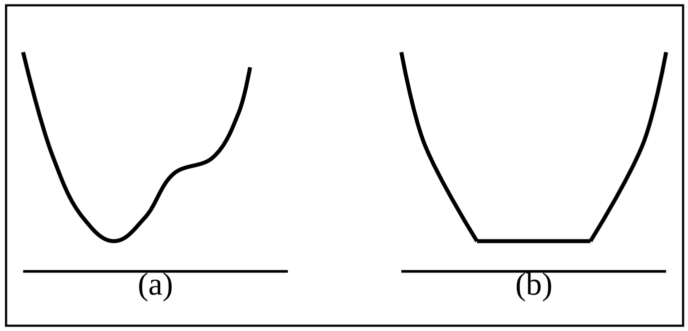

Figura 1.10: (a)Fuertemente cuasiconvexa. (b)No es fuertemente cuasiconvexa.

Supongamos, por contradicción, que existe $\widehat{\mathbf{x}} \in S$ tal que $f(\widehat{\mathbf{x}}) \leq f(\mathbf{x}_0)$ y $\widehat{\mathbf{x}} \neq \mathbf{x}_0$. Como f es fuertemente cuasiconvexa se sigue que

$$f(\lambda\widehat{\mathbf{x}} + (1-\lambda)\mathbf{x}_0) < \text{máx}\{f(\widehat{\mathbf{x}}), f(\mathbf{x}_0)\} = f(\mathbf{x}_0) \text{ para todo } \lambda \in (0,1).$$

Pero para λ suficientemente pequeño, $\lambda\widehat{\mathbf{x}} + (1-\lambda)\mathbf{x}_0 \in S \cap B(\mathbf{x}_0, \varepsilon)$ y, por lo tanto, la desigualdad anterior contradice el hecho de que $\mathbf{x}_0$ sea un óptimo local. Concluyendo así que $\mathbf{x}_0$ es la única solución óptima global. ■

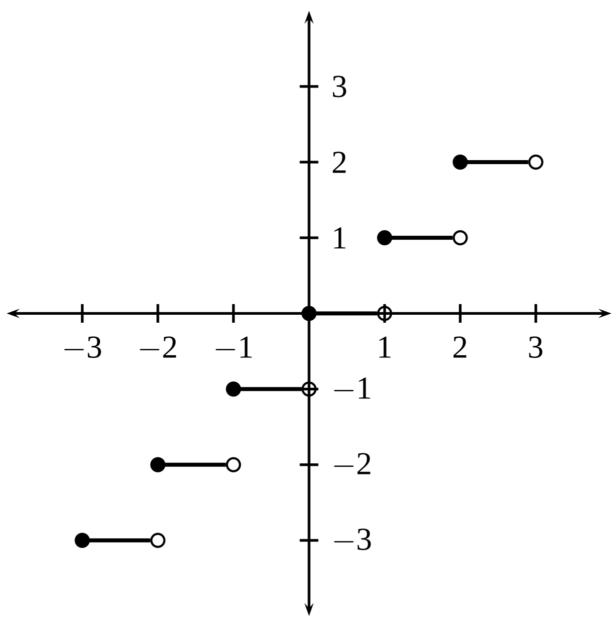

Figura 1.11: Ejemplo de una función cuasiconvexa.

Antes de pasar a definir otro tipo de funciones, tenemos el siguiente ejemplo. Sea $f : \mathbb{R} \to \mathbb{Z}$ la función parte entera $f(x) = [x]$, donde

$$[x] = \text{máx}\{n \in \mathbb{Z} \mid n \leq x\}.$$

La función $f(x)$ está ilustrada en la Figura1.11 y es una función cuasiconvexa que no es fuertemente cuasiconvexa ni convexa.

Funciones pseudoconvexas

Definición 1.40. Sean $S \subset \mathbb{R}^n$ un conjunto abierto no vacío y $f : S \to \mathbb{R}$ una función diferenciable en S. La función f se dice *pseudoconvexa* si

$$\forall \mathbf{x}_1, \mathbf{x}_2 \in S \ \text{ con } \ \nabla f(\mathbf{x}_1) \cdot (\mathbf{x}_2 - \mathbf{x}_1) \geq 0 \ \text{ se tiene que } \ f(\mathbf{x}_1) \leq f(\mathbf{x}_2).$$

Equivalentemente,

$$\text{si } \ f(\mathbf{x}_2) < f(\mathbf{x}_1) \ \text{ entonces } \nabla f(\mathbf{x}_1) \cdot (\mathbf{x}_2 - \mathbf{x}_1) < 0.$$

La función f se dice *estrictamente pseudoconvexa* si $\forall \mathbf{x}_1$, $\mathbf{x}_2 \in S$, $\mathbf{x}_1 \neq \mathbf{x}_2$, que cumplen $\nabla f(\mathbf{x}_1) \cdot (\mathbf{x}_2 - \mathbf{x}_1) \geq 0$ se tiene que $f(\mathbf{x}_1) < f(\mathbf{x}_2)$. Equivalentemente, si $\forall \mathbf{x}_1$, $\mathbf{x}_2 \in S$, $\mathbf{x}_1 \neq \mathbf{x}_2$, tal que $f(\mathbf{x}_2) \leq f(\mathbf{x}_1)$, entonces $\nabla f(\mathbf{x}_1) \cdot (\mathbf{x}_2 - \mathbf{x}_1) < 0$.

La función f se dice *pseudocóncava* (*estrictamente pseudocóncava*) si $-f$ es pseudoconvexa (estrictamente pseudoconvexa).

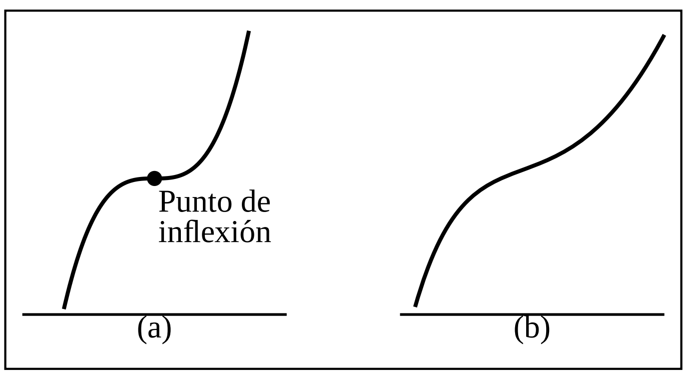

Figura 1.12: (a)No es pseudoconvexa. (b)Pseudoconvexa.

En la Figura 1.12 podemos observar un ejemplo de una función pseudoconvexa y un ejemplo de una función que no lo es. La función que tiene un punto de inflexión no es pseudoconvexa porque si consideramos a x_1 como el punto de inflexión y tomamos $x_2 < x_1$, en x_1 el gradiente de la función es cero, pero no se cumple que $f(x_1) \leq f(x_2)$. Así que se puede verificar, a partir de la deinición, que si f es pseudoconvexa y $\nabla f(\mathbf{x}_0) = \mathbf{0}$, entonces $\mathbf{x}_0$ es un mínimo global de f.

Por otra parte, $f(x) = \arctan(x)$ es un ejemplo de una función que es pseudoconvexa pero no es convexa. Esta función se ilustra en la Figura1.13. También se puede observar que de las definiciones de convexidad y pseudoconvexidad se sigue que toda función convexa, bajo el supuesto de diferenciabilidad, es pseudoconvexa.

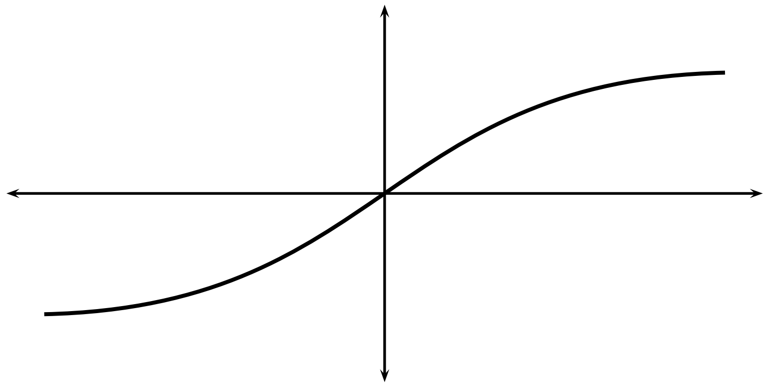

Figura 1.13: Ejemplo de una función pseudoconvexa.

Teorema 1.41. *Sean $S \subset \mathbb{R}^n$ un conjunto convexo abierto no vacío y $f : S \to \mathbb{R}$ una función diferenciable estrictamente pseudoconvexa, entonces f es fuertemente cuasiconvexa.*

Demostración.
Supongamos, por contradicción, que existen $\mathbf{x}_1$, $\mathbf{x}_2 \in S$, $\mathbf{x}_1 \neq \mathbf{x}_2$, y $\lambda \in (0,1)$ tal que

$$\text{máx}\{f(\mathbf{x}_1), f(\mathbf{x}_2)\} \leq f(\mathbf{x}),$$

donde $\mathbf{x} = \lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$. Por la pseudoconvexidad estricta de f se tiene que $f(\mathbf{x}_1) \leq f(\mathbf{x})$, entonces $\nabla f(\mathbf{x}) \cdot (\mathbf{x}_1 - \mathbf{x}) < 0$ y, por lo tanto,

$$\nabla f(\mathbf{x}) \cdot (\mathbf{x}_1 - \mathbf{x}_2) < 0. \tag{1.18}$$

Análogamente, como $f(\mathbf{x}_2) \leq f(\mathbf{x})$, entonces

$$\nabla f(\mathbf{x}) \cdot (\mathbf{x}_2 - \mathbf{x}_1) < 0. \tag{1.19}$$

Contradicción, pues las desigualdades (1.18) y (1.19) no son compatibles. Por lo tanto f es fuertemente cuasiconvexa. ■

1.6

Correspondencias

Definición 1.42. Dado un conjunto $A \subseteq \mathbb{R}^n$, una *correspondencia* $\mathfrak{F} : A \to \mathbb{R}^k$ es una regla que asigna un conjunto $\mathfrak{F}(\mathbf{x}) \subseteq \mathbb{R}^k$ para todo $\mathbf{x} \in A$.

Cabe señalar que una función es una caso especial de correspondencia. Para discutir la idea de continuidad de correspondencias, necesitamos las siguientes definiciones.

Definición 1.43. Sean $A \subset \mathbb{R}^n$ y $Y \subset \mathbb{R}^k$, la gráfica de la correspondencia $\mathfrak{F} : A \to Y$ es el conjunto

$$\{(\mathbf{x}, \mathbf{y}) \in A \times Y \mid \mathbf{y} \in \mathfrak{F}(\mathbf{x})\}.$$

Definición 1.44. Sean $A \subseteq \mathbb{R}^n$ y $Y \subseteq \mathbb{R}^k$ un conjunto cerrado, la correspondencia $\mathfrak{F} : A \to Y$ tiene una *gráfica cerrada* si para cualesquiera dos sucesiones $\mathbf{x}_m \to \mathbf{x} \in A$ y $\mathbf{y}_m \to \mathbf{y}$, con $\mathbf{x}_m \in A$ y $\mathbf{y}_m \in \mathfrak{F}(\mathbf{x}_m)$ para todo m, se tiene $\mathbf{y} \in \mathfrak{F}(\mathbf{x})$.

Sea

$$f(x) = \begin{cases} 1, & 0 \leq x \leq 2; \\ \frac{1}{x-2}, & 2 < x \leq 4. \end{cases}$$

En la Figura 1.14 se ilustra la gráfica de la función antes definida. Note que $f(x)$ tiene una gráfica cerrada.

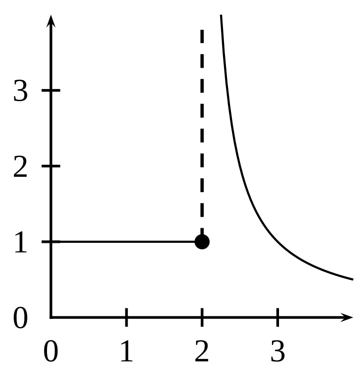

Figura 1.14: Gráfica de $f(x)$.

Definición 1.45. Sean $A \subset \mathbb{R}^n$ y un conjunto cerrado $Y \subset \mathbb{R}^k$, la correspondencia $\mathfrak{F} : A \to Y$ es *superiormente hemicontinua* (shc) si tiene una gráfica cerrada y las imágenes de conjuntos compactos son acotados, esto es, para todo conjunto compacto $B \subset A$ el conjunto

$$\mathfrak{F}(B) = \{\mathbf{y} \in Y \mid \mathbf{y} \in \mathfrak{F}(\mathbf{x}) \text{ para algún } \mathbf{x} \in B\}$$

es acotado.

Sea

$$g(x) = \begin{cases} 1, & 0 \leq x \leq 2; \\ \frac{1}{x-1.5}, & 2 \leq x \leq 4. \end{cases}$$

En la Figura 1.15 se ilustra la gráfica de $g(x)$, que es una correspondencia shc. Notemos que la función del ejemplo anterior no es superiormente hemicontinua ya que la imagen del intervalo $[\frac{3}{2}, \frac{5}{2}]$ no es acotada.

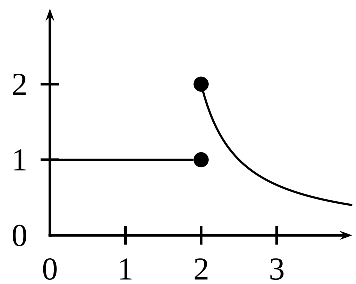

Figura 1.15: Gráfica de $g(x)$.

Teorema 1.46. *Sean $A \subseteq \mathbb{R}^n$ y $Y \subseteq \mathbb{R}^k$ un conjunto cerrado, entonces la función $f : A \to Y$ es superiormente hemicontinua si y sólo si es continua.*

La prueba del teorema anterior se puede consultar en Mas-Colell (1995).

Definición 1.47. Sean $A \subset \mathbb{R}^n$ y $Y \subset \mathbb{R}^k$ un conjunto compacto, la correspondencia $\mathfrak{F} : A \to Y$ es inferiormente hemicontinua si para toda sucesión $\mathbf{x}_m \to \mathbf{x} \in A$ con $\mathbf{x}_m \in A$ para todo m, y para todo $\mathbf{y} \in \mathfrak{F}(\mathbf{x})$, podemos encontrar una sucesión $\mathbf{y}_m \to \mathbf{y}$ y un $M \in \mathbb{N}$ tal que $\mathbf{y}_m \in \mathfrak{F}(\mathbf{x}_m)$ para todo $m > M$.

Finalmente, cuando una correspondencia es superior e inferiormente hemicontinua, decimos que es *continua*.

Capítulo 2

Programación no lineal

En este capítulo se establecen las condiciones para optimizar el problema:

$$\begin{array}{cc} \text{Maximizar} & f(\mathbf{x}) \\ \text{sujeto a} & \mathbf{x} \in S \end{array}$$

donde f es una función convexa y S es un conjunto convexo. La mayoría de los problemas económicos de optimización son de esta naturaleza ya que el hecho de limitar la función objetivo y las restricciones a un molde lineal, en ocasiones, puede ser inconveniente; pues es más apropiada una formulación no lineal que una lineal para describir el problema a tratar.

2.1

Conceptos básicos

A lo largo de esta sección presentamos algunos conceptos de programación lineal que serán de utilidad para el desarrollo del presente capítulo.

Definición 2.1. Sea V un espacio vectorial. Se dice que los vectores $\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_m} \in V$ son *linealmente dependientes* si existen escalares $a_1, a_2, \ldots, a_m \in \mathbb{R}$, no todos cero, tales que

$$a_1\mathbf{x_1} + a_2\mathbf{x_2} + \ldots + a_m\mathbf{x_m} = \mathbf{0}.$$

En caso contrario se dice que los vectores son *linealmente independientes*.

Definición 2.2. Dado $\mathbf{c} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$ y $b \in \mathbb{R}$ se define el *semiespacio inferior*, denotado por $S_-(\mathbf{c}; b)$, como

$$S_-(\mathbf{c}; b) = \{\mathbf{x} \in \mathbb{R}^n \mid \langle \mathbf{c}, \mathbf{x} \rangle \leq b\}.$$

Así mismo se define el *semiespacio superior*, $S_+(\mathbf{c};b)$, como

$$S_+(\mathbf{c};b) = \{\mathbf{x} \in \mathbb{R}^n \mid \langle \mathbf{c},\mathbf{x}\rangle \geq b\}.$$

Por otra parte, la intersección de $S_+(\mathbf{c};b)$ y $S_-(\mathbf{c};b)$ determina un *hiperplano*:

$$H(\mathbf{c};b) = S_+(\mathbf{c};b) \cap S_-(\mathbf{c};b).$$

En la siguiente figura se ilustran los conjuntos antes definidos.

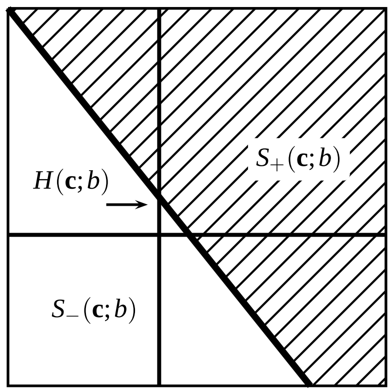

Figura 2.1: Ilustración de $S_-(\mathbf{c};b)$, $S_+(\mathbf{c};b)$ y $H(\mathbf{c};b)$.

Definición 2.3. Un conjunto no vacío $C \subseteq \mathbb{R}^n$ es llamado *cono* si $\mathbf{x} \in C$ implica que $\lambda\mathbf{x} \in C$ para todo $\lambda \geq 0$. Si además, C es un conjunto convexo, C es llamado *cono convexo*.

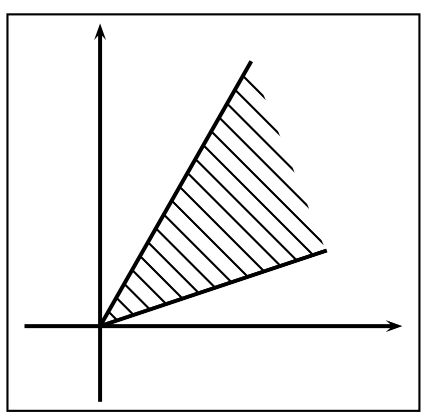

Figura 2.2: Cono convexo

En la Figura 2.2 se ilustra un cono convexo. Los siguientes teoremas sólo se utilizarán para las demostraciones de los Teoremas 2.6 y 2.7, para mayores detalles consultar Bazaraa (1993).

Teorema 2.4. *Sean $S \subseteq \mathbb{R}^n$ un conjunto convexo cerrado no vacío y $\mathbf{y} \notin S$, entonces existen un vector no cero $\mathbf{p}$ y un escalar α tal que $\alpha < \mathbf{p} \cdot \mathbf{y}$ y $\mathbf{p} \cdot \mathbf{x} \leq \alpha$ para cada $\mathbf{x} \in S$.*

Teorema 2.5. *Sean S_1 y S_2 conjuntos convexos no vacíos en $\mathbb{R}^n$ y supongamos que $S_1 \cap S_2 = \emptyset$, entonces existe un hiperplano que separa a S_1 y S_2; es decir, existe un vector no cero $\mathbf{p} \in \mathbb{R}^n$ tal que*

$$\sup\{\mathbf{p} \cdot \mathbf{x} \mid \mathbf{x} \in S_2\} \leq \text{ínf}\{\mathbf{p} \cdot \mathbf{x} \mid \mathbf{x} \in S_1\}.$$

Los dos teoremas que se presentan a continuación serán utilizados para probar algunos resultados importantes que se expondrán más adelante. Por comodidad de notación, sólo en los Teoremas 2.6 y 2.7, un vector $\mathbf{x} \in \mathbb{R}^n$ se denotará por una matriz de n filas y una columna.

Teorema 2.6. ***Teorema de Farkas***. *Sean $\mathbf{A}$ una matriz $m \times n$ y $\mathbf{c} \in \mathbb{R}^n$ un vector, entonces a lo más uno de los siguientes sistemas tiene solución:*

$$Sistema1 \quad \mathbf{Ax} \leq \mathbf{0} \;\; y \;\; \mathbf{c}^t\mathbf{x} > 0 \quad para\ algún \;\; \mathbf{x} \in \mathbb{R}^n.$$

$$Sistema2 \quad \mathbf{A}^t\mathbf{y} = \mathbf{c} \;\; y \;\; \mathbf{y} \geq \mathbf{0} \quad para\ algún \;\; \mathbf{y} \in \mathbb{R}^m.$$

Demostración.

Supongamos que el $Sistema2$ tiene solución, es decir, existe $\mathbf{y} \geq \mathbf{0}$ tal que $\mathbf{A}^t\mathbf{y} = \mathbf{c}$. Sea $\mathbf{x}$ tal que $\mathbf{Ax} \leq \mathbf{0}$, entonces $\mathbf{c}^t\mathbf{x} = \mathbf{y}^t\mathbf{Ax} \leq 0$. Por lo tanto, el $Sistema1$ no tiene solución.

Ahora supongamos que el $Sistema2$ no tiene solución. Definamos el siguiente conjunto

$$S = \{\mathbf{x} \mid \mathbf{x} = \mathbf{A}^t\mathbf{y}, \;\; \mathbf{y} \geq \mathbf{0}\}.$$

Note que S es un conjunto convexo cerrado y que $\mathbf{c} \notin S$. Por el Teorema 2.4, existen un vector $\mathbf{p} \in \mathbb{R}^n$ y un escalar α tal que $\alpha < \mathbf{p}^t\mathbf{c}$ y $\mathbf{p}^t\mathbf{x} \leq \alpha$ para todo $\mathbf{x} \in S$. Como $\mathbf{0} \in S$, entonces $\alpha \geq 0$ y, por lo tanto, $\mathbf{p}^t\mathbf{c} > 0$; también se cumple que $\alpha \geq \mathbf{p}^t\mathbf{A}^t\mathbf{y} = \mathbf{y}^t\mathbf{Ap}$ para todo $\mathbf{y} \geq \mathbf{0}$. Como $\mathbf{y} \geq \mathbf{0}$ puede ser tomado arbitrariamente grande, la última desigualdad implica que $\mathbf{Ap} \leq \mathbf{0}$. Hemos construido un vector $\mathbf{p} \in \mathbb{R}^n$ tal que $\mathbf{Ap} \leq \mathbf{0}$ y $\mathbf{p}^t\mathbf{c} > 0$; por lo tanto, el $Sistema1$ tiene una solución. ■

Teorema 2.7. ***Teorema de Gordan***. *Sea $\mathbf{A}$ una matriz $m \times n$, entonces sólo uno de los siguientes sistemas tiene solución:*

$$Sistema1 \quad \mathbf{Ax} < \mathbf{0} \quad para\ algún \;\; \mathbf{x} \in \mathbb{R}^n.$$

$$Sistema2 \quad \mathbf{A}^t\mathbf{p} = \mathbf{0} \;\; y \;\; \mathbf{p} \geq \mathbf{0} \quad para\ algún\ vector \;\; \mathbf{p} \in \mathbb{R}^m \ no\ cero.$$

Demostración.

Primero probaremos que si el $Sistema1$ tiene una solución $\tilde{\mathbf{x}}$, entonces no podemos tener una solución para $\mathbf{A}^t\mathbf{p} = \mathbf{0}$, $\mathbf{p} \geq \mathbf{0}$, $\mathbf{p}$ no cero. Supongamos, por contradicción, que existe una solución $\tilde{\mathbf{p}}$ para el $Sistema2$. Como $\mathbf{A}\tilde{\mathbf{x}} < \mathbf{0}$, $\tilde{\mathbf{p}} \geq \mathbf{0}$, y $\mathbf{p} \neq \mathbf{0}$, entonces tenemos que $\tilde{\mathbf{p}}^t\mathbf{A}\tilde{\mathbf{x}} < 0$, es decir, $\tilde{\mathbf{x}}^t\mathbf{A}^t\tilde{\mathbf{p}} < 0$. Pero por suposición $\mathbf{A}^t\tilde{\mathbf{p}} = 0$, entonces $\tilde{\mathbf{x}}^t\mathbf{A}^t\tilde{\mathbf{p}} = 0$ y llegamos a una contradicción. Por lo tanto, el $Sistema2$ no tiene solución.

Ahora supongamos que el $Sistema1$ no tiene solución. Consideremos los siguientes conjuntos,

$$S_1 = \{\mathbf{z} \mid \mathbf{z} = \mathbf{A}\mathbf{x},\ \ \mathbf{x} \in \mathbb{R}^n\}$$
$$S_2 = \{\mathbf{z} \mid \mathbf{z} < \mathbf{0}\}.$$

Note que S_1 y S_2 son conjuntos convexos no vacíos tales que $S_1 \cap S_2 = \emptyset$, entonces, por el Teorema 2.5, existe un hiperplano que separa a S_1 y S_2; es decir, existe un vector $\mathbf{p}$, no cero, tal que

$$\mathbf{p}^t\mathbf{z} \leq \mathbf{p}^t\mathbf{A}\mathbf{x} \text{ para todo } \mathbf{x} \in \mathbb{R}^n \text{ y } \mathbf{z} \in \overline{S_2}.$$

Como cada componente de $\mathbf{z}$ puede ser un número negativo arbitrariamente grande, se sigue que $\mathbf{p} \geq \mathbf{0}$. También, haciendo $\mathbf{z} = \mathbf{0}$, se debe cumplir que $\mathbf{p}^t\mathbf{A}\mathbf{x} \geq 0$ para cada $\mathbf{x} \in \mathbb{R}^n$. Haciendo $\mathbf{x} = -\mathbf{A}^t\mathbf{p}$, se sigue que $-\|\mathbf{A}^t\mathbf{p}\|^2 \geq 0$ y entonces $\mathbf{A}^t\mathbf{p} = \mathbf{0}$. Por lo tanto el $Sistema2$ tiene una solución. ■

2.2

Optimización de problemas sin restricciones

La forma general de un problema sin restricciones es: minimizar o maximizar una función $f(\mathbf{x})$ sin ninguna restricción sobre el vector $\mathbf{x}$. Los problemas sin restricciones no son muy utilizados en aplicaciones prácticas, pero empezaremos estudiando este tipo de problemas porque las condiciones de optimalidad para problemas con restricciones surgen como una extensión lógica del análisis que se hará en esta sección.

2.2.1

Condiciones necesarias de primer orden

Dado $\mathbf{x} \in \mathbb{R}^n$ queremos determinar, de ser posible, si $\mathbf{x}$ es un óptimo local o global de una función f. Para ello necesitamos caracterizar el punto óptimo, y el hecho de que f sea diferenciable nos ayuda a obtener dicha caracterización.

Teorema 2.8. *Sea* $f : \mathbb{R}^n \to \mathbb{R}$ *una función diferenciable en* $\mathbf{x}_0$*. Si existe un vector* $\mathbf{d}$ *tal que* $\nabla f(\mathbf{x}_0) \cdot \mathbf{d} < 0$*, entonces existe* $\delta > 0$ *tal que* $f(\mathbf{x}_0 + \lambda \mathbf{d}) < f(\mathbf{x}_0)$ *para cada* $\lambda \in (0, \delta)$*. El vector* $\mathbf{d}$ *es llamado dirección descendente de* f *en* $\mathbf{x}_0$*.*

Demostración.

Por la diferenciabilidad de f en $\mathbf{x}_0$ tenemos

$$f(\mathbf{x}_0 + \lambda \mathbf{d}) = f(\mathbf{x}_0) + \lambda \nabla f(\mathbf{x}_0) \cdot \mathbf{d} + \lambda \|\mathbf{d}\| E(\mathbf{x}_0; \lambda \mathbf{d})$$

donde $E(\mathbf{x}_0; \lambda \mathbf{d}) \to 0$ cuando $\lambda \to 0$. Ordenando los términos y dividiendo por λ tenemos

$$\frac{f(\mathbf{x}_0 + \lambda \mathbf{d}) - f(\mathbf{x}_0)}{\lambda} = \nabla f(\mathbf{x}_0) \cdot \mathbf{d} + \|\mathbf{d}\| E(\mathbf{x}_0; \lambda \mathbf{d}).$$

Tomando el límite cuando $\lambda \to 0$ y como $\nabla f(\mathbf{x}_0) \cdot \mathbf{d} < 0$ y $E(\mathbf{x}_0; \lambda \mathbf{d}) \to 0$ cuando $\lambda \to 0$, entonces existe $\delta > 0$ tal que

$$\nabla f(\mathbf{x}_0) \cdot \mathbf{d} + \|\mathbf{d}\| E(\mathbf{x}_0; \lambda \mathbf{d}) < 0 \ \text{ para todo } \lambda \in (0, \delta).$$

De esta forma se tiene que $f(\mathbf{x}_0 + \lambda \mathbf{d}) < f(\mathbf{x}_0)$ para cada $\lambda \in (0, \delta)$. ■

Corolario 2.9. *Sea* $f : \mathbb{R}^n \to \mathbb{R}$ *una función diferenciable en* $\mathbf{x}_0$*. Si* $\mathbf{x}_0$ *es un mínimo local, entonces* $\nabla f(\mathbf{x}_0) = \mathbf{0}$*.*

Demostración.

Supongamos, por contradicción, que $\nabla f(\mathbf{x}_0) \neq 0$ y sea $\mathbf{d} = -\nabla f(\mathbf{x}_0)$. Entonces se cumple que $\nabla f(\mathbf{x}_0) \cdot \mathbf{d} = -\|\nabla f(\mathbf{x}_0)\|^2 < 0$ y, por el Teorema 2.8, existe $\delta > 0$ tal que

$$f(\mathbf{x}_0 + \lambda \mathbf{d}) < f(\mathbf{x}_0) \ \text{ para todo } \lambda \in (0, \delta).$$

Esto contradice el hecho de que $\mathbf{x}_0$ sea un mínimo local. ■

2.2.2

Condiciones suficientes

Las condiciones discutidas en la sección anterior son condiciones necesarias, es decir, se deben satisfacer para toda solución óptima local; pero un punto que satisfaga dichas condiciones no necesariamente es un mínimo local. El Teorema 2.10 establece que la condición necesaria, $\nabla f(\mathbf{x}_0) = 0$, es también suficiente para que $\mathbf{x}_0$ sea un mínimo global de f, si dicha función es pseudoconvexa en $\mathbf{x}_0$.

Teorema 2.10. *Sea $f : \mathbb{R}^n \to \mathbb{R}$ una función pseudoconvexa en $\mathbf{x}_0$, entonces $\mathbf{x}_0$ es un mínimo global si y sólo si $\nabla f(\mathbf{x}_0) = 0$.*

Demostración.
($\Rightarrow$ Supongamos que $\mathbf{x}_0$ es un mínimo global, por el Corolario 2.9 tenemos que $\nabla f(\mathbf{x}_0) = 0$.

$\Leftarrow$) Supongamos que $\nabla f(\mathbf{x}_0) = 0$, entonces $\nabla f(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0) = 0$ para todo $\mathbf{x} \in \mathbb{R}^n$. Por la pseudoconvexidad de f en $\mathbf{x}_0$ se sigue que

$$f(\mathbf{x}_0) \leq f(\mathbf{x}) \text{ para todo } \mathbf{x} \in \mathbb{R}^n.$$

Completando la prueba. ■

2.3

Optimización de problemas con restricciones

En esta sección analizaremos, en primer término, una condición necesaria para optimizar el problema: minimizar $f(\mathbf{x})$ sujeto a $\mathbf{x} \in S$. Después veremos que el conjunto S es la región factible de un problema de programación no lineal de la forma: minimizar $f(\mathbf{x})$ sujeta a $g(\mathbf{x}) \leq 0$ y $\mathbf{x} \in X$.

2.3.1

Condiciones geométricas

Desarrollaremos una condición necesaria para optimizar el problema: minimizar $f(\mathbf{x})$ sujeto a $\mathbf{x} \in S$ utilizando el cono de direcciones factibles que definimos en seguida.

Definición 2.11. Sean $S \subseteq \mathbb{R}^n$ un conjunto no vacío y $\mathbf{x}_0 \in \overline{S}$. El *cono de direcciones factibles* de S en $\mathbf{x}_0$, denotado por D, está definido como

$$D = \{\mathbf{d} \mid \mathbf{d} \neq \mathbf{0} \text{ y } \mathbf{x}_0 + \lambda \mathbf{d} \in S\}$$

para toda $\lambda \in (0, \delta)$ y para algún $\delta > 0$. Cada vector no cero $\mathbf{d} \in D$ es llamado una *dirección factible*.

De la Definición 2.11, es claro que al hacer un pequeño movimiento de $\mathbf{x}_0$ a lo largo de un vector $\mathbf{d} \in D$ obtenemos puntos factibles; pero podemos decir aún más, por el Teorema 2.8, si $\nabla f(\mathbf{x}_0) \cdot \mathbf{d} < 0$, entonces $\mathbf{d}$ es una mejor dirección de movimiento; es decir, empezando en $\mathbf{x}_0$, un pequeño movimiento a lo largo de $\mathbf{d}$ reducirá el valor de f.

En el siguiente teorema probaremos que si $\mathbf{x}_0$ es un mínimo local y si $\nabla f(\mathbf{x}_0) \cdot \mathbf{d} < 0$, entonces $\mathbf{d} \notin D$; es decir, una condición necesaria para un óptimo local es que cada mejor dirección de movimiento no sea una dirección factible.

Teorema 2.12. *Consideremos el problema de minimizar $f(\mathbf{x})$ sujeto a $\mathbf{x} \in S$, donde $f : \mathbb{R}^n \to \mathbb{R}$ y $S \subseteq \mathbb{R}^n$ es un conjunto no vacío. Supongamos que f es diferenciable en $\mathbf{x}_0 \in \overline{S}$. Si $\mathbf{x}_0$ es un óptimo local, entonces $F_0 \cap D = \emptyset$, donde*

$$F_0 = \{\mathbf{d} \mid \nabla f(\mathbf{x}_0) \cdot \mathbf{d} < 0\}$$

y D es el cono de direcciones factibles de S en $\mathbf{x}_0$.

Demostración.
Supongamos, por contradicción, que existe un vector $\mathbf{d} \in F_0 \cap D$; entonces, por el Teorema 2.8, existe $\delta_1 > 0$ tal que

$$f(\mathbf{x}_0 + \lambda \mathbf{d}) < f(\mathbf{x}_0) \ \text{ para cada } \ \lambda \in (0, \delta_1). \tag{2.1}$$

Más aún, por la Definición 2.11, existe $\delta_2 > 0$ tal que

$$(\mathbf{x}_0 + \lambda \mathbf{d}) \in S \ \text{ para cada } \ \lambda \in (0, \delta_2). \tag{2.2}$$

De esta forma, (2.1) y (2.2) contradicen el hecho de que $\mathbf{x}_0$ sea un óptimo local de f. Por lo tanto $F_0 \cap D = \emptyset$. ■

Ahora especificamos la región factible, S, de la siguiente forma:

$$S = \{\mathbf{x} \in X \mid g_i(\mathbf{x}) \leq 0\}$$

donde $g_i : \mathbb{R}^n \to \mathbb{R}$ para $i = 1, \ldots, m$ y $X \subseteq \mathbb{R}^n$ es un conjunto abierto no vacío.

Esto nos lleva a la forma general de un problema de programación no lineal con rectricciones de desigualdad:

Problema 2.13.

$$\begin{array}{lll} \text{Minimizar} & f(\mathbf{x}) & \\ \text{sujeto a} & g_i(\mathbf{x}) \leq 0 & \text{para } i = 1, \ldots, m \\ & \mathbf{x} \in X & \end{array}$$

Recordemos que una condición necesaria para que $\mathbf{x}_0$ sea un óptimo local es que $F_0 \cap D = \emptyset$, donde F_0 es un semiespacio abierto definido en términos del vector gradiente $\nabla f(\mathbf{x}_0)$ y D es el cono de direcciones factibles, el cual no necesariamente está definido en términos de los gradientes de las funciones que determinan la región factible S. En el siguiente teorema veremos que podemos definir un cono abierto G_0 definido en términos de los gradientes de las funciones que determinan la región factible que se satisfacen con igualdad en $\mathbf{x}_0$, tal que $G_0 \subset D$. Como $F_0 \cap D = \emptyset$ se debe cumplir en $\mathbf{x}_0$ y como $G_0 \subset D$, entonces $F_0 \cap G_0 = \emptyset$ también es una condición necesaria para que $\mathbf{x}_0$ sea un óptimo local.

Teorema 2.14. *Sean $g_i : \mathbb{R}^n \to \mathbb{R}$ para $i = 1, \ldots, m$ y $X \subseteq \mathbb{R}^n$ un conjunto abierto no vacío. Considere el Problema 2.13, sea $\mathbf{x}_0$ un punto factible y sea $I = \{i \mid g_i(\mathbf{x}_0) = 0\}$. Más aún, supongamos que f y g_i con $i \in I$ son diferenciables en $\mathbf{x}_0$, y que g_i con $i \notin I$ es continua en $\mathbf{x}_0$. Si $\mathbf{x}_0$ es un óptimo local, entonces $F_0 \cap G_0 = \emptyset$, donde*

$$
\begin{aligned}
F_0 &= \{\mathbf{d} \mid \nabla f(\mathbf{x}_0) \cdot \mathbf{d} < 0\} \\
G_0 &= \{\mathbf{d} \mid \nabla g_i(\mathbf{x}_0) \cdot \mathbf{d} < 0 \ \text{ para todo } \ i \in I\}.
\end{aligned}
$$

Demostración.

Sea $\mathbf{d} \in G_0$, como $\mathbf{x}_0 \in X$ y X es un conjunto abierto, existe $\delta_1 > 0$ tal que

$$\mathbf{x}_0 + \lambda \mathbf{d} \in X \ \text{ para } \ \lambda \in (0, \delta_1). \tag{2.3}$$

Como $g_i(\mathbf{x}_0) < 0$ y g_i es continua en $\mathbf{x}_0$ para $i \notin I$, existe $\delta_2 > 0$ tal que

$$g_i(\mathbf{x}_0 + \lambda \mathbf{d}) < 0 \ \text{ para } \ \lambda \in (0, \delta_2) \text{ y para } i \notin I. \tag{2.4}$$

Por último, como $\mathbf{d} \in G_0$ se tiene que $\nabla g_i(\mathbf{x}_0) \cdot \mathbf{d} < 0$ para cada $i \in I$; entonces, por el Teorema 2.8, existe $\delta_3 > 0$ tal que

$$g_i(\mathbf{x}_0 + \lambda \mathbf{d}) < g_i(\mathbf{x}_0) = 0 \ \text{ para } \ \lambda \in (0, \delta_3) \text{ y para } i \in I. \tag{2.5}$$

De (2.3), (2.4) y (2.5), se sigue que los puntos de la forma $\mathbf{x}_0 + \lambda \mathbf{d}$ son puntos factibles para el Problema 2.13 para todo $\lambda \in (0, \delta)$, donde $\delta = \text{mín}(\delta_1, \delta_2, \delta_3)$. De aquí se tiene que $\mathbf{d} \in D$, el cono de direcciones factibles en $\mathbf{x}_0$. De esta forma, hemos probado que si $\mathbf{d} \in G_0$ entonces $\mathbf{d} \in D$, es decir, $G_0 \subset D$. Como $\mathbf{x}_0$ es un óptimo local del Problema 2.13, por el Teorema 2.12, $F_0 \cap D = \emptyset$. Además, como $G_0 \subset D$ se sigue que $F_0 \cap G_0 = \emptyset$. ∎

El siguiente es un ejemplo ilustrativo sobre una aplicación del Teorema 2.14.

Ejemplo 2.15. Consideremos el siguiente problema.

$$
\begin{array}{ll}
\text{Minimizar} & (x_1 - 3)^2 + (x_2 - 2)^2 \\
\text{sujeto a} & x_1^2 + x_2^2 \leq 5 \\
 & x_1 + x_2 \leq 3 \\
 & x_1 \geq 0 \\
 & x_2 \geq 0
\end{array}
$$

En este caso, hacemos

$$g_1(\mathbf{x}) = x_1^2 + x_2^2 - 5$$
$$g_2(\mathbf{x}) = x_1 + x_2 - 3$$
$$g_3(\mathbf{x}) = -x_1$$
$$g_4(\mathbf{x}) = -x_2$$

y $X = \mathbb{R}^2$. Sea $\mathbf{x}_0 = (\frac{9}{5}, \frac{6}{5})$. Note que la única restricción que se cumple con igualdad es g_2 y que

$$\nabla f(\mathbf{x}_0) = (\frac{-12}{5}, \frac{-8}{5}) \quad \text{y} \quad \nabla g_2(\mathbf{x}_0) = (1,1).$$

Los conjuntos F_0 y G_0, trasladados del origen al punto $(\frac{9}{5}, \frac{6}{5})$ por facilidad, se ilustran el la Figura 2.3. Como $F_0 \cap G_0 \neq \emptyset$, $\mathbf{x}_0 = (\frac{9}{5}, \frac{6}{5})$ no es una solución óptima local.

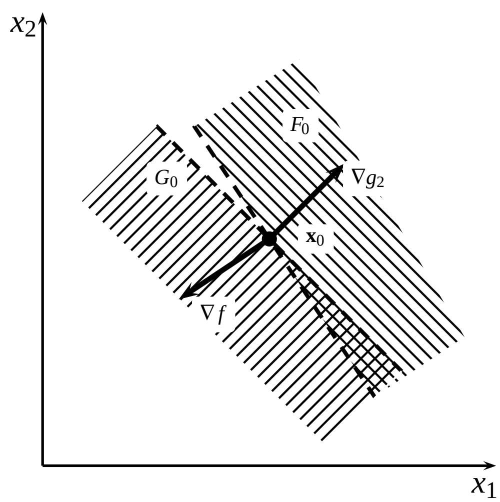

Figura 2.3: $F_0 \cap G_0 \neq \emptyset$

Ahora consideremos el punto $\mathbf{x}_1 = (2,1)$, note que en esta ocasión las restricciones que se cumplen con igualdad son g_1 y g_2. Los gradientes en este punto son:

$$\nabla f(\mathbf{x}_1) = (-2,-2), \quad \nabla g_1(\mathbf{x}_1) = (4,2) \quad \text{y} \quad \nabla g_2(\mathbf{x}_1) = (1,1).$$

Los conjuntos F_0 y G_0 se ilustran el la Figura 2.4; observe que $F_0 \cap G_0 = \emptyset$. Note que el Teorema 2.14 establece una condición necesaria, por lo tanto no garantiza que $\mathbf{x}_1 = (2,1)$ sea un punto óptimo. Sólo podemos concluir que $\mathbf{x}_1$ es un candidato a óptimo.

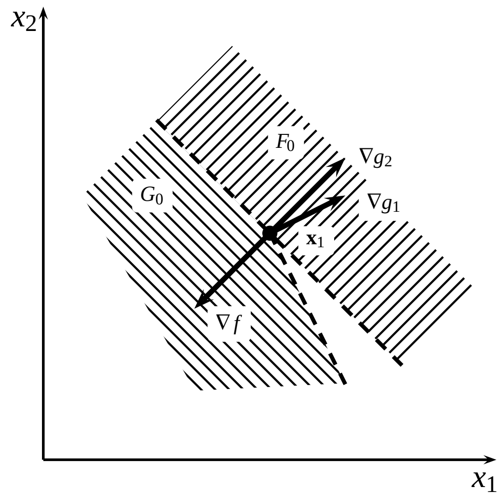

Figura 2.4: $F_0 \cap G_0 = \emptyset$.

2.3.2

Condiciones de Fritz John

En esta parte analizaremos las condiciones de optimalidad necesarias establecidas por Fritz John, las cuales están en términos de los gradientes de la función objetivo y de las funciones que determinan la región factible.

Teorema 2.16. *Sean X un conjunto abierto no vacío en $\mathbb{R}^n$, $f:\mathbb{R}^n \to \mathbb{R}$ y $g_i:\mathbb{R}^n \to \mathbb{R}$ para $i = 1,\dots,m$. Considere el Problema 2.13, sea $\mathbf{x}_0$ un punto factible y sea $I = \{i \mid g_i(\mathbf{x}_0) = 0\}$. Más aún, supongamos que f y g_i con $i \in I$ son diferenciables en $\mathbf{x}_0$, y que g_i con $i \notin I$ son continuas en $\mathbf{x}_0$. Si $\mathbf{x}_0$ es un óptimo local, entonces existen escalares u_0 y u_i para $i \in I$ tales que*

$$\begin{array}{rcll} u_0\nabla f(\mathbf{x}_0) + \sum_{i\in I} u_i \nabla g_i(\mathbf{x}_0) & = & \mathbf{0} & \\ u_0, u_i & \geq & 0 & \textit{para } i \in I \\ (u_0, \mathbf{u}_I) & \neq & (0,\mathbf{0}). & \end{array} \tag{2.6}$$

donde $\mathbf{u}_I$ es el vector cuyas componentes son u_i para $i \in I$. Más aún, si las funciones g_i con $i \notin I$ también son diferenciables en $\mathbf{x}_0$, entonces las condiciones de Fritz John pueden ser escritas de la siguiente forma:

$$\begin{array}{rcll} u_0\nabla f(\mathbf{x}_0) + \sum_{i=1}^{m} u_i \nabla g_i(\mathbf{x}_0) & = & \mathbf{0} & \\ u_i g_i(\mathbf{x}_0) & = & 0 & \textit{para } i = 1,\dots,m \\ u_0, u_i & \geq & 0 & \textit{para } i = 1,\dots,m \\ (u_0, \mathbf{u}) & \neq & (0,\mathbf{0}). & \end{array} \tag{2.7}$$

donde $\mathbf{u}$ *es el vector cuyas componentes son* u_i *para* $i = 1, \ldots, m$.

Demostración.
Como $\mathbf{x}_0$ es un óptimo local del Problema 2.13, por el Teorema 2.14, no existe un vector $\mathbf{d}$ tal que $\nabla f(\mathbf{x}_0) \cdot \mathbf{d} < 0$ y $\nabla g_i(\mathbf{x}_0) \cdot \mathbf{d} < 0$ para cada $i \in I$.

Sea $\mathbf{A}$ la matriz cuyas filas son $\nabla f(\mathbf{x}_0)$ y $\nabla g_i(\mathbf{x}_0)$ con $i \in I$. La condición establecida en el Teorema 2.14 es equivalente a afirmar que el sistema $\mathbf{A}\mathbf{d}^t < \mathbf{0}$ es inconsistente. Por el Teorema 2.7, existe un vector columna no cero $\mathbf{p} \geq \mathbf{0}$, tal que $\mathbf{A}^t\mathbf{p} < \mathbf{0}$.

Denotando las componentes de $\mathbf{p}$ por u_0 y u_i para $i \in I$, obtenemos la prueba para (2.6); la prueba de (2.7) se obtiene de la misma forma que la anterior, sólo que en esta ocasión hacemos $u_i = 0$ para $i \notin I$. ■

En las condiciones de Fritz John, los escalares u_0 y u_i, para $i = 1, \ldots, m$, son llamados *multiplicadores de Lagrange*. La condición $u_i g_i(\mathbf{x}_0) = 0$, para $i = 1, \ldots, m$, es llamada *condición de holgura complementaria*; se requiere que $u_i = 0$ si $g_i(\mathbf{x}_0) < 0$ y $u_i \geq 0$ si $g_i(\mathbf{x}_0) = 0$.

Los ejemplos que se exponen a continuación son para ilustrar el Teorema 2.16.

Ejemplo 2.17. Consideremos el siguiente problema.

$$\begin{array}{lr} \text{Minimizar} & (x_1-3)^2+(x_2-2)^2 \\ \text{sujeto a} & x_1^2+x_2^2 \leq 5 \\ & x_1+2x_2 \leq 4 \\ & -x_1 \leq 0 \\ & -x_2 \leq 0 \end{array}$$

La región factible de este problema se ilustra en la Figura 2.5. Ahora probaremos que las condiciones de Fritz John se satisfacen en el punto óptimo $\mathbf{x}_0 = (2,1)$. Observe que el conjunto de restricciones que se cumplen con igualdad en $\mathbf{x}_0$ es $I = \{1,2\}$, entonces los multiplicadores de Lagrange u_3 y u_4 asociados con $-x_1 \leq 0$ y $-x_2 \leq 0$, respectivamente, son iguales a cero. Tenemos que

$$\nabla f(\mathbf{x}_0) = (-2,-2), \quad \nabla g_1(\mathbf{x}_0) = (4,2) \quad \text{y} \quad \nabla g_2(\mathbf{x}_0) = (1,2).$$

Haciendo $u_0 = 3$, $u_1 = 1$ y $u_2 = 2$ se satisfacen las condiciones de Fritz John, puesto que tenemos un vector no cero $(u_0, u_1, u_2) \geq \mathbf{0}$ que satisface lo siguiente:

$$u_0(-2,-2)+u_1(4,2)+u_2(1,2) = (0,0).$$

Para fines ilustrativos, vamos a checar si las condiciones de Fritz John se satisfacen en el punto $\mathbf{x}_1 = (0,0)$, que no es un óptimo. En esta ocasión, $I = \{3,4\}$ y $u_1 = u_2 = 0$. Tenemos que

$$\nabla f(\mathbf{x}_1) = (-6,-4), \quad \nabla g_3(\mathbf{x}_1) = (-1,0) \quad \text{y} \quad \nabla g_4(\mathbf{x}_1) = (0,-1).$$

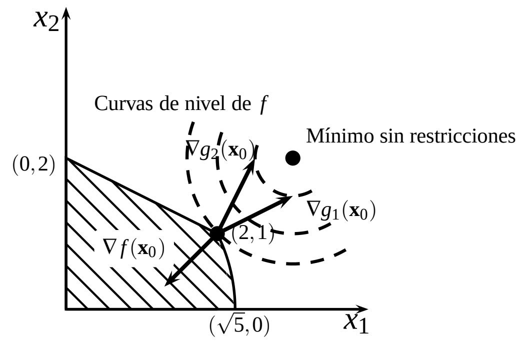

Figura 2.5: Ilustración del Ejemplo 2.17.

Note que

$$u_0(-6,-4)+u_3(-1,0)+u_4(0,-1)=(0,0)$$

se satisface si y sólo si $u_3=-6u_0$ y $u_4=-4u_0$. Si $u_0>0$, entonces $u_3,u_4<0$ y llegamos a una contradicción; si $u_0=0$, entonces $u_3=u_4=0$, lo cual contradice el hecho de que (u_0,u_3,u_4) es un vector no cero. Por lo tanto, las condiciones de Fritz John no se satisfacen en el punto $\mathbf{x}_1=(0,0)$, lo cual prueba que el origen no es un punto óptimo.

Ejemplo 2.18. Consideremos el siguiente problema de Kuhn y Tucker [1951].

$$\begin{array}{lr} \text{Minimizar} & -x_1 \\ \text{sujeto a} & x_2-(1-x_1)^3 \le 0 \\ & -x_2 \le 0 \end{array}$$

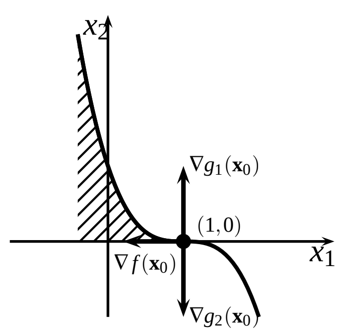

Figura 2.6: Ilustración del Ejemplo 2.18.

La región factible de este problema se ilustra en la Figura 2.6. Ahora probaremos que las condiciones de Fritz John se satisfacen en el punto óptimo $\mathbf{x}_0=(1,0)$. Observe que el conjunto de restricciones que se cumplen con igualdad es $I=\{1,2\}$. Tenemos que

$$\nabla f(\mathbf{x}_0)=(-1,0), \quad \nabla g_1(\mathbf{x}_0)=(0,1) \quad \text{y} \quad \nabla g_2(\mathbf{x}_0)=(0,-1).$$

Entonces

$$u_0(-1,0)+u_1(0,1)+u_2(0,-1)=(0,0)$$

se cumple si $u_0 = 0$ y $u_1 = u_2 = \alpha$, donde α es un escalar positivo. Por lo tanto, las condiciones de Fritz John se satisfacen en $\mathbf{x}_0$.

Ejemplo 2.19. Consideremos el siguiente problema.

$$\begin{array}{cr} \text{Minimizar} & -x_1 \\ \text{sujeto a} & x_1 + x_2 - 1 \leq 0 \\ & -x_2 \leq 0 \end{array}$$

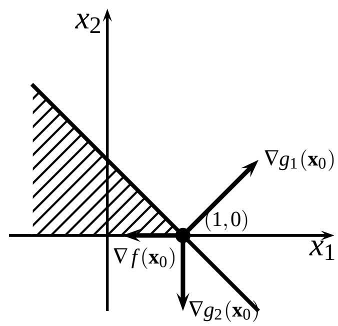

Figura 2.7: Ilustración del Ejemplo 2.19.

La región factible de este problema se ilustra en la Figura 2.7 y el punto óptimo es $\mathbf{x}_0 = (1,0)$. Observe que el conjunto de restricciones que se cumplen con igualdad es $I = \{1,2\}$ y que

$$\nabla f(\mathbf{x}_0) = (-1,0), \quad \nabla g_1(\mathbf{x}_0) = (1,1) \quad \text{y} \quad \nabla g_2(\mathbf{x}_0) = (0,-1).$$

Las condiciones de Fritz John se satisfacen haciendo $u_0 = u_1 = u_2 = \alpha$, donde α es cualquier escalar positivo.

2.3.3

Condiciones de Kuhn-Tucker

Se observa que si $u_0 = 0$, entonces las condiciones de Fritz John no utilizan información alguna perteneciente al gradiente de la función objetivo; únicamente establecen que existe una combinación lineal no trivial de los gradientes de las funciones g_i que se cumplen con igualdad, que suma cero. Por lo tanto, cuando $u_0 = 0$, las condiciones de Fritz John no son de utilidad para localizar un punto óptimo; de esta forma, son de mayor interés los casos en que $u_0 > 0$.

Kuhn y Tucker desarrollaron unas condiciones necesarias que son precisamente las condiciones de Fritz John con la propiedad adicional de que $u_0 > 0$. Las condiciones que se imponen a las restricciones para garantizar que $u_0 > 0$ son llamadas *cualificación de restricciones*. El siguiente teorema establece las condiciones de Kuhn-Tucker bajo la cualificación de restricciones que establece que los gradientes de las funciones g_i que se cumplen con igualdad son linealmente independientes.

Teorema 2.20. ***Condiciones necesarias de Kuhn-Tucker.*** *Sean X un conjunto abierto no vacío en $\mathbb{R}^n$, $f:\mathbb{R}^n \to \mathbb{R}$ y $g_i:\mathbb{R}^n \to \mathbb{R}$, para $i = 1,\ldots,m$. Considere el Problema 2.13, sea $\mathbf{x}_0$ un punto factible y sea $I = \{i \mid g_i(\mathbf{x}_0) = 0\}$. Supongamos que f y g_i con $i \in I$ son diferenciables en $\mathbf{x}_0$, y que g_i con $i \notin I$ son continuas en $\mathbf{x}_0$. Más aún, supongamos que $\nabla g_i(\mathbf{x}_0)$ para $i \in I$ son linealmente independientes. Si $\mathbf{x}_0$ es un óptimo local entonces existen escalares u_i para $i \in I$ tales que*

$$\begin{aligned} \nabla f(\mathbf{x}_0) + \sum_{i\in I} u_i \nabla g_i(\mathbf{x}_0) &= \mathbf{0} \\ u_i &\geq 0 \qquad \textit{para } i \in I. \end{aligned} \tag{2.8}$$

Además de las suposiciones hechas anteriormente, si las funciones g_i con $i \notin I$ también son diferenciables en $\mathbf{x}_0$, entonces las condiciones de Kuhn-Tucker pueden ser escritas de la siguiente forma

$$\begin{aligned} \nabla f(\mathbf{x}_0) + \sum_{i=1}^{m} u_i \nabla g_i(\mathbf{x}_0) &= \mathbf{0} \\ u_i g_i(\mathbf{x}_0) &= 0 \qquad \textit{para } i = 1,\ldots,m \\ u_i &\geq 0 \qquad \textit{para } i = 1,\ldots,m. \end{aligned} \tag{2.9}$$

Demostración.
Por el Teorema 2.16 existen escalares u_0 y $\hat{u}_i$ para $i \in I$, no todos iguales a cero, tales que

$$\begin{aligned} u_0 \nabla f(\mathbf{x}_0) + \sum_{i\in I} \hat{u}_i \nabla g_i(\mathbf{x}_0) &= \mathbf{0} \\ u_0, \hat{u}_i &\geq 0 \qquad \text{para } i \in I. \end{aligned} \tag{2.10}$$

Note que $u_0 > 0$, pues si $u_0 = 0$, el sistema (2.10) contradice el hecho de que $\nabla g_i(\mathbf{x}_0)$ para $i \in I$ son linealmente independientes. Haciendo $u_i = \dfrac{\hat{u}_i}{u_0}$ obtenemos la prueba para (2.8). La prueba de (2.9) se obtiene haciendo $u_i = 0$ para $i \notin I$. ■

Al igual que en las condiciones de Fritz John, los escalares u_i, para $i = 1,\ldots,m$, son llamados multiplicadores de Lagrange y la condición $u_i g_i(\mathbf{x}_0) = 0$ para $i = 1,\ldots,m$ es llamada condición de holgura complementaria.

Interpretación geométrica de las condiciones de Kuhn-Tucker

Note que cualquier vector de la forma $\sum_{i\in I} u_i \nabla g_i(\mathbf{x}_0)$, donde $u_i \geq 0$ para $i \in I$, pertenece al cono generado por los gradientes de las funciones g_i que se cumplen con igualdad. Las condiciones de Kuhn-Tucker, $-\nabla f(\mathbf{x}_0) = \sum_{i \in I} u_i \nabla g_i(\mathbf{x}_0)$ y $u_i \geq 0$ para $i \in I$, pueden ser interpretadas como que $-\nabla f(\mathbf{x}_0)$ pertenece al cono antes mencionado.

El Teorema 2.21 establece que, bajo condiciones razonables de convexidad, las condiciones de Kuhn-Tucker son también suficientes.

Teorema 2.21. ***Condiciones suficientes de Kuhn-Tucker.*** *Sean X un conjunto abierto no vacío en $\mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$ y $g_i : \mathbb{R}^n \to \mathbb{R}$ para $i = 1, \ldots, m$. Considere el Problema 2.13, sea $\mathbf{x}_0$ un punto factible y sea $I = \{i \mid g_i(\mathbf{x}_0) = 0\}$. Supongamos que f es una función pseudoconvexa en $\mathbf{x}_0$ y que g_i es cuasiconvexa y diferenciable en $\mathbf{x}_0$ para cada $i \in I$. Más aún, supongamos que las condiciones necesarias de Kuhn-Tucker se satisfacen en $\mathbf{x}_0$; es decir, existen escalares no negativos u_i para $i \in I$ tales que*

$$\nabla f(\mathbf{x}_0) + \sum_{i \in I} u_i \nabla g_i(\mathbf{x}_0) = \mathbf{0}.$$

Entonces $\mathbf{x}_0$ es una solución óptima global del Problema 2.13.

Demostración.

Sea $\mathbf{x}$ un punto factible del Problema 2.13, entonces, para $i \in I$, se cumple que $g_i(\mathbf{x}) \leq g_i(\mathbf{x}_0)$, ya que $g_i(\mathbf{x}) \leq 0$ y $g_i(\mathbf{x}_0) = 0$. Por la cuasiconvexidad de g_i en $\mathbf{x}_0$, se sigue que

$$g_i[\mathbf{x}_0 + \lambda(\mathbf{x} - \mathbf{x}_0)] = g_i[\lambda \mathbf{x} + (1-\lambda)\mathbf{x}_0] \leq \text{máx}\{g_i(\mathbf{x}), g_i(\mathbf{x}_0)\} = g_i(\mathbf{x}_0)$$

para todo $\lambda \in (0,1)$. Esto implica que g_i no incrementa su valor al moverse de $\mathbf{x}_0$ a lo largo de la dirección $\mathbf{x} - \mathbf{x}_0$, entonces, por el Teorema 2.8, debemos tener $\nabla g_i(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0) \leq 0$. Multiplicando por u_i y sumando sobre I obtenemos

$$\left[\sum_{i \in I} u_i \nabla g_i(\mathbf{x}_0)\right] \cdot (\mathbf{x} - \mathbf{x}_0) \leq 0.$$

Pero como

$$\nabla f(\mathbf{x}_0) + \sum_{i \in I} u_i \nabla g_i(\mathbf{x}_0) = \mathbf{0},$$

se sigue que $\nabla f(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0) \geq 0$. Entonces, por la pseudoconvexidad de f en $\mathbf{x}_0$, debemos tener $f(\mathbf{x}_0) \leq f(\mathbf{x})$. ∎

Note que si f y g_i son convexas en $\mathbf{x}_0$, entonces se tendrá también que serán pseudoconvexa y cuasiconvexa en $\mathbf{x}_0$, respectivamente, y por lo tanto las condiciones de Kuhn-Tucker son suficientes. Por otra parte, si la convexidad en un punto es reemplazada por convexidad global, las condiciones de Kuhn-Tucker son también suficientes.

2.3.4

Cualificación de restricciones

Hasta este momento hemos probado que optimalidad local implica que $F_0 \cap G_0 = \emptyset$, con lo que se deducen las condiciones de Fritz John; y bajo la suposición de cualificación de restricciones de independencia lineal se obtienen las condiciones de Kuhn-Tucker. En esta sección obtendremos las condiciones de Kuhn-Tucker directamente, sin tener que probar primero las condiciones de Fritz John. Como se demostrará en el Teorema 2.23, una condición necesaria para un óptimo local es que $F_0 \cap T = \emptyset$, donde T es el cono de tangentes definido enseguida. Usando la cualificación de restricciones $T = G'$, donde G' se definirá en el Teorema 2.24, $F_0 \cap G' = \emptyset$. Finalmente, usando el Teorema 2.6, obtendremos las condiciones de Kuhn-Tucker.

Definición 2.22. Sean S un conjunto no vacío en $\mathbb{R}^n$ y $\mathbf{x}_0 \in \overline{S}$. El *cono de tangentes* de S en $\mathbf{x}_0$, denotado por T, es el conjunto de todas las direcciones $\mathbf{d}$ tales que

$$\mathbf{d} = \lim_{k\to\infty} \lambda_k(\mathbf{x}_k - \mathbf{x}_0),$$

donde $\lambda_k > 0$, $\mathbf{x}_k \in S$ para todo k y $\mathbf{x}_k \to \mathbf{x}_0$.

De la definición anterior, es claro que $\mathbf{d}$ está en el cono de tangentes si existe una sucesión factible $\{\mathbf{x}_k\}$ que converja a $\mathbf{x}_0$ tal que las direcciones de las coordenadas de $\mathbf{x}_k - \mathbf{x}_0$ converjan a $\mathbf{d}$. La Figura 2.8 ilustra dos ejemplos de conos de tangentes.

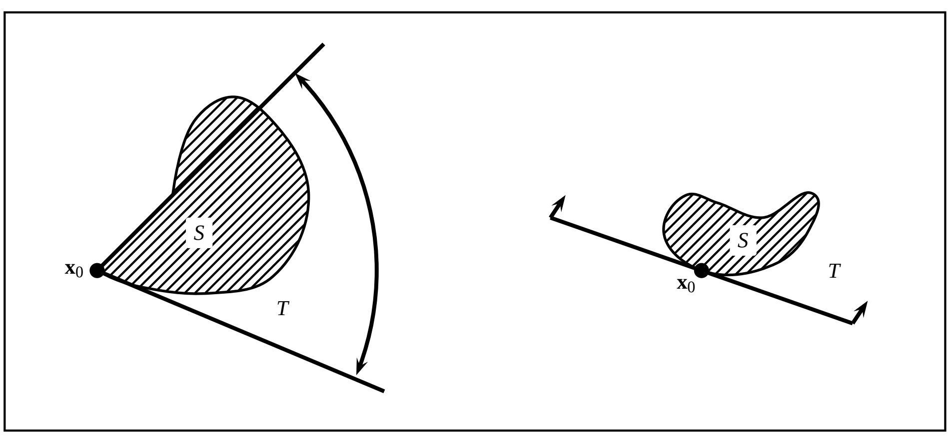

Figura 2.8: Ejemplos de conos de tangentes.

Teorema 2.23. *Sean S un conjunto no vacío en $\mathbb{R}^n$ y $\mathbf{x}_0 \in S$. Supongamos que $f : \mathbb{R}^n \to \mathbb{R}$ es diferenciable en $\mathbf{x}_0$. Si $\mathbf{x}_0$ es un óptimo local del problema de minimizar $f(\mathbf{x})$ sujeto a $\mathbf{x} \in S$,* *entonces* $F_0 \cap T = \emptyset$, *donde*

$$F_0 = \{\mathbf{d} \mid \nabla f(\mathbf{x}_0) \cdot \mathbf{d} < 0\}$$

y T *es el cono de tangentes de* S *en* $\mathbf{x}_0$.

Demostración.

Sea $\mathbf{d} \in T$, es decir, $\mathbf{d} = \lim_{k\to\infty} \lambda_k(\mathbf{x}_k - \mathbf{x}_0)$, donde $\lambda_k > 0$, $\mathbf{x}_k \in S$ para todo k y $\mathbf{x}_k \to \mathbf{x}_0$. Como f es diferenciable en $\mathbf{x}_0$, tenemos que

$$f(\mathbf{x}_k) - f(\mathbf{x}_0) = \nabla f(\mathbf{x}_0) \cdot (\mathbf{x}_k - \mathbf{x}_0) + \|\mathbf{x}_k - \mathbf{x}_0\| E(\mathbf{x}_0, \mathbf{x}_k - \mathbf{x}_0) \tag{2.11}$$

donde $E(\mathbf{x}_0, \mathbf{x}_k - \mathbf{x}_0) \to 0$ cuando $\mathbf{x}_k \to \mathbf{x}_0$. Como $\mathbf{x}_0$ es un óptimo local, para k suficientemente grande, tenemos $f(\mathbf{x}_0) \leq f(\mathbf{x}_k)$; y de (2.11) tenemos que

$$\nabla f(\mathbf{x}_0) \cdot (\mathbf{x}_k - \mathbf{x}_0) + \|\mathbf{x}_k - \mathbf{x}_0\| E(\mathbf{x}_0, \mathbf{x}_k - \mathbf{x}_0) \geq 0.$$

Multiplicando por $\lambda_k > 0$ y tomando el límite cuando $k \to \infty$, la desigualdad anterior implica que $\nabla f(\mathbf{x}_0) \cdot \mathbf{d} \geq 0$. Hemos probado que $\mathbf{d} \in T$ implica que $\nabla f(\mathbf{x}_0) \cdot \mathbf{d} \geq 0$ y de aquí se sigue que $F_0 \cap T = \emptyset$. ∎

Teorema 2.24. ***Teorema de Kuhn-Tucker (Condiciones necesarias).*** *Sean* X *un conjunto no vacío en* $\mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$ *y* $g_i : \mathbb{R}^n \to \mathbb{R}$ *para* $i = 1, \ldots, m$. *Considere el Problema 2.13, sea* $\mathbf{x}_0$ *un punto factible y sea* $I = \{i \mid g_i(\mathbf{x}_0) = 0\}$. *Supongamos que* f *y* g_i *con* $i \in I$ *son diferenciables en* $\mathbf{x}_0$. *Más aún, supongamos que la cualificación de restricciones* $T = G'$ *se satisface, donde* T *es el cono de tangentes de la región factible en* $\mathbf{x}_0$ *y*

$$G' = \{\mathbf{d} \mid \nabla g_i(\mathbf{x}_0) \cdot \mathbf{d} \leq 0 \;\; para \;\; i \in I\}.$$

Si $\mathbf{x}_0$ *es una solución óptima local, entonces existen escalares no negativos* u_i *para* $i \in I$ *tales que*

$$\nabla f(\mathbf{x}_0) + \sum_{i\in I} u_i \nabla g_i(\mathbf{x}_0) = \mathbf{0}.$$

Demostración.

Por el Teorema 2.23, $F_0 \cap T = \emptyset$. Por suposición $T = G'$, así que $F_0 \cap G' = \emptyset$; es decir, el siguiente sistema no tiene solución.

$$\nabla f(\mathbf{x}_0) \cdot \mathbf{d} < 0 \qquad \nabla g_i(\mathbf{x}_0) \cdot \mathbf{d} \leq 0 \quad \text{para } i \in I.$$

A partir del Teorema 2.6 se sigue el resultado. ∎

2.3.5

Problemas con restricciones lineales

El siguiente lema establece que si las restricciones son lineales, entonces se satisface el Teorema 2.24. Esto también implica que las condiciones de Kuhn-Tucker son siempre necesarias para problemas con restricciones lineales, independientemente de que la función objetivo sea lineal o no.

Lema 2.25. *Sean* $\mathbf{A}$ *una matriz* $m \times n$*,* $\mathbf{b}$ *y* $\mathbf{x}$ *vectores columna* $m \times 1$ *y*

$$S = \{\mathbf{x} \mid \mathbf{Ax} \leq \mathbf{b}\}.$$

Supongamos que $\mathbf{x}_0 \in S$ *es tal que* $\mathbf{A}_1\mathbf{x} = \mathbf{b}_1$ *y* $\mathbf{A}_2\mathbf{x} < \mathbf{b}_2$*, donde* $\mathbf{A}^t = (\mathbf{A}_1^t, \mathbf{A}_2^t)$ *y* $\mathbf{b}^t = (\mathbf{b}_1^t, \mathbf{b}_2^t)$*; entonces* $T = G'$*, donde* T *es el cono de tangentes de* S *en* $\mathbf{x}_0$ *y* $G' = \{\mathbf{d} \mid \mathbf{A}_1\mathbf{d} \leq \mathbf{0}\}$.

Demostración.
Si $\mathbf{A}_1$ es la matriz cero, $G' = \mathbb{R}^n$. Más aún, $\mathbf{x}_0 \in int(S)$ y por lo tanto $T = \mathbb{R}^n$. De esta forma $T = G'$.

Ahora supongamos que la matriz $\mathbf{A}_1$ tiene al menos una entrada diferente de cero. Sea $\mathbf{d} \in T$, es decir, $\mathbf{d} = \lim\limits_{k\to\infty} \lambda_k(\mathbf{x}_k - \mathbf{x}_0)$, donde $\mathbf{x}_k \in S$ y $\lambda_k > 0$, para todo k. Entonces

$$\mathbf{A}_1(\mathbf{x}_k - \mathbf{x}_0) \leq \mathbf{b}_1 - \mathbf{b}_1 = \mathbf{0}. \tag{2.12}$$

Multiplicando (2.12) por $\lambda_k > 0$ y tomando límite cuando $k \to \infty$ se tiene que $\mathbf{A}_1\mathbf{d} \leq \mathbf{0}$, entonces llegamos a que $\mathbf{d} \in G'$ y $T \subset G'$. Ahora queremos probar que $G' \subset T$. Sea $\mathbf{d} \in G'$, es decir, $\mathbf{A}_1\mathbf{d} \leq \mathbf{0}$. Como $\mathbf{A}_2\mathbf{x}_0 < \mathbf{b}_2$, existe $\delta > 0$ tal que $\mathbf{A}_2(\mathbf{x}_0 + \lambda\mathbf{d}) < \mathbf{b}_2$ para todo $\lambda \in (0, \delta)$. Más aún, como $\mathbf{A}_1\mathbf{x}_0 = \mathbf{b}_1$ y $\mathbf{A}_1\mathbf{d} \leq \mathbf{0}$, entonces $\mathbf{A}_1(\mathbf{x}_0 + \lambda\mathbf{d}) \leq \mathbf{b}_1$ para todo $\lambda > 0$. Entonces $\mathbf{x}_0 + \lambda\mathbf{d} \in S$ para todo $\lambda \in (0, \delta)$; lo cual implica que $\mathbf{d} \in T$, es decir, $G' \subset T$. Por lo tanto $T = G'$. ■

Notemos que el Problema 2.13 también puede ser escrito de la siguiente forma:

Problema 2.26.

$$\begin{array}{lll} \text{Minimizar} & f(\mathbf{x}) & \\ \text{sujeto a} & g_i(\mathbf{x}) \leq c_i & \text{para } i = 1, \ldots, m \\ & \mathbf{x} \in X & \end{array}$$

donde c_i es constante para toda i. Denotemos los parámetros c_i's del Problema 2.26 por medio del vector $\overline{\mathbf{c}} = (\overline{c}_1, \ldots, \overline{c}_m)$, al cual llamaremos *vector parámetro*. Pero observemos que el valor de c_i puede presentar variaciones para cualquier i. Supongamos que para $\mathbf{c} \in \mathbb{R}^m$, el Problema 2.26 tiene una solución $\overline{\mathbf{x}}(\mathbf{c})$ y denotemos el valor de esta solución por $v(\mathbf{c}) = f(\overline{\mathbf{x}}(\mathbf{c}))$. Probaremos más adelante que, bajo ciertas condiciones, $v(\mathbf{c})$ depende continuamente del vector parámetro $\mathbf{c}$.

Consideremos el siguiente problema de optimización más general. Sea $f : \mathbb{R}^n \to \mathbb{R}$ una función dada,

Problema 2.27.

$$\begin{array}{cc} \text{Minimizar} & f(\mathbf{x}) \\ \text{sujeto a} & \mathbf{x} \in C(\mathbf{q}) \end{array}$$

donde $C(\mathbf{q})$ es un *conjunto restricción* no vacío y $\mathbf{q} = (q_1, \ldots, q_s)$ pertenece a un conjunto de parámetros $Q \subset \mathbb{R}^s$. Supongamos que f es continua y que $C(\mathbf{q})$ es compacto para toda $\mathbf{q} \in Q$, entonces, por el Corolario 1.23, que el Problema 2.27 tiene al menos una solución. Denotemos por $x(\mathbf{q}) \subset C(\mathbf{q})$ al conjunto de soluciones correspondiente a $\mathbf{q}$ y por $v(\mathbf{q})$ al valor máximo asociado; es decir, $v(\mathbf{q}) = f(\mathbf{x})$ para algún $\mathbf{x} \in x(\mathbf{q})$. El Teorema 2.28 establece la continuidad de $x(\cdot)$ y $v(\cdot)$ (Mas-Colell,1995).

Teorema 2.28. *Supongamos que la correspondencia de restricción $C : Q \to \mathbb{R}^n$ es continua y que f es una función continua, entonces la correspondencia maximizadora $x : Q \to \mathbb{R}^n$ es superiormente hemicontinua y la función valor $v : Q \to \mathbb{R}$ es continua.*

Capítulo 3

Elementos de microeconomía

En este capítulo se desarrolla una parte de la teoría de la elección del consumidor, con el fin de que el lector se familiarice con algunos conceptos económicos que son de importancia para el desarrollo del presente trabajo.

Notación. En este capítulo la notación $\mathbf{y} << \mathbf{x}$, con $\mathbf{y}, \mathbf{x} \in \mathbb{R}^n$, significa que $y_i < x_i$ para $i = 1, 2, \ldots, n$.

3.1 Las preferencias del consumidor

Cuando entramos en una tienda nos encontramos con una gran cantidad de bienes que podríamos comprar, pero como nuestros recursos financieros son limitados no podemos comprar todo lo que queremos. Por esta razón, observamos los precios de los diversos bienes y compramos los que, dados nuestros recursos, mejor se ajustan a nuestras necesidades y deseos. A la mayoría de las personas les gustaría aumentar la cantidad o calidad de los bienes que consumen, sin embargo, consumen menos de lo que desean porque su ingreso restringe o limita su gasto.

El problema al que se enfrenta el consumidor es elegir sus niveles de consumo de varios bienes y servicios que están disponibles en un determinado mercado. A estos bienes y servicios les llamaremos *artículos de consumo*. Por simplicidad, asumiremos que el número de artículos de consumo es finito e igual a n.

Un vector de artículos de consumo, $\mathbf{x} = (x_1, \ldots, x_n)$, es una lista de las cantidades de diferentes artículos de consumo y puede ser visto como un punto en $\mathbb{R}^n$, el *espacio de artículos de consumo*. La i-ésima entrada de éste vector representa la cantidad consumida del artículo de consumo i y el vector $\mathbf{x}$ representa un nivel de consumo de un individuo. Nos referiremos a dicho vector como una *cesta de consumo*. Por ejemplo, si $n = 3$ y mis artículos de consumo son botellas de agua, blusas y libretas; entonces $\mathbb{R}^3$ es el espacio de consumo y las cestas $\mathbf{x_1} = (1,2,3)$ y $\mathbf{x_2} = (3,1,2)$ representan los niveles de consumo de una botella de agua, dos blusas y tres libretas; y tres botellas de agua, una blusa y dos libretas, respectivamente.

Por otra parte, las elecciones de consumo están limitadas por un número de restricciones físicas; por ejemplo, es imposible consumir una cantidad negativa de un bien, tal como pan o agua. Formalmente, el conjunto de consumo, X, es un subespacio del espacio de consumo $\mathbb{R}^n$, es decir, $\mathrm{X} \subset \mathbb{R}^n$. El conjunto X consta de todas las posibles cestas de bienes que el individuo podría consumir dadas las restricciones físicas impuestas por su entorno.

Consideremos los siguientes ejemplos para el caso $n = 2$.

1. La Figura 3.1 representa los niveles de consumo posibles de pan y ocio en un día, ambos niveles deben ser no negativos y consumir más de 24 horas de ocio al día es imposible.

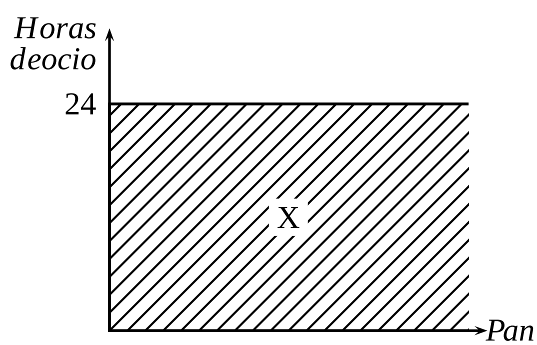

Figura 3.1: Ejemplo de un espacio de consumo X.

2. Supongamos que un individuo requiere un mínimo de 4 rebanadas de pan al día para sobrevivir y que hay dos tipos de pan, blanco e integral. La Figura 3.2 representa esta situación en la que el individuo puede combinar los dos tipos de pan de modo que cubra con sus necesidades básicas.

Para el desarrollo de este trabajo se tomará la clase más simple de un conjunto de consumo, el conjunto de todas las cestas de consumo cuyas entradas son no negativas, es decir,

$$\mathrm{X} = \mathbb{R}^n_+ = \{\mathbf{x} \in \mathbb{R}^n \mid x_i \geq 0 \ \text{ para } \ i = 1, \ldots, n\}.$$

Supondremos que dadas dos cestas de consumo cualesquiera en X, el consumidor puede ordenarlas según su atractivo, es decir, vamos a suponer que el individuo tiene preferencias sobre las cestas de consumo en X. Si el consumidor prefiere una cesta a otra, significa que elegirá la que prefiere, si tiene posibilidad de hacerlo. Por lo tanto, las decisiones del consumidor dependen de sus preferencias por los bienes.

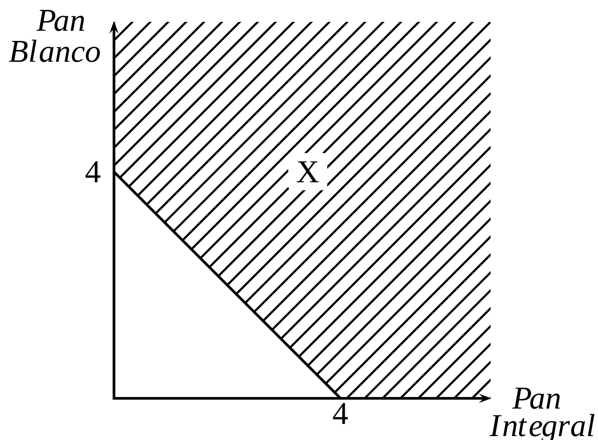

Figura 3.2: Ejemplo de un espacio de consumo X.

Definición 3.1. Definimos una *relación de preferencias* sobre el conjunto de consumo X, denotada por el símbolo $\precsim$, de la siguiente manera: $\mathbf{y} \precsim \mathbf{x}$ quiere decir que el consumidor considera que la cesta de consumo $\mathbf{x}$ es al menos tan buena como la cesta $\mathbf{y}$.

De esta relación se derivan las siguientes:

1. Cuando el consumidor prefiere estrictamente la cesta $\mathbf{x}$ a la cesta $\mathbf{y}$, $\mathbf{y} \prec \mathbf{x}$, esto ocurre si y sólo si $\mathbf{y} \precsim \mathbf{x}$ pero no $\mathbf{x} \precsim \mathbf{y}$.
2. Cuando el consumidor es indiferente entre las dos cestas de consumo, $\mathbf{x} \sim \mathbf{y}$, esto sucede si y sólo si $\mathbf{x} \precsim \mathbf{y}$ y $\mathbf{y} \precsim \mathbf{x}$.

3.1.1

Supuestos sobre las preferencias

Cuando $\precsim$ cumple la propiedad de orden total queremos decir que un individuo puede comparar dos cestas cualesquiera de su conjunto de consumo X y decidir cuál prefiere. Para que el individuo pueda elegir alguna cesta, antes debe reflexionar sobre su decisión; así que este supuesto también nos dice que el individuo sólo toma decisiones meditadas.

Por otra parte, el supuesto de que $\precsim$ es transitiva hace que las preferencias del consumidor sean coherentes y que no se ciclen; por ejemplo, si un individuo considera que una rebanada de melón, m, es al menos tan buena como una naranja, n, y ésta es al menos tan buena como una fresa, f, ($f \precsim n \precsim m$); no puede decir que una fresa es al menos tan buena como una rebanada de melón ($m \not\precsim f \precsim n \precsim m$).

De esta forma, llegamos a la siguiente definición.

Definición 3.2. La *relación de preferencias* $\precsim$ es *racional* si cumple las siguientes propiedades:

i) **Orden total.** Se cumple $\mathbf{y} \precsim \mathbf{x}$ o $\mathbf{x} \precsim \mathbf{y}$ o ambas, $\forall \mathbf{x}, \mathbf{y} \in \mathrm{X}$.

ii) **Transitividad.** Si $\mathbf{y} \precsim \mathbf{x}$ y $\mathbf{z} \precsim \mathbf{y}$, entonces $\mathbf{z} \precsim \mathbf{x}$, $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathrm{X}$.

La siguiente definición establece que si tenemos una cesta que contiene al menos la misma cantidad de todos los bienes que otra, es al menos tan buena como ésta.

Definición 3.3. La *relación de preferencias* $\precsim$ se dice ***monótona*** si $y_i \leq x_i\ \forall i$, entonces $\mathbf{y} \precsim \mathbf{x}$.

También podemos plasmar la siguiente idea mediante una definición, una cesta que contenga al menos la misma cantidad de todos los bienes que otra, y más de alguno de ellos, es estrictamente preferida a ésta.

Definición 3.4. La *relación de preferencias* $\precsim$ se dice ***monótona fuerte*** si $y_i \leq x_i\ \forall i$ y $\mathbf{x} \neq \mathbf{y}$, implica $\mathbf{y} \prec \mathbf{x}$.

El principio de insaciabilidad local es muy utilizado en economía, éste se basa en la idea de que al estar hablando de bienes, suponemos que cuanto más, mejor. Este principio plantea que el consumidor no se sacia con ninguna de las cestas de bienes viables, en el sentido de que una cesta que tiene más de cualquier bien es preferible a una cesta que tenga menos de ese bien.

Principio de insaciabilidad local. Dado cualquier $\mathbf{x} \in \mathrm{X}$ y cualquier $\varepsilon > 0$, existe $\mathbf{y} \in \mathrm{X}$ con $|\mathbf{x} - \mathbf{y}| < \varepsilon$ tal que $\mathbf{x} \prec \mathbf{y}$.

3.1.2

La utilidad

Los individuos son capaces de ordenar todas las situaciones posibles de la menos a la más deseable y las situaciones más deseables reportan mayor utilidad que las menos deseables. La *utilidad* es una medida subjetiva de la satisfacción o la felicidad que reporta a un consumidor una cesta de consumo.

Para el análisis económico es conveniente representar el comportamiento del consumidor mediante una función de utilidad. Una función de utilidad es un instrumento que asigna un valor numérico a cada elemento en X de tal forma que las cestas que son más preferidas tengan un número más alto que las menos preferidas. La función de utilidad tiene un carácter ordinal, únicamente refleja el orden de las preferencias del individuo.

Definición 3.5. Una función $u : \mathrm{X} \to \mathbb{R}$ es una *función de utilidad* que representa la relación de preferencias $\precsim$ si se cumple lo siguiente

$$\mathbf{y} \precsim \mathbf{x} \Leftrightarrow u(\mathbf{y}) \leq u(\mathbf{x}) \quad \forall \mathbf{x}, \mathbf{y} \in \mathrm{X}.$$

Dado que sólo importa la ordenación de las cestas de bienes, si $u(\mathbf{x})$ representa una relación de preferencias $\precsim$ y $f : \mathbb{R} \to \mathbb{R}$ es una función monótona creciente, entonces $f(u(\mathbf{x}))$ representará exactamente las mismas preferencias, ya que $f(u(\mathbf{y})) \leq f(u(\mathbf{x}))$ si y sólo si $u(\mathbf{y}) \leq u(\mathbf{x})$.

Proposición 3.6. *Una relación de preferencias $\precsim$ puede ser representada mediante una función de utilidad sólo si $\precsim$ es racional.*

Demostración.
Demostraremos que si existe una función de utilidad $u(\mathbf{x})$ que represente las preferencias $\precsim$, entonces $\precsim$ debe cumplir las propiedades de orden total y transitividad.

Primero probaremos que la relación de preferencias debe cumplir la propiedad de orden total. Sabemos que la función u va de X a $\mathbb{R}$, entonces para cualquiera $\mathbf{x}, \mathbf{y} \in \mathrm{X}$ se deberá tener $u(\mathbf{y}) \leq u(\mathbf{x})$ o $u(\mathbf{x}) \leq u(\mathbf{y})$ o ambas. Ahora, como u es una función de utilidad que representa las preferencias $\precsim$, de acuerdo con la Definición 3.5, entonces se tiene $\mathbf{y} \precsim \mathbf{x}$ o $\mathbf{x} \precsim \mathbf{y}$ o ambas. Por lo tanto se cumple la propiedad de orden total.

Ahora probaremos que se cumple la propiedad de transitividad. Supongamos que $\mathbf{y} \precsim \mathbf{x}$ y $\mathbf{z} \precsim \mathbf{y}$, como u representa las preferencias debemos tener $u(\mathbf{y}) \leq u(\mathbf{x})$ y $u(\mathbf{z}) \leq u(\mathbf{y})$, lo cual implica $u(\mathbf{z}) \leq u(\mathbf{x})$; y por lo tanto tenemos $\mathbf{z} \precsim \mathbf{x}$. Demostrando así que se cumple la propiedad de transitividad. ■

Pero una relación de preferencias racional no siempre puede ser representada mediante una función de utilidad, para ilustrar esto veamos el siguiente ejemplo.

Ejemplo 3.7. Preferencias lexicográficas. Por simplicidad, supongamos que $\mathrm{X} = \mathbb{R}^2_+$. Definimos las preferencias lexicográficas $\precsim^{\mathbb{L}}$ de la siguiente manera:

$$(x,y) \precsim^{\mathbb{L}} (x',y') \Leftrightarrow x < x' \text{ o si } x = x' \text{ y } y \leq y',$$

análogamente

$$(x,y) \prec^{\mathbb{L}} (x',y') \Leftrightarrow x < x' \text{ o si } x = x' \text{ y } y < y'.$$

Es decir, una cesta es al menos tan buena como otra si tiene más o igual cantidad del bien uno y en caso de tener igual, tiene más o igual cantidad del bien dos. De esta forma, para el consumidor es más importante el bien uno que el bien dos.

En la Figura 3.3 podemos observar que las preferencias lexicográficas nos dan la relación $D \prec^{\mathbb{L}} C \prec^{\mathbb{L}} B \prec^{\mathbb{L}} A$. Las preferencias lexicográficas son racionales pero a pesar de ello no se pueden representar mediante una función de utilidad.

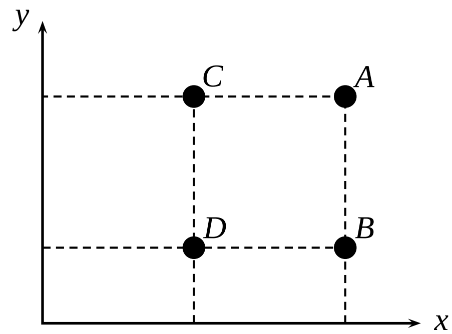

Figura 3.3: Preferencias lexicográficas

Supongamos, por contradicción, que existe una función $u : \mathrm{X} \to \mathbb{R}$ que representa $\precsim^{\mathbb{L}}$. Tomemos dos cantidades positivas del segundo bien, y_1 y y_2, tales que $y_2 < y_1$. Sea $x_1 \in \mathbb{R}_+$ arbitrario, entonces tenemos $(x_1, y_2) \prec^{\mathbb{L}} (x_1, y_1)$ y como u representa dichas preferencias, se tiene que $u(x_1, y_2) < u(x_1, y_1)$. Ahora, como $\mathbb{Q}$ es denso en $\mathbb{R}$, existe $r(x_1) \in \mathbb{Q}$ tal que

$$u(x_1, y_2) < r(x_1) < u(x_1, y_1).$$

Por otra parte, debido al carácter lexicográfico de las preferencias, el tener $x_2 < x_1$ implica $r(x_2) < r(x_1)$, ya que

$$u(x_2, y_2) < r(x_2) < u(x_2, y_1) < u(x_1, y_2) < r(x_1) < u(x_1, y_1).$$

Entonces, para cada $x \in \mathbb{R}_+$ existe un $r(x) \in \mathbb{Q}$ distinto, es decir, tenemos una función inyectiva $r : \mathbb{R}_+ \to \mathbb{Q}$, lo cual es imposible porque $\mathbb{Q}$ es numerable y $\mathbb{R}_+$ no lo es. Por lo tanto, no existe una función de utilidad que represente las preferencias lexicográficas.

Necesitamos que la relación de preferencias cumpla una propiedad más para asegurar la existencia de una función de utilidad, la propiedad de continuidad. Dicha propiedad nos dice que las preferencias del consumidor no pueden exhibir “brincos”, es decir, que la dirección de preferencia no se puede revertir al pasar al límite de una sucesión de comparaciones, además nos dice que el consumidor es capaz de distinguir pequeñas diferencias entre las cestas y tomar decisiones. Para comparar cestas que tienen pequeñas diferencias necesitamos decidir qué tan cerca está una de la otra, así que tenemos que incorporar la idea de distancia.

Definición 3.8. La *relación de preferencias* $\precsim$ es *continua* si se preserva bajo límites, es decir, para cualquier par de sucesiones $\{(\mathbf{x}_n, \mathbf{y}_n)\}_{n\in\mathbb{N}}$ tales que $\mathbf{y}_n \precsim \mathbf{x}_n$ $\forall n$, $\lim_{n\to\infty} \mathbf{x}_n = \mathbf{x}$ y $\lim_{n\to\infty} \mathbf{y}_n = \mathbf{y}$, se cumple $\mathbf{y} \precsim \mathbf{x}$.

Una forma equivalente de establecer la idea de continuidad es decir que los conjuntos

$$\{\mathbf{x} \in \mathrm{X} : \mathbf{y} \precsim \mathbf{x}\} \quad y \quad \{\mathbf{x} \in \mathrm{X} : \mathbf{x} \precsim \mathbf{y}\}$$

son cerrados, $\forall \mathbf{y} \in \mathrm{X}$.

El que la relación de preferencias $\precsim$ cumpla la propiedad de continuidad es suficiente para que exista una función de utilidad que la represente, de hecho nos garantiza la existencia de una función de utilidad continua. Esto se establece en el siguiente teorema.

Teorema 3.9. *Supongamos que la relación de preferencias $\precsim$ es racional y continua en X, entonces existe una función de utilidad $u(\mathbf{x})$ que representa la relación de preferencias $\precsim$.*

Demostración.
Probaremos la proposición para el caso de $X = \mathbb{R}^n_+$ y una relación de preferencias monótona. Sean

$$Z = \{\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_1 = \ldots = x_n\}$$

y $\mathbf{e}$ el vector unidad en $\mathbb{R}^n_+$,

$$\mathbf{e} = (1, \ldots, 1).$$

Observemos que $\alpha\mathbf{e} \in Z$ para todo $\alpha \geq 0$. Además, para todo $\mathbf{x} \in \mathbb{R}^n_+$, la monotonicidad de la relación de preferencias implica que $\mathbf{0} \precsim \mathbf{x}$; notemos también que para cualquier $\tilde{\alpha}$ tal que $\mathbf{x} << \tilde{\alpha}\mathbf{e}$, entonces $\mathbf{x} \precsim \tilde{\alpha}\mathbf{e}$. La construcción de una función de utilidad para el caso $n = 2$ está ilustrada en la Figura 3.4.

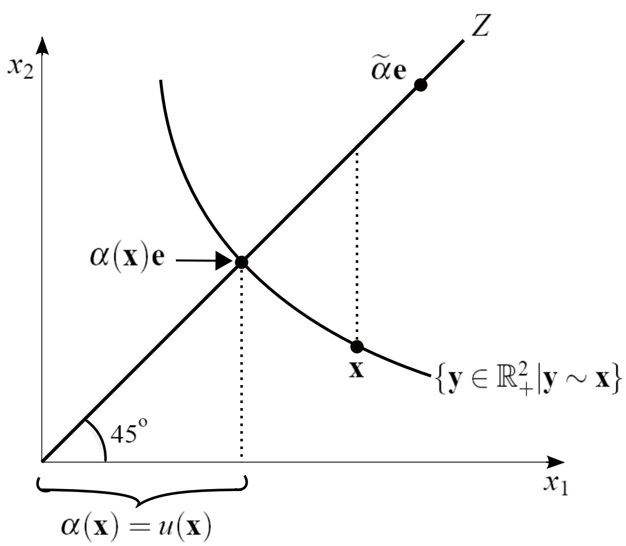

Figura 3.4: Construcción de una función de utilidad.

Utilizando el hecho de que la relación de preferencias sea monótona y continua, vamos a probar que existe un único valor $\alpha(\mathbf{x}) \in [0, \tilde{\alpha}]$ tal que $\alpha(\mathbf{x})\mathbf{e} \sim \mathbf{x}$.

Por continuidad, los conjuntos

$$A_+ = \{\alpha \in \mathbb{R}_+ \mid \mathbf{x} \precsim \alpha\mathbf{e}\}$$

y

$$A_- = \{\alpha \in \mathbb{R}_+ \mid \alpha\mathbf{e} \precsim \mathbf{x}\}$$

son no vacíos y cerrados. Debido a que la relación de preferencias cumple la propiedad de orden total, $\mathbb{R}_+ \subset (A_+ \cup A_-)$. Por otra parte, como A_+ y A_- son conjuntos cerrados no vacios y $\mathbb{R}$ es conexo, entonces $A_+ \cap A_- \neq \emptyset$; por lo tanto existe un escalar α tal que $\alpha\mathbf{e} \sim \mathbf{x}$. Más aún, por monotonicidad, $\alpha_1\mathbf{e} \prec \alpha_2\mathbf{e}$ siempre que $\alpha_1 < \alpha_2$; entonces, existe un único escalar que satisface $\alpha\mathbf{e} \sim \mathbf{x}$, a dicho escalar lo denotamos por $\alpha(\mathbf{x})$.

Hasta este punto $\alpha(\mathbf{x})$ es nuestra función de utilidad, es decir, asignamos un valor de utilidad $u(\mathbf{x}) = \alpha(\mathbf{x})$ para todo $\mathbf{x}$. Este nivel de utilidad está representado en la Figura 3.4 por el conjunto $\{\mathbf{y} \in \mathbf{R}^2 \mid \mathbf{y} \sim \mathbf{x}\}$. Ahora necesitamos probar que la función de utilidad en realidad representa las preferencias $\precsim$ y que es una función continua.

El hecho de que $\alpha(\mathbf{x})$ represente las preferencias $\precsim$ se sigue de su construcción. Vamos a probar que $\mathbf{y} \precsim \mathbf{x} \Leftrightarrow \alpha(\mathbf{y}) \leq \alpha(\mathbf{x})$.

$\Leftarrow$) Supongamos que $\alpha(\mathbf{y}) \leq \alpha(\mathbf{x})$. Por monotonicidad, se tiene que $\alpha(\mathbf{y})\mathbf{e} \precsim \alpha(\mathbf{x})\mathbf{e}$; como $\mathbf{x} \sim \alpha(\mathbf{x})\mathbf{e}$ y $\mathbf{y} \sim \alpha(\mathbf{y})\mathbf{e}$ entonces $\mathbf{y} \precsim \mathbf{x}$.

$\Rightarrow$) Supongamos que $\mathbf{y} \precsim \mathbf{x}$, entonces $\alpha(\mathbf{y})\mathbf{e} \sim \mathbf{y} \precsim \mathbf{x} \sim \alpha(\mathbf{x})\mathbf{e}$ y por monotonicidad se debe cumplir que $\alpha(\mathbf{y}) \leq \alpha(\mathbf{x})$.

Ahora vamos a probar que $\alpha(\mathbf{x})$ es una función continua para todo $\mathbf{x}$, es decir, para toda sucesión $\{\mathbf{x}_n\}_{n\in\mathbb{N}}$ tal que $\text{lím}_{n\to\infty} \mathbf{x}_n = \mathbf{x}$, entonces $\text{lím}_{n\to\infty} \alpha(\mathbf{x}_n) = \alpha(\mathbf{x})$.

Consideremos una sucesión $\{\mathbf{x}_n\}_{n\in\mathbb{N}}$ tal que $\text{lím}_{n\to\infty} \mathbf{x}_n = \mathbf{x}$. Notemos que la sucesión dada por $\{\alpha(\mathbf{x}_n)\}_{n\in\mathbb{N}}$ debe tener una subsucesión convergente. Por monotonicidad se tiene que para cualquier $\varepsilon > 0$, $\alpha(\mathbf{x}')$ está en un subconjunto compacto de $\mathbb{R}_+$, $[\alpha_0, \alpha_1]$, para todo $\mathbf{x}'$ tal que $\|\mathbf{x}' - \mathbf{x}\| \leq \varepsilon$ (ver Figura 3.5). Como $\{\mathbf{x}_n\}_{n\in\mathbb{N}}$ converge a $\mathbf{x}$, existe $N \in \mathbb{N}$ tal que $\alpha(\mathbf{x}_n) \in [\alpha_0, \alpha_1]$ para todo $n > N$. Pero cualquier sucesión infinita que esté en un conjunto compacto debe tener una subsucesión convergente.

Ahora falta demostrar que todas las subsucesiones convergentes de $\{\alpha(\mathbf{x}_n)\}_{n=1}^{\infty}$ convergen a $\alpha(\mathbf{x})$. Supongamos, por contradicción, que existe una función $m(\cdot)$ estrictamente creciente que asigna a cada entero positivo n un entero positivo $m(n)$ y para la cual, la subsucesión $\{\alpha(\mathbf{x}_{m(n)})\}_{n\in\mathbb{N}}$ converge a $\alpha' \neq \alpha(\mathbf{x})$.

Primero probaremos que si suponemos $\alpha(\mathbf{x}) < \alpha'$ llegamos a una contradicción. Supongamos $\alpha(\mathbf{x}) < \alpha'$, por monotonicidad tenemos que $\alpha(\mathbf{x})\mathbf{e} \prec \alpha'\mathbf{e}$. Sea $\widehat{\alpha} = \frac{1}{2}[\alpha' + \alpha(\mathbf{x})]$. El punto $\widehat{\alpha}\mathbf{e}$ es el punto medio que está sobre Z entre $\alpha'\mathbf{e}$ y $\alpha(\mathbf{x})\mathbf{e}$ (ver Figura 3.5). Por monotonicidad se cumple que $\alpha(\mathbf{x})\mathbf{e} \prec \widehat{\alpha}\mathbf{e}$. Por otra parte, como $\alpha(\mathbf{x}_{m(n)}) \to \alpha' > \widehat{\alpha}$, existe $\overline{N} \in \mathbb{N}$ tal que para todo $n > \overline{N}$ se tiene que $\alpha(\mathbf{x}_{m(n)}) > \widehat{\alpha}$. Ahora, para todos esos n's se cumple $\mathbf{x}_{m(n)} \sim \alpha(\mathbf{x}_{m(n)})\mathbf{e}$ y por monotonicidad $\widehat{\alpha}\mathbf{e} \prec \alpha(\mathbf{x}_{m(n)})\mathbf{e}$. Ya que la relación de preferencias es continua, entonces

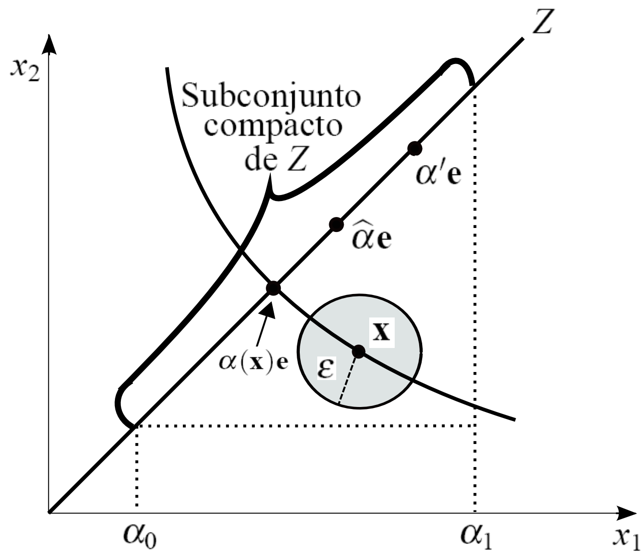

Figura 3.5: Prueba de que la función de utilidad construida es continua.

$\widehat{\alpha}\mathbf{e} \precsim \mathbf{x}$, y como $\mathbf{x} \sim \alpha(\mathbf{x})\mathbf{e}$, se tiene que $\widehat{\alpha}\mathbf{e} \precsim \alpha(\mathbf{x})\mathbf{e}$; lo cual es una contradicción pues teníamos que $\alpha(\mathbf{x})\mathbf{e} \prec \widehat{\alpha}\mathbf{e}$. El argumento para descartar el caso en el que $\alpha' < \alpha(\mathbf{x})$ es similar.

Por lo tanto, ya que todas las subsucesiones convergentes de $\{\alpha(\mathbf{x}_n)\}_{n\in\mathrm{N}}$ convergen a $\alpha(\mathbf{x})$, tenemos que $\lim_{n\to\infty} \alpha(\mathbf{x}_n) = \alpha(\mathbf{x})$, completando la prueba. ■

Capítulo 4

Un modelo de la asignación del tiempo de los niños a educación y trabajo

Este capítulo tiene una fuerte intersección con Acosta (2008), ya que dicho trabajo se enfoca en el análisis de una parte de la solución del modelo que se discute en esta tesis. De manera evidente, el planteamiento del problema es el mismo. Asimismo, el presente capítulo está basado ampliamente en el libro *El Trabajo Infantil en México 1984-2000* de H. Robles.

El análisis de las decisiones entre estudio y trabajo para los niños se puede abordar a través de distintos enfoques económicos. Estos enfoques pueden ser clasificados de acuerdo al carácter de la educación y el altruismo de los padres.

Por una parte, el bienestar de los padres puede depender o no de la riqueza de sus hijos, independientemente de que éstos los ayuden económicamente en el futuro. Consideremos dos tipos de padres:

Padre altruista: aquel padre cuyo bienestar es influido directamente por la felicidad de sus hijos a lo largo de su ciclo de vida.

Padre no altruista: aquel padre que sea indiferente a la situación en que se encuentren sus hijos ahora y en el futuro.

Descartamos el caso de padres malintencionados que incrementan su felicidad a costa del bienestar y desarrollo de sus hijos.

Por otra parte, la educación formal puede ser vista como inversión en capital humano o como un bien de consumo.

El enfoque que considera a la educación como inversión está basado en la teoría del capital humano. Dicha teoría fue desarrollada, entre otros, por el economista estadounidense Gary Becker. De la manera más simple, podemos definir al capital humano como los conocimientos, las aptitudes y la experiencia en las que intencionadamente los individuos invierten recursos con objeto de disfrutar de un flujo de beneficios futuros; estos beneficios pueden ser económicos o de otra índole. El capital humano puede incrementarse invirtiendo en educación, atención a la salud y capacitación laboral, principalmente. Esta teoría deduce que es más conveniente que los individuos dediquen parte de su tiempo a adquirir educación formal a edades tempranas, ya que en esa etapa de la vida sus costos de oportunidad son muy bajos o nulos. Al considerar esto último, podemos decir que los padres tienen la posibilidad de maximizar la remuneración del capital humano de sus hijos, al educarlos desde pequeños; simplemente porque entre más jóvenes terminen ellos de educarse, recibirán por un periodo más extenso dicha remuneración. De esta forma, la teoría del capital humano concluye que un proceso eficiente para maximizar el ingreso de los individuos consiste en la adquisición de educación formal durante los periodos de niñez y adolescencia (Becker, 1993).

La perspectiva del capital humano considera que las personas, al invertir en sí mismas, además de perseguir un bienestar presente, también buscan un rendimiento futuro de cualquier índole. En este trabajo se estudiará la adquisición de educación como una inversión en capital humano y se analizarán los determinantes de la inversión en educación suponiendo que los beneficios futuros se reducen únicamente a los económicos.

Cabe destacar que la adquisición de educación es analizada también desde otra perspectiva, considerándola como un gasto en un bien de consumo. En esta perspectiva, las personas se educan porque la obtención de conocimientos, habilidades y aptitudes son satisfactorias en sí mismas. El enfoque que considera a la educación como un bien de consumo plantea que los padres educan a sus hijos porque ésta es un bien que influye positivamente en el desarrollo social y cognitivo de los niños. Este punto de vista, económicamente equivale a atribuirle a la educación las características de un bien normal de consumo duradero. Se dice que es un bien normal porque al aumentar el ingreso, aumenta el consumo de la misma; de consumo porque el objetivo de la educación es satisfacer la demanda final del individuo sin intentar producir otros bienes o servicios, contrario al caso anterior; y por último, se considera duradero porque le retribuye al individuo, durante un tiempo relativamente largo, un flujo de servicios.

El esquema siguiente ilustra los tres modelos generales en los que pueden ser ordenados los estudios económicos sobre la asignación del tiempo de los niños a educación y trabajo. Se excluye el caso en el que los padres son no altruistas y la educación es vista como un bien de consumo, pues como la educación influye positivamente en el desarrollo de los niños, entonces los padres que educan a sus hijos tienen que ser altruistas.

	Altruismo de los padres	
Carácter de la educación	Altruistas	No altruistas
Inversión	Posible	Posible
Consumo	Posible	No posible

Es importante notar que las relaciones intrafamiliares pueden influir en las decisiones de asignación del tiempo de los niños; para simplificar esta realidad se tienen algunas hipótesis sobre las relaciones existentes en el núcleo familiar. Bergstrom (1997), distingue dos esquemas teóricos de la familia:

Modelos unitarios: son aquellos en los cuales las decisiones económicas de los miembros de la familia se analizan a partir de una sola función de utilidad, que corresponde a la persona que toma las decisiones.

Modelos plurales: son aquellos donde los miembros de la unidad familiar desarrollan interacciones estratégicas para la toma de decisiones económicas de los recursos del hogar.

En este estudio se toma el enfoque que considera a la educación como una inversión en capital humano, asumiendo a los padres altruistas. Además sólo es admitido un esquema familiar sin conflictos, por lo que optaremos por un modelo unitario.

4.1

Planteamiento del problema.

El presente trabajo considera que la educación es capital humano y que los beneficios futuros de ésta, son económicos. El modelo microeconómico se centra en los usos del tiempo del menor. Es un modelo heurístico, unitario y dinámico de los usos del tiempo de un niño. Es un modelo dinámico porque distingue explícitamente en los tiempos de inversión y de recolección de beneficios.

La teoría del capital humano asume que el proveer de educación a los hijos es una forma de transferirles riqueza (Becker,1993). El objetivo de los padres es transferirle la mayor cantidad de riqueza humana y no humana al menor, a modo de maximizar su bienestar futuro; pues debido al supuesto de padres altruistas, éstos serán más felices mientras mayor bienestar tenga el niño en su infancia y en su vida adulta. Esto lleva a los padres a tratar de proporcionarle a su hijo las mejores condiciones de vida posibles; otorgándole tiempo para descansar, invirtiendo en su educación y transfiriéndole activos en vida o a través de herencias en forma óptima.

El modelo se desarrollará considerando los siguientes supuestos:

i) Se supone que el padre puede influir en el bienestar futuro de su hijo mediante la asignación de una parte del tiempo del menor a su educación formal, y/o dándole transferencias monetarias cuando alcance la madurez.

ii) Suponemos dos periodos, el primero de ellos es la **infancia**, en esta etapa los padres cuidan de sus hijos y asignan el tiempo de éstos a educación y, eventualmente, a trabajar en el mercado laboral o en la producción económica o doméstica de la unidad familiar. El segundo es la **madurez**, en donde los padres envejecen y abandonan el mercado laboral. Ellos viven de sus ahorros y, eventualmente, transfieren activos económicos a sus hijos.

iii) Sin pérdida de generalidad, las decisiones laborales de los padres se consideran exógenas, suponiendo que el padre trabaja lo más posible.

iv) La dirección de las transferencias es de padres a hijos.

v) Existe perfecta previsión de todos los precios, incluida la tasa de interés.

A continuación explicaremos las hipótesis económicas comprendidas en el modelo para poder plantearlo totalmente.

4.1.1

Función de utilidad del padre

Como estamos bajo el esquema de un modelo unitario, las decisiones económicas de los miembros del hogar se analizan a partir de una sola función de utilidad, correspondiente a la persona que toma las decisiones. A dicha persona la llamaremos padre.

De acuerdo a la perspectiva de la teoría del capital humano, es natural dividir el análisis del modelo en dos periodos, en el primero se invierte en la educación del niño y en el segundo se obtienen los beneficios de dicha inversión. De esta forma, se tendrá una función de utilidad del padre para cada periodo.

La felicidad del padre estará directamente influenciada por la felicidad de su hijo, ya que aquel es altruista. En el primer periodo el niño es feliz al consumir tiempo en ocio, el cual denotaremos por L, esto lleva al padre a considerar al ocio del menor como una de sus variables de su función de utilidad. Además, el tiempo dedicado al ocio y al trabajo limitan el tiempo que el menor podría dedicar a su educación formal. Por otra parte, es razonable considerar que la felicidad del padre depende del bienestar su familia; dicho bienestar puede ser contemplado a través del consumo familiar de bienes y servicios. Al consumo familiar del primer periodo lo denotaremos por C_1; notemos que el gasto en C_1 limita las posibilidades de que el menor adquiera educación formal. De esta forma, $U(C_1, L)$ representará la función de utilidad del padre del primer periodo.

Cuando el hijo alcance la madurez, su felicidad dependerá, principalmente, del nivel de riqueza que posea. La adquisición de esta riqueza puede ser por medio de la remuneración a su capital humano formado durante su niñez, denotado por τH; o posiblemente reciba algún tipo de transferencias, las cuales están representadas por Tr. El parámetro τ y la variable H representan el precio de una unidad de capital humano del individuo y el nivel de capital humano, respectivamente. Análogamente al primer periodo, el bienestar paterno dependerá del consumo familiar del segundo periodo, denotado por C_2. De esta manera, $V(C_2, \tau H, Tr)$ representará la función de utilidad paterna del segundo periodo.

Debido a que se tienen dos funciones de utilidad en diferentes periodos de tiempo, es necesario incluir un factor de descuento temporal, al que llamaremos β. Ya que el propósito de β es comparar funciones de naturaleza subjetiva, él mismo tiene dicha característica, pues para cada uno de nosotros un objeto hoy no nos retribuye la misma utilidad que en el futuro. Por ejemplo, si les preguntáramos a dos personas de diferentes recursos económicos, donde una tiene un nivel de riqueza mayor que la otra, cuál sería su utilidad de tener una unidad monetaria adicional hoy y cuál la de obtenerla en el futuro; seguramente la primera respondería que en el futuro, pues es muy probable que en el presente cuente con los recursos económicos suficientes para satisfacer sus necesidades y le es más atractiva la posibilidad de tener dicha riqueza adicional en el futuro. En cambio, podríamos pensar que para la otra persona es más benéfico contar con esos recursos en el presente, pues es muy probable que los recursos con que cuenta no sean suficientes para cubrir sus necesidades. De esta manera el parámetro β asociado a la persona que posee un menor nivel de riqueza será más pequeño que en comparación con el parámetro β asociado a la persona que posee un menor mayor nivel de riqueza.

Finalmente, la función de utilidad paterna para los dos periodos está dada por:

$$U(C_1, L) + \beta V(C_2, \tau H, Tr). \tag{4.1}$$

4.1.2

Función de producción de capital humano

En general se puede definir a una función de producción como aquella que muestra la relación entre la cantidad total de producción y la cantidad de insumo utilizado. En particular, la producción de capital humano del menor depende del tiempo que éste asigne al estudio. Así, es razonable suponer una función de producción de capital humano para el menor.

Una característica importante que cumplen las funciones de producción, y por ende la función de producción de capital humano, es la establecida por el principio de rendimientos marginales decrecientes. Veamos un sencillo ejemplo antes de enunciar dicho principio.

Imaginemos que queremos vender galletas y que para producirlas se necesitan batidoras, hornos, trabajadores, harina, azúcar, aromatizantes, colorantes, etcétera. Además, supongamos que tenemos una cantidad fija de batidoras, hornos, harina, azúcar, aromatizantes y colorantes; y que estamos interesados en analizar cómo varía la producción de galletas variando simplemente el número de trabajadores. La siguiente tabla muestra el número de trabajadores, en la primer columna; la producción de galletas por hora, en la segunda columna y la tercer columna indica el aumento que experimenta la producción de galletas al contratar un trabajador adicional; a lo que llamamos producto marginal.

Número de trabajadores	**Producto total**	**Producto marginal**
0	0	-
1	10	10
2	30	20
3	60	30
4	80	20
5	95	15
6	105	10
7	110	5

Como se puede observar en la tabla, el segundo trabajador tiene un producto marginal de 20 galletas, el tercero de 30, y así sucesivamente. Notemos que a niveles bajos de contratación de trabajadores, el producto marginal es positivo y creciente, esto es, el contratar un trabajador adicional aumenta la producción. Sin embargo, llega el momento en el que seguir contratando trabajadores provoca una producción adicional positiva pero decreciente; en el ejemplo este fenómeno se presenta a partir de 3 trabajadores. Este hecho es bastante razonable, ya que a medida que aumenta su número, los trabajadores adicionales tienen que compartir el equipo y trabajar en un lugar más saturado. Por lo tanto, a medida que se contratan más trabajadores, cada empleado adicional contribuye menos a la producción total de galletas.

Análogamente, si en lugar de variar el número de trabajadores variamos el número de hornos, manteniendo todos los demás insumos constantes, se presentará una situación similar a la descrita. Este comportamiento es conocido como el principio de rendimientos marginales decrecientes, el cual establece que a partir de una cierta cantidad de insumo, si la cantidad de dicho insumo aumenta, entonces la producción aumenta pero a un ritmo más y más débil; manteniendo todos los demás factores constantes. En otras palabras, llegará un momento en el cual los incrementos de la producción marginal resultantes del aumento de algún insumo serán decrecientes.

Este principio puede ser formalizado de la siguiente manera:

Sea $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ una función de producción que cumple con el *principio de rendimientos marginales decrecientes* para el i-ésimo insumo, entonces, manteniéndose todo lo demás constante, a partir de un momento $D_i f > 0$ y $D_{i,i} f < 0$; es decir, f es cóncava como función de la i-ésima variable. En el caso de que la productividad marginal de las primeras unidades consumidas del i-ésimo insumo sea positiva y creciente, entonces $D_i f > 0$ y $D_{i,i} f > 0$, es decir, f es convexa como función de la i-ésima variable.

Podemos apreciar el principio de rendimientos marginales decrecientes en la función de producción de capital humano del menor como sigue: para las primeras unidades de tiempo de estudio del niño, la producción marginal será positiva y creciente, en otras palabras, la cantidad de capital humano se incrementará conforme el menor aumente su tiempo de estudio; sin embargo, a partir de un cierto nivel de tiempo asignado al estudio, al aumentar más unidades de éste, el capital humano del menor se incrementará pero en menor cantidad. Por dar un ejemplo simple, supongamos que un individuo ha decidido estudiar todo un día alguna materia en específico, muy probablemente las primeras horas de estudio serán muy productivas, pero conforme avance el día el rendimiento del individuo será menor, pues seguramente llegará un momento en el que éste se sature de información.

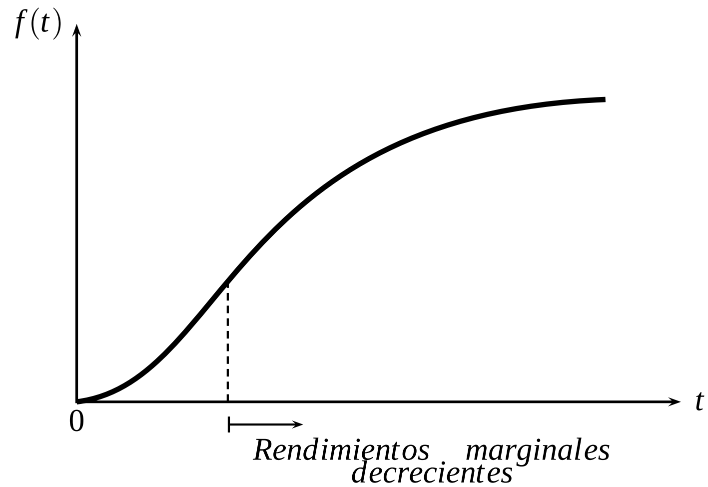

Figura 4.1: Función de producción de capital humano.

La Figura 4.1 ilustra el comportamiento general de la función de producción de capital humano del menor, la cual denotaremos por $f(t)$, donde t es el tiempo dedicado a su educación formal. Así, el nivel de capital humano formado durante la niñez, H, estará determinado por la función de producción del menor, $H = f(t)$. Es importante notar que para un nivel de estudio nulo no habrá producción de capital humano, es decir, $f(0) = 0$. Cabe señalar que si el propósito es maximizar ésta función, la parte de interés es la región cóncava. [1]

4.1.3

Concavidad de las funciones de utilidad

Debido a que cada persona tiene distintos gustos y preferencias, que cambian incluso de momento a momento, es muy difícil medir la utilidad; pese a esto, la utilidad suele observar

[1] Por sentido común podemos ver que la función de producción de capital humano del menor no sólo depende de t, pero se asumirá así, ya que el objeto de estudio del modelo es la asignación del tiempo del menor. Esto se puede consultar en el Anexo.

un comportamiento marginal decreciente. Esta propiedad es mejor conocida como el *principio de utilidad marginal decreciente*, el cual establece que, más allá de la diversidad de los gustos individuales, la satisfacción lograda mediante el consumo de un bien aumenta con el incremento del consumo, pero tal aumento de satisfacción se produce a un ritmo cada vez más débil, de tal manera que se presenta una saturación progresiva. Para ilustrar esto, pensemos que tenemos un antojo muy grande de comer pastel y vamos a una pastelería para merendar. La primera rebanada de pastel nos producirá mucha satisfacción; la segunda ya no nos gustará tanto. Si seguimos comiendo más rebanadas llegará un momento en el que nos sentiremos saciados y la satisfacción obtenida al comer una rebanada adicional será casi nula.

La noción de marginalidad es un término económico utilizado en el sentido de "adicional". La *utilidad marginal* indica la utilidad adicional que reporta el consumo de una unidad más de un bien o servicio.

El principio de utilidad marginal decreciente puede ser integrado a las funciones de utilidad asumiéndolas cóncavas, esto se puede justificar de la siguiente manera. Para cualquier valor de C_1, la primera unidad de consumo de ocio, L, genera una utilidad muy grande, y conforme se van consumiendo más unidades de ocio, la utilidad va aumentando pero en menor cuantía que la anterior. Claramente, este comportamiento puede ser descrito mediante una función cóncava. Por lo que si fijamos C_1 y vemos a U como función de L solamente, la podemos considerar cóncava. De igual forma, si fijamos L y vemos a U únicamente como función de C_1, también la podemos considerar cóncava.

Mediante un argumento análogo al anterior, podemos asumir a la función V cóncava como función únicamente de una de sus variables, manteniendo las otras dos constantes.

Si consideramos a $U(C_1,L)$ y $V(C_2,\tau f(t),Tr)$ como funciones de todas sus variables, podría suceder que éstas no fueran cóncavas. Sin embargo, más adelante se impondrán algunas hipótesis económicas sobre los argumentos de estas funciones con las cuales se asegura la concavidad de U y V.

4.1.4

Hipótesis económicas sobre las variables de las funciones

Los bienes C_1, C_2 y L poseen la característica de ser *bienes necesarios*, pues no es posible que la familia no consuma ningún tipo de bienes y servicios, o que el menor no cuente con un momento de descanso. Esta propiedad se incorpora al modelo mediante la positividad estricta de los niveles óptimos de C_1, C_2 y L en la solución. Esto se garantiza con las condiciones siguientes.

$$\lim_{C_1 \to 0} \frac{\partial U(C_1, L)}{\partial C_1} = +\infty \qquad \forall L > 0 \tag{4.2}$$

$$\lim_{L \to 0} \frac{\partial U(C_1, L)}{\partial L} = +\infty \qquad \forall C_1 > 0 \tag{4.3}$$

$$\lim_{C_2 \to 0} \frac{\partial V(C_2, \tau H, Tr))}{\partial C_2} = +\infty \qquad \forall H, Tr \geq 0. \tag{4.4}$$

La interpretación económica de la ecuación (4.2) es la siguiente: para todo nivel de ocio, el consumo inicial de una cantidad de C_1 cercana a cero genera una utilidad arbitrariamente grande; lo cual es muy razonable, pues imaginemos que una persona no ha bebido ningún tipo de líquido durante todo el día, y al anochecer se le ofrece un vaso de agua, de manera que al tomar el primer sorbo su satisfacción será enorme. Observemos que el valor óptimo de C_1 no puede ser cero, ya que el nivel de utilidad más grande es el generado por el consumo de la primera unidad. La argumentación anterior se mantiene para las ecuaciones (4.3) y (4.4).

Otra hipótesis económica utilizada en el modelo, que no contraviene el *principio de no saciedad*, es que consumos adicionales de C_1 y C_2, cuando los niveles de éstos son arbitrariamente grandes, producen únicamente cambios marginales en la utilidad muy pequeños y que tienden a cero. Analíticamente se cumple

$$\lim_{C_1 \to +\infty} \frac{\partial U(C_1, L)}{\partial C_1} = 0 \qquad \forall L > 0 \tag{4.5}$$

y

$$\lim_{C_2 \to +\infty} \frac{\partial V(C_2, \tau H, Tr))}{\partial C_2} = 0 \qquad \forall H, Tr \geq 0. \tag{4.6}$$

La ecuación (4.5) establece que para cualquier valor de L, el comportamiento de la función de utilidad U en el infinito es asintótico. Esto quiere decir que cuando se haya consumido una gran cantidad de C_1, al consumir otra unidad más, la utilidad generada por esta unidad adicional es casi nula. La interpretación anterior es la misma para la ecuación (4.6).

Finalmente, se tiene una hipótesis sobre la existencia de un nivel de ocio $L = T$ que sacia U, donde T es la dotación de tiempo del niño durante su niñez. Esta hipótesis es razonable ya que si asignáramos todo nuestro tiempo disponible a ocio, con seguridad podemos decir que en algún momento nos sentiríamos aburridos y con deseos de realizar alguna otra actividad. Un ejemplo muy simple que puede ilustrar esta hipótesis es suponer que un niño ha distribuido todo su tiempo en actividades de entretenimiento, divirtiéndose con sus juguetes y los de sus amigos, yendo al cine, viendo televisión, jugando algunos juegos de mesa, etcétera. Seguramente, después de un tiempo todas estas actividades podrían aburrirle, y le sería atractivo dedicar parte de su tiempo a alguna otra tarea.

Esta hipótesis se puede formular de la siguiente manera:

$$\lim_{L \to T} \frac{\partial U(C_1, L)}{\partial L} = 0 \qquad \forall C_1 > 0. \tag{4.7}$$

Cabe hacer notar que esta condición permite que el menor dedique tiempo a otras actividades en su niñez, a estudiar, por ejemplo.

4.1.5

Separabilidad de las funciones de utilidad

Asumiremos los siguientes supuestos para facilitar el análisis del modelo.

i) Las decisiones de ocio son separables de las decisiones de consumo.

ii) Las decisiones sobre las transferencias e inversión en capital humano son separables de las decisiones de consumo.

iii) Las transferencias y el valor del acervo de capital humano son sustitutos perfectos.

Recordemos que el objeto de estudio en el modelo es el tiempo del niño, así que la suposición que se tiene en el primer inciso sólo es para enfocarnos en el tiempo de ocio y simplificar el análisis de éste. Este supuesto es poco intuitivo, ya que se pierde la relación de sustitución existente entre consumo y ocio, dicha relación refleja la disponibilidad del individuo a intercambiar consumo por ocio para mantener el mismo nivel de satisfacción.

La explicación del segundo supuesto es análoga a la anterior. Así que la función de utilidad del padre (4.1) puede ser reescrita como

$$U_p(C_1) + U_h(L) + \beta[V_p(C_2) + V_h(\tau f(t), Tr)],$$

donde los subíndices p y h representan al padre y al hijo, respectivamente.

Por otra parte, la satisfacción del deseo del padre por transferirle riqueza a su hijo puede ser obtenida a través de la inversión en capital humano y de las transferencias que eventualmente le pueda dar, ya que ambas opciones tienen la finalidad de que el hijo cuente con una buena calidad de vida en el futuro. La tercera hipótesis refleja esta idea, así que tenemos:

$$V_h(\tau f(t), Tr) = V_h(\tau f(t) + Tr).$$

La función $V_h(\tau f(t)+Tr)$ es una composición de funciones, $V_h \circ g$, donde $V_h : \mathbb{R} \to \mathbb{R}$ y $g : \mathbb{R}^2 \to \mathbb{R}$ definida como $g(t,Tr) = \tau f(t)+Tr$. Observemos que g es suma de funciones cóncavas, por lo que ésta también es una función cóncava. Luego, la función $V_h(\tau f(t)\ +Tr)$ es composición de funciones cóncavas, por lo que ella misma cumple dicha propiedad.

Por último, tomando en cuenta la discusión hecha hasta este punto, podemos considerar a las funciones $U_p(C_1), U_h(L), V_p(C_2), V_h(\tau f(t)+Tr)$ y $f(t)$ no sólo cóncavas, sino estrictamente cóncavas crecientes y diferenciables. Las consideramos de esta forma con el fin de asegurar la existencia de una única solución y la aplicación de algunos resultados matemáticos para resolver el problema que se planteará en este modelo. Por consiguiente, la función de utilidad del padre es suma de funciones estrictamente cóncavas crecientes y diferenciables, así que ella misma tiene dichas propiedades.

La función de utilidad del padre para los dos períodos está dada por:

$$U_p(C_1)+U_h(L)+\beta[V_p(C_2)+V_h(\tau f(t)+Tr)],$$

donde

$$\begin{aligned}
C_1 :=&\ \text{consumo familiar durante el primer periodo.}\\
C_2 :=&\ \text{consumo familiar durante el segundo periodo.}\\
L :=&\ \text{ocio del menor.}\\
t :=&\ \text{tiempo de estudio del menor.}\\
Tr :=&\ \text{transferencias que los padres dan a sus hijos en su vida adulta.}\\
\tau :=&\ \text{precio de una unidad de capital humano del individuo.}\\
\beta :=&\ \text{factor de descuento temporal.}
\end{aligned}$$

El precio de una unidad de capital humano del individuo y el factor de descuento temporal son parámetros de la función de utilidad.

4.1.6

Restricciones del modelo

La primera restricción que consideraremos es conocida como restricción presupuestal, la cual muestra todas las cestas de consumo posibles que puede permitirse el consumidor con un ingreso dado; ya que las decisiones de elección de bienes y servicios por parte del consumidor dependen de su presupuesto total. Para este caso, la restricción presupuestal es la siguiente:

$$\boxed{p_1C_1+\omega(L+t)+\frac{(p_2C_2+Tr)}{1+r} \leq W+\omega T,}$$

donde

$$
\begin{aligned}
p_1 &:= \text{ precios de los bienes de consumo en la infancia.}\\
p_2 &:= \text{ precios de los bienes de consumo en la madurez.}\\
\omega &:= \text{ tasa salarial del hijo.}\\
r &:= \text{ tasa de interés.}\\
t &:= \text{ tiempo asignado a la formación de capital humano.}\\
W &:= \text{ riqueza familiar.}
\end{aligned}
$$

Esta restricción indica todos los gastos de la familia durante los dos periodos, así como el ingreso que posee para solventar dichos gastos. La expresión p_1C_1 denota el gasto en el consumo familiar del primer periodo, p_2C_2 y Tr los gastos en el consumo familiar del segundo periodo y en transferencias, respectivamente; los últimos dos gastos están considerados en el futuro, así que es necesario traerlos al presente, por lo que dividimos $p_2C_2 + Tr$ entre $1 + r$. Estos gastos no deben exceder el presupuesto familiar, el cual está formado por toda la riqueza de la familia, como por ejemplo, el ingreso laboral del padre; el ingreso no laboral de la familia, que es aquel dinero proveniente de herencias y/o la lotería; en caso de que el menor trabaje, el ingreso laboral de éste, $\omega(T-L-t)$, donde ω representa la tasa salarial del menor; propiedades (valuadas monetariamente), etcétera. Denotaremos por W a toda la riqueza familiar excepto el ingreso laboral del menor, y nos referiremos a ella simplemente como riqueza familiar o como ingreso no laboral del menor.

La siguiente restricción surge de manera muy natural, pues establece que el tiempo que el menor dedica a estudio, ocio y, de ser necesario, al trabajo, no debe rebasar el tiempo total del que dispone, así tenemos

$$\boxed{t+L \leq T.}$$

La tercera restricción establece que el tiempo que el menor asigna a su educación formal está acotado por la solución de un modelo básico de inversión en capital humano. Este modelo considera que el único costo de oportunidad del individuo es el tiempo que podría dedicar a trabajar en lugar de estudiar, asumiendo que el individuo no enfrenta restricciones crediticias de ningún tipo y que las decisiones de inversión en capital humano son independientes de las decisiones de ocio y consumo.

De esta forma, el tiempo óptimo de educación formal de un niño en edad escolar, está dado por la solución al siguiente problema:

Problema 4.1.

$$
\begin{aligned}
\text{Maximizar} \quad & \tau f(t) - \omega(1+r)t\\
\text{sujeto a} \quad & 0 \leq t \leq T
\end{aligned}
$$

donde

$$\begin{aligned}
\tau :=& \text{ precio de la renta por una unidad de capital humano.}\\
t :=& \text{ tiempo asignado a la formación de capital humano.}\\
f(t) :=& \text{ función de producción de capital humano del menor.}\\
\omega :=& \text{ tasa salarial del hijo.}\\
r :=& \text{ tasa de interés.}
\end{aligned}$$

El Problema 4.1 consiste en optimizar los beneficios del individuo, maximizando el ingreso que obtendrá por su formación de capital humano, denotado por $\tau f(t)$, menos el costo generado por estudiar, que es el ingreso que obtendría si en lugar de estudiar trabajara, $\omega(1+r)t$.

Analíticamente el Problema 4.1 es maximizar una función real estrictamente cóncava creciente y diferenciable sobre un conjunto compacto, por lo que alcanza su valor óptimo. Sea t_s dicho valor óptimo, entonces debe satisfacer:

$$\tau f'(t_s) = \omega(1+r). \tag{4.8}$$

El nivel de estudio t_s puede ser interpretado como el tiempo óptimo económico para la educación en mercados que permitan la perfecta financiación de ésta, es claro que este tiempo óptimo sólo depende de la productividad del menor. De manera evidente, cuando el menor se enfrenta a mercados imperfectos en la financiación de la educación, el tiempo asignado a su educación debe satisfacer:

$$\boxed{t \le t_s.}$$

Por último, suponemos que las transferencias de padres a hijos es no negativa al igual que el tiempo de estudio del menor:

$$\boxed{0 \le Tr} \quad \text{y} \quad \boxed{0 \le t.}$$

Por lo tanto, el problema enfrentado por el padre es maximizar su función de utilidad, sujeta a la restricción presupuestal, las restricciones en el uso del tiempo del menor y la no negatividad de las transferencias paternas. Analíticamente el problema es el siguiente:

Sea $\mathscr{P} : \mathbb{R}^5 \to \mathbb{R}$ la función definida como

$$\mathscr{P}(C_1, L, C_2, t, Tr) = U_p(C_1) + U_h(L) + \beta\left[V_p(C_2) + V_h(\tau f(t) + Tr)\right],$$

la cual llamaremos *función objetivo*. Sean $g_i : \mathbb{R}^5 \to \mathbb{R}$ para $i = 1 \ldots 5$ las siguientes funciones:

$$\begin{aligned}
g_1(C_1, L, C_2, t, Tr) &= W + \omega(T - L - t) - p_1 C_1 - \frac{(p_2 C_2 + Tr)}{1+r}\\
g_2(C_1, L, C_2, t, Tr) &= T - t - L\\
g_3(C_1, L, C_2, t, Tr) &= t_s - t\\
g_4(C_1, L, C_2, t, Tr) &= Tr\\
g_5(C_1, L, C_2, t, Tr) &= t
\end{aligned}$$

donde t_s es solución del Problema 4.1.

Así, el problema al que se enfrenta el padre es:

Problema 4.2.

$$\begin{array}{cll} \text{Maximizar} & \mathscr{P}(C_1, L, C_2, t, Tr) & \\ \text{sujeto a} & g_i(C_1, L, C_2, t, Tr) \geq 0 & \text{para } i = 1, \ldots, 5. \end{array}$$

Las diferentes soluciones de asignación del tiempo del menor a ocio, tiempo de estudio y trabajo propuestas por Robles al Problema 4.2 (Robles, 2000), están descritas en función del nivel de riqueza del hogar y de la factibilidad del tiempo de estudio t_s, que se discutirá más adelante. El objetivo de este trabajo es estudiar una parte de las soluciones del Problema 4.2 propuestas por Robles para el caso en el que el tiempo de estudio t_s no es factible; realizando un análisis tanto analítico como microeconómico sobre el comportamiento de las ecuaciones que resultan al aplicar el Teorema de Kuhn-Tucker al Problema 4.2.

Capítulo 5

Solución del problema

En el presente capítulo se da solución al problema y se exponen dos proposiciones, una de las cuales establece la existencia de una cota superior para el tiempo óptimo de estudio que resuelve el Problema 4.2, y la otra establece que la asignación de los recursos familiares es económicamente eficiente. La cota antes mencionada, junto con otros parámetros, determinarán distintos modos de asignación eficiente del tiempo del niño. A lo largo de este capítulo se denotarán con asterisco a los valores óptimos de las variables.

Proposición 5.1. *Para cualquier nivel de transferencias Tr fijo el tiempo óptimo para educarse, t^*, que resuelve el Problema 4.2, está acotado por el valor $T - Lr$, determinado por la solución a la ecuación:*

$$U_h'(Lr) = \tau\beta V_h'(\tau f(T-Lr)+Tr)f'(T-Lr). \tag{5.1}$$

Demostración.

El tiempo $T - Lr$ es solución de un problema de optimización más sencillo que el que se pretende resolver, en el cual se considera que la única variable de decisión es el tiempo de ocio del niño y que éste sólo distribuye su tiempo entre educación y ocio. De esta manera, las decisiones de inversión en capital humano son dependientes de las decisiones de ocio y, por el contrario, independientes de las decisiones de consumo. Esto puede interpretarse de la siguiente forma: el padre sólo está interesado en saber el mejor uso del tiempo de su hijo de tal forma que maximice su felicidad en ambos periodos; por lo tanto, C_1 y C_2 no serán variables de estudio y las transferencias pasarán a ser uno de los parámetros del problema.

Al aplicar estos supuestos para representar las decisiones de asignación del tiempo de un niño, los usos óptimos del tiempo de educación formal y ocio están dados por la solución del siguiente problema:

Problema 5.2.

$$\begin{array}{cc} \text{Maximizar} & U(L)+\beta\left[V(\tau f(T-L)+Tr)\right] \\ \text{sujeto a} & 0 \le L \le T \end{array}$$

Las características de las funciones del primer y segundo periodo son las mismas que las que poseen las funciones involucradas en el Problema 4.2, por lo tanto el Problema 5.2 consiste en optimizar una función real estrictamente cóncava creciente y diferenciable en un compacto, por lo que ésta alcanza un máximo. Luego, las cantidades óptimas de ocio, Lr, y de estudio, $T-Lr$, deben satisfacer la siguiente ecuación:

$$U_h'(Lr) = \tau\beta V_h'(\tau f(T-Lr)+Tr)f'(T-Lr).$$

El valor $T-Lr$ puede interpretarse como el tiempo óptimo para educarse por parte de un niño de un hogar que sólo toma en cuenta la utilidad marginal del ocio y el valor marginal que los padres atribuyen a la educación, en otras palabras, entendemos por $T-Lr$ como el tiempo máximo de estudio de un modelo que sólo toma en cuenta la felicidad del menor al consumir ocio y el bienestar de los padres por saber que su hijo percibirá una remuneración por su capital humano. Nos referiremos a las soluciones Lr y $T-Lr$ como los tiempos óptimos de gustos.

Debido a la concavidad estricta de las funciones $U_h(L)$ y $V_h(\tau f(T-L)+Tr)$, sus derivadas, $U_h'(L)$ y $\tau\beta V_h'(\tau f(T-L)+Tr)f'(T-L)$, son funciones estrictamente decrecientes y positivas con respecto a L y $t=T-L$, respectivamente.

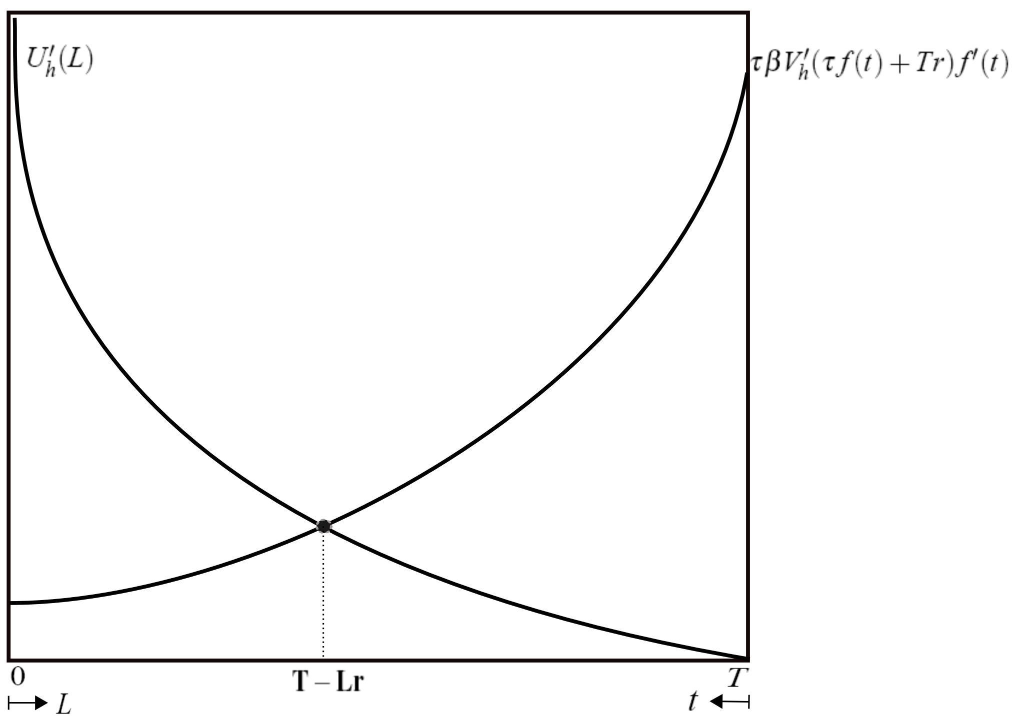

Figura 5.1: Geometría de la solución del Problema 5.2.

La Figura 5.1 ilustra la determinación del valor $T-Lr$. El eje horizontal representa el intervalo $[0,T]$. El ocio se mide de izquierda a derecha, mientras que el tiempo para educarse se mide en sentido opuesto. Las funciones $U_h'(L)$ y $\tau\beta V_h'(\tau f(t)+Tr)f'(t)$ representan la utilidad marginal del ocio del niño y el valor presente del valor subjetivo marginal paterno del tiempo dedicado a educarse por parte del menor, respectivamente. Así, el valor $T-Lr$ queda determinado por la intersección de ambas curvas.

Notemos que para cualquier nivel de transferencias la geometría de la solución es básicamente la misma, ya que al aumentar las transferencias la curva $\tau\beta V_h'(\tau f(t)+Tr)f'(t)$ se desplaza hacia la derecha. Sea $T-Lr_0$ la solución de la ecuación (5.1) para un nivel de transferencias cero, es decir, satisface

$$U_h'(Lr_0)=\tau\beta V_h'(\tau f(T-Lr_0))f'(T-Lr_0); \tag{5.2}$$

entonces $T-Lr_0$ siempre se encontrará a la izquierda de $T-Lr_1$, donde éste último denota la solución de la ecuación (5.1) para un nivel de transferencias $Tr_1>0$ (ver Figura 5.2).

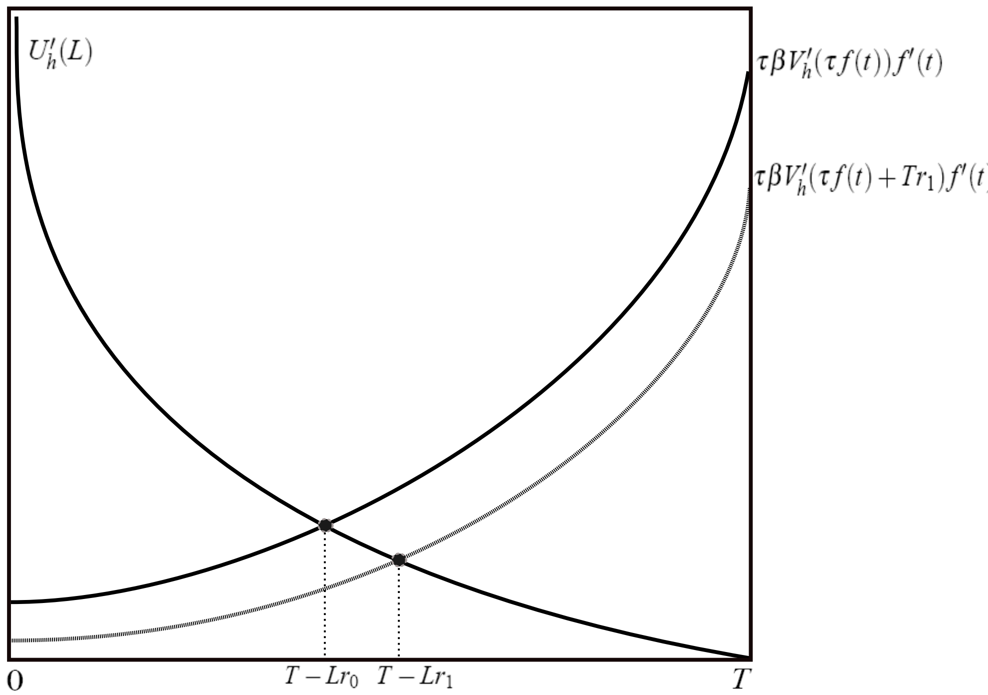

Figura 5.2: Determinación de $T-Lr$ dependiendo del nivel de transferencias.

Es claro que cuando el menor se enfrenta a un problema más complicado, en el que se consideran como variables los consumos familiares en ambos periodos, las transferencias, así como la posibilidad de que el menor dedique parte de su tiempo a trabajar; el tiempo asignado a su educación está acotado por el valor $T-Lr$ asociado a un nivel de transferencias Tr. ∎

Una vez discutido lo anterior es posible explicar cuándo se considerará al tiempo de estudio t_s como factible y cuándo no. Dado que el tiempo de estudio del menor está acotado por dos valores, t_s y $T-Lr$, donde éste último varía de acuerdo al nivel de transferencias, diremos que t_s es factible o no dependiendo de que se cumpla $t_s \leq T-Lr_0$ o $T-Lr_0 < t_s$, respectivamente. El primer caso sucede cuando el periodo de la niñez es suficiente para alcanzar el nivel óptimo económico t_s. Por el contrario, si la niñez cubre un periodo muy corto, es probable que t_s

exceda dicho periodo. Este fenómeno ocurre cuando los niños se casan a edades tempranas o se incorporan de tiempo completo al mercado laboral, abandonando la escuela. Tal situación es más frecuente en las áreas rurales de los países pobres o en zonas indígenas. Evidentemente la factibilidad de t_s depende de diversos factores económicos, tales como: el precio de una unidad de capital humano, la valoración que los padres dan a la educación, la productividad de la función de capital humano y el costo de oportunidad del niño.

Retomando el problema general enfrentado por el padre, observemos que dar solución al Problema 4.2 es lo mismo que resolver el siguiente:

Problema 5.3.

$$\begin{array}{lll} \text{Minimizar} & -\mathscr{P}(C_1,L,C_2,t,Tr) & \\ \text{sujeto a} & g_i(C_1,L,C_2,t,Tr) \leq 0 & \text{para } i=1,\ldots,5. \end{array}$$

Donde

$$\begin{aligned}
\mathscr{P}(C_1,L,C_2,t,Tr) &= U_p(C_1)+U_h(L)+\beta\,[V_p(C_2)+V_h(\tau f(t)+Tr)] \\
g_1(C_1,L,C_2,t,Tr) &= p_1C_1+\frac{(p_2C_2+Tr)}{1+r}-W-\omega(T-L-t) \\
g_2(C_1,L,C_2,t,Tr) &= t+L-T \\
g_3(C_1,L,C_2,t,Tr) &= -t \\
g_4(C_1,L,C_2,t,Tr) &= -Tr \\
g_5(C_1,L,C_2,t,Tr) &= t-t_s
\end{aligned}$$

El objetivo es resolver una parte del Problema 5.3 cuando se considera que el nivel óptimo económico de estudio no es factible, pero antes será necesario obtener las ecuaciones que resultan de aplicar el Teorema de Kuhn-Tucker a dicho problema.

Primero verificaremos que la función objetivo, $-\mathscr{P}$, y las restricciones, g_i's, cumplen las condiciones para aplicar el Teorema 2.21. Estas condiciones exigen que $-\mathscr{P}$ sea pseudoconvexa y diferenciable en el óptimo $\mathbf{x}_0$, que las g_i's sean cuasiconvexas y diferenciables en $\mathbf{x}_0$ para todo $i \in I$, donde $I = \{i \mid g_i(\mathbf{x}_0)=0\}$ y que la cualificación de restricción $T=G'$ se cumpla. La función objetivo $\mathscr{P}(C_1,L,C_2,t,Tr)$ es estrictamente cóncava creciente y diferenciable, luego $-\mathscr{P}$ es estrictamente convexa y por lo tanto pseudoconvexa. Ya que las g_i's son lineales, entonces son diferenciables y convexas, y por lo tanto cuasiconvexas; además, por el Lema 2.25, podemos asegurar que éstas cumplen con la cualificación de restricciones $T=G'$.

Ya que hemos visto que se satisfacen las condiciones, el Teorema 2.21 establece que si el punto $\mathbf{x}_0 = (C_1^*,L^*,C_2^*,t^*,Tr^*)$ es una solución óptima local del Problema 5.3, entonces existen escalares no negativos λ_i para $i \in I$ tales que:

$$
\begin{aligned}
\nabla(-\mathscr{P}(\mathbf{x}_0)) + \sum_{i=1}^{5} \lambda_i \nabla g_i(\mathbf{x}_0) &= \mathbf{0} \\
\lambda_i g_i(\mathbf{x}_0) &= 0 \qquad \text{para } i = 1, \ldots, 5 \\
\lambda_i &\geq 0 \qquad \text{para } i = 1, \ldots, 5.
\end{aligned} \tag{5.3}
$$

Observemos que la primera expresión del sistema (5.3) es el gradiente del siguiente Lagrangiano igualado a cero:

$$
\begin{aligned}
\mathscr{L} = \quad & -(U_p(C_1) + U_h(L) + \beta\,[V_p(C_2) + V_h(\tau f(t) + Tr)]) \\
& + \lambda_1(p_1 C_1 + \frac{(p_2 C_2 + Tr)}{1+r} - W - \omega(T - L - t)) + \lambda_2(t + L - T) \\
& + \lambda_3(-t) + \lambda_4(-Tr) + \lambda_5(t - t_s).
\end{aligned}
$$

Del sistema (5.3) se derivan las siguientes ecuaciones, las cuales son las condiciones de Kuhn-Tucker del Problema 5.3.

$$U'_p(C_1) = \lambda_1 P_1 \tag{5.4}$$

$$U'_h(L) = \lambda_1 \omega + \lambda_2 \tag{5.5}$$

$$\beta V'_p(C_2) = \lambda_1 \frac{P_2}{1+r} \tag{5.6}$$

$$\tau \beta V'_h(\tau f(t) + Tr) f'(t) = \lambda_1 \omega + \lambda_2 - \lambda_3 + \lambda_5 \tag{5.7}$$

$$\beta V'_h(\tau f(t) + Tr) = \frac{\lambda_1}{1+r} - \lambda_4 \tag{5.8}$$

$$\lambda_1 \left(W + \omega(T - L - t) - P_1 C_1 - \frac{P_2 C_2 + Tr}{1+r} \right) = 0 \tag{5.9}$$

$$\lambda_2(T - L - t) = 0 \tag{5.10}$$

$$\lambda_3 t = 0 \tag{5.11}$$

$$\lambda_4 Tr = 0 \tag{5.12}$$

$$\lambda_5(t_s - t) = 0 \tag{5.13}$$

Robles (2000) obtiene las siguientes soluciones para el caso en que t_s no es factible mediante un análisis microeconómico.

"***Proposición 4***. Si t_s no es factible, la solución del sistema (5.4)-(5.13) es compatible con cuatro distintos modos en que se asignan el tiempo del menor y la transferencia de activos no humanos. Estos cuatro modos se determinan por tres valores críticos del ingreso no laboral del menor, W_1, W_2 y W_3 (con $0 < W_1 < W_2 < W_3$), tal que:

i) el tiempo del niño es asignado a trabajar y al ocio. Las transferencias de otros activos no humanos es nula ($t^* = 0$ y $Tr^* = 0$) para $0 < W \leq W_1$;

ii) el tiempo del niño es dedicado a ocio, participación laboral y educación. Las transferencias en otros activos es cero ($t^* > 0$, $R^* = T - L^* - t^* > 0$ y $Tr^* = 0$) para $W_1 < W < W_2$;

iii) el tiempo del menor es asignado a las cotas superiores de ocio y educación, y las transferencias no humanas son cero ($t^* = T - Lr_0$, $L^* = Lr_0$ y $Tr^* = 0$) para $W_2 \leq W \leq W_3$;

iv) el tiempo del menor es asignado a las cotas superiores de ocio y educación; las transferencias de otros activos son estrictamente positivas ($t^* = T - Lr$, $L^* = Lr$ y $Tr^* > 0$) para $W_3 < W$."

Las primeras dos soluciones son explicadas en Acosta (2008). Dentro de este trabajo nos enfocaremos en el análisis detallado, así como en la representación gráfica de las últimas dos soluciones. Utilizando el resultado establecido en el Teorema 2.28 asumiremos que las soluciones son funciones continuas respecto a los parámetros.

Antes de discutir la solución del problema, probaremos la siguiente proposición, la cual establece que la asignación de los recursos familiares entre la inversión del capital humano del menor y otras transferencias de riqueza, es económicamente eficiente. En otras palabras, la posibilidad de que el nivel de transferencias óptimo sea positivo sólo se da cuando el menor haya alcanzado la cota superior del tiempo de estudio, $T - Lr_0$. De acuerdo con esta hipótesis, el padre persigue invertir en la formación de capital humano del niño antes de transferirle algún otro tipo de riqueza.

Proposición 5.4. *Si t_s no es factible y el nivel óptimo de transferencias, Tr^*, es positivo, entonces el tiempo óptimo de estudio es igual a su cota superior, $t^* = T - Lr$.*

Demostración. Por hipótesis $Tr^* > 0$ y t_s no es factible, por lo que se satisface $\lambda_4^* = 0$ y $\lambda_5^* = 0$.

De la ecuación (5.8) se tiene que

$$(1+r)\beta V_h'(\tau f(t) + Tr) = \lambda_1,$$

multiplicando por ω tenemos

$$\omega(1+r)\beta V_h'(\tau f(t) + Tr) = \lambda_1 \omega,$$

y considerando la ecuación (4.8) se cumple

$$\tau\beta V_h'(\tau f(t) + Tr) f'(t_s) = \lambda_1 \omega. \tag{5.14}$$

Primero demostraremos que $t^* > 0$. Supongamos, por contradicción, que $t = 0$. Por la restricción de los usos del tiempo del menor tendríamos que $T - L \geq 0$, pero probaremos que esta última desigualdad implica una contradicción.

Supongamos que $T-L>0$, por la condición (5.10) se tendría que $\lambda_2=0$. Notemos que

$$\lambda_1\omega-\lambda_3\leq\lambda_1\omega,$$

entonces considerando las ecuaciones (5.7) y (5.14) se sigue que

$$\tau\beta V_h'(Tr)f'(0)\leq\tau\beta V_h'(Tr)f'(t_s),$$

lo que implica

$$f'(0)\leq f'(t_s);$$

pero como $t_s>0$ y f' es una función estrictamente decreciente se tiene $f'(t_s)<f'(0)$, llegando a una contradicción.

Ahora supongamos que $T-L=0$, por lo que $T=L$. Considerando la ecuación (5.5), las características de λ_1 y λ_2, así como la hipótesis económica impuesta al modelo que establece que el nivel de ocio $L=T$ sacia U (ver ecuación (4.7)), se tiene que el término del lado izquierdo de la ecuación (5.5) es cero, mientras que el otro término es estrictamente positivo, así que

$$U_h'(T)\neq\lambda_1\omega+\lambda_2;$$

por lo que no es posible que $T-L$ sea igual a cero.

Por lo tanto podemos concluir que $t^*>0$ y por la condición (5.11), el valor de $\lambda_3^*=0$. Ahora probaremos que $t^*=T-Lr$.

Reescribiendo la ecuación (5.7)

$$\tau\beta V_h'(\tau f(t)+Tr)f'(t)-\lambda_1\omega=\lambda_2, \tag{5.15}$$

y substituyendo (5.14) en (5.15) se tiene

$$\begin{aligned}\tau\beta V_h'(\tau f(t)+Tr)f'(t)-\tau\beta V_h'(\tau f(t)+Tr)f'(t_s)&=\lambda_2\\ \tau\beta V_h'(\tau f(t)+Tr)[f'(t)-f'(t_s)]&=\lambda_2.\end{aligned}$$

Probaremos que λ_2 no puede ser cero, pues si fuese así se tendría

$$\tau\beta V_h'(\tau f(t)+Tr)[f'(t)-f'(t_s)]=0,$$

lo que implica que

$$f'(t)=f'(t_s),$$

es decir, que t tendría que ser igual a t_s, pero esto no es posible ya que t_s no es factible.

Por lo tanto $\lambda_2^*>0$ y por la condición (5.10) se tiene que $T-L=t$, es decir, $t^*=T-Lr$ para un nivel de transferencias $Tr^*>0$. ■

Antes que nada, es conveniente explorar la geometría involucrada en la asignación del tiempo del menor. Consideremos el valor $\lambda_1^*\omega$, a la variable λ_1^* se le llama utilidad marginal del dinero[1], esta variable representa el incremento del bienestar de los padres a lo largo de su ciclo de vida si el hogar dispone de un peso adicional. De esta forma, $\lambda_1^*\omega$ puede ser interpretado como el valor paterno de la tasa salarial del menor o, dicho de otra forma, como la valuación paterna de una unidad de tiempo del menor. Al graficar dicho valor junto con la geometría de la solución del Problema 5.2, ilustrado en la Figura 5.3, se observa que la recta $\lambda_1^*\omega$ interseca las gráficas de las funciones $U_h'(L)$ y $\tau\beta V_h'(\tau f(t)+Tr)f'(t)$. Esta geometría es suficiente para determinar la asignación del tiempo del menor.

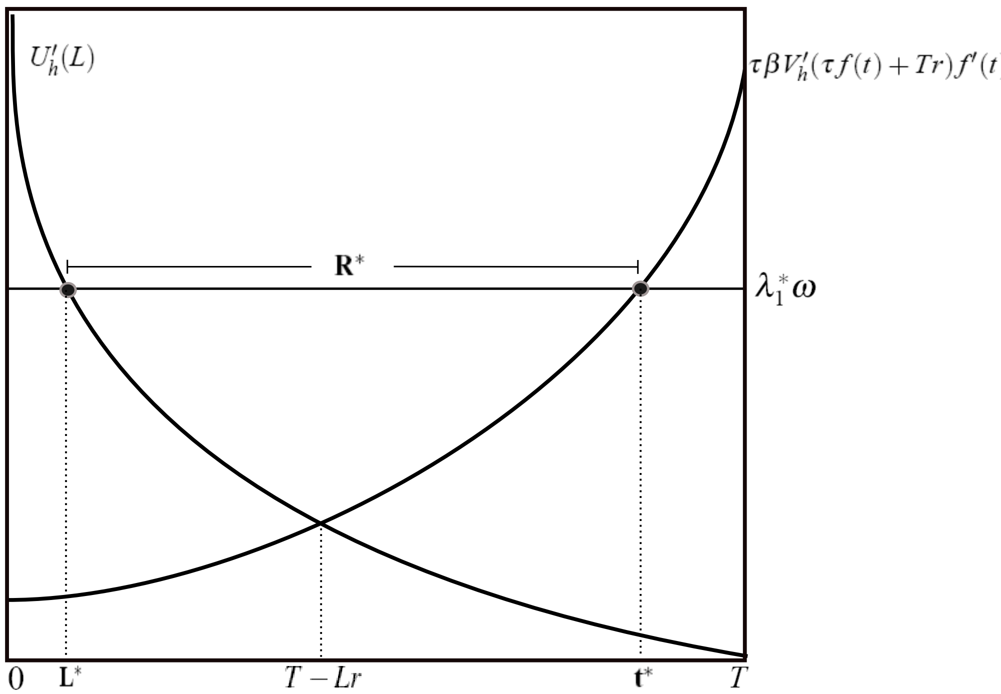

Figura 5.3: Asignación del tiempo del menor.

Para el caso ilustrado en la Figura 5.3, los padres asignan el tiempo del niño a todos sus usos posibles: ocio, trabajo y educación. El tiempo óptimo de estudio, t^*, se determina por las itersecciones de $\lambda_1^*\omega$ y $\tau\beta V_h'(\tau f(t)+Tr)f'(t)$; el ocio óptimo del niño está dado por L^*, que se determina por las intersecciones de $\lambda_1^*\omega$ y $U_h'(L)$. La oferta laboral del niño está dada por el complemento de esas cantidades, $R^* = T - L^* - t^*$, que en la Figura 5.3 está determinado por el tiempo entre L^* y t^* y denotada por R^*.

Naturalmente, la asignación del tiempo del menor está en función de la riqueza familiar, ya que el valor de λ_1^* depende de dicho nivel de riqueza; esto se tiene ya que al aumentar sin límite la riqueza familiar, la utilidad marginal del dinero decrece continuamente a cero, lo cual se

[1]Para mayores detalles consultar el Anexo.

puede deducir a partir de la ecuación (5.4). Si la riqueza crece sin límite, es claro que el consumo aumenta de la misma forma, por lo que la utilidad marginal del consumo tiende a cero. Así, en la ecuación (5.4), al disminuir la utilidad marginal del consumo y al ser P_1 fijo, λ_1^* tiene que decrecer. Con esto se puede observar que λ_1 es una función continua estrictamente decreciente con respecto a W, así que $\lambda_1(W)$ denotará el valor de λ_1 asociado a una familia que posee un nivel de riqueza W. Dicho valor satisface

$$\lim_{W\to 0^+} \lambda_1(W) = +\infty \qquad \text{y} \qquad \lim_{W\to +\infty} \lambda_1(W) = 0^+.$$

Esto implica que la valuación paterna de la tasa salarial del menor tienda a infinito cuando W tienda a cero y que tienda a cero cuando W crezca sin límite, es decir,

$$\lim_{W\to 0^+} \lambda_1(W)\omega = +\infty \qquad \text{y} \qquad \lim_{W\to +\infty} \lambda_1(W)\omega = 0^+.$$

Por otra parte, notemos que debido a la no factibilidad del tiempo de estudio t_s se tiene que en el óptimo siempre se cumple $\lambda_5^* = 0$, por las condiciones de Kuhn-Tucker. Por otra parte, al considerar la ecuación (5.4), se puede concluir que λ_1^* siempre es positivo ya que P_1 es fijo y, por hipótesis, $U'(C_1)$ es positiva para cualquier nivel de consumo. De esta forma, de la condición (5.9), se deduce que la restricción presupuestal se cumple con igualdad; que es lo que se esperaba, ya que por el supuesto de insaciabilidad local se tiene que el individuo tiende a gastar todo su ingreso disponible en los bienes y servicios que desea.

Definimos W_1, W_2 y W_3 como los niveles de riqueza tales que:

$$\lambda_1(W_1) = \frac{\tau\beta V_h'(0) f'(0)}{\omega} \tag{5.16}$$

$$\lambda_1(W_2) = \frac{\tau\beta V_h'(\tau f(T-Lr_0)) f'(T-Lr_0)}{\omega} \tag{5.17}$$

$$\lambda_1(W_3) = (1+r)\beta V_h'(\tau f(T-Lr_0)) \tag{5.18}$$

A continuación obtendremos la tercera solución para el caso en el que t_s no es factible.

$\boxed{\textbf{Solución 3}}$

El tiempo del menor es asignado a las cotas superiores de ocio y educación, y las transferencias no humanas son cero ($t^* = T - Lr_0$*,* $L^* = Lr_0$*,* $R^* = 0$ *y* $Tr^* = 0$*) para* $W_2 \le W \le W_3$*.*

Enseguida se realiza el análisis correspondiente para verificar que esta es una asignación óptima del tiempo del menor para niveles de riqueza W tales que $W_2 \le W \le W_3$.

Como λ_1 es una función estrictamente decreciente con respecto a la riqueza, se tiene

$$\lambda_1(W_3) \leq \lambda_1(W) \leq \lambda_1(W_2),$$

multiplicando por ω

$$\lambda_1(W_3)\omega \leq \lambda_1(W)\omega \leq \lambda_1(W_2)\omega. \tag{5.19}$$

Substituyendo (5.17) y (5.18) en (5.19)

$$\omega(1+r)\beta V_h'(\tau f(T-Lr_0)) \leq \lambda_1(W)\omega \leq \tau\beta V_h'(\tau f(T-Lr_0))f'(T-Lr_0). \tag{5.20}$$

Sabemos que el tiempo óptimo de estudio es positivo, ya que en la segunda solución propuesta por Robles (2000) el tiempo óptimo de estudio se encuentra en el intervalo $(0, T-Lr_0)$,[2] y debido a la continuidad de las soluciones, podemos descartar la posibilidad de que t^* sea cero. De esta forma, por la condición (5.11), $\lambda_3^* = 0$.

Demostraremos que $t^* = T - Lr_0$ suponiendo por contradicción que $t < T - Lr_0$. Luego, como V_h' y f' son funciones estrictamente decrecientes y, por la Proposición 5.4, $Tr = 0$, se tiene

$$\tau\beta V_h'(\tau f(T-Lr_0))f'(T-Lr_0) < \tau\beta V_h'(\tau f(t))f'(t). \tag{5.21}$$

Despejando $\lambda_1(W)\omega$ de (5.7), en el óptimo se debe satisfacer

$$\tau\beta V_h'(\tau f(t))f'(t) - \lambda_2 = \lambda_1(W)\omega,$$

substituyendo la igualdad anterior en la segunda desigualdad de (5.20) se tiene

$$\tau\beta V_h'(\tau f(t))f'(t) - \lambda_2 \leq \tau\beta V_h'(\tau f(T-Lr_0))f'(T-Lr_0)$$

y por (5.21)

$$\tau\beta V_h'(\tau f(T-Lr_0))f'(T-Lr_0) - \lambda_2 < \tau\beta V_h'(\tau f(T-Lr_0))f'(T-Lr_0).$$

Esto implica que λ_2 tendría que ser estrictamente positivo, pero si esto sucede, por la condición (5.10), se cumpliría $R = 0$, es decir, $t = T - L$. Ya que el nivel de transferencias es nulo, entonces $t = T - Lr_0$, pero t no puede tomar este valor porque por hipótesis $t < T - Lr_0$, llegando así a una contradicción.

Por lo tanto, $t^* = T - Lr_0$ y $L^* = Lr_0$, de esta forma $R^* = 0$ y, por la condición (5.10), $\lambda_2^* \geq 0$.

Ahora demostraremos que $Tr^* = 0$. Al igual que antes procedemos por contradicción suponiendo que $Tr > 0$, así, por (5.12), se tendría que $\lambda_4 = 0$.

Multiplicando la ecuación (5.8) por $\omega(1+r)$ obtenemos

$$\omega(1+r)\beta V_h'(\tau f(T-Lr_0) + Tr) = \lambda_1(W)\omega,$$

[2] Dicha solución es analizada en la Tesis "La asignación del tiempo de un niño a educación y trabajo: un análisis microeconómico I" realizada por Berencie Acosta Rivera.

substituyendo esta última igualdad en la primera desigualdad de (5.20) se tiene

$$\omega(1+r)\beta V_h'(\tau f(T-Lr_0)) \leq \omega(1+r)\beta V_h'(\tau f(T-Lr_0)+Tr) \tag{5.22}$$
$$V_h'(\tau f(T-Lr_0)) \leq V_h'(\tau f(T-Lr_0)+Tr) \tag{5.23}$$

pero como $\tau f(T-Lr_0) < \tau f(T-Lr_0)+Tr$ y V_h' es una función estrictamente decreciente se debe cumplir que

$$V_h'(\tau f(T-Lr_0)+Tr) < V_h'(\tau f(T-Lr_0)),$$

contradiciendo la desigualdad (5.23). Por lo tanto el nivel óptimo de transferencias es nulo, $Tr^* = 0$ y $\lambda_4^* \geq 0$.

La geometría de esta solución está representada en la Figura 5.4.

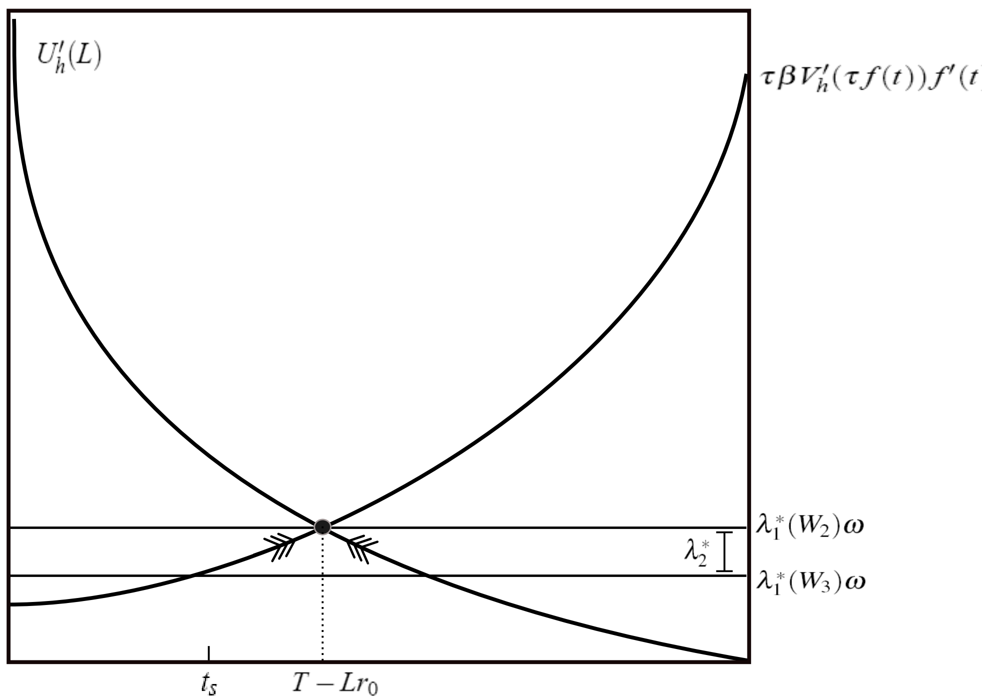

Figura 5.4: Geometría de la solución para niveles de riqueza $W \in [W_2, W_3]$.

En seguida se discutirá el análisis de la última solución propuesta por Robles (2000).

Solución 4

El tiempo del menor es asignado a las cotas superiores de ocio y educación; las transferencias de otros activos son estrictamente positivas ($t^* = T-Lr$, $L^* = Lr$, $R^* = 0$ y $Tr^* > 0$) para $W_3 < W$.

En lo que sigue, nos ocuparemos de demostrar que esta es una asignación óptima del tiempo del menor para niveles de riqueza W tales que $W_3 < W$.

Como λ_1 es una función estrictamente decreciente con respecto a la riqueza, se tiene

$$\lambda_1(W) < \lambda_1(W_3),$$

multiplicando por ω

$$\lambda_1(W)\omega < \lambda_1(W_3)\omega, \tag{5.24}$$

substituyendo (5.18) en (5.24) tenemos

$$\lambda_1(W)\omega < \beta\omega(1+r)V_h'(\tau f(T-Lr_0)). \tag{5.25}$$

Primero demostraremos que el nivel óptimo de transferencias es positivo, $Tr^* > 0$. De la ecuación (5.8) se tiene

$$\beta V_h'(\tau f(T-Lr_0)+Tr)+\lambda_4 = \frac{\lambda_1(W)}{1+r},$$

multiplicando por $\omega(1+r)$ tenemos

$$\omega(1+r)\beta V_h'(\tau f(T-Lr_0)+Tr)+\lambda_4\omega(1+r) = \lambda_1(W)\omega.$$

Sustituyendo la igualdad anterior en (5.25) se obtiene

$$\omega(1+r)\beta V_h'(\tau f(T-Lr_0)+Tr)+\lambda_4\omega(1+r) < \omega(1+r)\beta V_h'(\tau f(T-Lr_0)),$$

pero notemos que si las transferencias fueran cero se tendría

$$\omega(1+r)\beta V_h'(\tau f(T-Lr_0))+\lambda_4\omega(1+r) < \omega(1+r)\beta V_h'(\tau f(T-Lr_0)),$$

pero λ_4 es no negativo, llegando a una contradicción. Por lo tanto, $Tr^* > 0$ y $\lambda_4^* = 0$.

Observemos que como $Tr^* > 0$, la curva $\tau\beta V_h'(\tau f(t))+Tr)f'(t)$ se desplaza hacia la izquierda. Este desplazamiento provoca que las curvas $U_h'(L)$ y $\tau\beta V_h'(\tau f(t))+Tr)f'(t)$ determinen un nuevo punto de intersección, $T-Lr$; el cual es la nueva cota superior del tiempo de estudio del menor.

Con respecto al nivel óptimo de trabajo, R, demostraremos que éste es cero. Para verificar lo anterior procederemos por contradicción, supongamos que el menor trabaja; entonces, por las condiciones de Kuhn-Tucker, el valor óptimo de λ_2 es cero. De esta manera, de la ecuación (5.5) se sigue que

$$U_h'(L) = \lambda_1\omega,$$

recordemos que para niveles de riqueza suficientemente grandes, el valor paterno de la tasa salarial del menor decrece continuamente a cero, provocando que $U_h'(L)$ decrezca de la misma forma.

Por otra parte, una de las hipótesis económicas del modelo establece que existe un nivel de ocio $L = T$ el cual sacia U_h (ver ecuación (4.7)), por lo que se tiene que $U_h'(L)$ tiende a cero si y sólo si L tiende a T; pero esto no es posible debido a la hipótesis de que el menor trabaja, llegando a una contradicción. Por lo tanto $R^* = 0$, es decir, $t^* = T - L^*$, el menor sólo distribuye su tiempo entre educación y ocio.

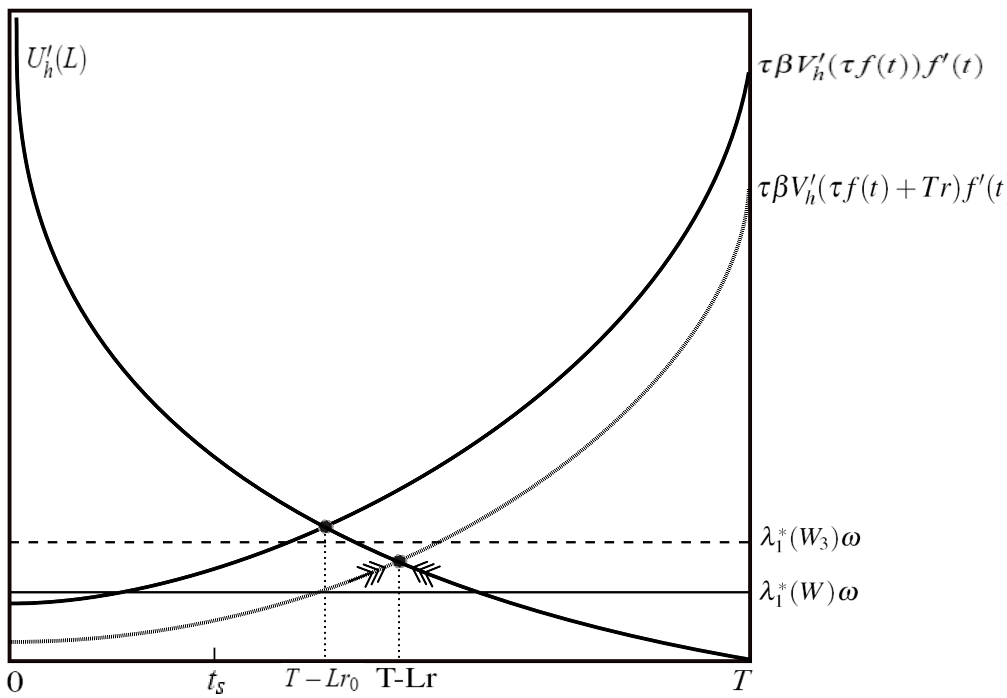

Figura 5.5: Representación de la solución para niveles de riqueza familiar W tales que $W_3 < W$.

Por último, probaremos que los tiempo óptimos de ocio y educación están dados por las nuevas cotas superiores de los mismos, determinadas por la intersección de las curvas $U_h'(L)$ y $\tau\beta V_h'(\tau f(t)) + Tr)f'(t)$ con un nivel óptimo de transferencias positivo, es decir, $L^* = Lr$ y $t^* = T - Lr$. Para probar lo anterior supongamos, por contradicción, que $L < Lr$. Esto implica que

$$T - Lr < T - L,$$

es decir,

$$T - Lr < t$$

pero esto no es posible ya que $T - Lr$ es cota superior del tiempo de estudio del menor. De esta forma, $L^* = Lr$ y como ya probamos que el menor no trabaja, $t^* = T - Lr$.

La Figura 5.5 ilustra la geometría de esta solución. Se puede apreciar que la valuación paterna de una unidad de tiempo del menor, $\lambda_1^*(W)\omega$, se encuentra por debajo del punto de intersección de ambas curvas; pues se cumple que

$$U_h'(Lr^*) = \tau\beta V_h'(\tau f(T-Lr^*)+Tr^*)f'(T-Lr^*) \tag{5.26}$$
$$= \tau\frac{\lambda_1^*}{1+r}f'(T-Lr^*) \tag{5.27}$$
$$> \tau\frac{\lambda_1^*}{1+r}f'(t_s). \tag{5.28}$$

La igualdad (5.26) se tiene por las ecuaciones (5.5) y (5.7) evaluadas en el óptimo. Sustituyendo la ecuación (5.8) evaluada en el óptimo en el segundo término de la primera igualdad, obtenemos (5.27). Por último, la desigualdad (5.28) se debe a que $T-Lr^* < t_s$. Pero $\tau f'(t_s)$ es igual a $\omega(1+r)$, pues t_s es solución del Problema 4.1. Obteniendo

$$\lambda_1^*(W)\omega < U_h'(Lr^*).$$

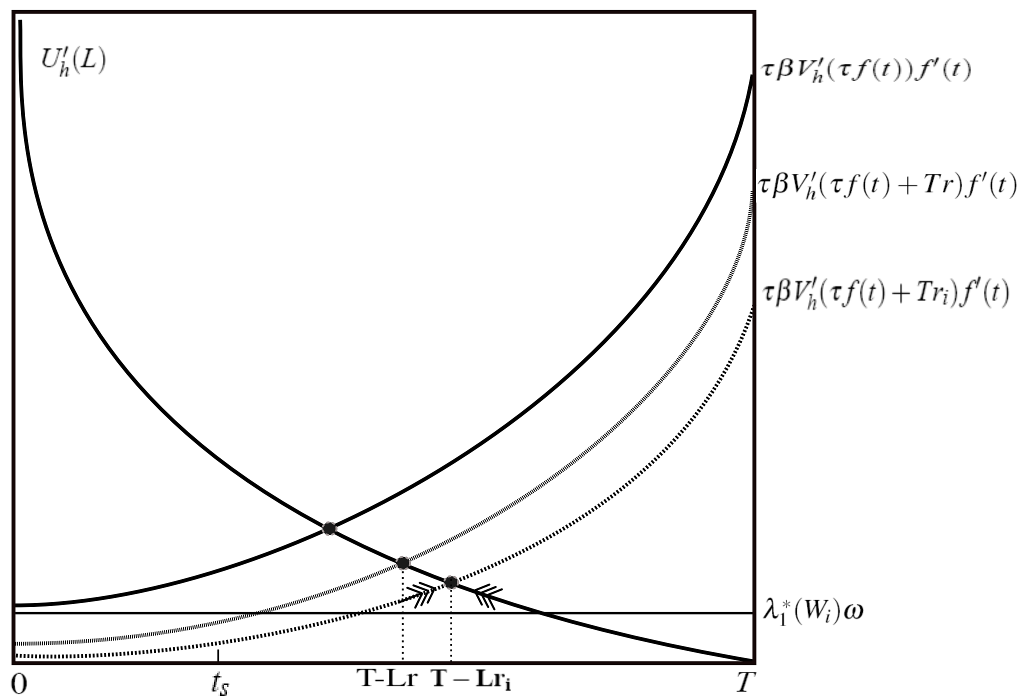

Figura 5.6: Disminución del tiempo óptimo de estudio al aumentar las transferencias.

Es primordial señalar que conforme la riqueza familiar aumente continuamente a partir del nivel W_3, las transferencias aumentarán y el tiempo óptimo de estudio disminuirá, ambos continuamente. Asimismo, el nivel óptimo de educación se alcanzará en la cota superior de gustos, $T-Lr_i$, determinada por un nivel de transferencias $Tr_i^* > 0$ (ver Figura 5.6). Además $\lambda_1^*\omega$ también se encuentra por debajo del punto de intersección de ambas curvas, pues en general se satisface que $\lambda_1^*(W_i)\omega < U_h'(Lr_i^*)$, donde W_i es un nivel de riqueza familiar tal que $W_3 < W_i$.

Capítulo 6

Conclusiones

Se distinguen dos tipos de conclusiones. Las primeras que se presentan están ligadas directamente al objeto de esta tesis, las segundas son de carácter general.

El trabajo se complementa con el trabajo de Acosta (2008) para discutir con mayor detalle las soluciones de uno de los dos casos del modelo de asignación del tiempo de los niños propuesto por Robles (2000). Este modelo plantea básicamente que el fenómeno del trabajo infantil puede ser explicado desde una perspectiva económica sin considerar comportamientos malintencionados por parte de padres que busquen su propio beneficio a costa de sus hijos. Basta suponer que el tiempo de los niños es un activo económico para la familia y que las decisiones sobre la asignación de su tiempo disponible a educación o al trabajo dependen de la riqueza familiar, de parámetros económicos y del altruismo de los padres.

Tanto el trabajo de Acosta (2008) como el presente, discuten el caso en que el periodo de la infancia de los niños es relativamente corto, de modo que no permite alcanzar las soluciones óptimas predichas por el modelo básico de inversión en capital humano. Esto no resta generalidad a los resultados, pues las soluciones del caso faltante se obtienen de manera similar.

Mientras Acosta (2008) discute las primeras dos soluciones, cuando el nivel de riqueza familiar es suficientemente bajo de tal forma que existe el trabajo infantil, en esta tesis se considera que dicha riqueza es suficientemente grande como para excluir esta posibilidad, permitiendo alcanzar dos posibles soluciones. En la primera, los niveles de riqueza son tales que el menor no trabaja y alcanza una de sus cotas superiores del tiempo de estudio. En esta solución la única transmisión de riqueza de padres a hijos es mediante la inversión en capital humano. En la segunda, los niveles de riqueza son tales que el menor no trabaja, estudia lo máximo posible y los padres heredan al hijo otros activos físicos, además de la inversión en capital humano. Cabe destacar que en esta última solución, de manera paradójica, el menor disminuye su tiempo de estudio mientras que simultáneamente incrementa el tiempo dedicado al ocio. Además, el nivel óptimo de transferencias es creciente conforme el nivel de riqueza familiar se incrementa.

En el modelo analizado se tiene que el nivel de riqueza familiar es uno de los principales determinantes de la asignación del tiempo de un niño. En esta línea, estudios recientes muestran que más alla de las características de las escuelas, el gasto destinado por parte del gobierno a la educación, la preparación con que cuenten los maestros o el tiempo que un niño pase en un salón de clases; la familia y el entorno sociocultural en el que se desenvuelva el menor son los principales factores que determinan el nivel de conocimiento que puede adquirir un idividuo. Las características sociodemográficas de la familia, como el número de individuos que integran el núcleo familiar, el nivel educativo y el ingreso de los padres; forman parte del llamado entorno familiar.

Entre lo futuros desarrollos de este trabajo se contempla actualizar el análisis econométrico que ofrece Robles (2000).

En cuanto a las conclusiones generales, se puede mencionar que en este trabajo se incluyen los prerrequisitos teóricos de economía y matemáticas necesarios para resolver modelos microeconómicos utilizando técnicas de optimización con restricciones. Durante el proceso de elaboración de esta tesis, fue necesario hacer una revisión de la teoría antes mencionada debido a que el programa de estudios de la Licenciatura en Matemáticas Aplicadas de la Universidad Autónoma del Estado de Hidalgo, no contempla algunos de estos temas. Por lo que la elaboración de este trabajo permitió ampliar la formación curricular que ofrece dicha licenciatura, especialmente en las partes de microeconomía y optimización no lineal.

Esta tesis muestra que es posible realizar investigación con las herramientas que provee la Licenciatura en Matemáticas Aplicadas integrando equipos multidisciplinarios, donde alumnos de los últimos semestres participen en grupos lidereados por investigadores de distintas disciplinas.

Anexo

Función de producción de capital humano

Siguiendo a Ghez y Becker (1975), la función de producción de capital humano, representada por $H(t)$, tiene la siguiente forma aditiva.

Sea la función $H : \mathbb{R} \to \mathbb{R}$ definida como sigue:

$$H(t) = f(t; g, \varepsilon) + (1 - \delta)H_1$$

donde

$t :=$ tiempo que el menor dedica a su educación formal.

Parámetros:

$g :=$ gastos gubernamentales en educación que incrementan el capital humano del menor.

$\varepsilon :=$ dotaciones genéticas y culturales del menor.

$H_1 :=$ dotación de capital humano adquirido por el menor.

$\delta :=$ depreciación del acervo de capital humano.

Podemos comprender la forma de $H(t)$ a través del siguiente ejemplo. Imaginemos que la función de producción de capital humano de un niño está dada por $f(t; g, \varepsilon)$, donde t es el tiempo de estudio durante su educación secundaria. Obviamente la cantidad de capital humano del menor depende de los conocimientos que adquirió durante su formación primaria; a lo que llamamos H_1. Pero es muy probable que el menor no recuerde todos los conocimientos que aprendió, es decir, su capital humano sufrió una depreciación; dicha depreciación se puede representar como $(1 - \delta)H_1$. De esta forma, $H(t)$ es la suma de los conocimientos adquiridos durante su educación secundaria y primaria, $H = f(t; g, \varepsilon) + (1 - \delta)H_1$.

Como $H(t)$ es una función de producción de capital humano, cumple con el comportamiento general descrito en la Subsección 4.1.2, veremos que $H(t)$ mantiene este comportamiento al variar cada uno de sus parámetros.

Supongamos que el gobierno siempre ha destinado una cantidad g para invertir en educación, pero esta ocasión ese presupuesto aumentará a g_1. De esta forma, se tendrá la oportunidad de enriquecer el acervo bibliográfico en las bibliotecas públicas, mejorar la infraestructura de las escuelas, dotar de computadoras a un gran número de planteles, impartir buenos cursos para tener maestros mejor capacitados, etcétera. Por lo que la probabilidad de que las clases que reciba el menor sean más productivas es muy alta, así como el que tenga acceso a mayor información para incrementar y enriquecer sus conocimientos. En cambio, si en lugar de que el gobierno hubiese incrementado el presupuesto para educación, hubiera decidido disminuirlo a g_2, posiblemente el menor no tenga la oportunidad de consultar una bibliografía más extensa, ni tenga acceso a una computadora y tal vez las técnicas de enseñanza de sus profesores no sean tan buenas. Entonces, para un determinado tiempo t^*, si el gasto gubernamental en educación es g_1, el capital humano del niño indudablemente será mayor que si el gasto fuera g. Por lo que un aumento de g a g_1 causará una modificación en H de tal forma que la gráfica resultante cumple que $\forall t > 0$, $H'(t;g) < H'(t;g_1)$. Algo semejante ocurre si el gobierno decide disminuir el gasto a g_2. Ahora para un cierto tiempo t^* dedicado al estudio, las posibilidades de que el individuo tenga acceso a libros, a computadoras y a un mejor ambiente para su educación, se reducen innegablemente. Por lo que una disminución de g a g_2 causará una modificación en H, de tal forma que la gráfica resultante cumple que $\forall t > 0$, $H'(t;g_2) < H'(t;g)$ (Figura 6.1).

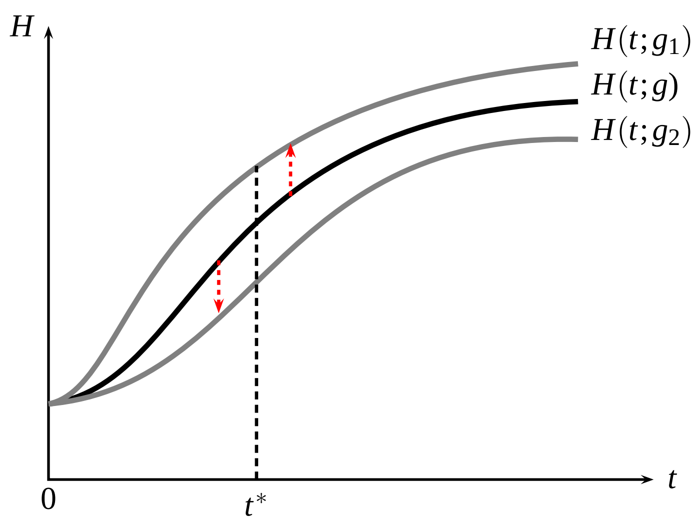

Figura 6.1: Función H cuando $g_2 < g < g_1$.

Los cambios en $H(t)$ al variar el parámetro ε son similares a los descritos anteriormente. Supongamos que cierto niño tiene una dotación genética y cultural ε y su función de producción de capital humano esta dada por $H(t;\varepsilon)$, mientras que otro cuenta con una dotación genética y cultural ε_1, mayor que ε. Sin duda, el niño que cuenta con la dotación ε_1 puede adquirir una mayor cantidad de capital humano, ya que seguramente se desenvuelve en un ambiente más apropiado para el desarrollo de la educación, además de contar con un coeficiente intelectual más elevado, aunado a que muy probablemente se desarrolla dentro de un núcleo familiar en el cual es de gran importancia el hecho de que el menor adquiera educación formal. Así que un aumento (disminución) de ε a ε_1 (ε_2) causará una modificación en H de tal forma que la gráfica resultante cumple que $\forall t > 0$, $H'(t;\varepsilon) < H'(t;\varepsilon_1)$ $(H'(t;\varepsilon_2) < H'(t;\varepsilon))$. Observemos que la variación de este parámetro provoca modificaciones en $H(t)$ similares a las anteriores, representadas en la Figura 6.1.

Para finalizar, analizaremos las variaciones provocadas por δ. Imaginemos que dos niños terminan a la par su educación primaria y uno de ellos inicia inmediatamente su educación secundaria, mientras que el otro se incorpora al mercado laboral y después de un tiempo decide continuar con sus estudios. Supongamos que sus funciones de producción de capital humano son $H(t;\delta) = f(t) + (1-\delta)H_1$ y $H(t;\delta_1) = f(t) + (1-\delta_1)H_1$, respectivamente, donde t es el tiempo dedicado al estudio durante su educación secundaria y H_1 es el capital humano adquirido durante su formación primaria. Naturalmente, la depreciación que sufre el capital humano del primer niño es significativamente menor que la que sufre el niño que ingresa tiempo después a la secundaria, $\delta < \delta_1$. De esta forma, la función $H(t;\delta_1)$ se encontrará por debajo de $H(t;\delta)$, ilustrado en la Figura 6.2.

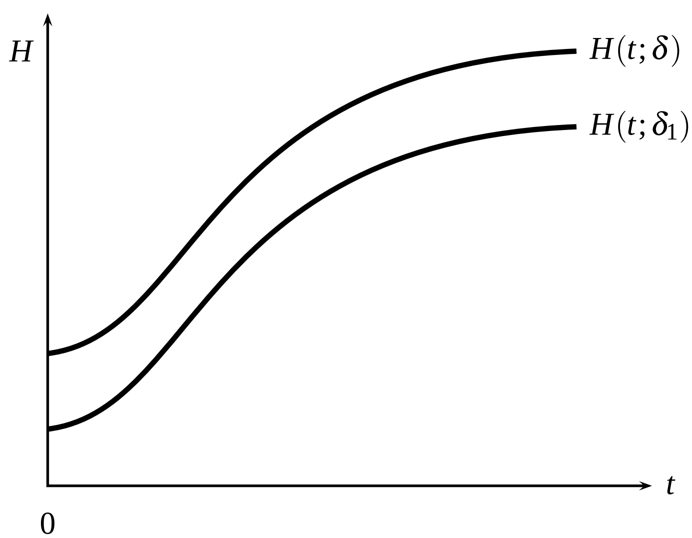

Figura 6.2: Función H cuando $\delta < \delta_1$.

Para el análisis del modelo sólo consideraremos $H(t) = f(t)$, pues debido a que el tiempo de estudio se encuentra dentro del periodo de infancia del menor, $H_1 = 0$, y el último sumando se anula; con respecto a los parámetros g y ε hemos justificado que podemos excluirlos, ya que sólo desplazan a $H(t)$ sin alterar el comportamiento esencial de ésta.

Utilidad marginal del dinero

A la variable λ_1 se le llama utilidad marginal del dinero, pues supongamos que nos enfrentamos al siguiente problema, el cual es el más sencillo de optimización de elección de un consumidor.

Problema 6.1.

$$\begin{array}{cc} \text{Maximizar} & \mathscr{U}(x_1,x_2,\ldots,x_n) \\ \text{sujeto a} & p_1x_1+\ldots+p_nx_n-I=0 \end{array}$$

donde x_i y p_i representan la cantidad y el precio del i-ésimo bien, respectivamente, I el ingreso total y $\mathscr{U}$ una función función de utilidad estrictamente cóncava creciente y diferenciable. Denotemos como $g(x_1,\ldots,x_n)$ a la única restricción del Problema 6.1, aplicando los Teoremas 2.21 y 2.24, si $x_0 = (x_1^*,x_2^*,\ldots,x_n^*)$ es una solución óptima local del Problema 6.1, entonces existen escalares no negativos λ_i para $i\in I$ tales que

$$\begin{aligned} \nabla-\mathscr{U}(x_0)+\lambda\nabla g(x_0)&=0 \\ \lambda g(x_0)&=0 \\ \lambda&\geq 0 \end{aligned} \tag{6.1}$$

Del sistema (6.1) se derivan las siguientes ecuaciones:

$$\begin{aligned} -\mathscr{U}_{x_1}+\lambda p_1&=0 \\ \vdots&=\vdots \\ -\mathscr{U}_{x_n}+\lambda p_n&=0 \\ \lambda(p_1x_1+\ldots+p_nx_n-I)&=0 \end{aligned}$$

El Teorema 2.28 establece que las soluciones son funciones continuas de los parámetros, por lo que en el óptimo

$$\frac{\partial\mathscr{U}(x_0(I,p))}{\partial I}=\mathscr{U}_{x_1}\frac{\partial x_1}{\partial I}+\ldots+\mathscr{U}_{x_n}\frac{\partial x_n}{\partial I}$$

Luego, derivando la restricción parcialmente con respecto al ingreso obtenemos:

$$p_1\partial x_1+\ldots+p_n\partial x_n=\partial I$$

Observemos que del sistema de ecuaciones $\lambda=\dfrac{\mathscr{U}_{x_i}}{p_i}$, por lo tanto

$$\frac{\partial\mathscr{U}(x_0(I,p))}{\partial I}=\lambda p_1\frac{\partial x_1}{\partial I}+\ldots+\lambda p_n\frac{\partial x_n}{\partial I} \tag{6.2}$$

$$=\lambda\left(\frac{p_1\partial x_1+\ldots+p_n\partial x_n}{\partial I}\right) \tag{6.3}$$

$$=\lambda. \tag{6.4}$$

Es decir, la utilidad marginal del dinero es igual a λ la cual representa el incremento del bienestar del consumidor, si éste dispone de un peso adicional. Por lo tanto, en términos de nuestro problema λ_1 representa el incremento del bienestar de los padres a lo largo de su ciclo de vida, si el hogar dispone de un peso adicional.

Bibliografía

[1] ACOSTA, BERENICE (2008). *La Asignación del Tiempo de un Niño a Educación y Trabajo: Un Modelo Microeconómico I*. Tesis de Licenciatura. Universidad Autónoma del Estado de Hidalgo.

[2] APOSTOL, TOM M. (1992). *Calculus Volumen 2*. Reverté.

[3] BAZAARA, M., H. SHERALI Y C. SHETTY (1993). *Nonlinear Programming: Theory and Algorithms*. 2.ed. Nueva York, Wiley.

[4] BECKER, GARY (1993). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. 3.ed. Chicago, The University of Chicago Press.

[5] BERGSTROM, T. (1997). *"A Survey of Theories of the Family"*. En: Rosenzweig, M. y O. Stark, ed., op. cit. vol. IA..

[6] DIEUDONNÉ, J. (1969). *Foundations of Modern Analysis*. Academic Press, Inc..

[7] GHEZ, G. Y G. BECKER (1975). *The Allocation of Time and Goods Over the Life Cycle*. Nueva York, National Bureau of Economic Research.

[8] MAS-COLELL, A., MICHAEL D. WHINSTON Y JERRY R. GREEN (1995). *Microeconomic Theory*. Oxford University Press, Inc..

[9] ROBLES, HÉCTOR V. (2000). *A Microeconomic Analysis of Child Labor Participation and Education: The Case of Mexico, 1984-1996*. Ph. D. Thesis in Agricultural Economics and Demography. The Pennsylvania State University.

[10] ROBLES, HÉCTOR V. (2004). *El Trabajo Infantil en México 1984-2000*. Cuernavaca: UNAM, Centro Regional de Investigaciones Multidisciplinarias.

[11] SPIVAK, MICHAEL (1965). *Calculus on Manifolds*. Wesley Publishing Company.

[12] VARIAN, HAL R. (1992). *Microeconomic Analysis*. W.W. Norton & Company, Inc..