



# **Cálculo II**

## **Notas de Clase**

**Lorena Zogaib**

**Agosto 1, 2016**

# Contenido

|  |            |
|--|------------|
| <b>Contenido</b>   | <b>2</b>   |
| <b>Prólogo</b>   | <b>4</b>   |
| <b>1 El Espacio <math>\mathbb{R}^n</math></b>                              | <b>5</b>   |
| 1.1 Vectores   | 5          |
| 1.2 Curvas paramétricas. Vector tangente a una curva paramétrica           | 29         |
| 1.3 Rectas en el espacio. Segmento de recta                                | 37         |
| 1.4 Planos e hiperplanos   | 43         |
| 1.5 Conjuntos abiertos, cerrados, acotados, compactos, convexos            | 53         |
| <b>2 Funciones de varias variables</b>                                     | <b>69</b>  |
| 2.1 Dominio e imagen. Representación geométrica                            | 69         |
| 2.2 Conjuntos de nivel   | 72         |
| 2.3 Superficies cuadráticas y otras superficies                            | 75         |
| 2.4 Límites y continuidad  | 83         |
| <b>3 Diferenciación</b>  | <b>91</b>  |
| 3.1 Derivadas parciales. Interpretación geométrica                         | 91         |
| 3.2 Diferenciabilidad. Linealización y diferenciales                       | 98         |
| 3.3 Regla de la cadena   | 103        |
| 3.4 Diferenciación implícita   | 107        |
| 3.5 Derivada direccional y vector gradiente. Recta normal y plano tangente | 112        |
| 3.6 Funciones homogéneas. Teorema de Euler                                 | 122        |
| <b>4 Funciones cóncavas y cuasicóncavas</b>                                | <b>131</b> |
| 4.1 Polinomio de Taylor de orden 2. Matriz hessiana                        | 131        |

|   |            |
|---|------------|
| 4.2 Funciones cóncavas y funciones convexas   | 137        |
| 4.3 Funciones cuasicóncavas y funciones cuasiconvexas   | 143        |
| <b>5 Optimización</b>   | <b>154</b> |
| 5.1 Optimización libre. Criterio del Hessiano   | 154        |
| 5.1.1 Condiciones necesarias de primer orden  | 156        |
| 5.1.2 Condiciones suficientes de segundo orden  | 159        |
| 5.2 Optimización con restricciones de igualdad. Multiplicadores de Lagrange                           | 165        |
| 5.2.1 Condiciones necesarias de primer orden. Significado del multiplicador de Lagrange               | 166        |
| 5.2.2 Condiciones suficientes de segundo orden  | 173        |
| 5.2.3 El caso multidimensional  | 175        |
| 5.2.4 Cualificación de las restricciones: ¿cuándo falla el método de los multiplicadores de Lagrange? | 178        |
| 5.3 Optimización con restricciones de desigualdad. Condiciones de Kuhn-Tucker                         | 182        |
| 5.3.1 Problemas de maximización   | 183        |
| 5.3.2 Problemas de minimización   | 204        |
| 5.3.3 Cualificación de las restricciones: ¿cuándo fallan las condiciones de Kuhn-Tucker?              | 210        |
| 5.4 Teorema de la envolvente  | 212        |
| 5.4.1 Optimización libre  | 213        |
| 5.4.2 Optimización restringida  | 225        |
| <b>6 Temas selectos de cálculo avanzado</b>   | <b>231</b> |
| 6.1 Funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$   | 231        |
| 6.2 Regla de la cadena en el caso general   | 238        |
| 6.3 Teorema general de la función implícita   | 239        |
| 6.4 Teorema del punto fijo  | 244        |
| <b>A Cónicas</b>  | <b>247</b> |
| <b>B Teoremas de concavidad para funciones en <math>\mathbb{R}^n</math></b>                           | <b>251</b> |
| <b>Bibliografía</b>   | <b>258</b> |

# Prólogo

Este documento constituye un material de apoyo para el curso de Cálculo II para las carreras de Economía y Dirección Financiera en el ITAM. Se trata de una recopilación de mis notas de clase, con el fin de agilizar la discusión de los temas en el aula. El material se presenta en estricto apego al orden del temario vigente, aunque es discutido bajo un enfoque personal y en un lenguaje un tanto coloquial.

Estas notas no pretenden sustituir la lectura de la bibliografía seleccionada para el curso. Están basadas en el material extraído precisamente de esos textos, así como de documentos y libros escritos por mis colegas y amigos del Departamento de Matemáticas del ITAM. En particular, tomé prestados varios conceptos y ejemplos del Documento de Trabajo *Matemáticas IV*, elaborado por Guillermo Pastor. Para algunos temas de optimización, me basé en el libro *Métodos Dinámicos en Economía: Otra Búsqueda del Tiempo Perdido*, de Héctor Lomelí y Beatriz Rumbos. Muy especialmente, quiero expresar mi gratitud y gran admiración por Knut Sydsaeter, de la Universidad de Oslo, quien fue el autor de una colección maravillosa de textos de matemáticas para economistas. De él aprendí mucho, aunque nunca tuve el privilegio de conocerlo. Estuve a punto de hacerlo, en un taller de matemáticas que él iba a impartir en México. Desafortunadamente, Sydsaeter falleció en un accidente en octubre de 2012, faltando una semana para su visita a este país.

Se espera que el estudiante resuelva una gran variedad de ejercicios, que no han sido incluidos en este documento debido a su extensión. Al respecto, el estudiante puede utilizar el Documento de Trabajo *Cálculo II, Cuaderno de Ejercicios*, Lorena Zogaib, Departamento de Matemáticas, ITAM, agosto 1 de 2016.

Agradezco todas las sugerencias y correcciones que he recibido de mis colegas y varias generaciones de estudiantes. Me han enriquecido mucho los comentarios de mis amigos Carmen López y Ramón Espinosa. Igualmente importantes han sido las observaciones de las varias generaciones de alumnos que han consultado estas notas. Especialmente, estoy muy agradecida con Francisco Contreras Marroquín, quien estudió Ciencia Política en el ITAM, por sus valiosas aportaciones en relación con el capítulo de Optimización.

De antemano ofrezco una disculpa al lector por los errores y omisiones que encuentre en este texto. Siempre serán bienvenidas las correcciones y cualquier comentario que me hagan llegar.

Lorena Zogaib

# Capítulo 1

## El Espacio $\mathbb{R}^n$

### 1.1 Vectores

Considera los precios  $p_1, p_2, \dots, p_n$  de  $n$  bienes. Conviene representar este conjunto de precios por

$$(p_1, p_2, \dots, p_n) \quad \text{o bien} \quad \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{pmatrix}$$

Un conjunto ordenado de números como éste, que se caracteriza no sólo por los elementos que lo constituyen sino por el orden en que están colocados, se llama un *vector* o *n-vector*. Nota que un vector con una sola componente es un simple número real, también denominado un *escalar*. La representación del lado izquierdo se conoce como *vector renglón*, mientras que la del lado derecho es un *vector columna*. Por lo general utilizaremos la representación de vector renglón a lo largo de este texto, con excepción de algunos temas de los capítulos 3 y 6.

Hay varias maneras cortas de designar el vector de precios  $(p_1, p_2, \dots, p_n)$ , por ejemplo,

$$\vec{p} = (p_1, p_2, \dots, p_n), \quad \mathbf{p} = (p_1, p_2, \dots, p_n), \quad \bar{p} = (p_1, p_2, \dots, p_n), \dots$$

La primera de éstas,  $\vec{p}$ , utiliza una flechita encima del nombre del vector, y esto está relacionado con su significado geométrico, como discutiremos un poco más adelante en esta sección. Aquí adoptaremos precisamente esa notación para designar cualquier  $n$ -vector arbitrario  $(a_1, a_2, \dots, a_n)$ , es decir,

$$\vec{a} = (a_1, a_2, \dots, a_n), \quad \vec{a} \in \mathbb{R}^n.$$

Los números  $a_1, a_2, \dots, a_n$  se llaman las *componentes escalares* del vector  $\vec{a}$ , y decimos que  $a_i$  es la *i-ésima componente* de  $\vec{a}$ . La notación  $\vec{a} \in \mathbb{R}^n$  indica que cada una de las  $n$  componentes del vector  $\vec{a}$  es un escalar en el campo de los reales,  $\mathbb{R}$ .

## Operaciones con vectores

**Definición.** Dos vectores  $\vec{a}$  y  $\vec{b}$  son *iguales* o *equivalentes* si todas sus componentes son iguales. En ese caso, escribimos

$$\vec{a} = \vec{b}.$$

Si el número de componentes, su valor numérico o su distribución son diferentes, decimos que  $\vec{a} \neq \vec{b}$ .

Ejemplo:

Sean  $\vec{x} = (x, y, z)$  y  $\vec{a} = (-1, 0, 3)$ . Se tiene entonces que  $\vec{x} = \vec{a}$  si y sólo si  $x = -1$ ,  $y = 0$  y  $z = 3$ .

**Definición.** Sean  $\vec{a} = (a_1, a_2, \dots, a_n)$ ,  $\vec{b} = (b_1, b_2, \dots, b_n) \in \mathbb{R}^n$  y  $\beta \in \mathbb{R}$ .

a) El *producto del escalar*  $\beta$  con el vector  $\vec{a}$  es el vector  $\beta \vec{a} \in \mathbb{R}^n$ , dado por

$$\begin{aligned}\beta \vec{a} &= \beta(a_1, a_2, \dots, a_n) \\ &= (\beta a_1, \beta a_2, \dots, \beta a_n).\end{aligned}$$

b) La *suma* de los vectores  $\vec{a}$  y  $\vec{b}$  es el vector  $\vec{a} + \vec{b} \in \mathbb{R}^n$ , dado por

$$\begin{aligned}\vec{a} + \vec{b} &= (a_1, a_2, \dots, a_n) + (b_1, b_2, \dots, b_n) \\ &= (a_1 + b_1, a_2 + b_2, \dots, a_n + b_n).\end{aligned}$$

Ejemplo:

Sean  $\vec{a} = (3, -2, 5)$  y  $\vec{b} = (-3, 0, 3)$ . Así,

$$\begin{aligned}-2\vec{a} &= -2(3, -2, 5) = (-6, 4, -10), \\ \vec{a} + \vec{b} &= (3, -2, 5) + (-3, 0, 3) = (0, -2, 8).\end{aligned}$$

**Definición.** Sean  $\vec{a} = (a_1, a_2, \dots, a_n)$ ,  $\vec{b} = (b_1, b_2, \dots, b_n) \in \mathbb{R}^n$ . La *resta* o *diferencia* de  $\vec{a}$  con  $\vec{b}$  es el vector  $\vec{a} - \vec{b} \in \mathbb{R}^n$ , dado por

$$\begin{aligned}\vec{a} - \vec{b} &= \vec{a} + (-1)\vec{b} \\ &= (a_1 - b_1, a_2 - b_2, \dots, a_n - b_n).\end{aligned}$$

## 1.1 Vectores

Ejemplo:

Sean  $\vec{a} = (3, -2, 5)$  y  $\vec{b} = (-3, 0, 3)$ . Así,

$$\vec{a} - \vec{b} = (3, -2, 5) - (-3, 0, 3) = (6, -2, 2).$$

**Definición.** Para cada  $\vec{a} \in \mathbb{R}^n$  la diferencia  $\vec{a} - \vec{a}$  es el *vector nulo* o *vector cero*  $\vec{0}$ , dado por

$$\vec{0} = (0, 0, \dots, 0).$$

Nota que

$$\vec{a} - \vec{b} = \vec{0} \Leftrightarrow \vec{a} = \vec{b}.$$

**Definición.** Si  $\vec{a}_1, \vec{a}_2, \dots, \vec{a}_m \in \mathbb{R}^n$  y  $\beta_1, \beta_2, \dots, \beta_m \in \mathbb{R}$ , entonces el  $n$ -vector

$$\beta_1 \vec{a}_1 + \beta_2 \vec{a}_2 + \dots + \beta_m \vec{a}_m$$

se conoce como una *combinación lineal* de los vectores  $\vec{a}_1, \vec{a}_2, \dots, \vec{a}_m$ .

Ejemplo:

Sean  $\vec{a} = (3, -2, 5)$  y  $\vec{b} = (-3, 0, 3)$ . Así,

$$3\vec{a} - 5\vec{b} = 3(3, -2, 5) - 5(-3, 0, 3) = (9, -6, 15) + (15, 0, -15) = (24, -6, 0).$$

### Reglas de adición de vectores y multiplicación por escalares

Si  $\vec{a}, \vec{b}, \vec{c} \in \mathbb{R}^n$  y  $\alpha, \beta \in \mathbb{R}$ , entonces

1.  $(\vec{a} + \vec{b}) + \vec{c} = \vec{a} + (\vec{b} + \vec{c})$
2.  $\vec{a} + \vec{b} = \vec{b} + \vec{a}$
3.  $\vec{a} + \vec{0} = \vec{0} + \vec{a} = \vec{a}$
4.  $\vec{a} + (-\vec{a}) = (-\vec{a}) + \vec{a} = \vec{0}$
5.  $(\alpha + \beta) \vec{a} = \alpha \vec{a} + \beta \vec{a}$
6.  $\alpha (\vec{a} + \vec{b}) = \alpha \vec{a} + \alpha \vec{b}$
7.  $\alpha (\beta \vec{a}) = (\alpha \beta) \vec{a}$
8.  $1 \vec{a} = \vec{a}$

## Capítulo 1 El Espacio $\mathbb{R}^n$

Ejemplo:

Dados  $\vec{a}, \vec{b} \in \mathbb{R}^n$  halla un vector  $\vec{x} \in \mathbb{R}^n$  tal que  $3\vec{x} + 2\vec{a} = 5\vec{b}$ .

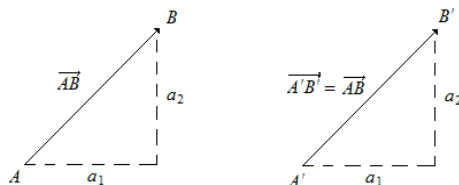
Usando las reglas anteriores, se tiene

$$\begin{aligned} 3\vec{x} + 2\vec{a} + (-2\vec{a}) &= 5\vec{b} + (-2\vec{a}) \\ 3\vec{x} + \vec{0} &= 5\vec{b} - 2\vec{a} \\ 3\vec{x} &= 5\vec{b} - 2\vec{a} \\ \left(\frac{1}{3}\right) 3\vec{x} &= \left(\frac{1}{3}\right) (5\vec{b} - 2\vec{a}) \\ 1\vec{x} &= \frac{5}{3}\vec{b} - \frac{2}{3}\vec{a} \\ \vec{x} &= \frac{5}{3}\vec{b} - \frac{2}{3}\vec{a}. \end{aligned}$$

### Interpretación geométrica de los vectores en el plano $\mathbb{R}^2$

La palabra vector proviene del latín y significa transporte. Por esa razón, un vector se asocia con un desplazamiento. Podemos describir ese desplazamiento en el plano  $xy$  por la distancia dirigida  $a_1$  que se mueve en la dirección del eje  $x$  y por la distancia dirigida  $a_2$  que se mueve en la dirección del eje  $y$ . Entendemos por distancia dirigida al hecho de que  $a_1 > 0$  si se desplaza hacia la derecha del punto inicial y  $a_1 < 0$  si se desplaza hacia la izquierda. Similarmente, se tiene  $a_2 > 0$  si el desplazamiento es hacia arriba o  $a_2 < 0$  si es hacia abajo.

Geométricamente, esta translación se puede visualizar como una flecha o segmento de recta dirigido de un punto  $A$  a otro punto  $B$ , que denotamos por  $\vec{AB}$ . Si desplazamos la flecha paralelamente a sí misma, de tal manera que su nuevo origen sea  $A'$  y el nuevo destino  $B'$ , la flecha resultante  $\vec{A'B'}$  describirá el mismo desplazamiento, porque sus componentes  $x$  y  $y$  siguen siendo  $a_1$  y  $a_2$ , respectivamente.



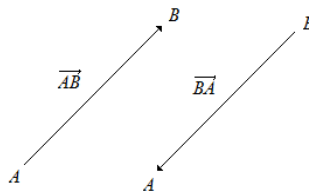
De esta manera, se tiene que

$$\vec{AB} = \vec{A'B'}.$$

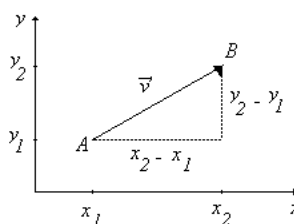


## 1.1 Vectores

Así, desde el punto de vista geométrico, decimos que dos vectores son iguales o equivalentes si tienen la misma dirección y longitud (dados por las mismas componentes  $a_1$  y  $a_2$ ). En consecuencia, es claro que  $\overrightarrow{AB} \neq \overrightarrow{BA}$ .



**Definición.** Dados dos puntos  $A(x_1, y_1)$  y  $B(x_2, y_2)$  del plano  $\mathbb{R}^2$ , el vector  $\vec{v}$  que va de  $A$  hacia  $B$  es el vector  $\vec{v} = \overrightarrow{AB} = (x_2 - x_1, y_2 - y_1)$ .



Ejemplos:

- Si  $A(1, 1)$ ,  $B(2, -1)$  y  $\vec{v} = \overrightarrow{AB}$  es el vector que va de  $A$  a  $B$ , entonces  

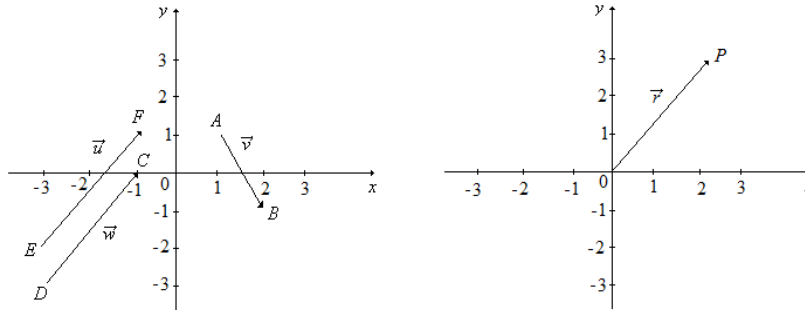
$$\vec{v} = (2 - 1, -1 - 1) = (1, -2).$$
- Si  $C(-1, 0)$ ,  $D(-3, -3)$  y  $\vec{w} = \overrightarrow{DC}$  es el vector que va de  $D$  a  $C$ , entonces  

$$\vec{w} = (-1 - (-3), 0 - (-3)) = (2, 3).$$
- Si  $E(-3, -2)$ ,  $F(-1, 1)$  y  $\vec{u} = \overrightarrow{EF}$  es el vector que va de  $E$  a  $F$ , entonces  

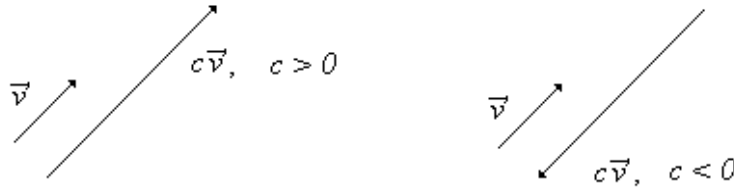
$$\vec{u} = (-1 - (-3), 1 - (-2)) = (2, 3).$$

Observamos que los vectores  $\vec{w}$  y  $\vec{u}$  son iguales, ya que poseen las mismas componentes, a pesar de tener asociados diferentes puntos de origen y destino. De hecho, los vectores  $\vec{w}$  y  $\vec{u}$  son también iguales al vector  $\vec{r} = \overrightarrow{OP}$  que va del origen de coordenadas  $O(0, 0)$  al punto  $P(2, 3)$ . A un vector que sale del origen se le conoce como *vector de posición* o *representante principal*.

## Capítulo 1 El Espacio $\mathbb{R}^n$

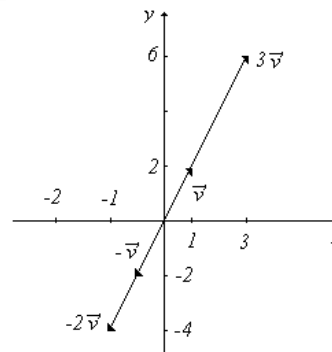


De acuerdo con la definición, la multiplicación de un vector  $\vec{v}$  por un escalar  $c$  es un nuevo vector,  $c\vec{v}$ , cuyas componentes son las componentes de  $\vec{v}$  multiplicadas cada una por el factor  $c$ . Geométricamente, el vector  $c\vec{v}$  es paralelo a  $\vec{v}$ , con las siguientes características: i) si  $c > 0$ , entonces  $c\vec{v}$  es un vector en el mismo sentido que  $\vec{v}$ , y si  $c < 0$ , su sentido es opuesto, ii) si  $|c| > 1$ , entonces  $c\vec{v}$  es un vector de magnitud mayor que  $\vec{v}$ , y si  $|c| < 1$ , su magnitud es menor.



Por ejemplo, si  $\vec{v} = (1, 2)$ , entonces

- a)  $3\vec{v} = 3(1, 2) = (3, 6)$
- b)  $-\vec{v} = -(1, 2) = (-1, -2)$
- c)  $-2\vec{v} = -2(1, 2) = (-2, -4)$



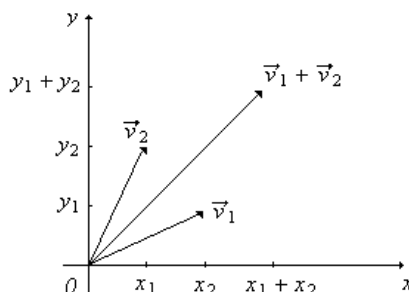
**Definición.** Dos vectores no nulos  $\vec{a}$  y  $\vec{b}$  son *paralelos* si son múltiplos entre sí, es decir, si existe un escalar  $\alpha \neq 0$  tal que  $\vec{b} = \alpha \vec{a}$ . Que los vectores  $\vec{a}$  y  $\vec{b}$  sean paralelos se denota por  $\vec{a} \parallel \vec{b}$ .

## 1.1 Vectores

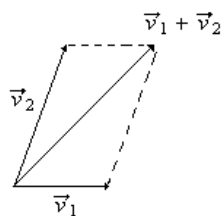
Ejemplo:

El vector  $\vec{v}_1 = \left(\frac{1}{5}, -\frac{3}{5}\right)$  es paralelo al vector  $\vec{v}_2 = \left(-\frac{1}{3}, 1\right)$ , ya que  $\vec{v}_1 = -\frac{3}{5}\vec{v}_2$ .

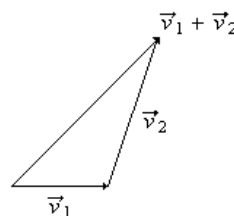
Por otra parte, sabemos de la definición que la suma  $\vec{v}_1 + \vec{v}_2$  de dos vectores  $\vec{v}_1$  y  $\vec{v}_2$  es la suma de sus componentes, como se muestra en la siguiente figura.



El vector suma  $\vec{v}_1 + \vec{v}_2$  se construye más fácilmente a partir del método del paralelogramo, así como el método del triángulo, ilustrados en la siguiente figura.

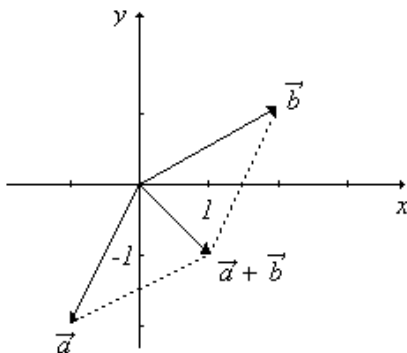


método del  
paralelogramo



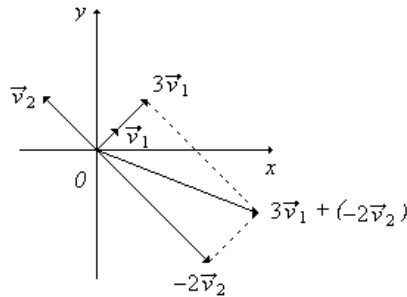
método del  
triángulo

La siguiente figura muestra la suma de los vectores  $\vec{a} = (-1, -2)$  y  $\vec{b} = (2, 1)$ , dada por  $\vec{a} + \vec{b} = (-1, -2) + (2, 1) = (-1 + 2, -2 + 1) = (1, -1)$ .

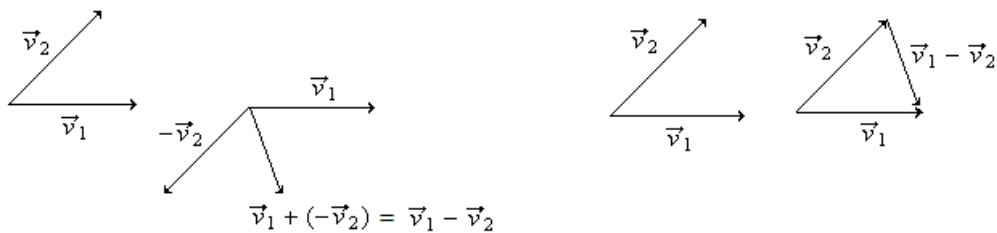


## Capítulo 1 El Espacio $\mathbb{R}^n$

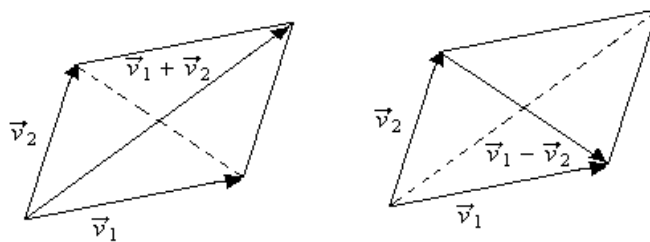
Asimismo, combinando las dos operaciones anteriores, se puede construir cualquier combinación lineal de vectores en  $\mathbb{R}^2$ . Por ejemplo, para  $\vec{v}_1 = (1, 1)$  y  $\vec{v}_2 = (-2, 3)$ , la siguiente figura muestra la combinación lineal  $-2\vec{v}_2 + 3\vec{v}_1 = -2(-2, 3) + 3(1, 1) = (4, -6) + (3, 3) = (7, -3)$ .



Por último, sabemos que el vector resta  $\vec{v}_1 - \vec{v}_2$  se construye como la suma de vectores  $\vec{v}_1 + (-\vec{v}_2)$ , como se muestra en las figuras de la izquierda. Esto equivale a decir que la resta  $\vec{v}_1 - \vec{v}_2$  es el vector que une las “puntas” de los vectores  $\vec{v}_1$  y  $\vec{v}_2$ , en dirección de  $\vec{v}_2$  hacia  $\vec{v}_1$ , como se muestra en las figuras de la derecha.

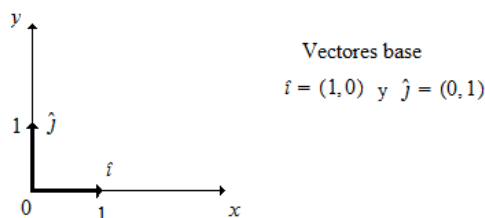


Un resultado muy útil es que la suma  $\vec{v}_1 + \vec{v}_2$  y la resta  $\vec{v}_1 - \vec{v}_2$  pueden asociarse con las diagonales del paralelogramo formado al unir los vectores  $\vec{v}_1$  y  $\vec{v}_2$ , de la manera que se muestra en la siguiente figura.



## 1.1 Vectores

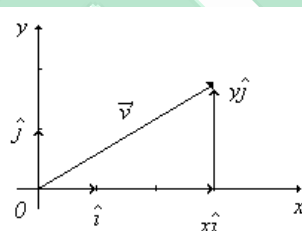
Existe una representación alternativa para los vectores en el plano, utilizando los llamados *vectores base*. Estos últimos son vectores en términos de los cuales podemos expresar cualquier otro vector. En el caso de  $\mathbb{R}^2$  se necesitan dos vectores base, no paralelos, para poder generar cualquier vector en el plano. La base más simple es la base canónica, dada por los vectores  $\hat{i} = (1, 0)$  y  $\hat{j} = (0, 1)$ , mostrados en la siguiente figura.



**Definición.** Cualquier vector  $\vec{v} = (x, y) \in \mathbb{R}^2$  se puede expresar como

$$\vec{v} = x\hat{i} + y\hat{j},$$

con  $\hat{i} = (1, 0)$  y  $\hat{j} = (0, 1)$ .

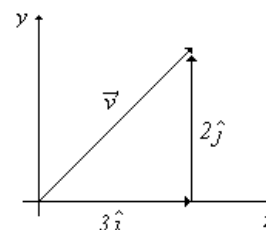


**Definición.** Si  $\vec{v} = x\hat{i} + y\hat{j}$ , entonces los vectores  $x\hat{i}$  y  $y\hat{j}$  son los *vectores componentes* o *componentes vectoriales* de  $\vec{v}$  en las direcciones  $\hat{i}$  y  $\hat{j}$ , respectivamente. Los números  $x$  y  $y$  son las *componentes escalares* de  $\vec{v}$  en las direcciones  $\hat{i}$  y  $\hat{j}$ , respectivamente.

Ejemplo:

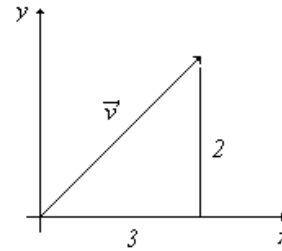
Por ejemplo, si  $\vec{v} = 3\hat{i} + 2\hat{j}$ , entonces

- i)  $3\hat{i}$  es la componente vectorial de  $\vec{v}$  en la dirección  $\hat{i}$
- ii)  $2\hat{j}$  es la componente vectorial de  $\vec{v}$  en la dirección  $\hat{j}$



iii) 3 es la componente escalar de  $\vec{v}$  en la dirección  $\hat{i}$

iv) 2 es la componente escalar de  $\vec{v}$  en la dirección  $\hat{j}$

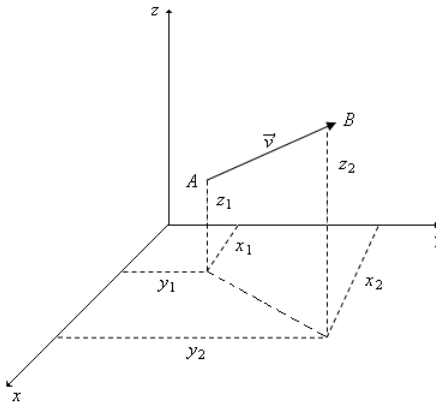


### Interpretación geométrica de los vectores en el espacio $\mathbb{R}^3$

Los resultados anteriores pueden extenderse fácilmente para vectores en el espacio  $\mathbb{R}^3$ , como se presenta a continuación.

**Definición.** Dados dos puntos  $A(x_1, y_1, z_1)$  y  $B(x_2, y_2, z_2)$  en el espacio, el vector  $\vec{v}$  que va de  $A$  a  $B$  es el vector

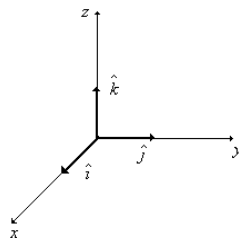
$$\vec{v} = \overrightarrow{AB} = (x_2 - x_1, y_2 - y_1, z_2 - z_1).$$



Por otra parte, la base canónica en  $\mathbb{R}^3$  son los vectores

$$\hat{i} = (1, 0, 0), \quad \hat{j} = (0, 1, 0), \quad \hat{k} = (0, 0, 1),$$

mostrados en la siguiente figura.



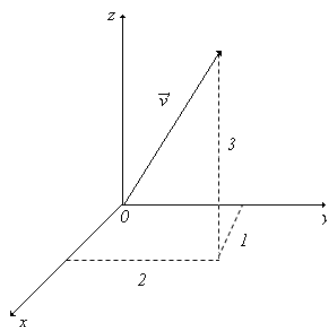
## 1.1 Vectores

En términos de estos vectores base, cualquier vector  $\vec{v} = (x, y, z)$  en  $\mathbb{R}^3$  puede expresarse como

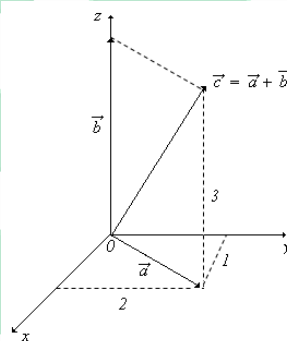
$$\vec{v} = x\hat{i} + y\hat{j} + z\hat{k}.$$

Ejemplos:

1. Dibuja el vector  $\vec{v} = \hat{i} + 2\hat{j} + 3\hat{k}$



2. Dibuja los vectores  $\vec{a} = \hat{i} + 2\hat{j}$ ,  $\vec{b} = 3\hat{k}$  y  $\vec{c} = \vec{a} + \vec{b}$ .



### Norma de un vector en $\mathbb{R}^n$

Como ya se mencionó, una de las dos características de un vector es su longitud, también conocida como su *norma*. Para vectores en  $\mathbb{R}^n$  la norma se determina a partir del teorema de Pitágoras generalizado, como se define a continuación.

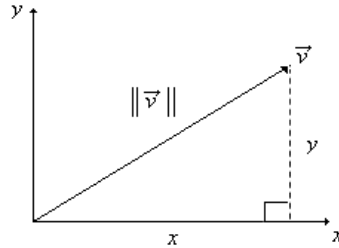
**Definición.** La *norma*, o *magnitud*, de un vector  $\vec{v} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$  es el número real no negativo  $\|\vec{v}\|$  dado por

$$\|\vec{v}\| = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}.$$

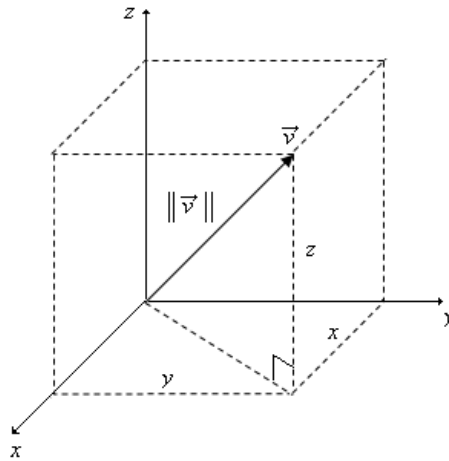
## Capítulo 1 El Espacio $\mathbb{R}^n$

En particular, se tienen los siguientes casos:

i) si  $\vec{v} = x\hat{i} + y\hat{j} \in \mathbb{R}^2$ , entonces  $\|\vec{v}\| = \sqrt{x^2 + y^2}$ .



ii) si  $\vec{v} = x\hat{i} + y\hat{j} + z\hat{k} \in \mathbb{R}^3$ , entonces  $\|\vec{v}\| = \sqrt{x^2 + y^2 + z^2}$ .



Cabe señalar que la norma de un vector  $\vec{v} = x\hat{i}$  en  $\mathbb{R}$  es simplemente su valor absoluto, ya que  $\|\vec{v}\| = \sqrt{x^2} = |x|$ .

Ejemplos:

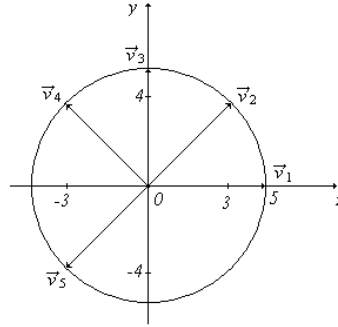
1. Si  $\vec{u} = \hat{i} + \hat{j}$ , entonces  $\|\vec{u}\| = \sqrt{1^2 + 1^2} = \sqrt{2}$ .
2. Si  $\vec{v} = -3\hat{i} + 4\hat{j}$ , entonces  $\|\vec{v}\| = \sqrt{(-3)^2 + 4^2} = \sqrt{25} = 5$ .
3. Si  $\vec{w} = -\hat{i} + 2\hat{j} - 3\hat{k}$ , entonces  $\|\vec{w}\| = \sqrt{(-1)^2 + 2^2 + (-3)^2} = \sqrt{14}$ .



## 1.1 Vectores

Nota que existe una infinidad de vectores con una misma norma dada. Por ejemplo, todos los siguientes vectores poseen norma 5.

$$\begin{aligned}\vec{v}_1 &= 5\hat{i} \\ \vec{v}_2 &= 3\hat{i} + 4\hat{j} \\ \vec{v}_3 &= 5\hat{j} \\ \vec{v}_4 &= -3\hat{i} + 4\hat{j} \\ \vec{v}_5 &= -3\hat{i} - 4\hat{j}\end{aligned}$$



Las cabezas de todos estos vectores describen una circunferencia de radio 5.

### Propiedades de la norma

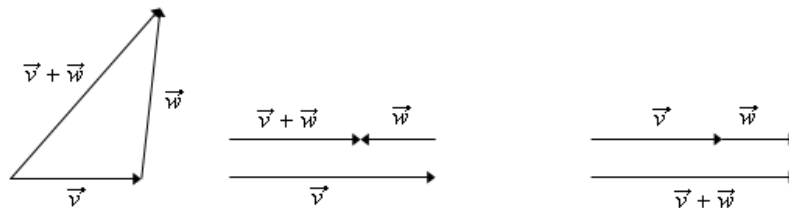
Sean  $\vec{v}, \vec{w} \in \mathbb{R}^n$  y sea  $c \in \mathbb{R}$ . Entonces,

- a)  $\|\vec{v}\| \geq 0$
- b)  $\|\vec{v}\| = 0 \Leftrightarrow \vec{v} = \vec{0}$
- c)  $\|c\vec{v}\| = |c| \|\vec{v}\|$
- d)  $\|\vec{v} + \vec{w}\| \leq \|\vec{v}\| + \|\vec{w}\|$  Desigualdad del triángulo

Las propiedades a) y b) establecen que la norma de un vector es no negativa, y sólo es cero si  $\vec{v}$  es el vector nulo. La propiedad c) establece que se preserva la escala al calcular la norma del múltiplo de un vector; así, por ejemplo, la expresión

$$\| -3\vec{v} \| = |-3| \|\vec{v}\| = 3 \|\vec{v}\|$$

establece que la norma  $\| -3\vec{v} \|$  del triple de un vector  $\vec{v}$ , es el triple de su norma,  $3 \|\vec{v}\|$ . Por último, la propiedad d), o desigualdad del triángulo, establece que la hipotenusa de un triángulo mide menos que la suma de sus catetos (figuras de la izquierda) y sólo es igual a la suma de estos cuando son paralelos (figura de la derecha).



$$\|\vec{v} + \vec{w}\| < \|\vec{v}\| + \|\vec{w}\|$$

$$\|\vec{v} + \vec{w}\| = \|\vec{v}\| + \|\vec{w}\|$$

## Capítulo 1 El Espacio $\mathbb{R}^n$

Así, por ejemplo, de acuerdo con la propiedad d), para vectores arbitrarios  $\vec{v}, \vec{w}$ , se tiene

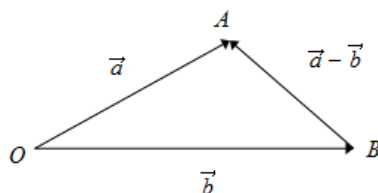
$$\begin{aligned}\|3\vec{v} - 5\vec{w}\| &\neq \|3\vec{v}\| + \|5\vec{w}\|, \\ \|3\vec{v} - 5\vec{w}\| &\neq \|3\vec{v}\| - \|5\vec{w}\|.\end{aligned}$$

**Definición.** La *distancia euclidiana*  $d(A, B)$  entre dos puntos  $A(a_1, a_2, \dots, a_n)$  y  $B(b_1, b_2, \dots, b_n)$  en  $\mathbb{R}^n$  está dada por

$$d(A, B) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + \dots + (b_n - a_n)^2}.$$

Esta es una generalización del teorema de Pitágoras en  $n$  dimensiones. Más generalmente, si denotamos por  $\vec{a} = (a_1, a_2, \dots, a_n)$  al vector  $\vec{OA}$  que va del origen al punto  $A(a_1, a_2, \dots, a_n)$  y por  $\vec{b} = (b_1, b_2, \dots, b_n)$  al vector  $\vec{OB}$  que va del origen al punto  $B(b_1, b_2, \dots, b_n)$ , entonces la distancia  $d(A, B)$  entre los puntos  $A$  y  $B$  es la norma del vector que los une, es decir,

$$d(A, B) = d(\vec{a}, \vec{b}) = \|\vec{b} - \vec{a}\| = \|\vec{a} - \vec{b}\|.$$



### Propiedades de la distancia euclidiana

Sean  $\vec{a}, \vec{b}, \vec{c} \in \mathbb{R}^n$ . Entonces,

- a)  $d(\vec{a}, \vec{b}) \geq 0$
- b)  $d(\vec{a}, \vec{b}) = 0 \Leftrightarrow \vec{a} = \vec{b}$
- c)  $d(\vec{a}, \vec{b}) = d(\vec{b}, \vec{a})$  Simetría
- d)  $d(\vec{a}, \vec{b}) + d(\vec{b}, \vec{c}) \geq d(\vec{a}, \vec{c})$  Desigualdad del triángulo

A un conjunto  $X$  con una función distancia  $d : X \times X \rightarrow \mathbb{R}$  que satisface estas propiedades se le llama *espacio métrico*. En particular, el espacio  $\mathbb{R}^n$  es un espacio métrico.

## 1.1 Vectores

**Definición.** Un *vector unitario*  $\hat{u}$  es un vector con norma igual a 1, es decir,  $\|\hat{u}\| = 1$ .

En el caso de  $\mathbb{R}^2$ , los vectores unitarios son todos aquellos que pueden dibujarse dentro de una circunferencia de radio 1 y centro en el origen, como es el caso de los siguientes vectores.

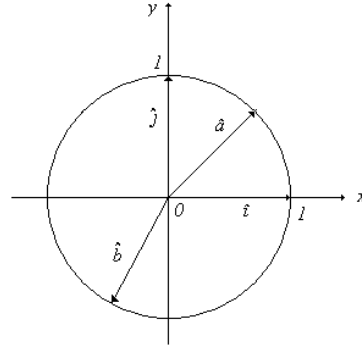
$$\hat{i}$$

$$\hat{j}$$

$$\hat{a} = \frac{1}{\sqrt{2}}\hat{i} + \frac{1}{\sqrt{2}}\hat{j}$$

$$\hat{b} = -\frac{3}{5}\hat{i} - \frac{4}{5}\hat{j}$$

$$\|\hat{i}\| = \|\hat{j}\| = \|\hat{a}\| = \|\hat{b}\| = 1.$$



Como la norma de un vector unitario es, por definición, siempre igual a 1, su única característica importante es su dirección. De ahí que los vectores unitarios son conocidos también como *vectores de dirección*. En el caso particular de un vector unitario en el plano,  $\hat{u} \in \mathbb{R}^2$ , su dirección se define como el ángulo  $\theta$  que éste determina con el eje  $x$ , medido en la dirección contraria al giro de las manecillas del reloj. Así, cualquier vector unitario en  $\mathbb{R}^2$ ,

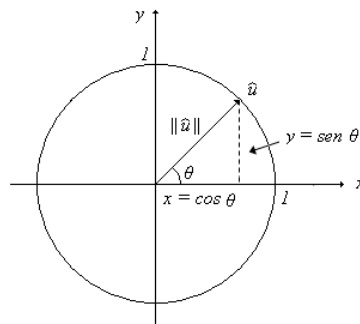
$$\hat{u} = x\hat{i} + y\hat{j},$$

donde  $\|\hat{u}\| = \sqrt{x^2 + y^2} = 1$ , puede expresarse como

$$\hat{u} = \cos \theta \hat{i} + \operatorname{sen} \theta \hat{j},$$

en donde se ha utilizado que

$$\cos \theta = \frac{x}{1} = x, \quad \operatorname{sen} \theta = \frac{y}{1} = y.$$



## Capítulo 1 El Espacio $\mathbb{R}^n$

Ejemplos:

1. Determina la dirección del vector unitario  $\hat{a} = \frac{1}{\sqrt{2}}\hat{i} + \frac{1}{\sqrt{2}}\hat{j}$ .

En este caso,  $\cos \theta = \frac{1}{\sqrt{2}}$  y  $\operatorname{sen} \theta = \frac{1}{\sqrt{2}}$ . Por lo tanto,

$$\theta = \cos^{-1} \left( \frac{1}{\sqrt{2}} \right) = \operatorname{sen}^{-1} \left( \frac{1}{\sqrt{2}} \right) = \frac{\pi}{4},$$

en donde  $\cos^{-1} x$  y  $\operatorname{sen}^{-1} x$  denotan “ángulo cuyo coseno es” y “ángulo cuyo seno es”, que son las funciones inversas de las funciones coseno y seno.

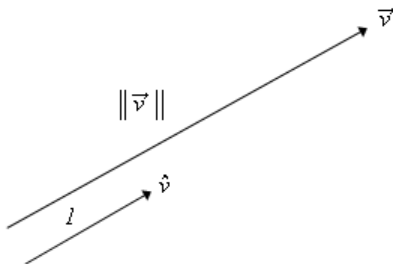
2. Determina la dirección del vector unitario  $\hat{b} = -\hat{i}$ .

En este caso,  $\cos \theta = -1$  y  $\operatorname{sen} \theta = 0$ . Por lo tanto,

$$\theta = \cos^{-1}(-1) = \operatorname{sen}^{-1}(0) = \pi.$$

Cualquier vector no nulo,  $\vec{v} \in \mathbb{R}^n$ , puede escribirse siempre en términos del vector unitario  $\hat{v}$  que apunta en la misma dirección que  $\vec{v}$ , de acuerdo con

$$\vec{v} = \|\vec{v}\| \hat{v}.$$



De esta manera, el vector unitario  $\hat{v}$  del vector no nulo  $\vec{v} \neq \vec{0}$  está dado por el cociente

$$\hat{v} = \frac{\vec{v}}{\|\vec{v}\|}.$$

Ejemplos:

1. Calcula el vector unitario  $\hat{a}$  del vector  $\vec{a} = -3\hat{i} + 4\hat{j}$ .

Como  $\|\vec{a}\| = \sqrt{(-3)^2 + 4^2} = 5$ , por lo tanto,

$$\hat{a} = \frac{\vec{a}}{\|\vec{a}\|} = \frac{-3\hat{i} + 4\hat{j}}{5} = -\frac{3}{5}\hat{i} + \frac{4}{5}\hat{j}.$$

## 1.1 Vectores

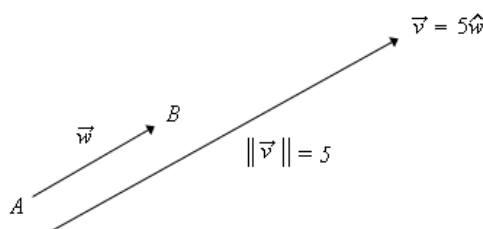
2. Calcula el vector unitario  $\hat{x}$  del vector  $\vec{x} = \hat{i} - 2\hat{j} + 3\hat{k}$ .

Como  $\|\vec{x}\| = \sqrt{1^2 + (-2)^2 + 3^2} = \sqrt{14}$ , por lo tanto,

$$\hat{x} = \frac{\vec{x}}{\|\vec{x}\|} = \frac{\hat{i} - 2\hat{j} + 3\hat{k}}{\sqrt{14}} = \frac{1}{\sqrt{14}}\hat{i} - \frac{2}{\sqrt{14}}\hat{j} + \frac{3}{\sqrt{14}}\hat{k}.$$

3. Encuentra un vector  $\vec{v}$  con magnitud (norma) igual a 5 y que tenga la misma dirección que el vector  $\vec{w}$  que va del punto  $A(-1, 2, 1)$  al punto  $B(-2, 0, 3)$ .

De acuerdo con el enunciado es claro que  $\vec{v} = 5\hat{w}$ , con  $\hat{w}$  el vector de dirección de  $\vec{w} = \overrightarrow{AB}$ .



Para calcular  $\hat{w}$ , notamos primero que

$$\begin{aligned}\vec{w} &= \overrightarrow{AB} = ((-2) - (-1), 0 - 2, 3 - 1) = (-1, -2, 2) \\ &= -\hat{i} - 2\hat{j} + 2\hat{k}.\end{aligned}$$

Como  $\|\vec{w}\| = \sqrt{(-1)^2 + (-2)^2 + 2^2} = 3$ , por lo tanto,

$$\hat{w} = \frac{\vec{w}}{\|\vec{w}\|} = \frac{-\hat{i} - 2\hat{j} + 2\hat{k}}{3} = -\frac{1}{3}\hat{i} - \frac{2}{3}\hat{j} + \frac{2}{3}\hat{k}.$$

De este modo,

$$\vec{v} = 5\hat{w} = 5 \left( -\frac{1}{3}\hat{i} - \frac{2}{3}\hat{j} + \frac{2}{3}\hat{k} \right) = -\frac{5}{3}\hat{i} - \frac{10}{3}\hat{j} + \frac{10}{3}\hat{k}.$$

Además del producto  $c\vec{v}$  de un vector  $\vec{v}$  por un escalar  $c$  existen otros dos productos importantes que involucran vectores. Uno de ellos, conocido como el *producto punto* o *producto escalar*, reviste de gran utilidad para este curso, por lo que se define a continuación. El otro producto se conoce como el *producto cruz* o *producto vectorial*, pero se omitirá en estas notas.

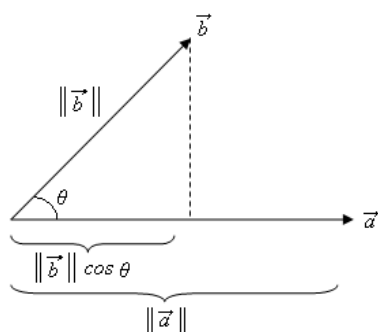
## Producto punto

**Definición.** El *producto escalar* o *producto punto*,  $\vec{a} \cdot \vec{b}$ , de dos vectores  $\vec{a}$  y  $\vec{b}$  en el plano  $\mathbb{R}^2$ , o en el espacio  $\mathbb{R}^3$ , es el escalar

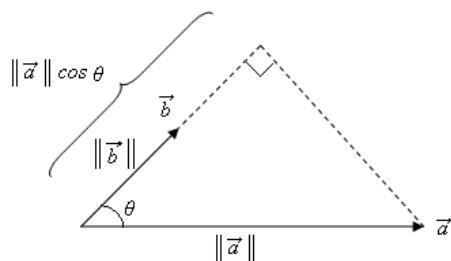
$$\vec{a} \cdot \vec{b} = \|\vec{a}\| \|\vec{b}\| \cos \theta,$$

donde  $\theta$  es el ángulo entre  $\vec{a}$  y  $\vec{b}$ , con  $0 \leq \theta \leq \pi$ .

Observa que  $\vec{a} \cdot \vec{b}$  no es un vector, sino un escalar. Geométricamente,  $\vec{a} \cdot \vec{b}$  representa el producto de la norma de cualquiera de los dos vectores por la componente del otro vector en la dirección de éste, como se muestra en las siguientes figuras.



$$\vec{a} \cdot \vec{b} = \|\vec{a}\| (\|\vec{b}\| \cos \theta)$$



$$\vec{a} \cdot \vec{b} = \|\vec{b}\| (\|\vec{a}\| \cos \theta)$$

La siguiente tabla resume algunos casos especiales.

|                          | $\theta$        | $\vec{a} \cdot \vec{b} = \ \vec{a}\  \ \vec{b}\  \cos \theta$ |              |
|--------------------------|-----------------|---|--------------|
| Vectores paralelos       | 0               | $\ \vec{a}\  \ \vec{b}\ $                                     | valor máximo |
| Vectores perpendiculares | $\frac{\pi}{2}$ | 0   |              |
| Vectores antiparalelos   | $\pi$           | $-\ \vec{a}\  \ \vec{b}\ $                                    | valor mínimo |

Observa que, de acuerdo con la definición, el producto punto es conmutativo, es decir,

$$\vec{a} \cdot \vec{b} = \vec{b} \cdot \vec{a}.$$

## 1.1 Vectores

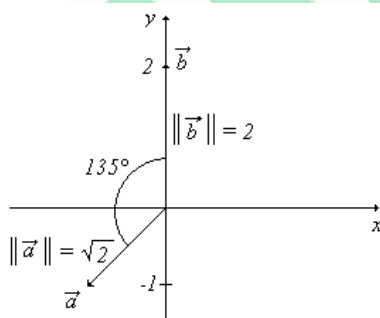
**Definición.** Decimos que dos vectores no nulos  $\vec{a}$  y  $\vec{b}$  son *perpendiculares u ortogonales* si y sólo si  $\vec{a} \cdot \vec{b} = 0$ . Que los vectores  $\vec{a}$  y  $\vec{b}$  sean perpendiculares se denota por  $\vec{a} \perp \vec{b}$ .

Ejemplos:

1. Calcula el producto punto de los vectores  $\vec{a} = -\hat{i} - \hat{j}$  y  $\vec{b} = 2\hat{j}$  en  $\mathbb{R}^2$ .

Sabemos que  $\|\vec{a}\| = \sqrt{2}$  y  $\|\vec{b}\| = 2$ . De la figura se observa que el ángulo entre  $\vec{a}$  y  $\vec{b}$  es  $135^\circ$ , es decir,  $\theta = \frac{3\pi}{4}$ . Así,

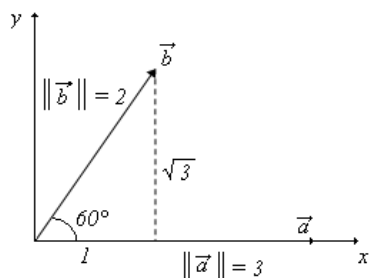
$$\vec{a} \cdot \vec{b} = \|\vec{a}\| \|\vec{b}\| \cos\left(\frac{3\pi}{4}\right) = (\sqrt{2})(2) \left(-\frac{1}{\sqrt{2}}\right) = -2.$$



2. Calcula el producto punto de los vectores  $\vec{a} = 3\hat{i}$  y  $\vec{b} = \hat{i} + \sqrt{3}\hat{j}$  en  $\mathbb{R}^2$ .

Sabemos que  $\|\vec{a}\| = 3$  y  $\|\vec{b}\| = 2$ . De la figura se observa que el ángulo entre  $\vec{a}$  y  $\vec{b}$  es  $60^\circ$ . Así,

$$\vec{a} \cdot \vec{b} = \|\vec{a}\| \|\vec{b}\| \cos\left(\frac{\pi}{3}\right) = (3)(2) \left(\frac{1}{2}\right) = 3.$$

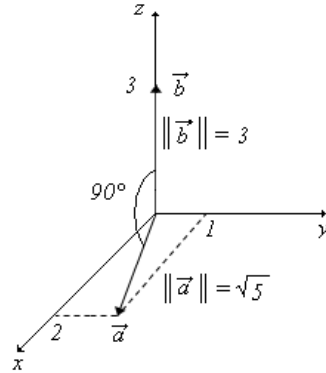


## Capítulo 1 El Espacio $\mathbb{R}^n$

3. Calcula el producto punto de los vectores  $\vec{a} = 2\hat{i} + \hat{j}$  y  $\vec{b} = 3\hat{k}$  en  $\mathbb{R}^3$ .

Sabemos que  $\|\vec{a}\| = \sqrt{5}$  y  $\|\vec{b}\| = 3$ . De la figura se observa que el ángulo entre  $\vec{a}$  y  $\vec{b}$  es  $90^\circ$ , es decir,  $\theta = \frac{\pi}{2}$ . Así,

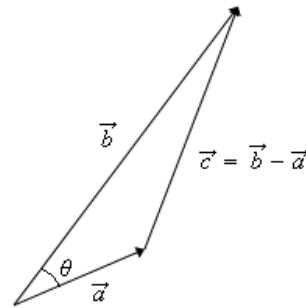
$$\vec{a} \cdot \vec{b} = \|\vec{a}\| \|\vec{b}\| \cos\left(\frac{\pi}{2}\right) = (\sqrt{5})(3)(0) = 0.$$



En general, la expresión  $\vec{a} \cdot \vec{b} = \|\vec{a}\| \|\vec{b}\| \cos \theta$  puede resultar poco práctica para calcular el producto punto de  $\vec{a}$  y  $\vec{b}$ , ya que requiere conocer el ángulo  $\theta$  entre  $\vec{a}$  y  $\vec{b}$ . Por esta razón, a continuación desarrollaremos una expresión alternativa para calcular  $\vec{a} \cdot \vec{b}$  a partir de las componentes de estos vectores, que suele ser la información que se tiene disponible.

Para este fin consideramos dos vectores  $\vec{a}$  y  $\vec{b}$ , así como su vector diferencia,  $\vec{c} = \vec{b} - \vec{a}$ . Estos tres vectores determinan un triángulo, cuyos catetos están relacionados entre sí por la ley de los cosenos, dada por

$$\|\vec{c}\|^2 = \|\vec{a}\|^2 + \|\vec{b}\|^2 - 2\|\vec{a}\| \|\vec{b}\| \cos \theta,$$





## 1.1 Vectores

en donde  $\theta$  denota el ángulo entre  $\vec{a}$  y  $\vec{b}$ . Nota que esta igualdad se reduce al teorema de Pitágoras en el caso particular  $\theta = \pi/2$ . El término  $\|\vec{a}\| \|\vec{b}\| \cos \theta$  en el lado derecho es precisamente el producto punto  $\vec{a} \cdot \vec{b}$  entre  $\vec{a}$  y  $\vec{b}$ , es decir,

$$\|\vec{c}\|^2 = \|\vec{a}\|^2 + \|\vec{b}\|^2 - 2\vec{a} \cdot \vec{b},$$

de modo que

$$\vec{a} \cdot \vec{b} = \frac{\|\vec{a}\|^2 + \|\vec{b}\|^2 - \|\vec{c}\|^2}{2}.$$

Para el caso particular de vectores  $\vec{a} = a_1\hat{i} + a_2\hat{j}$  y  $\vec{b} = b_1\hat{i} + b_2\hat{j}$  en  $\mathbb{R}^2$ , el vector  $\vec{c} = \vec{b} - \vec{a}$  está dado por  $\vec{c} = (b_1 - a_1)\hat{i} + (b_2 - a_2)\hat{j}$ , de modo que

$$\vec{a} \cdot \vec{b} = \frac{(a_1^2 + a_2^2) + (b_1^2 + b_2^2) - [(b_1 - a_1)^2 + (b_2 - a_2)^2]}{2}.$$

Desarrollando cuadrados en el numerador es posible simplificar varios términos, quedando simplemente,

$$\vec{a} \cdot \vec{b} = a_1b_1 + a_2b_2.$$

De esta manera, el cálculo de  $\vec{a} \cdot \vec{b}$  se reduce a multiplicar término a término las componentes escalares de  $\vec{a}$  y  $\vec{b}$ . Similarmente, es posible demostrar que en el caso de vectores  $\vec{a} = a_1\hat{i} + a_2\hat{j} + a_3\hat{k}$  y  $\vec{b} = b_1\hat{i} + b_2\hat{j} + b_3\hat{k}$  en  $\mathbb{R}^3$  el producto punto está dado por

$$\vec{a} \cdot \vec{b} = a_1b_1 + a_2b_2 + a_3b_3.$$

El resultado anterior puede extenderse muy fácilmente para cualesquiera dos vectores en  $\mathbb{R}^n$ , como se enuncia en el siguiente teorema.

**Teorema.** Dados dos vectores  $\vec{a} = (a_1, a_2, \dots, a_n)$  y  $\vec{b} = (b_1, b_2, \dots, b_n)$  en  $\mathbb{R}^n$ , su producto punto  $\vec{a} \cdot \vec{b}$  es el escalar

$$\vec{a} \cdot \vec{b} = a_1b_1 + a_2b_2 + \dots + a_nb_n.$$

Ejemplos:

1. Calcula  $\vec{x} \cdot \vec{y}$ , si  $\vec{x} = (-1, -3, 0)$  y  $\vec{y} = (2, 1, -3)$ .

En este caso,

$$\vec{x} \cdot \vec{y} = (-1)(2) + (-3)(1) + (0)(-3) = -5.$$

## Capítulo 1 El Espacio $\mathbb{R}^n$

2. Demuestra que los vectores  $\vec{u} = 2\hat{i} + 3\hat{j}$  y  $\vec{v} = -6\hat{i} + 4\hat{j}$  son perpendiculares entre sí.

Como

$$\vec{u} \cdot \vec{v} = (2)(-6) + (3)(4) = 0,$$

por lo tanto  $\vec{u} \perp \vec{v}$ .

3. Escribe el ingreso  $I = p_1q_1 + p_2q_2 + \cdots + p_nq_n$  como un producto punto de vectores.

El ingreso  $I$  puede expresarse como  $I = \vec{p} \cdot \vec{q}$ , donde  $\vec{p} = (p_1, p_2, \dots, p_n)$  es el vector de precios y  $\vec{q} = (q_1, q_2, \dots, q_n)$  es el vector de cantidades.

4. Sean  $\vec{a} = \hat{i} + 2\hat{j} + 3\hat{k}$  y  $\vec{b} = 4\hat{i} - \hat{j} + \hat{k}$ . Calcula el vector  $\vec{v} = \frac{\|\vec{a} - \vec{b}\|}{\|\vec{a} - \vec{b}\|} (\vec{a} \cdot \vec{b}) \hat{a}$ .

Por una parte, como  $\vec{a} - \vec{b} = -3\hat{i} + 3\hat{j} + 2\hat{k}$ , por lo tanto,

$$\|\vec{a} - \vec{b}\| = \sqrt{9 + 9 + 4} = \sqrt{22}.$$

Por otra parte,

$$\vec{a} \cdot \vec{b} = 4 - 2 + 3 = 5.$$

Por último, como  $\|\vec{a}\| = \sqrt{1 + 4 + 9} = \sqrt{14}$ , por lo tanto

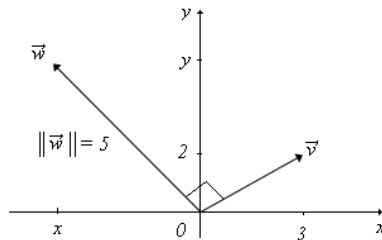
$$\hat{a} = \frac{1}{\sqrt{14}} (\hat{i} + 2\hat{j} + 3\hat{k}).$$

De esta manera,

$$\begin{aligned} \vec{v} &= \frac{\|\vec{a} - \vec{b}\|}{\|\vec{a} - \vec{b}\|} (\vec{a} \cdot \vec{b}) \hat{a} = (\sqrt{22}) (5) \left( \frac{1}{\sqrt{14}} (\hat{i} + 2\hat{j} + 3\hat{k}) \right) \\ &= \left( 5 \sqrt{\frac{11}{7}} \right) \hat{i} + \left( 10 \sqrt{\frac{11}{7}} \right) \hat{j} + \left( 15 \sqrt{\frac{11}{7}} \right) \hat{k}. \end{aligned}$$

5. Encuentra un vector  $\vec{w} \in \mathbb{R}^2$  que tenga norma 5 y sea perpendicular a  $\vec{v} = 3\hat{i} + 2\hat{j}$ .

Sea  $\vec{w} = x\hat{i} + y\hat{j}$  el vector que buscamos, representado en la siguiente figura.



## 1.1 Vectores

De acuerdo con el enunciado,  $\vec{w}$  debe satisfacer las siguientes dos condiciones

$$\begin{aligned}\|\vec{w}\| &= \sqrt{x^2 + y^2} = 5, \\ \vec{w} \cdot \vec{v} &= (3)(x) + (2)(y) = 0.\end{aligned}$$

De la segunda condición se tiene  $y = -3x/2$ , que sustituido en la primera condición implica

$$\sqrt{x^2 + \frac{9}{4}x^2} = 5.$$

De este modo,  $x = \pm \frac{10}{\sqrt{13}}$ . Así, existen dos vectores  $\vec{w}_1$  y  $\vec{w}_2$  que satisfacen las condiciones del problema,  $\vec{w}_1 = \frac{10}{\sqrt{13}}\hat{i} - \frac{15}{\sqrt{13}}\hat{j}$  y  $\vec{w}_2 = -\frac{10}{\sqrt{13}}\hat{i} + \frac{15}{\sqrt{13}}\hat{j}$ .

### Propiedades del producto escalar

Para todos  $\vec{a}, \vec{b}, \vec{c}, \vec{d} \in \mathbb{R}^n$  y  $\alpha \in \mathbb{R}$  se cumplen las siguientes propiedades:

1.  $\vec{a} \cdot \vec{a} \geq 0$  y  $\vec{a} \cdot \vec{a} = 0$  si y sólo si  $\vec{a} = \vec{0}$ .
2.  $\vec{a} \cdot \vec{b} = \vec{b} \cdot \vec{a}$
3.  $(\alpha \vec{a}) \cdot \vec{b} = \vec{a} \cdot (\alpha \vec{b}) = \alpha (\vec{a} \cdot \vec{b})$
4.  $\vec{a} \cdot (\vec{b} + \vec{c}) = \vec{a} \cdot \vec{b} + \vec{a} \cdot \vec{c}$
5.  $(\vec{a} + \vec{b}) \cdot (\vec{c} + \vec{d}) = \vec{a} \cdot \vec{c} + \vec{a} \cdot \vec{d} + \vec{b} \cdot \vec{c} + \vec{b} \cdot \vec{d}$
6.  $|\vec{a} \cdot \vec{b}| \leq \|\vec{a}\| \|\vec{b}\|$  Desigualdad de Cauchy-Schwarz

De la propiedad 6 se sigue un resultado interesante. Para ello, reescribimos la desigualdad de Cauchy-Schwarz de la siguiente manera

$$\begin{aligned}|\vec{a} \cdot \vec{b}| &\leq \|\vec{a}\| \|\vec{b}\| \\ -\|\vec{a}\| \|\vec{b}\| &\leq \vec{a} \cdot \vec{b} \leq \|\vec{a}\| \|\vec{b}\| \\ -\|\vec{a}\| \|\vec{b}\| &\leq \|\vec{a}\| \|\vec{b}\| \cos \theta \leq \|\vec{a}\| \|\vec{b}\| \\ -1 &\leq \cos \theta \leq 1,\end{aligned}$$

que permite generalizar el concepto de ángulo entre dos vectores en el espacio  $\mathbb{R}^n$ , como lo establece la siguiente definición.

**Definición.** El ángulo  $\theta$  entre dos vectores no nulos  $\vec{a}, \vec{b} \in \mathbb{R}^n$  está dado por

$$\theta = \cos^{-1} \left( \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \|\vec{b}\|} \right), \quad 0 \leq \theta \leq \pi.$$

## Capítulo 1 El Espacio $\mathbb{R}^n$

Ejemplos:

1. Encuentra el ángulo entre los vectores  $\vec{x} = \hat{i} + \hat{j}$  y  $\vec{y} = \hat{i}$ .

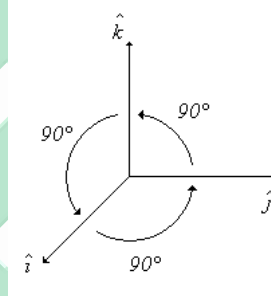
Como  $\|\vec{x}\| = \sqrt{2}$ ,  $\|\vec{y}\| = 1$  y  $\vec{x} \cdot \vec{y} = 1$ , por lo tanto  $\theta = \cos^{-1} \left( \frac{1}{\sqrt{2}} \right) = \frac{\pi}{4}$ .

2. Encuentra el ángulo entre los vectores  $\vec{x} = -\hat{i} - 2\hat{j}$  y  $\vec{y} = 2\hat{i} - \hat{j} + 3\hat{k}$ .

Como  $\|\vec{x}\| = \sqrt{5}$ ,  $\|\vec{y}\| = \sqrt{14}$  y  $\vec{x} \cdot \vec{y} = 0$ , por lo tanto  $\theta = \cos^{-1} \left( \frac{0}{\sqrt{5}\sqrt{14}} \right) = \frac{\pi}{2}$ . Concluimos entonces que  $\vec{x} \perp \vec{y}$ .

Por otra parte, en relación con el concepto de perpendicularidad, y en vista que los vectores base canónicos en  $\mathbb{R}^3$  son todos unitarios, es decir,  $\|\hat{i}\| = \|\hat{j}\| = \|\hat{k}\| = 1$ , se obtiene el siguiente resultado:

$$\begin{aligned}\hat{i} \cdot \hat{i} &= \hat{j} \cdot \hat{j} = \hat{k} \cdot \hat{k} = 1 \\ \hat{i} \cdot \hat{j} &= \hat{j} \cdot \hat{k} = \hat{k} \cdot \hat{i} = 0\end{aligned}$$



Utilizando este resultado, junto con la propiedad 5 del producto escalar, podemos llevar a cabo una diversidad de manipulaciones algebraicas. Por ejemplo, sin hacer uso de la ley de los cosenos podemos demostrar que  $\vec{a} \cdot \vec{b} = a_1b_1 + a_2b_2 + a_3b_3$ , para  $\vec{a}$  y  $\vec{b}$  en  $\mathbb{R}^3$ , de la siguiente manera:

$$\begin{aligned}\vec{a} \cdot \vec{b} &= (a_1\hat{i} + a_2\hat{j} + a_3\hat{k}) \cdot (b_1\hat{i} + b_2\hat{j} + b_3\hat{k}) \\ &= a_1b_1(\hat{i} \cdot \hat{i}) + a_1b_2(\hat{i} \cdot \hat{j}) + a_1b_3(\hat{i} \cdot \hat{k}) \\ &\quad + a_2b_1(\hat{j} \cdot \hat{i}) + a_2b_2(\hat{j} \cdot \hat{j}) + a_2b_3(\hat{j} \cdot \hat{k}) \\ &\quad + a_3b_1(\hat{k} \cdot \hat{i}) + a_3b_2(\hat{k} \cdot \hat{j}) + a_3b_3(\hat{k} \cdot \hat{k}) \\ &= a_1b_1 + a_2b_2 + a_3b_3.\end{aligned}$$

Asimismo, tomando en cuenta que el producto punto de un vector  $\vec{a}$  consigo mismo está dado por  $\vec{a} \cdot \vec{a} = \|\vec{a}\| \|\vec{a}\| \cos 0 = \|\vec{a}\| \|\vec{a}\| (1) = \|\vec{a}\|^2$ , se tiene

$$\|\vec{a}\| = \sqrt{\vec{a} \cdot \vec{a}}.$$

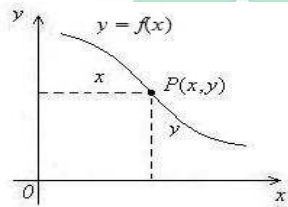
## 1.2 Curvas paramétricas. Vector tangente a una curva paramétrica

Así, por ejemplo,

$$\begin{aligned}\|2\vec{u} - 3\vec{v}\|^2 &= (2\vec{u} - 3\vec{v}) \cdot (2\vec{u} - 3\vec{v}) \\ &= 4(\vec{u} \cdot \vec{u}) - 6(\vec{u} \cdot \vec{v}) - 6(\vec{v} \cdot \vec{u}) + 9(\vec{v} \cdot \vec{v}) \\ &= 4\|\vec{u}\|^2 - 12(\vec{u} \cdot \vec{v}) + 9\|\vec{v}\|^2.\end{aligned}$$

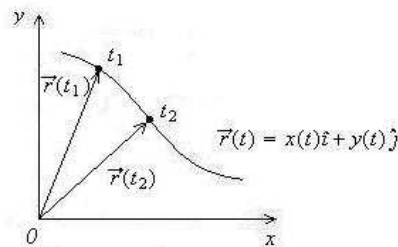
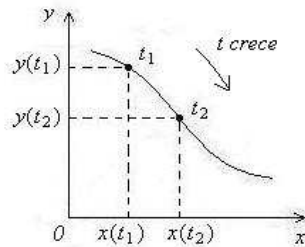
## 1.2 Curvas paramétricas. Vector tangente a una curva paramétrica

Una manera frecuente de definir una curva en el plano  $\mathbb{R}^2$  es la *representación cartesiana*, en donde la curva es el conjunto de puntos  $P(x, y)$  que satisfacen una ecuación de la forma  $y = f(x)$ .



Existen otras maneras para representar una curva en  $\mathbb{R}^2$ , que pueden resultar más convenientes que la cartesiana, dependiendo del tipo de simetrías de la curva o la naturaleza de sus posibles aplicaciones. Aquí nos interesa la llamada *representación paramétrica*, que además de proporcionar una información más detallada que en la forma cartesiana, puede extenderse fácilmente al caso general de curvas en  $\mathbb{R}^n$ .

La representación paramétrica de una curva en el plano  $\mathbb{R}^2$  expresa las coordenadas  $x$  y  $y$  de cada punto de la curva como funciones de una tercer variable, digamos  $t$ , que juega el papel de variable exógena o parámetro. Al ir cambiando de valores el parámetro  $t$ , se van generando nuevos puntos  $(x(t), y(t))$  de la curva, como se muestra en las siguientes figuras.



## Capítulo 1 El Espacio $\mathbb{R}^n$

La figura de la izquierda muestra la evolución de cada una de las coordenadas  $x(t)$  y  $y(t)$  al incrementarse  $t$ . La figura de la derecha presenta el mismo razonamiento pero en un lenguaje vectorial, considerando para cada  $t$  la evolución del vector de posición

$$\vec{r}(t) = x(t)\hat{i} + y(t)\hat{j}.$$

**Definición.** Una *curva paramétrica* o *trayectoria* es una función vectorial,  $\vec{r} : S \subset \mathbb{R} \rightarrow \mathbb{R}^n$ , que a cada número  $t \in S$  le asigna un único vector  $\vec{r}(t) \in \mathbb{R}^n$ .

De acuerdo con nuestra discusión anterior, en el caso del plano  $\mathbb{R}^2$  una curva paramétrica se representa mediante una función vectorial  $\vec{r} : \mathbb{R} \rightarrow \mathbb{R}^2$ , de la forma

$$\vec{r}(t) = x(t)\hat{i} + y(t)\hat{j},$$

en donde  $x$  y  $y$  son funciones del parámetro  $t$  en  $\mathbb{R}$ . Similarmente, en el caso del espacio  $\mathbb{R}^3$  una curva paramétrica se representa mediante una función vectorial  $\vec{r} : \mathbb{R} \rightarrow \mathbb{R}^3$ , de la forma

$$\vec{r}(t) = x(t)\hat{i} + y(t)\hat{j} + z(t)\hat{k},$$

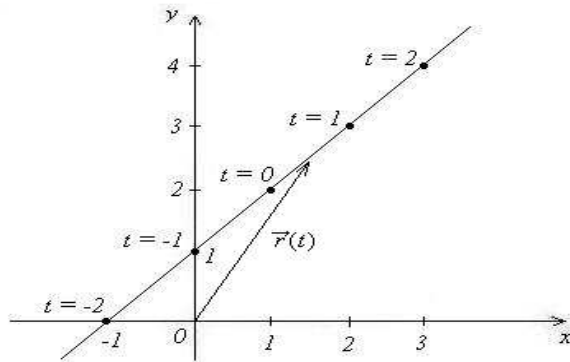
en donde  $x, y$  y  $z$  son funciones del parámetro  $t$  en  $\mathbb{R}$ . Un argumento similar se sigue para curvas en  $\mathbb{R}^n$ ,  $n \geq 4$ . Cabe mencionar, por último, que la parametrización de una curva no es única, como se ilustra en el ejemplo 2 a continuación.

Ejemplos:

1. Identifica la curva  $\vec{r}(t) = x(t)\hat{i} + y(t)\hat{j}$  en  $\mathbb{R}^2$ , con

$$\begin{aligned} x(t) &= 1 + t \\ y(t) &= 2 + t, \quad t \in \mathbb{R}. \end{aligned}$$

Asignando diferentes valores al parámetro  $t$  se obtiene la recta mostrada en la figura.



## 1.2 Curvas paramétricas. Vector tangente a una curva paramétrica

Efectivamente, al eliminar el parámetro  $t$  en el sistema  $x = 1 + t$ ,  $y = 2 + t$ , se obtiene la ecuación cartesiana de esta curva, dada por la recta

$$y = x + 1.$$

2. Identifica la curva  $\vec{r}(s) = x(s)\hat{i} + y(s)\hat{j}$  en  $\mathbb{R}^2$ , con

$$\begin{aligned} x(s) &= 1 - s \\ y(s) &= 2 - s, \quad s \in \mathbb{R}. \end{aligned}$$

Eliminando el parámetro  $s$  en el sistema  $x(s) = 1 - s$ ,  $y(s) = 2 - s$ , se obtiene la ecuación cartesiana  $y = x + 1$ , de modo que se trata de la misma curva que en ejemplo 1.

3. Identifica la curva  $\vec{r}(\theta) = x(\theta)\hat{i} + y(\theta)\hat{j}$  en  $\mathbb{R}^2$ , con

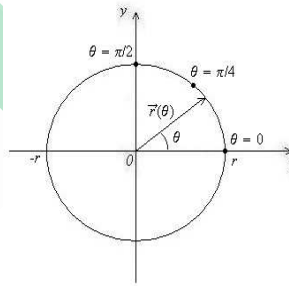
$$\begin{aligned} x(\theta) &= r \cos \theta \\ y(\theta) &= r \operatorname{sen} \theta, \quad 0 \leq \theta < 2\pi, \quad r > 0 \text{ constante.} \end{aligned}$$

Aquí no es fácil eliminar el parámetro  $\theta$  mediante métodos algebraicos. En lugar de esto, conviene utilizar identidades trigonométricas, de la siguiente manera.

Tomando en cuenta que  $\cos^2 \theta + \operatorname{sen}^2 \theta = 1$ , se tiene  $\left(\frac{x}{r}\right)^2 + \left(\frac{y}{r}\right)^2 = 1$ . Así, la ecuación cartesiana de la curva en este caso corresponde a la circunferencia

$$x^2 + y^2 = r^2,$$

como se muestra en la figura. Ahí se ilustra cómo se van generando los puntos de esta curva a medida que va cambiando el parámetro  $\theta$ .

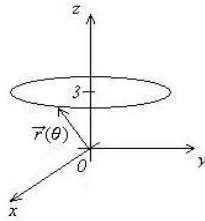


4. Identifica la curva  $\vec{r}(\theta) = x(\theta)\hat{i} + y(\theta)\hat{j} + z(\theta)\hat{k}$  en  $\mathbb{R}^3$ , con

$$\begin{aligned} x(\theta) &= \cos \theta \\ y(\theta) &= \operatorname{sen} \theta \\ z(\theta) &= 3, \quad 0 \leq \theta < 2\pi. \end{aligned}$$

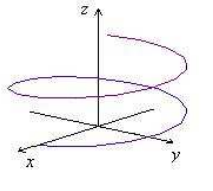
Para la curva  $\vec{r}(\theta) = \cos \theta \hat{i} + \operatorname{sen} \theta \hat{j} + 3 \hat{k}$ ,  $0 \leq \theta < 2\pi$ , las primeras dos componentes describen una circunferencia, mientras que la tercera permanece

constante (igual a 3). Así, la curva correspondiente es una circunferencia que está elevada 3 unidades en el eje vertical.



5. Identifica la curva  $\vec{r}(\theta) = \cos \theta \hat{i} + \sin \theta \hat{j} + \theta \hat{k}$  en  $\mathbb{R}^3$ , con  $0 \leq \theta < \infty$ .

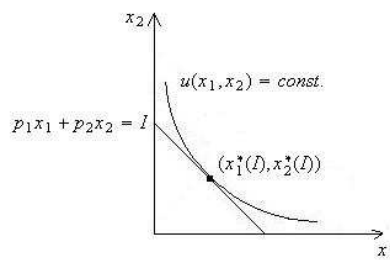
Para esta curva, las primeras dos componentes describen una circunferencia, mientras que la tercera se incrementa continuamente de manera lineal. La curva obtenida se conoce como hélice (espiral), como se ilustra en la figura.



6. Como una aplicación a economía, considera el problema de maximización de la utilidad  $u(x_1, x_2)$  correspondiente a una canasta  $(x_1, x_2)$  de dos bienes, con precios fijos  $p_1$  y  $p_2$ . Si se dispone de un ingreso  $I$ , se tendrá una restricción presupuestal dada por  $p_1x_1 + p_2x_2 = I$ . Esto nos lleva a un problema de optimización restringida, de la forma

$$\begin{aligned} &\text{maximizar} && u(x_1, x_2) \\ &\text{sujeto a} && p_1x_1 + p_2x_2 = I. \end{aligned}$$

Como veremos en el capítulo 5, el óptimo  $(x_1^*, x_2^*)$  de este problema ocurre en el punto de tangencia de la recta presupuestal  $p_1x_1 + p_2x_2 = I$  con alguna curva de indiferencia de la función  $u$ , lo que se conoce como la condición de equimarginalidad.



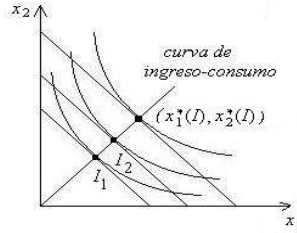


## 1.2 Curvas paramétricas. Vector tangente a una curva paramétrica

De esta manera, la canasta óptima depende del nivel de ingreso  $I$ , es decir,

$$(x_1^*, x_2^*) = (x_1^*(I), x_2^*(I)).$$

Aquí el ingreso  $I$  es un parámetro que al cambiar de valor hace que el punto óptimo  $(x_1^*(I), x_2^*(I))$  se mueva a lo largo de distintas curvas de indiferencia. La trayectoria que sigue la canasta óptima como función del parámetro  $I$  se conoce como *curva de ingreso-consumo* o *senda de expansión del consumo*.



La curva de ingreso-consumo es la curva paramétrica  $\vec{r} : \mathbb{R} \rightarrow \mathbb{R}^2$  que para cada valor del ingreso  $I \in \mathbb{R}^+$  le asigna una canasta óptima  $\vec{r} \in \mathbb{R}^2$ , dada por

$$\vec{r}(I) = x_1^*(I) \hat{i} + x_2^*(I) \hat{j}.$$

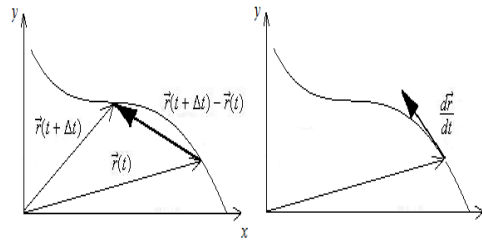
Como una curva paramétrica  $\vec{r}(t)$  es función del parámetro  $t$ , tiene sentido preguntarse sobre su razón de cambio o derivada,  $d\vec{r}/dt$ , con respecto al parámetro  $t$ . Para ello, primero necesitaríamos definir los conceptos de límite y continuidad, cuya definición formal omitiremos aquí.

**Definición.** Sea  $\vec{r}(t)$  una función vectorial, con  $\vec{r} : S \subset \mathbb{R} \rightarrow \mathbb{R}^n$ . La *derivada* de  $\vec{r}(t)$  con respecto a  $t$  es la función vectorial  $d\vec{r}/dt$  dada por

$$\frac{d\vec{r}(t)}{dt} = \lim_{\Delta t \rightarrow 0} \frac{\vec{r}(t + \Delta t) - \vec{r}(t)}{\Delta t},$$

cuando este límite existe.

Como se ilustra en la siguiente figura, de esta definición se sigue que el vector  $d\vec{r}/dt$  es un *vector tangente* a la curva  $\vec{r}(t)$ , para cada  $t$ .



Observa que el vector tangente  $d\vec{r}/dt$  no necesariamente es perpendicular a la curva  $\vec{r}(t)$  en cada valor del parámetro  $t$ .

El cálculo de la derivada  $d\vec{r}/dt$  es muy sencillo. Por ejemplo, para una función vectorial  $\vec{r}(t) = f(t)\hat{i} + g(t)\hat{j} + h(t)\hat{k}$  en  $\mathbb{R}^3$ , se tiene

$$\begin{aligned}\frac{d\vec{r}(t)}{dt} &= \lim_{\Delta t \rightarrow 0} \frac{\vec{r}(t + \Delta t) - \vec{r}(t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{\left(f(t + \Delta t)\hat{i} + g(t + \Delta t)\hat{j} + h(t + \Delta t)\hat{k}\right) - \left(f(t)\hat{i} + g(t)\hat{j} + h(t)\hat{k}\right)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{f(t + \Delta t) - f(t)}{\Delta t} \hat{i} + \lim_{\Delta t \rightarrow 0} \frac{g(t + \Delta t) - g(t)}{\Delta t} \hat{j} + \lim_{\Delta t \rightarrow 0} \frac{h(t + \Delta t) - h(t)}{\Delta t} \hat{k} \\ &= \frac{df(t)}{dt} \hat{i} + \frac{dg(t)}{dt} \hat{j} + \frac{dh(t)}{dt} \hat{k},\end{aligned}$$

siempre y cuando  $f, g$  y  $h$  sean todas funciones diferenciables de  $t$ .

**Teorema.** Sea  $\vec{r}(t) = f_1(t)\hat{e}_1 + f_2(t)\hat{e}_2 + \dots + f_n(t)\hat{e}_n$  una función vectorial en  $\mathbb{R}^n$ , con  $(f_1, f_2, \dots, f_n) : \mathbb{R} \rightarrow \mathbb{R}$  derivables y  $\hat{e}_1, \dots, \hat{e}_n$  la base canónica en  $\mathbb{R}^n$ . La derivada de  $\vec{r}(t)$  con respecto a  $t$  es la función vectorial  $d\vec{r}/dt$  dada por

$$\frac{d\vec{r}(t)}{dt} = \frac{df_1(t)}{dt} \hat{e}_1 + \frac{df_2(t)}{dt} \hat{e}_2 + \dots + \frac{df_n(t)}{dt} \hat{e}_n.$$

Ejemplos:

1. Encuentra la derivada de  $\vec{r}(t) = (te^{-3(t-1)})\hat{i} + (t \ln t)\hat{j}$ ,  $t > 0$ , en  $t = 1$ .

Para cada  $t > 0$  la derivada  $d\vec{r}(t)/dt$  es la función vectorial

$$\frac{d\vec{r}(t)}{dt} = (1 - 3t)e^{-3(t-1)}\hat{i} + (1 + \ln t)\hat{j}.$$

Así, en  $t = 1$  se tiene

$$\left. \frac{d\vec{r}(t)}{dt} \right|_{t=1} = -2\hat{i} + \hat{j}.$$

2. Encuentra un vector tangente a la circunferencia  $\vec{r}(\theta) = \cos \theta \hat{i} + \sin \theta \hat{j}$  en el punto correspondiente a  $\theta = 0$ . Ilustra con una figura.

Por una parte, el punto correspondiente a  $\theta = 0$  es

$$\vec{r}(0) = \hat{i}.$$

## 1.2 Curvas paramétricas. Vector tangente a una curva paramétrica

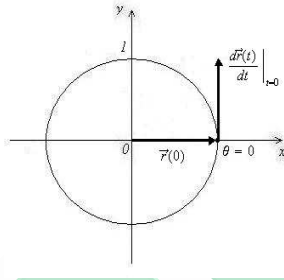
Por otra parte, la derivada  $d\vec{r}(\theta)/d\theta$  es la función vectorial

$$\frac{d\vec{r}(\theta)}{d\theta} = -\operatorname{sen}\theta \hat{i} + \cos\theta \hat{j},$$

que en  $\theta = 0$  es el vector

$$\left. \frac{d\vec{r}(\theta)}{d\theta} \right|_{\theta=0} = \hat{j}.$$

Por lo tanto, el vector tangente a la curva  $\vec{r}(\theta)$  en el punto  $\vec{r}(0) = \hat{i}$  es  $\vec{r}'(0) = \hat{j}$ , como se ilustra en la figura.



### Reglas de diferenciación de curvas paramétricas

Sean  $\vec{u} : \mathbb{R} \rightarrow \mathbb{R}^n$ ,  $\vec{v} : \mathbb{R} \rightarrow \mathbb{R}^n$  y  $\alpha : \mathbb{R} \rightarrow \mathbb{R}$  funciones diferenciables de  $t$ . Sean  $k \in \mathbb{R}$  y  $\vec{c} \in \mathbb{R}^n$  constantes. Entonces se cumplen las siguientes propiedades:

1.  $\frac{d\vec{c}}{dt} = \vec{0}$
2.  $\frac{d[k\vec{u}(t)]}{dt} = k \frac{d\vec{u}(t)}{dt}$
3.  $\frac{d[\vec{u}(t) + \vec{v}(t)]}{dt} = \frac{d\vec{u}(t)}{dt} + \frac{d\vec{v}(t)}{dt}$
4.  $\frac{d[\alpha(t)\vec{u}(t)]}{dt} = \alpha(t) \frac{d\vec{u}(t)}{dt} + \frac{d\alpha(t)}{dt} \vec{u}(t)$
5.  $\frac{d[\vec{u}(t) \cdot \vec{v}(t)]}{dt} = \vec{u}(t) \cdot \frac{d\vec{v}(t)}{dt} + \frac{d\vec{u}(t)}{dt} \cdot \vec{v}(t)$

Como una consecuencia de la regla 5 se sigue que si  $\vec{r}(t)$  es una función vectorial con norma constante,  $\|\vec{r}(t)\| = c$  ( $c$  constante), entonces

$$\vec{r} \cdot \frac{d\vec{r}}{dt} = 0.$$

Demostración:

Sea  $\vec{r}(t)$  una función vectorial tal que  $\|\vec{r}(t)\| = c$ , con  $c$  un real no negativo. Por lo tanto,

$$\begin{aligned}\|\vec{r}(t)\|^2 &= c^2 \\ \vec{r}(t) \cdot \vec{r}(t) &= c^2 \\ \frac{d[\vec{r}(t) \cdot \vec{r}(t)]}{dt} &= 0 \\ \vec{r}(t) \cdot \frac{d\vec{r}(t)}{dt} + \vec{r}(t) \cdot \frac{d\vec{r}(t)}{dt} &= 0 \\ 2 \vec{r}(t) \cdot \frac{d\vec{r}(t)}{dt} &= 0 \\ \vec{r}(t) \cdot \frac{d\vec{r}(t)}{dt} &= 0.\end{aligned}$$

En otras palabras, si la trayectoria  $\vec{r}(t)$  tiene norma constante, el vector de posición  $\vec{r}$  es ortogonal al vector tangente  $d\vec{r}/dt$ , para cada  $t$ .

Así, por ejemplo, para el caso de una trayectoria circular

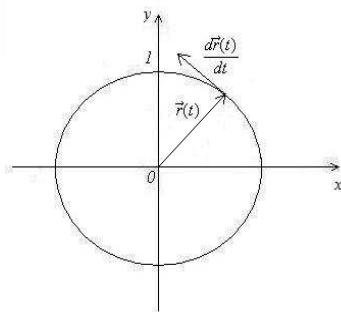
$$\vec{r}(t) = (\cos t) \hat{i} + (\sin t) \hat{j},$$

que siempre presenta norma constante

$$\|\vec{r}(t)\| = \sqrt{\cos^2 t + \sin^2 t} = 1,$$

se tiene

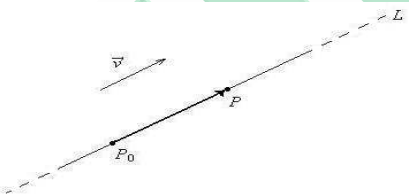
$$\begin{aligned}\vec{r}(t) \cdot \frac{d\vec{r}(t)}{dt} &= (\cos t \hat{i} + \sin t \hat{j}) \cdot (-\sin t \hat{i} + \cos t \hat{j}) \\ &= -\sin t \cos t + \sin t \cos t = 0.\end{aligned}$$



### 1.3 Rectas en el espacio. Segmento de recta

Estamos acostumbrados a escribir la ecuación de la recta como  $y = mx + b$ , donde  $m$  representa la pendiente o dirección de la recta y  $b$  su ordenada al origen. Sin embargo, esta forma para la ecuación de la recta sólo es válida para rectas en el plano  $\mathbb{R}^2$ . En el caso general de rectas en  $\mathbb{R}^n$  su ecuación ya no puede expresarse en términos de una sola pendiente, sino que es necesario tomar en cuenta la orientación de la recta en relación con cada uno de los  $n$  diferentes ejes coordenados (cosenos directores). Una manera sencilla de introducir la orientación es utilizando vectores, lo que nos llevará a una representación paramétrica para la recta, como se expone a continuación.

Para encontrar la ecuación de una recta  $L$  en el espacio general  $\mathbb{R}^n$  basta con proporcionar algún punto conocido  $P_0$  de la recta y un vector  $\vec{v}$  que sea paralelo al conjunto de puntos  $P$  de la recta.



La recta  $L$  es el lugar geométrico de todos los puntos  $P$  en  $\mathbb{R}^n$  tales que  $\overrightarrow{P_0P}$  es paralelo al vector de dirección  $\vec{v} \in \mathbb{R}^n$ , es decir,

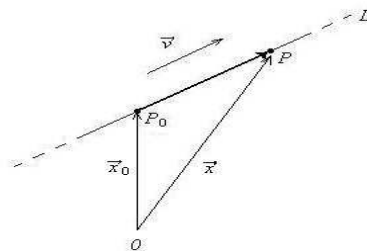
$$\overrightarrow{P_0P} \parallel \vec{v}.$$

Esto que implica que ambos vectores son múltiplos entre sí, de modo que existe algún escalar  $t \in \mathbb{R}$ , tal que

$$\overrightarrow{P_0P} = t\vec{v}.$$

Esta última ecuación puede expresarse de manera alternativa, introduciendo un origen de coordenadas,  $O$ , a partir del cual los puntos  $P_0$  y  $P$  están localizados por los vectores de posición

$$\vec{x}_0 = \overrightarrow{OP_0} \text{ y } \vec{x} = \overrightarrow{OP}.$$



## Capítulo 1 El Espacio $\mathbb{R}^n$

De esta manera, se tiene

$$\overrightarrow{P_0P} = \vec{x} - \vec{x}_0,$$

de modo que la ecuación de la recta se convierte en

$$\vec{x} - \vec{x}_0 = t\vec{v},$$

o, equivalentemente,

$$\vec{x} = \vec{x}_0 + t\vec{v}.$$

**Definición.** La ecuación vectorial paramétrica de la recta en  $\mathbb{R}^n$  que contiene al punto  $\vec{x}_0 \in \mathbb{R}^n$  y es paralela al vector no nulo  $\vec{v} \in \mathbb{R}^n$  es

$$\vec{x} = \vec{x}_0 + t\vec{v},$$

donde  $\vec{x} \in \mathbb{R}^n$  y  $t \in \mathbb{R}$ .

La ecuación vectorial paramétrica de una recta también puede escribirse en términos de sus componentes escalares. En el caso particular de una recta en  $\mathbb{R}^3$ , si  $\vec{v} = a\hat{i} + b\hat{j} + c\hat{k}$  denota el vector de dirección,  $\vec{x}_0 = x_0\hat{i} + y_0\hat{j} + z_0\hat{k}$  el punto conocido y  $\vec{x} = x\hat{i} + y\hat{j} + z\hat{k}$  el punto libre de la recta, la ecuación vectorial  $\vec{x} = \vec{x}_0 + t\vec{v}$  se convierte en

$$x\hat{i} + y\hat{j} + z\hat{k} = (x_0\hat{i} + y_0\hat{j} + z_0\hat{k}) + t(a\hat{i} + b\hat{j} + c\hat{k}) = (x_0 + at)\hat{i} + (y_0 + bt)\hat{j} + (z_0 + ct)\hat{k}.$$

Igualando término a término ambos lados de la ecuación se obtienen tres ecuaciones escalares, conocidas como las *ecuaciones paramétricas* de la recta.

**Definición.** Las *ecuaciones escalares paramétricas* de la recta en  $\mathbb{R}^3$  que contiene al punto  $P_0(x_0, y_0, z_0)$  y es paralela al vector no nulo  $\vec{v} = a\hat{i} + b\hat{j} + c\hat{k}$  son

$$x = x_0 + at, \quad y = y_0 + bt, \quad z = z_0 + ct, \quad t \in \mathbb{R}.$$

Similarmente, las ecuaciones escalares de una recta en  $\mathbb{R}^2$  son

$$x = x_0 + at, \quad y = y_0 + bt, \quad t \in \mathbb{R}.$$

Ejemplos:

1. Escribe la ecuación vectorial paramétrica de la recta en  $\mathbb{R}^2$  que contiene al punto  $\vec{x}_0 = \hat{i} + 2\hat{j}$  y es paralela al vector  $\vec{v} = \hat{i} + \hat{j}$ . Luego escribe las ecuaciones escalares paramétricas de esta recta.

La ecuación vectorial es  $\vec{x} = \vec{x}_0 + t\vec{v} = (\hat{i} + 2\hat{j}) + t(\hat{i} + \hat{j})$ , esto es

$$\vec{x} = (\hat{i} + 2\hat{j}) + t(\hat{i} + \hat{j}), \quad t \in \mathbb{R}.$$

### 1.3 Rectas en el espacio. Segmento de recta

Las ecuaciones escalares son

$$x = 1 + t, \quad y = 2 + t, \quad t \in \mathbb{R}.$$

Observa que ésta es la misma recta que la del ejemplo 1 de la sección 1.2.

2. Halla las ecuaciones escalares paramétricas de la recta en  $\mathbb{R}^3$  con la información dada:

a) Contiene al punto  $P(1, -2, 7)$  y es paralela al vector  $\vec{v} = 5\hat{i} + 3\hat{j} - \hat{k}$ .

En este caso, se tiene simplemente

$$x = 1 + 5t, \quad y = -2 + 3t, \quad z = 7 - t, \quad t \in \mathbb{R}.$$

b) Contiene al origen y es paralela al vector  $\vec{v} = 4\hat{i} - 3\hat{j}$ .

Como el origen es el punto  $O(0, 0, 0)$ , por lo tanto las ecuaciones son

$$x = 4t, \quad y = -3t, \quad z = 0, \quad t \in \mathbb{R}.$$

c) Contiene al punto  $Q(1, 2, 3)$  y es paralela al eje  $y$ .

Podemos tomar  $\vec{v} = \hat{j}$  (o cualquier múltiplo de éste), de modo que

$$x = 1, \quad y = 2 + t, \quad z = 3, \quad t \in \mathbb{R}.$$

3. Encuentra las ecuaciones escalares paramétricas de la recta que contiene los puntos  $A(-2, 1, 4)$  y  $B(-1, 0, 3)$ . Asimismo, proporciona algunos otros puntos contenidos en esta recta.

Podemos tomar, por ejemplo,  $\vec{v} = \overrightarrow{AB} = \hat{i} - \hat{j} - \hat{k}$ , y el punto conocido puede ser tanto  $A$  como  $B$ . Así, cualquiera de las siguientes respuestas es válida

$$x = -2 + t, \quad y = 1 - t, \quad z = 4 - t, \quad t \in \mathbb{R},$$

$$x = -1 + t, \quad y = -t, \quad z = 3 - t, \quad t \in \mathbb{R}.$$

Por otra parte, para obtener cualquiera de los puntos de esta recta basta con asignar valores arbitrarios al parámetro  $t$ . Así, por ejemplo, si en la primer respuesta tomamos  $t = 2$  obtenemos el punto  $P_1(0, -1, 2)$ , o bien, si tomamos  $t = -1$  generamos el punto  $P_2(-3, 2, 5)$ , etc. Nota que el punto  $A$  se obtiene cuando  $t = 0$ , y el punto  $B$ , cuando  $t = 1$ .

4. Encuentra las ecuaciones escalares paramétricas de la recta tangente a la curva  $\vec{r}(\alpha) = \alpha\hat{i} + \alpha^2\hat{j}$  en  $\mathbb{R}^2$ ,  $\alpha \in \mathbb{R}$ , en el punto con  $\alpha = 1$ .

Primero notamos que un punto conocido  $\vec{x}_0$  de la recta tangente es, precisamente, su punto de tangencia con la curva  $\vec{r}(\alpha)$  en  $\alpha = 1$ , es decir,

$$\vec{x}_0 = \vec{r}(1) = (1, 1).$$

## Capítulo 1 El Espacio $\mathbb{R}^n$

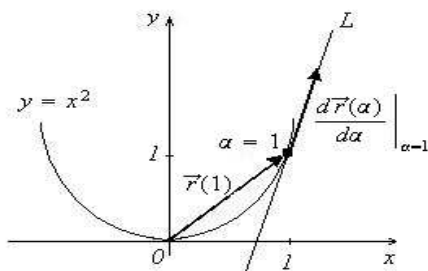
Por otra parte, sabemos que un vector tangente a la curva  $\vec{r}(\alpha)$  es  $d\vec{r}(\alpha)/d\alpha = \hat{i} + 2\alpha\hat{j}$ , para cada  $\alpha \in \mathbb{R}$ . Así, la dirección  $\vec{v}$  de la recta tangente a la curva en  $\alpha = 1$  puede tomarse como

$$\vec{v} = \left. \frac{d\vec{r}(\alpha)}{d\alpha} \right|_{\alpha=1} = (1, 2).$$

Así, las ecuaciones paramétricas de la recta tangente a  $\vec{r}(\alpha)$  en  $\alpha = 1$  son

$$\begin{aligned} x &= 1 + t \\ y &= 1 + 2t, \quad t \in \mathbb{R}. \end{aligned}$$

La siguiente figura muestra la curva  $\vec{r}(\alpha)$  y su recta tangente  $L$  en  $\alpha = 1$ . En este ejemplo, la curva paramétrica es la parábola  $y = x^2$ , como se deduce a partir de  $x = \alpha$  y  $y = \alpha^2$ .



5. Encuentra las ecuaciones paramétricas de los ejes de coordenadas en  $\mathbb{R}^3$ .

Como lo muestra la figura, una posible representación para las ecuaciones paramétricas de los ejes coordenados está dada por:

i) Eje  $x$ :  $O(0, 0, 0)$ ,  $\vec{v} = \hat{i}$

$$x = t, \quad y = 0, \quad z = 0, \quad t \in \mathbb{R}.$$

ii) Eje  $y$ :  $O(0, 0, 0)$ ,  $\vec{v} = \hat{j}$

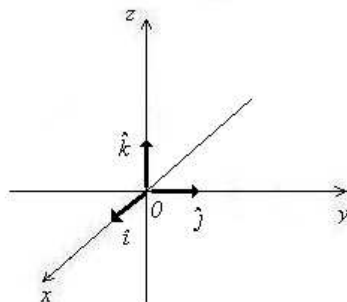
$$x = 0, \quad y = t, \quad z = 0, \quad t \in \mathbb{R}.$$

iii) Eje  $z$ :  $O(0, 0, 0)$ ,  $\vec{v} = \hat{k}$

$$x = 0, \quad y = 0, \quad z = t, \quad t \in \mathbb{R}.$$



### 1.3 Rectas en el espacio. Segmento de recta



Por último, como sucede con cualquier representación paramétrica, las ecuaciones paramétricas de la recta no admiten una única representación. Esto se debe a que cualquier punto de la recta puede seleccionarse como el punto conocido  $P_0$ , y que cualquier múltiplo del vector de dirección  $\vec{v}$  es también paralelo a la recta. Así, por ejemplo, la recta representada por las ecuaciones

$$\begin{aligned} x &= 1 + t \\ y &= 1 - t, \quad t \in \mathbb{R}, \end{aligned}$$

es la misma que la descrita por cualquiera de las siguientes ecuaciones:

$$\begin{aligned} x &= 2 + s & x &= u & x &= 1 - 3w \\ y &= -s, \quad s \in \mathbb{R}, & y &= 2 - u, \quad u \in \mathbb{R}, & y &= 1 + 3w, \quad w \in \mathbb{R}. \end{aligned}$$

Una forma alternativa de la ecuación de la recta, válida en general para rectas en  $\mathbb{R}^n$ , es la llamada *forma simétrica*, que se discute a continuación para el caso de rectas en  $\mathbb{R}^3$ . Para obtener la forma simétrica de la ecuación de la recta, se despeja el parámetro  $t$  en cada una de las tres ecuaciones  $x = x_0 + at$ ,  $y = y_0 + bt$ ,  $z = z_0 + ct$ , es decir,

$$t = \frac{x - x_0}{a}, \quad t = \frac{y - y_0}{b}, \quad t = \frac{z - z_0}{c},$$

y luego se igualan entre sí (claro está, suponiendo que  $a \neq 0$ ,  $b \neq 0$  y  $c \neq 0$ ), como se define a continuación.

**Definición.** La *forma simétrica de la ecuación de la recta en  $\mathbb{R}^3$*  que contiene al punto  $P_0(x_0, y_0, z_0)$  y es paralela al vector  $\vec{v} = a\hat{i} + b\hat{j} + c\hat{k}$ , con  $a \neq 0$ ,  $b \neq 0$  y  $c \neq 0$ , es

$$\frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c}.$$

## Capítulo 1 El Espacio $\mathbb{R}^n$

Por ejemplo, la forma simétrica de las ecuaciones

$$x = 1 + 3t, \quad y = 4t, \quad z = -5 - 2t, \quad t \in \mathbb{R},$$

está dada por

$$\frac{x-1}{3} = \frac{y}{4} = \frac{z+5}{-2}.$$

Nota que esta última no es una ecuación, sino más bien son tres ecuaciones, a saber,

$$\frac{x-1}{3} = \frac{y}{4}, \quad \frac{y}{4} = \frac{z+5}{-2} \quad \text{y} \quad \frac{x-1}{3} = \frac{z+5}{-2}.$$

Cuando alguna de las componentes del vector  $\vec{v}$  es igual a cero, es posible aún contar con una forma simétrica para la ecuación de la recta correspondiente, de la siguiente manera:

caso: forma simétrica:

$$a = 0 \quad \frac{y-y_0}{b} = \frac{z-z_0}{c}, \quad x = x_0$$

$$b = 0 \quad \frac{x-x_0}{a} = \frac{z-z_0}{c}, \quad y = y_0$$

$$c = 0 \quad \frac{x-x_0}{a} = \frac{y-y_0}{b}, \quad z = z_0$$

Vale la pena señalar que en el caso particular de rectas en  $\mathbb{R}^2$  la correspondiente forma simétrica,

$$\frac{x-x_0}{a} = \frac{y-y_0}{b},$$

puede reescribirse como

$$y = \frac{b}{a}(x-x_0) + y_0,$$

que es precisamente la ecuación punto-pendiente de la recta ( $m = b/a$ ), con la que seguramente estás familiarizado. No olvides, sin embargo, que este resultado sólo es válido para rectas en  $\mathbb{R}^2$ . Así, por ejemplo, para la recta

$$\begin{aligned} x &= 1 + 3t \\ y &= -2 - 5t, \quad t \in \mathbb{R}, \end{aligned}$$

cuya ecuación en su forma simétrica es

$$\frac{x-1}{3} = \frac{y+2}{-5},$$

se obtiene la ecuación cartesiana

$$y = -\frac{5}{3}x - \frac{1}{3}.$$

## 1.4 Planos e hiperplanos

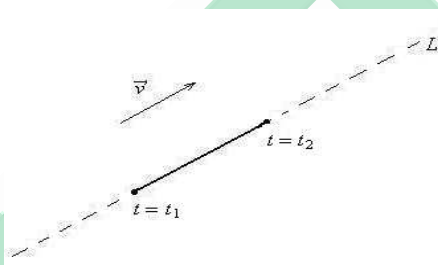
### Segmento de recta

Hemos visto ya que las ecuaciones paramétricas de una recta en el espacio contienen un parámetro libre,  $t \in \mathbb{R}$ . Cada vez que  $t$  toma un valor diferente en los reales, se genera un nuevo punto a lo largo de la recta infinita. Sin embargo, si en lugar de tener la condición  $t \in \mathbb{R}$ , el parámetro  $t$  se limitara a tomar valores dentro de un intervalo  $t_1 \leq t \leq t_2$  en los reales, entonces éste ya no generaría todos los puntos de la recta infinita, sino tan sólo un segmento de la recta.

**Definición.** Dada la recta  $L$  en  $\mathbb{R}^n$  que contiene al punto  $\vec{x}_0 \in \mathbb{R}^n$  y es paralela al vector no nulo  $\vec{v} \in \mathbb{R}^n$ , la ecuación

$$\vec{x} = \vec{x}_0 + t\vec{v}, \quad t_1 \leq t \leq t_2,$$

con  $t_1$  y  $t_2$  fijos, determina un *segmento de la recta*  $L$ .



Ejemplo:

Halla la ecuación del segmento de la recta que une los puntos  $P(-3, 2, -3)$  y  $Q(1, -1, 4)$ .

Lo más sencillo es definir el vector de dirección  $\vec{v}$  como

$$\vec{v} = \overrightarrow{PQ} = 4\hat{i} - 3\hat{j} + 7\hat{k}.$$

De esta manera, el segmento de recta que une a  $P$  y  $Q$  queda descrito por

$$x = -3 + 4t, \quad y = 2 - 3t, \quad z = -3 + 7t, \quad 0 \leq t \leq 1.$$

En efecto, cuando  $t = 0$  se obtiene el punto  $P$ , cuando  $t = 1$  se obtiene el punto  $Q$  y para  $0 < t < 1$  se generan todos los puntos intermedios entre  $P$  y  $Q$ .

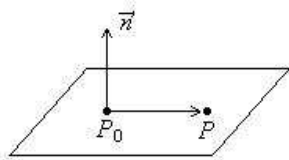
## 1.4 Planos e hiperplanos

Se trata de encontrar la ecuación del plano  $\pi$  en el espacio  $\mathbb{R}^3$  que contiene a un punto conocido  $P_0$  y es perpendicular a un vector normal no nulo,  $\vec{n}$ . En ese caso,  $\pi$  es el conjunto de puntos  $P$  para los cuales se cumple que

$$\overrightarrow{P_0P} \perp \vec{n}.$$

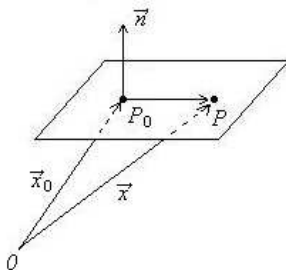
En otras palabras,

$$\vec{n} \cdot \overrightarrow{P_0P} = 0.$$



Introduciendo un origen de coordenadas,  $O$ , se puede definir los vectores  $\vec{x} = \overrightarrow{OP}$  y  $\vec{x}_0 = \overrightarrow{OP_0}$ , de modo que  $\overrightarrow{P_0P} = \vec{x} - \vec{x}_0$ . Así, la condición anterior se convierte en

$$\vec{n} \cdot (\vec{x} - \vec{x}_0) = 0.$$



**Definición.** La ecuación del plano que contiene al punto  $\vec{x}_0 \in \mathbb{R}^3$  y es perpendicular al vector no nulo  $\vec{n} \in \mathbb{R}^3$  es

$$\vec{n} \cdot (\vec{x} - \vec{x}_0) = 0.$$

La forma  $\vec{n} \cdot (\vec{x} - \vec{x}_0) = 0$  para la ecuación del plano puede reescribirse en términos más simples si se conocen las componentes de los vectores  $\vec{n}$  y  $\vec{x}_0$ . En efecto, si se sabe que  $\vec{n} = a\hat{i} + b\hat{j} + c\hat{k}$ ,  $\vec{x}_0 = x_0\hat{i} + y_0\hat{j} + z_0\hat{k}$  y  $\vec{x} = x\hat{i} + y\hat{j} + z\hat{k}$ , entonces

$$\vec{x} - \vec{x}_0 = (x - x_0)\hat{i} + (y - y_0)\hat{j} + (z - z_0)\hat{k}.$$

De esta manera, la ecuación del plano está dada por

$$\begin{aligned} \vec{n} \cdot (\vec{x} - \vec{x}_0) &= (a\hat{i} + b\hat{j} + c\hat{k}) \cdot ((x - x_0)\hat{i} + (y - y_0)\hat{j} + (z - z_0)\hat{k}) \\ &= a(x - x_0) + b(y - y_0) + c(z - z_0) = 0. \end{aligned}$$

**Definición.** La ecuación cartesiana del plano en  $\mathbb{R}^3$  que contiene al punto  $P_0(x_0, y_0, z_0)$  y es perpendicular al vector no nulo  $\vec{n} = a\hat{i} + b\hat{j} + c\hat{k}$  es

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0.$$

## 1.4 Planos e hiperplanos

Por ejemplo, la ecuación del plano que contiene al punto  $P(1, 0, -3)$  y es perpendicular al vector  $\vec{n} = 5\hat{i} + \hat{j} - 2\hat{k}$  se obtiene de

$$(5)(x - 1) + (1)(y - 0) + (-2)(z - (-3)) = 0.$$

Llevando a cabo las operaciones algebraicas correspondientes, esta ecuación se reduce a

$$5x + y - 2z = 11.$$

De acuerdo con el resultado anterior, la ecuación de un plano en  $\mathbb{R}^3$  siempre puede llevarse a la forma general

$$ax + by + cz = d,$$

donde  $a, b$  y  $c$  son las componentes del vector normal al plano, y  $d = ax_0 + by_0 + cz_0$  es una constante.

Ejemplos:

1. Proporciona tres puntos contenidos en el plano  $3x + 2y + 4z = 12$  en  $\mathbb{R}^3$ .

Los puntos se obtienen simplemente al encontrar tres valores  $x, y$  y  $z$  que satisfagan la ecuación  $3x + 2y + 4z = 12$ . Por ejemplo, están los puntos  $P_1(2, 3, 0)$ ,  $P_2(0, 0, 3)$  y  $P_3(0, -2, 4)$ .

2. Encuentra la ecuación cartesiana del plano que contiene a los puntos  $A(1, 1, 1)$ ,  $B(2, 1, 3)$  y  $C(3, 2, 1)$ .

El vector normal  $\vec{n}$  es perpendicular a cualesquiera dos vectores no paralelos en el plano. Por ejemplo, si se consideran los vectores  $\vec{AB} = \hat{i} + 2\hat{k}$  y  $\vec{AC} = 2\hat{i} + \hat{j}$ , y se define  $\vec{n} = x\hat{i} + y\hat{j} + z\hat{k}$ , se tiene

$$\vec{AB} \cdot \vec{n} = (1, 0, 2) \cdot (x, y, z) = 0$$

$$\vec{AC} \cdot \vec{n} = (2, 1, 0) \cdot (x, y, z) = 0$$

es decir,

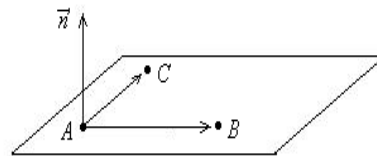
$$x + 2z = 0$$

$$2x + y = 0.$$

Tomando, por ejemplo,  $z = 1$ , obtenemos  $x = -2, y = 4, z = 1$ .

De esta manera,

$$\vec{n} = -2\hat{i} + 4\hat{j} + \hat{k}, \text{ o algún múltiplo de éste.}$$



## Capítulo 1 El Espacio $\mathbb{R}^n$

El punto conocido  $P_0(x_0, y_0, z_0)$  puede ser cualquiera de los tres puntos dados. Por ejemplo, si se considera el punto  $A(1, 1, 1)$  se llega a que la ecuación cartesiana del plano es

$$(-2)(x - 1) + (4)(y - 1) + (1)(z - 1) = 0,$$

o bien,

$$-2x + 4y + z = 3.$$

Nota que el resultado es independiente de la selección del punto, o si tomaste cualquier otro múltiplo del vector normal  $\vec{n}$ .

3. Encuentra la ecuación cartesiana del plano que contiene al punto  $P_0(2, 3, 1)$  y a la recta  $L: x = 1 + t, y = 1 - t, z = t, t \in \mathbb{R}$ .

Primero notamos que el punto  $P_0(2, 3, 1)$  no está contenido en la recta. Si lo estuviera, existiría una infinidad de planos que la contendrían. El vector normal  $\vec{n}$  al plano es un vector perpendicular a cualesquiera dos vectores  $\vec{v}_1$  y  $\vec{v}_2$  en el plano. Así, por ejemplo, podemos escoger  $\vec{v}_1 = \overrightarrow{P_0P} = -\hat{i} - 2\hat{j} - \hat{k}$  y  $\vec{v}_2 = \hat{i} - \hat{j} + \hat{k}$  (vector de dirección de la recta), como se muestra en la figura. Definiendo  $\vec{n} = x\hat{i} + y\hat{j} + z\hat{k}$ , se tiene

$$\vec{v}_1 \cdot \vec{n} = (-1, -2, -1) \cdot (x, y, z) = 0$$

$$\vec{v}_2 \cdot \vec{n} = (1, -1, 1) \cdot (x, y, z) = 0$$

es decir,

$$-x - 2y - z = 0$$

$$x - y + z = 0.$$

Tomando, por ejemplo,  $z = 1$ , obtenemos  $x = -1, y = 0, z = 1$ .

De esta manera,

$\vec{n} = -\hat{i} + \hat{k}$ , o algún múltiplo de éste.

Por tanto, la ecuación cartesiana del plano es

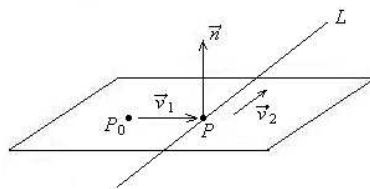
$$(-1)(x - 1) + (0)(y - 1) + (1)(z - 0) = 0,$$

o bien,

$$x - z = 1.$$

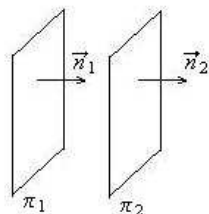
4. Encuentra la ecuación cartesiana del plano que contiene al punto  $P(1, 2, 3)$  y es paralelo al plano  $5x - 3y + 2z = 11$ .

Como lo muestra la figura, si un plano  $\pi_1$ , con vector normal  $\vec{n}_1$ , es paralelo a otro plano  $\pi_2$ , con vector normal  $\vec{n}_2$ , entonces los vectores  $\vec{n}_1$  y  $\vec{n}_2$  son



## 1.4 Planos e hiperplanos

también paralelos entre sí.



De este modo, el vector normal  $\vec{n}$  al plano que buscamos puede escogerse simplemente como  $\vec{n} = 5\hat{i} - 3\hat{j} + 2\hat{k}$ , que es el vector normal al plano  $5x - 3y + 2z = 11$ . Así, la ecuación del plano es

$$(5)(x - 1) + (-3)(y - 2) + (2)(z - 3) = 0,$$

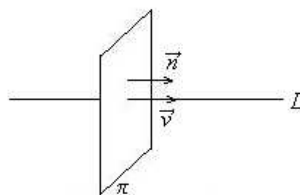
o bien,

$$5x - 3y + 2z = 5.$$

5. Encuentra la ecuación cartesiana del plano que contiene al punto  $P(1, 1, 1)$  y es normal a la recta

$$\frac{x - 1}{2} = y + 1 = -\frac{z}{2}.$$

Como lo muestra la figura, si un plano  $\pi$  con vector normal  $\vec{n}$  es perpendicular a una recta  $L$  con vector de dirección  $\vec{v}$ , entonces los vectores  $\vec{n}$  y  $\vec{v}$  son paralelos entre sí.



De este modo, el vector normal  $\vec{n}$  al plano que buscamos puede escogerse simplemente como  $\vec{n} = 2\hat{i} + \hat{j} - 2\hat{k}$ , que es el vector de dirección de la recta  $\frac{x - 1}{2} = y + 1 = -\frac{z}{2}$ . Así, la ecuación del plano es

$$(2)(x - 1) + (1)(y - 1) + (-2)(z - 1) = 0,$$

o bien,

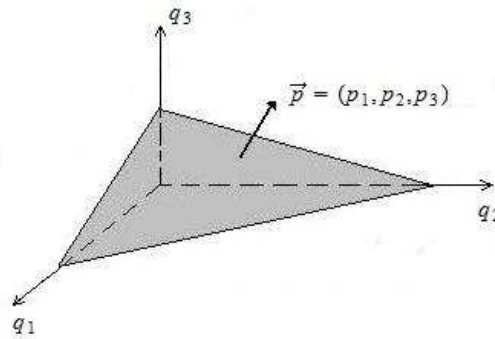
$$2x + y - 2z = 1.$$

6. Un lindo ejemplo de planos en economía es el de una restricción presupuestal, de la forma

$$p_1q_1 + p_2q_2 + p_3q_3 = I, \quad (p_1, p_2, p_3, I \text{ constantes})$$

## Capítulo 1 El Espacio $\mathbb{R}^n$

que representa un plano en el espacio de cantidades  $q_1q_2q_3$ , con vector normal dado por el vector de precios  $\vec{p} = (p_1, p_2, p_3)$ .



7. Otro ejemplo bonito es el de la ecuación de costo lineal

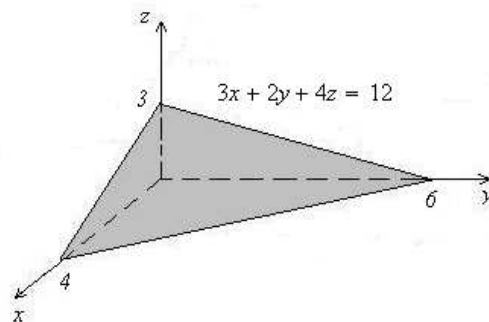
$$C = wL + rK,$$

para cada nivel de trabajo  $L$  y de capital  $K$ , con precios unitarios dados por el salario  $w$  y la tasa de interés  $r$ , respectivamente. Nota que la ecuación para el costo presenta la forma

$$wL + rK - C = 0,$$

que representa un plano en el espacio  $LKC$ , con vector normal  $\vec{n} = (w, r, -1)$ .

Por lo general, la gráfica de un plano de la forma  $ax + by + cz = d$  en  $\mathbb{R}^3$  se construye a partir de las intersecciones de éste con los ejes coordenados. Por ejemplo, la siguiente figura muestra la gráfica del plano  $3x + 2y + 4z = 12$ .



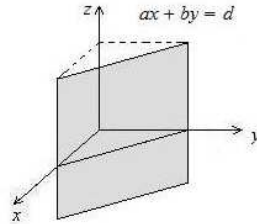
Una gráfica como la anterior presupone que los coeficientes  $a, b, c$  y  $d$  en la ecuación  $ax + by + cz = d$  son todos diferentes de cero. A continuación se muestra la gráfica de algunos casos especiales, en donde uno o varios de los coeficientes  $a, b$  o  $c$  pueda ser igual a cero.



## 1.4 Planos e hiperplanos

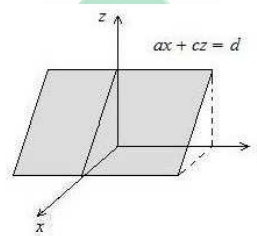
$$c = 0 \Rightarrow ax + by = d$$

( $z$  libre)



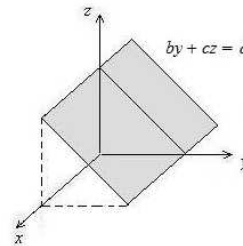
$$b = 0 \Rightarrow ax + cz = d$$

( $y$  libre)



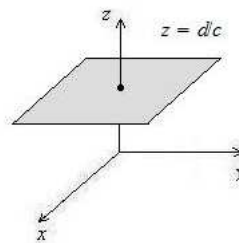
$$a = 0 \Rightarrow by + cz = d$$

( $x$  libre)



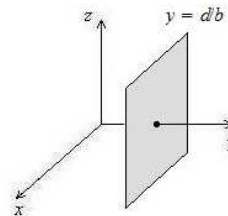
$$a = b = 0 \Rightarrow cz = d$$

( $x, y$  libres)

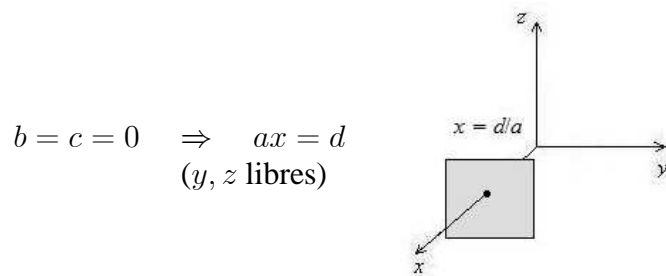


$$a = c = 0 \Rightarrow by = d$$

( $x, z$  libres)



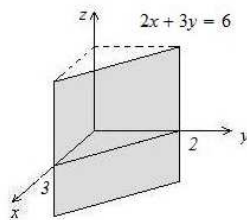
## Capítulo 1 El Espacio $\mathbb{R}^n$



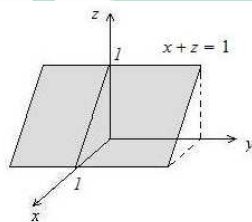
Ejemplos:

1. Dibuja los siguientes planos en  $\mathbb{R}^3$ :

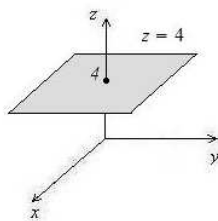
a)  $2x + 3y = 6$



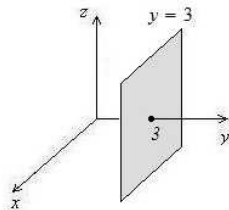
b)  $x + z = 1$



c)  $z = 4$



d)  $y = 3$

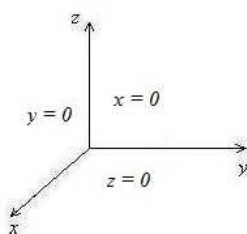


## 1.4 Planos e hiperplanos

2. Encuentra las ecuaciones de los planos coordenados en  $\mathbb{R}^3$ .

En cada caso, podemos tomar como punto conocido el origen  $0(0, 0, 0)$ , y como vector normal alguno de los vectores base, obteniendo

|            | vector normal       | ecuación del plano: |
|------------|---------------------|---------------------|
| plano $xy$ | $\vec{n} = \hat{k}$ | $z = 0$             |
| plano $yz$ | $\vec{n} = \hat{i}$ | $x = 0$             |
| plano $xz$ | $\vec{n} = \hat{j}$ | $y = 0$             |



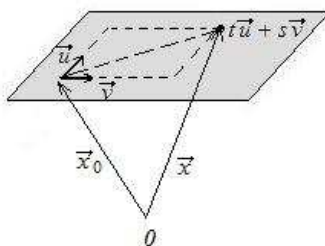
### Ecuación vectorial paramétrica del plano

Además de la representación cartesiana que ya vimos, la ecuación del plano también admite una representación paramétrica, que presentaremos muy brevemente.

**Definición.** La *ecuación vectorial paramétrica del plano* que contiene al punto  $\vec{x}_0$  y a los vectores no nulos  $\vec{u}$  y  $\vec{v}$  que no son paralelos es

$$\vec{x} = \vec{x}_0 + t\vec{u} + s\vec{v},$$

donde  $s, t \in (-\infty, \infty)$ .



## Capítulo 1 El Espacio $\mathbb{R}^n$

Para pasar de la ecuación cartesiana del plano a su ecuación paramétrica, se parametrizan dos de las tres variables,  $x, y$  o  $z$ , como se muestra a continuación.

Ejemplo:

Encuentra la ecuación paramétrica del plano  $x + 2y - z = 3$  en  $\mathbb{R}^3$ .

Simplemente podemos proponer la parametrización  $y = t$  y  $z = s$ , de modo que

$$\begin{aligned}x &= 3 - 2t + s \\y &= t \\z &= s, \quad t, s \in \mathbb{R}.\end{aligned}$$

Estas ecuaciones pueden expresarse en forma vectorial como

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 3 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} + s \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad t, s \in \mathbb{R},$$

que es de la forma  $\vec{x} = \vec{x}_0 + t\vec{u} + s\vec{v}$ , con  $\vec{x}_0 = 3\hat{i}$ ,  $\vec{u} = -2\hat{i} + \hat{j}$  y  $\vec{v} = \hat{i} + \hat{k}$ . Es fácil demostrar que los vectores  $\vec{u}$  y  $\vec{v}$  son perpendiculares al vector normal  $\vec{n} = \hat{i} + 2\hat{j} - \hat{k}$  del plano  $x + 2y - z = 3$ .

### Hiperplanos

La forma  $\vec{n} \cdot (\vec{x} - \vec{x}_0) = 0$  para la ecuación del plano no se limita al espacio tridimensional  $\mathbb{R}^3$ , sino que es válida para espacios  $\mathbb{R}^m$  de dimensión mayor ( $m > 3$ ). En este caso, al plano se le denomina *hiperplano*.

**Definición.** La *ecuación del hiperplano* que contiene al punto  $\vec{x}_0 \in \mathbb{R}^m$  y es perpendicular al vector no nulo  $\vec{n} \in \mathbb{R}^m$  es

$$\vec{n} \cdot (\vec{x} - \vec{x}_0) = 0.$$

En particular, si  $P_0(x_1^0, x_2^0, \dots, x_m^0)$  y  $\vec{n} = (a_1, a_2, \dots, a_m)$ , la ecuación cartesiana del hiperplano es

$$a_1(x_1 - x_1^0) + a_2(x_2 - x_2^0) + \dots + a_m(x_m - x_m^0) = 0.$$

## 1.5 Conjuntos abiertos, cerrados, acotados, compactos, convexos

Ejemplos:

1. La ecuación del hiperplano en  $\mathbb{R}^4$  que contiene al punto  $P_0(1, -2, 0, 3)$  y es normal al vector  $\vec{n} = (5, 2, 3, -1)$  está dada por

$$5(x_1 - 1) + 2(x_2 - (-2)) + 3(x_3 - 0) - (x_4 - 3) = 0,$$

es decir,

$$5x_1 + 2x_2 + 3x_3 - x_4 = -2.$$

2. Un ejemplo en economía está dado por el hiperplano presupuestal,

$$p_1x_1 + p_2x_2 + \cdots + p_nx_n = I,$$

o bien

$$\vec{p} \cdot \vec{x} = I,$$

cuyo vector normal es el vector de precios,  $\vec{p} = (p_1, p_2, \dots, p_n)$ .

## 1.5 Conjuntos abiertos, cerrados, acotados, compactos, convexos

Aquí presentaremos algunas nociones básicas de topología, necesarias para comprender el significado de los teoremas y conceptos que veremos más adelante en el curso, particularmente en el tema de optimización de funciones en varias variables.

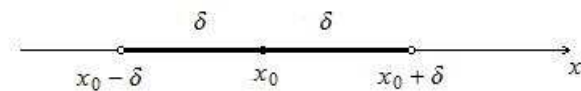
**Definición.** Dado un punto  $\vec{x}_0 \in \mathbb{R}^n$  y un número real  $\delta > 0$  la *vecindad*  $V_\delta(\vec{x}_0)$  con centro en  $\vec{x}_0$  y radio  $\delta$  es el conjunto de todos los puntos  $\vec{x} \in \mathbb{R}^n$  cuya distancia a  $\vec{x}_0$  es menor que  $\delta$ , es decir,

$$V_\delta(\vec{x}_0) = \{ \vec{x} \in \mathbb{R}^n \mid \|\vec{x} - \vec{x}_0\| < \delta \}.$$

Ejemplos:

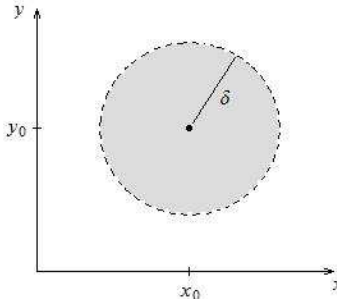
1. Una vecindad en  $\mathbb{R}$  es el conjunto  $V_\delta(x_0) = \{ x \in \mathbb{R} \mid |x - x_0| < \delta \}$ , que representa un intervalo abierto en los reales, con radio  $\delta$  y centro en  $x_0$ .

$$\begin{aligned} & |x - x_0| < \delta \\ \therefore -\delta & < x - x_0 < \delta \\ \therefore x_0 - \delta & < x < x_0 + \delta \end{aligned}$$

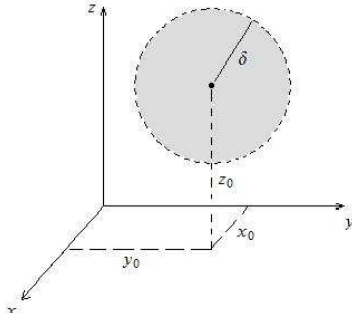


## Capítulo 1 El Espacio $\mathbb{R}^n$

2. Una vecindad en  $\mathbb{R}^2$  es el conjunto  $V_\delta(\vec{x}_0) = \{ \vec{x} \in \mathbb{R}^2 \mid \|\vec{x} - \vec{x}_0\| < \delta \}$ , que representa los puntos dentro de un círculo de radio  $\delta$  y centro en  $\vec{x}_0$ .

$$\begin{aligned} \|\vec{x} - \vec{x}_0\| &< \delta \\ \therefore \sqrt{(x - x_0)^2 + (y - y_0)^2} &< \delta \\ \therefore (x - x_0)^2 + (y - y_0)^2 &< \delta^2 \end{aligned}$$


3. Una vecindad en  $\mathbb{R}^3$  es el conjunto  $V_\delta(\vec{x}_0) = \{ \vec{x} \in \mathbb{R}^3 \mid \|\vec{x} - \vec{x}_0\| < \delta \}$ , que representa los puntos dentro de una esfera de radio  $\delta$  y centro en  $\vec{x}_0$ .

$$\begin{aligned} \|\vec{x} - \vec{x}_0\| &< \delta \\ \therefore \sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2} &< \delta \\ \therefore (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 &< \delta^2 \end{aligned}$$


Con base en los ejemplos anteriores, es claro por qué a una vecindad también se le llama *bola abierta*.

**Definición.** Sea  $A \subset \mathbb{R}^n$  y sea  $\vec{x}_0 \in \mathbb{R}^n$ . Decimos que:

- $\vec{x}_0$  es un *punto interior* de  $A$  si existe un número  $\delta > 0$  tal que la vecindad  $V_\delta(\vec{x}_0)$  está totalmente contenida en  $A$ .
- $\vec{x}_0$  es un *punto exterior* de  $A$  si existe un número  $\delta > 0$  tal que la vecindad  $V_\delta(\vec{x}_0)$  no contiene puntos de  $A$ .
- $\vec{x}_0$  es un *punto frontera* de  $A$  si para todo número  $\delta > 0$  la vecindad  $V_\delta(\vec{x}_0)$  contiene puntos de  $A$  y puntos fuera de  $A$ . Los puntos frontera de  $A$  pueden, o no, pertenecer a  $A$ .

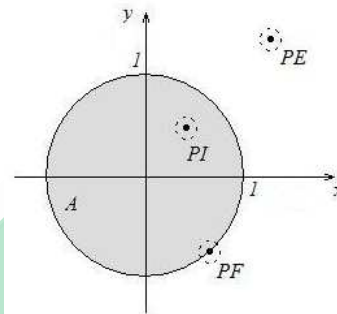
## 1.5 Conjuntos abiertos, cerrados, acotados, compactos, convexos

Por ejemplo, si  $A = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$ , entonces el conjunto de puntos interiores (PI), puntos exteriores (PE) y puntos frontera (PF) de  $A$  son los conjuntos:

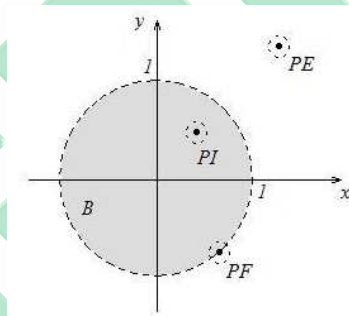
$$PI = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1\}$$

$$PE = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 > 1\}$$

$$PF = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$$

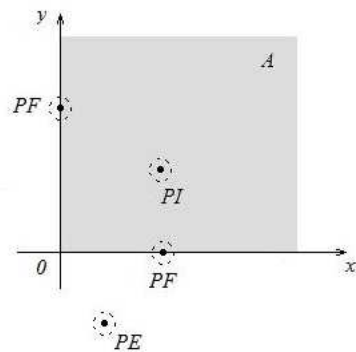


Nota que los conjuntos  $PI$ ,  $PE$  y  $PF$  anteriores se mantienen igual si en lugar de  $A$  se considera el conjunto  $B = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1\}$ .



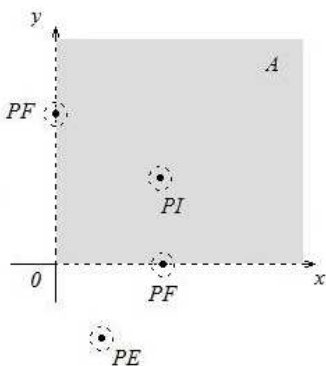
Ejemplos:

1.  $A = \mathbb{R}_+^2 = \{(x, y) \in \mathbb{R}^2 \mid x \geq 0 \text{ y } y \geq 0\}$ .

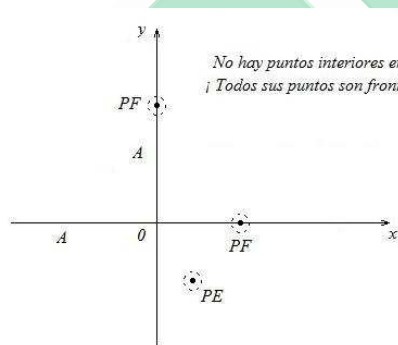


## Capítulo 1 El Espacio $\mathbb{R}^n$

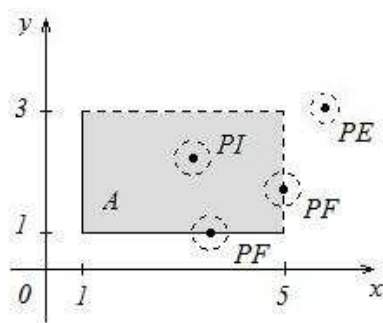
2.  $A = \mathbb{R}_{++}^2 = \{(x, y) \in \mathbb{R}^2 \mid x > 0 \text{ y } y > 0\}$ .



3.  $A = \{(x, y) \in \mathbb{R}^2 \mid xy = 0\}$ .



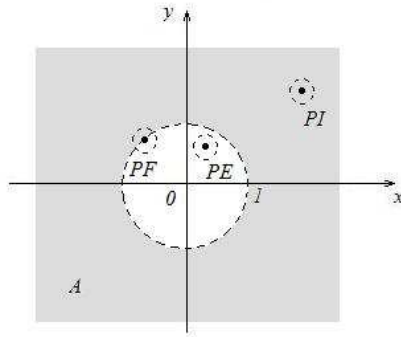
4.  $A = \{(x, y) \in \mathbb{R}^2 \mid 1 \leq x < 5 \text{ y } 1 \leq y < 3\}$ .



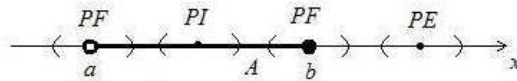


## 1.5 Conjuntos abiertos, cerrados, acotados, compactos, convexos

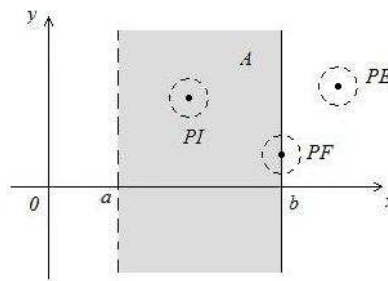
5.  $A = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 > 1\}$ .



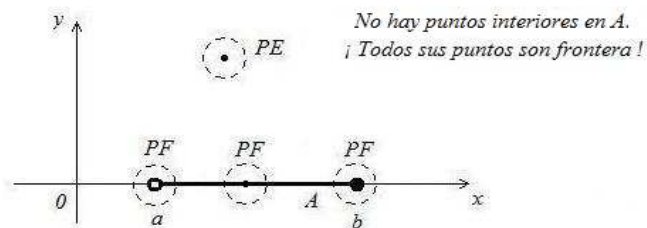
6.  $A = \{x \in \mathbb{R} \mid a < x \leq b\}$ .



7.  $A = \{(x, y) \in \mathbb{R}^2 \mid a < x \leq b\}$ .



8.  $A = \{(x, y) \in \mathbb{R}^2 \mid a < x \leq b, y = 0\}$ .



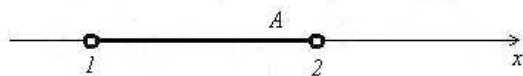
## Capítulo 1 El Espacio $\mathbb{R}^n$

**Definición.** Sea  $A \subset \mathbb{R}^n$ . Se dice que  $A$  es un *conjunto abierto* si  $A$  está formado exclusivamente por puntos interiores, es decir, si para todo  $\vec{x} \in A$  existe  $V_\delta(\vec{x})$  tal que  $V_\delta(\vec{x}) \subset A$ .

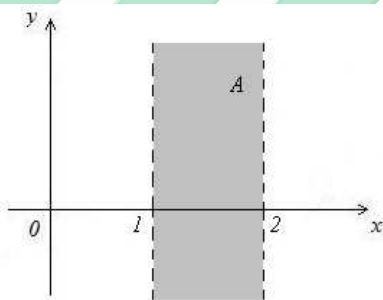
En otras palabras, se dice que  $A$  es un conjunto abierto cuando ninguno de sus puntos frontera pertenece a  $A$ .

Ejemplos:

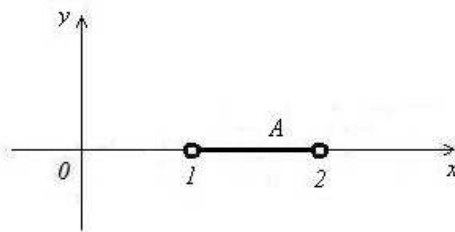
1.  $A = \{x \in \mathbb{R} \mid 1 < x < 2\}$  es abierto: sus puntos frontera son  $x = 1$  y  $x = 2$ , y ninguno de estos pertenece a  $A$ .



2.  $A = \{(x, y) \in \mathbb{R}^2 \mid 1 < x < 2\}$  es abierto: sus puntos frontera son todos aquellos sobre las rectas  $x = 1$  y  $x = 2$ , y ninguno de estos pertenece a  $A$ .

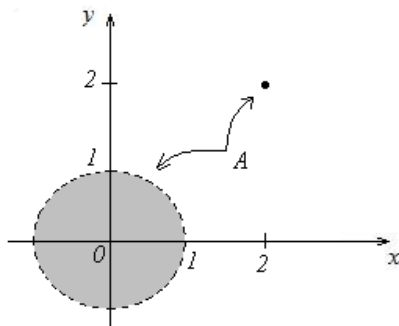


3.  $A = \{(x, y) \in \mathbb{R}^2 \mid 1 < x < 2, y = 0\}$  no es abierto: todos los puntos de  $A$  son puntos frontera.



## 1.5 Conjuntos abiertos, cerrados, acotados, compactos, convexos

4.  $A = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1\} \cup \{(2, 2)\}$  no es abierto:  $A$  contiene un punto frontera, que es el punto  $(2, 2)$ .



**Teorema.** a) La unión de conjuntos abiertos es un conjunto abierto. b) La intersección finita de conjuntos abiertos es un conjunto abierto.

En relación con el inciso b) de este último teorema es importante entender por qué se requiere que la intersección sea finita, y no infinita, para garantizar que el conjunto resultante de la unión sea un conjunto abierto. Para ello, considera como ejemplo el conjunto de intervalos  $I_n$  definidos por

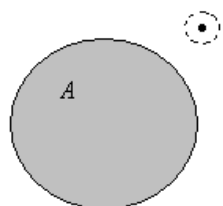
$$I_n = \left(-\frac{1}{n}, \frac{1}{n}\right),$$

para todo  $n \in \mathbb{N}$ . Es claro que cada  $I_n$  es un conjunto abierto; sin embargo la intersección de todos los conjuntos  $I_n$  es el conjunto

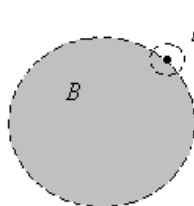
$$\bigcap_{n \in \mathbb{N}} I_n = I_1 \cap I_2 \cap \cdots \cap I_n = \{0\},$$

que no es un conjunto abierto (el único elemento del conjunto es 0, que es un punto frontera).

**Definición.** Sea  $A \subset \mathbb{R}^n$ . Se dice que  $A$  es un *conjunto cerrado* si para todo punto que no pertenece a  $A$  es posible encontrar una vecindad que no contenga puntos de  $A$ .



$A$  es cerrado



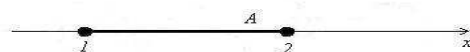
$B$  no es cerrado

La frontera de  $B$  no pertenece a  $B$ , pero su vecindad sí contiene puntos de  $B$

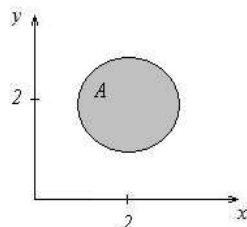
**Teorema.** Un conjunto es cerrado si y sólo si contiene a todos sus puntos frontera.

Ejemplos:

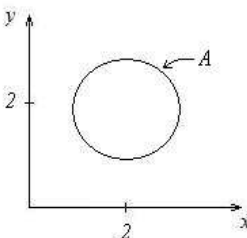
1.  $A = \{x \in \mathbb{R} \mid 1 \leq x \leq 2\}$  es cerrado: sus puntos frontera son  $x = 1$  y  $x = 2$ , y ambos pertenecen a  $A$ .



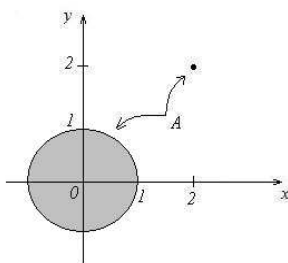
2.  $A = \{(x, y) \in \mathbb{R}^2 \mid (x - 2)^2 + (y - 2)^2 \leq 1\}$  es cerrado: sus puntos frontera son todos los puntos de la circunferencia, que pertenecen a  $A$ .



3.  $A = \{(x, y) \in \mathbb{R}^2 \mid (x - 2)^2 + (y - 2)^2 = 1\}$  es cerrado: todos sus puntos son frontera.

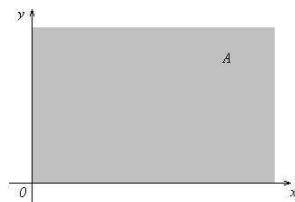


4.  $A = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\} \cup \{(2, 2)\}$  es cerrado:  $A$  contiene toda su frontera, que consiste en los puntos de la circunferencia, junto con el punto  $(2, 2)$ .



## 1.5 Conjuntos abiertos, cerrados, acotados, compactos, convexos

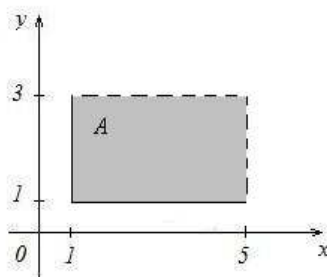
5.  $A = \mathbb{R}_+^2 = \{(x, y) \in \mathbb{R}^2 \mid x, y \geq 0\}$  es cerrado:  $A$  contiene a toda su frontera, que son los ejes coordenados, en su parte no negativa.



6.  $A = \{x \in \mathbb{R} \mid x \geq 2\}$  es cerrado:  $A$  contiene a toda su frontera, que es el punto  $x = 2$ .



No necesariamente un conjunto debe ser abierto o cerrado. Existen conjuntos que no son ni abiertos ni cerrados, como es el caso de  $A = \{(x, y) \in \mathbb{R}^2 \mid 1 \leq x < 5 \text{ y } 1 \leq y < 3\}$ , ya que éste contiene algunos de sus puntos frontera (de modo que no es abierto), pero no los contiene a todos ellos (de modo que no es cerrado).



**Teorema.** Un conjunto es cerrado si y sólo si su complemento es abierto.

A partir de este teorema se puede demostrar que existen dos (y sólo dos) conjuntos que son abiertos y cerrados a la vez, que son el conjunto  $\mathbb{R}^n$  y el conjunto vacío,  $\emptyset$ . Para ello, nota primero que  $\mathbb{R}^n$  es el complemento de  $\emptyset$ , y viceversa. El argumento es el siguiente. Por una parte,  $\mathbb{R}^n$  es abierto, ya que no contiene puntos frontera. En consecuencia,  $\emptyset$  es cerrado. Por otra parte,  $\emptyset$  es abierto, ya que no contiene puntos frontera (de hecho, no contiene ningún punto). En consecuencia,  $\mathbb{R}^n$  es cerrado.

**Teorema.** a) La intersección de conjuntos cerrados es un conjunto cerrado.  
b) La unión finita de conjuntos cerrados es un conjunto cerrado.

De acuerdo con el inciso b) de este teorema, sólo se puede asegurar que la unión de cerrados es un conjunto cerrado cuando el número de estos conjuntos es finito. El siguiente ejemplo ilustra cómo la unión infinita de conjunto cerrados puede resultar en un conjunto abierto. Considera el conjunto de intervalos  $I_n$  definidos por

$$I_n = [-n, n],$$

para todo  $n \in \mathbb{N}$ . Es claro que cada  $I_n$  es un conjunto cerrado. La unión de todos ellos es el conjunto

$$\bigcup_{n \in \mathbb{N}} I_n = I_1 \cup I_2 \cup \dots \cup I_n = \mathbb{R},$$

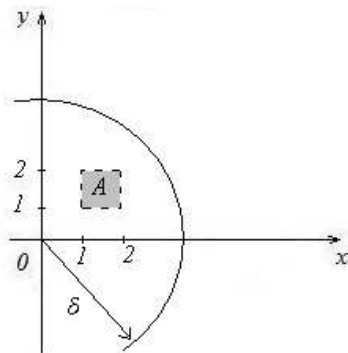
que es un conjunto abierto (el conjunto de los reales no contiene puntos frontera).

**Definición.** Un conjunto  $A \subset \mathbb{R}^n$  es un *conjunto acotado* si existe una vecindad con centro en el origen que contiene totalmente a  $A$ , es decir, si existe  $\delta > 0$  tal que  $A \subset V_\delta(\vec{0})$ .

En otras palabras, un conjunto es acotado si no contiene puntos arbitrariamente alejados del origen.

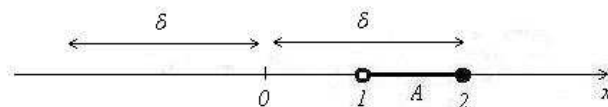
Ejemplos:

1.  $A = \{(x, y) \in \mathbb{R}^2 \mid 1 < x < 2, 1 < y < 2\}$  es acotado: cualquier vecindad  $V_\delta(\vec{0})$  de radio  $\delta > \sqrt{8}$  contiene totalmente los puntos de  $A$ .

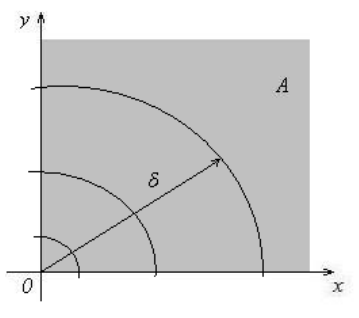


## 1.5 Conjuntos abiertos, cerrados, acotados, compactos, convexos

2.  $A = \{x \in \mathbb{R} \mid 1 < x \leq 2\}$  es acotado: cualquier vecindad  $V_\delta(0)$  de radio  $\delta > 2$  contiene totalmente los puntos de  $A$ .



3.  $A = \mathbb{R}_+^2 = \{(x, y) \in \mathbb{R}^2 \mid x \geq 0, y \geq 0\}$  no es un conjunto acotado, pero sí es cerrado.

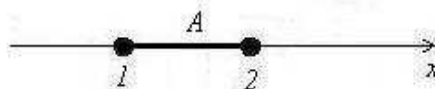


Los ejemplos anteriores muestran que un conjunto puede, o no, ser acotado, independientemente de si es abierto, cerrado o ninguno de estos.

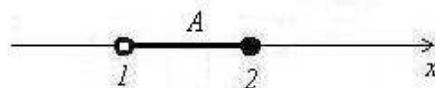
**Definición.** Un conjunto  $A \subset \mathbb{R}^n$  es un *conjunto compacto* si  $A$  es cerrado y acotado.

Ejemplos:

1.  $A = \{x \in \mathbb{R} \mid 1 \leq x \leq 2\}$  es compacto, ya que es cerrado y acotado.

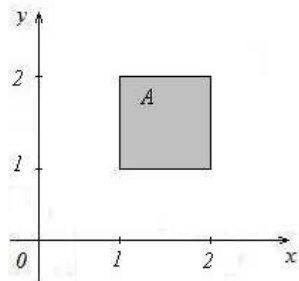


2.  $A = \{x \in \mathbb{R} \mid 1 < x \leq 2\}$  no es compacto, ya que es acotado, pero no cerrado.

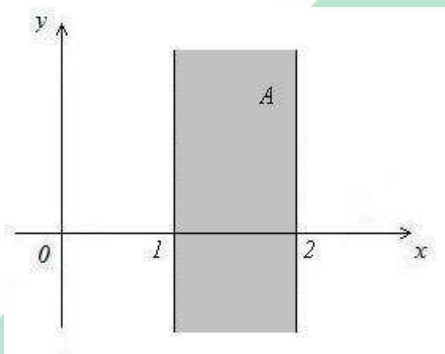


## Capítulo 1 El Espacio $\mathbb{R}^n$

3.  $A = \{(x, y) \in \mathbb{R}^2 \mid 1 \leq x \leq 2, 1 \leq y \leq 2\}$  es compacto, ya que es cerrado y acotado.



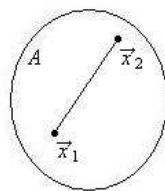
4.  $A = \{(x, y) \in \mathbb{R}^2 \mid 1 \leq x \leq 2\}$  no es compacto, ya que es cerrado, pero no acotado (la variable  $y$  es libre).



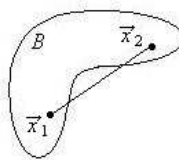
**Definición.** Un conjunto  $A \subset \mathbb{R}^n$  es un *conjunto convexo* si para cualquier par de puntos  $\vec{x}_1, \vec{x}_2 \in A$  el segmento de recta que los une también está en  $A$ , es decir, si

$$t\vec{x}_1 + (1-t)\vec{x}_2 \in A,$$

para todo  $0 \leq t \leq 1$ .



*A es convexo*



*B no es convexo*

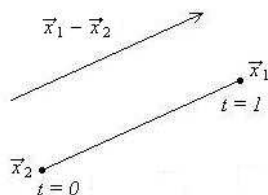
En esta definición, nota que la expresión

$$t\vec{x}_1 + (1-t)\vec{x}_2 = \vec{x}_2 + t(\vec{x}_1 - \vec{x}_2),$$



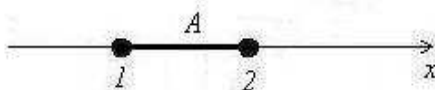
## 1.5 Conjuntos abiertos, cerrados, acotados, compactos, convexos

conocida como *combinación convexa*, es la ecuación paramétrica de la recta que contiene al punto  $\vec{x}_2$  y está en la dirección  $\vec{x}_1 - \vec{x}_2$ ; al limitar el dominio de  $t$ , entre 0 y 1, se obtiene el segmento de recta entre los puntos  $\vec{x}_1$  y  $\vec{x}_2$ .

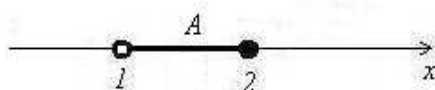


Ejemplos:

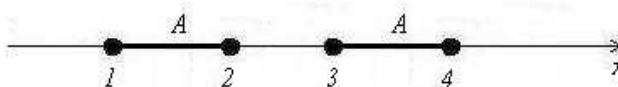
1.  $A = \{x \in \mathbb{R} \mid 1 \leq x \leq 2\}$  es convexo.



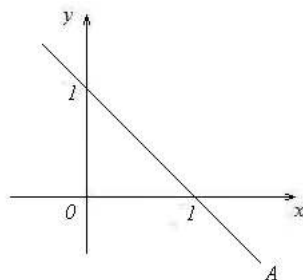
2.  $A = \{x \in \mathbb{R} \mid 1 < x \leq 2\}$  es convexo.



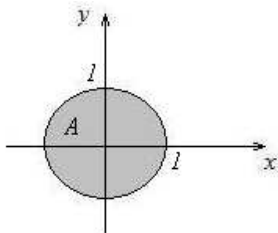
3.  $A = \{x \in \mathbb{R} \mid 1 \leq x \leq 2\} \cup \{x \in \mathbb{R} \mid 3 \leq x \leq 4\}$  no es convexo.



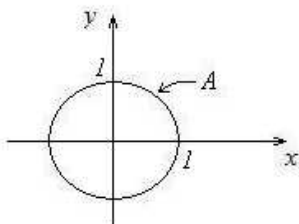
4.  $A = \{(x, y) \in \mathbb{R}^2 \mid x + y = 1\}$  es convexo.



5.  $A = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$  es convexo.

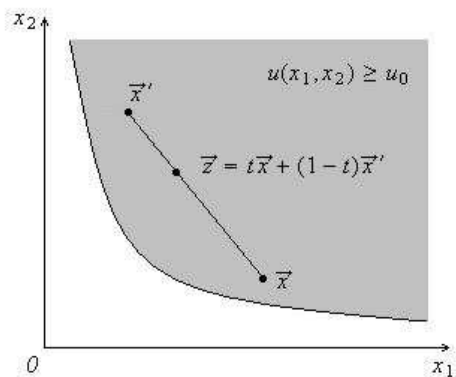


6.  $A = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$  no es convexo.



Los conjuntos convexos son muy importantes en economía. Por ejemplo, similarmente al ejemplo 4, tenemos que las canastas  $(x_1, x_2)$  en  $\mathbb{R}_+^2$  que satisfacen una restricción presupuestal de la forma  $I = p_1x_1 + p_2x_2$ , con  $I, p_1$  y  $p_2$  fijos, forman un conjunto convexo. Como un segundo ejemplo podemos considerar las preferencias de un consumidor, dadas por el conjunto  $P = \{(x_1, x_2) \in \mathbb{R}_+^2 \mid u((x_1, x_2)) \geq u_0\}$  de las canastas  $(x_1, x_2)$  que dan una utilidad  $u$  mayor o igual a un valor  $u_0$ . Si suponemos que  $P$  es convexo y  $\vec{x}, \vec{x}' \in P$ , entonces cualquier canasta intermedia  $\vec{z} = t\vec{x} + (1-t)\vec{x}'$ ,  $0 \leq t \leq 1$ , también dará una utilidad mayor o igual a  $u_0$ .

$$u(\vec{x}) \geq u_0 \text{ y } u(\vec{x}') \geq u_0 \Rightarrow u(\vec{z}) \geq u_0$$



## 1.5 Conjuntos abiertos, cerrados, acotados, compactos, convexos

En los siguientes ejemplos se presenta cómo demostrar formalmente que un conjunto es convexo.

Ejemplos:

1. Demuestra que el conjunto  $A = \{ (x, y) \in \mathbb{R}^2 \mid x + y = 1 \}$  es convexo.

Sean  $\vec{x}_1 = (x_1, y_1)$ ,  $\vec{x}_2 = (x_2, y_2) \in A$ . Por lo tanto,

$$x_1 + y_1 = 1 \text{ y } x_2 + y_2 = 1.$$

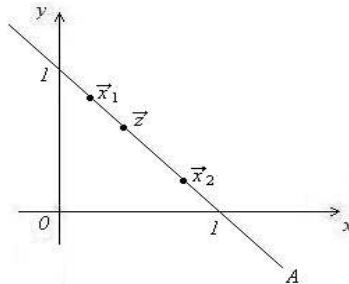
Sea  $\vec{z} = t\vec{x}_1 + (1-t)\vec{x}_2$ , con  $0 \leq t \leq 1$ , de modo que

$$\begin{aligned} \vec{z} &= t(x_1, y_1) + (1-t)(x_2, y_2) \\ &= (tx_1 + (1-t)x_2, ty_1 + (1-t)y_2) \\ &= (z_1, z_2). \end{aligned}$$

Así,

$$\begin{aligned} z_1 + z_2 &= tx_1 + (1-t)x_2 + ty_1 + (1-t)y_2 \\ &= t(x_1 + y_1) + (1-t)(x_2 + y_2) \\ &= t(1) + (1-t)(1) = 1, \end{aligned}$$

de donde concluimos que  $\vec{z} = (z_1, z_2) \in A$ . Por lo tanto,  $A$  es convexo.



2. Demuestra que el conjunto  $A = \{ (x, y) \in \mathbb{R}^2 \mid a \leq x \leq b \}$  es convexo.

Sean  $\vec{x}_1 = (x_1, y_1)$ ,  $\vec{x}_2 = (x_2, y_2) \in A$ . Por lo tanto,

$$a \leq x_1 \leq b \text{ y } a \leq x_2 \leq b.$$

Sea  $\vec{z} = t\vec{x}_1 + (1-t)\vec{x}_2$ , con  $0 \leq t \leq 1$ , de modo que

$$\begin{aligned} \vec{z} &= t(x_1, y_1) + (1-t)(x_2, y_2) \\ &= (tx_1 + (1-t)x_2, ty_1 + (1-t)y_2) \\ &= (z_1, z_2). \end{aligned}$$

Como  $t \geq 0$  y  $1-t \geq 0$ , por lo tanto

$$ta \leq tx_1 \leq tb \text{ y } (1-t)a \leq (1-t)x_2 \leq (1-t)b.$$

Sumando ambas expresiones tenemos

$$ta + (1-t)a \leq tx_1 + (1-t)x_2 \leq tb + (1-t)b,$$

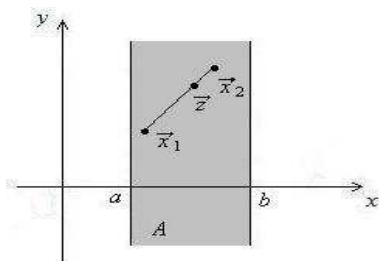
es decir,

$$a \leq tx_1 + (1-t)x_2 \leq b,$$

y, por lo tanto,

$$a \leq z_1 \leq b.$$

Así,  $\vec{z} = (z_1, z_2) \in A$ , de modo que  $A$  es convexo.



**Teorema.** La intersección de conjuntos convexos es un conjunto convexo.

Demostración:

Sean  $A$  y  $B$  dos conjuntos convexos. Si  $A \cap B = \emptyset$ , entonces  $A \cap B$  es convexo (el vacío es un conjunto convexo). Supongamos que  $A \cap B \neq \emptyset$  y sean  $\vec{x}, \vec{x}' \in A \cap B$ . Por lo tanto,  $\vec{x}, \vec{x}' \in A$  y  $\vec{x}, \vec{x}' \in B$ . Sea  $\vec{z} = t\vec{x} + (1-t)\vec{x}'$ , con  $0 \leq t \leq 1$ . Como  $A$  es convexo y  $\vec{x}, \vec{x}' \in A$ , por lo tanto

$$\vec{z} = t\vec{x} + (1-t)\vec{x}' \in A.$$

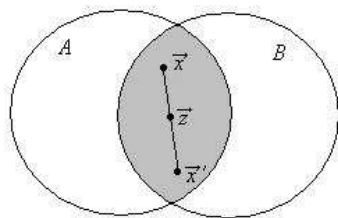
Como  $B$  es convexo y  $\vec{x}, \vec{x}' \in B$ , por lo tanto  $\vec{x}'$

$$\vec{z} = t\vec{x} + (1-t)\vec{x}' \in B.$$

Por lo tanto,

$$\vec{z} \in A \cap B,$$

de modo que  $A \cap B$  es convexo.



Por último, es importante señalar que la unión de conjuntos convexos no es un conjunto convexo, en general.

## Capítulo 2

### Funciones de varias variables

En este capítulo extenderemos la definición de función al caso de varias variables, presentando diversos conceptos relacionados, tales como el de conjuntos de nivel de la función. Posteriormente, presentaremos algunas superficies cuadráticas de interés. Concluiremos estudiando los conceptos de límite y continuidad.

#### 2.1 Dominio e imagen. Representación geométrica

**Definición.** Sea  $S \subset \mathbb{R}^n$ . Una *función real*, o *campo escalar*,  $f : S \rightarrow \mathbb{R}$ , es una regla de correspondencia que a cada elemento  $(x_1, x_2, \dots, x_n) \in S$  le asigna un único número  $w = f(x_1, x_2, \dots, x_n) \in \mathbb{R}$ . El conjunto  $S$  es el *dominio* de  $f$  y  $\mathbb{R}$  es el *contradominio* de  $f$ .

En la expresión  $w = f(x_1, x_2, \dots, x_n)$ , los elementos  $(x_1, x_2, \dots, x_n) \in S$  son las *variables independientes*, y  $w \in \mathbb{R}$  es la *variable dependiente*. Así, por ejemplo, para la función  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ , definida por  $f(x, y) = x^2 + y^2$ , las variables independientes son todas las parejas  $(x, y) \in \mathbb{R}^2$  y la variable dependiente es  $z \in \mathbb{R}$ , que depende de las anteriores a través de  $z = x^2 + y^2$ .

Cuando el dominio de una función  $f(x_1, x_2, \dots, x_n)$  no se especifica a priori, debe entenderse como tal al conjunto más grande de elementos  $(x_1, x_2, \dots, x_n) \in \mathbb{R}^n$  para los que  $f$  toma valores en  $\mathbb{R}$  (por ejemplo, que no se divida por cero o se extraiga la raíz cuadrada de un número negativo). A este conjunto se le conoce como el *dominio natural*  $D_f$  de  $f$ , dado por

$$D_f = \{ (x_1, x_2, \dots, x_n) \in \mathbb{R}^n \mid f(x_1, x_2, \dots, x_n) \in \mathbb{R} \}.$$

Por otra parte, la *imagen* o *rango*  $I_f$  de la función  $f$  es el conjunto de valores  $w$  en el contradominio,  $\mathbb{R}$ , obtenidos al aplicar la regla  $f$  a los elementos de  $D_f$ , es decir,

$$I_f = \{ w \in \mathbb{R} \mid w = f(x_1, x_2, \dots, x_n), \text{ para todo } (x_1, x_2, \dots, x_n) \in D_f \}.$$

## Capítulo 2 Funciones de varias variables

Ejemplos:

1. Sea  $z = f(x, y)$ , con  $f(x, y) = \frac{1}{x^2 + y^2}$  una función en  $\mathbb{R}^3$ . El dominio natural  $D_f$  se obtiene al pedir que el denominador sea diferente de cero ( $x^2 + y^2 \neq 0$ ). Así,

$$D_f = \{ (x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \neq 0 \} = \mathbb{R}^2 \setminus \{(0, 0)\}.$$

Como  $f$  sólo puede tomar valores positivos, entonces su imagen  $I_f$  es el conjunto

$$I_f = \{ z \in \mathbb{R} \mid z > 0 \} = \mathbb{R}^+.$$

2. Sea  $z = f(x, y)$ , con  $f(x, y) = -\frac{1}{\sqrt{9 - x^2 - y^2}}$  una función en  $\mathbb{R}^3$ . Para encontrar  $D_f$  pedimos que el denominador sea diferente de cero ( $9 - x^2 - y^2 \neq 0$ ) y el radicando sea no negativo ( $9 - x^2 - y^2 \geq 0$ ), es decir,  $9 - x^2 - y^2 > 0$ . Así,

$$D_f = \{ (x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 9 \}.$$

Como  $f$  sólo puede tomar valores negativos y no mayores que  $-1/3$ , entonces su imagen  $I_f$  es el conjunto

$$I_f = \{ z \in \mathbb{R} \mid z \leq -1/3 \}.$$

3. Sea  $z = f(x, y)$ , con  $f(x, y) = \ln(x + y)$  una función en  $\mathbb{R}^3$ . En este caso,

$$\begin{aligned} D_f &= \{ (x, y) \in \mathbb{R}^2 \mid x + y > 0 \} \\ I_f &= \{ z \in \mathbb{R} \} = \mathbb{R}. \end{aligned}$$

4. Sea  $w = f(x, y, z)$ , con  $f(x, y, z) = \frac{x \ln z}{y}$  una función en  $\mathbb{R}^4$ . En este caso,

$$\begin{aligned} D_f &= \{ (x, y, z) \in \mathbb{R}^3 \mid y \neq 0, z > 0 \} \\ I_f &= \{ w \in \mathbb{R} \} = \mathbb{R}. \end{aligned}$$

5. Sea  $w = f(x, y, z)$ , con  $f(x, y, z) = 1 - \sqrt{1 - x^2 - y^2}$  una función en  $\mathbb{R}^4$ . En este caso,

$$\begin{aligned} D_f &= \{ (x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 \leq 1 \} \\ I_f &= \{ w \in \mathbb{R} \mid 0 \leq w \leq 1 \}. \end{aligned}$$

En economía hay varios ejemplos de funciones, como las que se presentan a continuación.

## 2.1 Dominio e imagen. Representación geométrica

1. Las funciones de producción Cobb-Douglas,  $P : \mathbb{R}_+^2 \rightarrow \mathbb{R}$ , dadas por

$$P(L, K) = AL^\alpha K^{1-\alpha},$$

en donde  $P$  denota la producción,  $L$  el trabajo y  $K$  el capital, y donde  $A > 0$  y  $0 < \alpha < 1$  son constantes.

2. Una función de costo lineal,  $C : \mathbb{R}_+^2 \rightarrow \mathbb{R}$ , dada por

$$C(L, K) = wL + rK,$$

donde  $C$  denota el costo,  $L$  el trabajo y  $K$  el capital, y donde  $w > 0$  y  $r > 0$  denotan el salario y la tasa de interés, respectivamente.

3. Las funciones de utilidad,  $u : \mathbb{R}_{++}^n \rightarrow \mathbb{R}$ , dadas por

$$\begin{aligned} u(x_1, x_2, \dots, x_n) &= \alpha_1 \ln x_1 + \alpha_2 \ln x_2 + \dots + \alpha_n \ln x_n \\ &= \ln(x_1^{\alpha_1} \cdot x_2^{\alpha_2} \cdot \dots \cdot x_n^{\alpha_n}), \end{aligned}$$

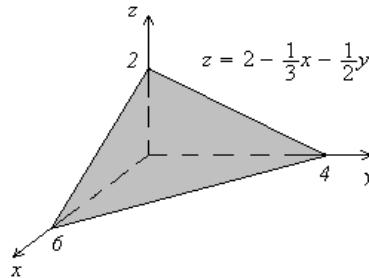
en donde  $u$  denota la utilidad para una canasta  $(x_1, x_2, \dots, x_n)$  de  $n$  bienes, con  $x_i > 0$ , y donde cada  $\alpha_i$  es constante, con  $0 < \alpha_i < 1$  y  $\alpha_1 + \alpha_2 + \dots + \alpha_n = 1$ .

Geoméricamente, la función  $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$  representa un objeto en  $\mathbb{R}^{n+1}$ . Si  $n = 1$ , la ecuación  $y = f(x)$  representa una *curva* en  $\mathbb{R}^2$ . Si  $n = 2$ , la ecuación  $z = f(x, y)$  representa una *superficie* en  $\mathbb{R}^3$ . Si  $n \geq 3$ , la ecuación  $w = f(x_1, x_2, \dots, x_n)$  representa una *hipersuperficie* en  $\mathbb{R}^{n+1}$  (sin representación gráfica).

Ejemplos:

1. La ecuación  $2x + 3y + 6z = 12$  puede pensarse como una función lineal  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ , dada por  $z = f(x, y) = 2 - \frac{1}{3}x - \frac{1}{2}y$ , cuya gráfica corresponde a un plano en  $\mathbb{R}^3$ .

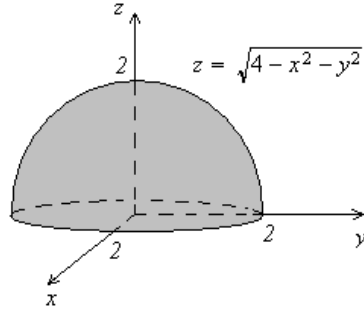
$$\begin{aligned} D_f &= \{ (x, y) \in \mathbb{R}^2 \} = \mathbb{R}^2 \\ I_f &= \{ z \in \mathbb{R} \} = \mathbb{R}. \end{aligned}$$



2. La ecuación  $x^2 + y^2 + z^2 = 4$ , con  $z \geq 0$ , puede pensarse como una función  $f : D_f \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ , dada por  $z = f(x, y) = \sqrt{4 - x^2 - y^2}$ , cuya gráfica corresponde a la parte superior de una esfera en  $\mathbb{R}^3$ .

$$D_f = \{ (x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 4 \}$$

$$I_f = \{ z \in \mathbb{R} \mid 0 \leq z \leq 2 \}.$$



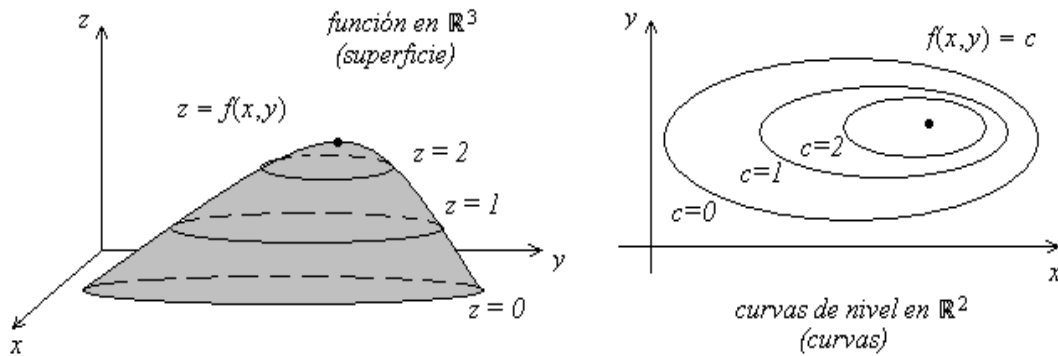
## 2.2 Conjuntos de nivel

**Definición.** Un *conjunto de nivel* de una hipersuperficie  $w = f(x_1, x_2, \dots, x_n)$  en  $\mathbb{R}^{n+1}$  es el conjunto de puntos  $(x_1, x_2, \dots, x_n) \in D_f$  tales que  $w$  toma un valor constante  $c$ , es decir,

$$f(x_1, x_2, \dots, x_n) = c.$$

En particular, si  $n = 2$  el conjunto es una *curva de nivel* y si  $n = 3$  es una *superficie de nivel*.

Los conjuntos de nivel pertenecen al mismo espacio que el dominio de la función. Así, si  $f$  está en  $\mathbb{R}^{n+1}$ , sus conjuntos de nivel están en  $\mathbb{R}^n$ .



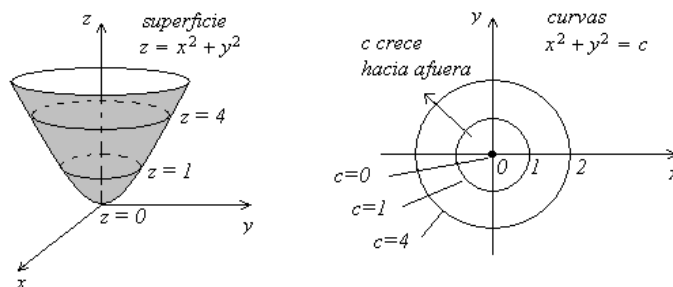


## 2.2 Conjuntos de nivel

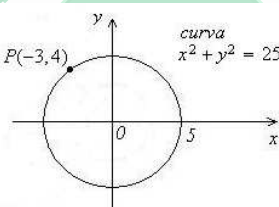
Ejemplos:

1. Identifica los conjuntos de nivel de la función  $z = f(x, y)$ , con  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  definida por  $f(x, y) = x^2 + y^2$ . ¿Cuál de estos contiene al punto  $P(-3, 4)$ ?

Como veremos en la sección 2.3, la superficie  $z = x^2 + y^2$  es un paraboloide en  $\mathbb{R}^3$ , según se ilustra en la figura de la izquierda. En la figura de la derecha se muestran algunas de sus curvas de nivel en  $\mathbb{R}^2$ , dadas por las circunferencias  $x^2 + y^2 = c$ , con centro en el origen y radio  $\sqrt{c}$ ,  $c \geq 0$ .

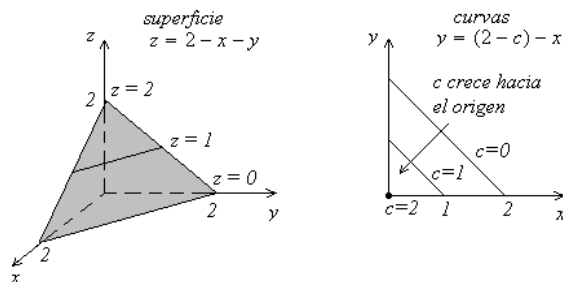


La curva de nivel de  $f$  que contiene a  $P(-3, 4)$  es tal que  $(-3)^2 + (4)^2 = 25 = c$ . Así, la curva de nivel buscada es  $x^2 + y^2 = 25$ .



2. Sea  $S = \{ (x, y) \in \mathbb{R}_+^2 \mid x + y \leq 2 \}$ . Identifica los conjuntos de nivel de la función  $z = f(x, y)$ , con  $f : S \rightarrow \mathbb{R}$  definida por  $f(x, y) = 2 - x - y$ .

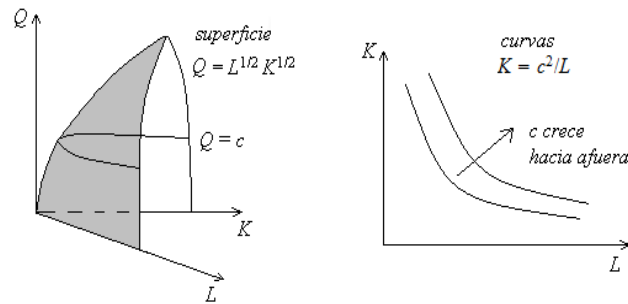
La superficie  $z = 2 - x - y$  representa la porción del plano  $x + y + z = 2$  correspondiente al primer octante de  $\mathbb{R}^3$ , como se muestra en la figura de la izquierda. En la figura de la derecha se muestran algunas de sus curvas de nivel en  $\mathbb{R}^2$ , dadas por los segmentos de recta  $x + y = 2 - c$ , con  $0 \leq c \leq 2$ .



## Capítulo 2 Funciones de varias variables

3. Identifica los conjuntos de nivel de la función de producción Cobb-Douglas  $Q = P(L, K)$ , con  $P : \mathbb{R}_+^2 \rightarrow \mathbb{R}$  definida por  $P(L, K) = L^{1/2} K^{1/2}$ .

La superficie  $Q = L^{1/2} K^{1/2}$  tiene la forma de una tienda de campaña en  $\mathbb{R}^3$ , como se ilustra en la figura de la izquierda. En la figura de la derecha se muestran algunas de sus curvas de nivel en  $\mathbb{R}^2$ , o *isocuantas*, que representan hipérbolas de la forma  $K = c^2/L$  en  $\mathbb{R}_+^2$ , con  $c > 0$ .

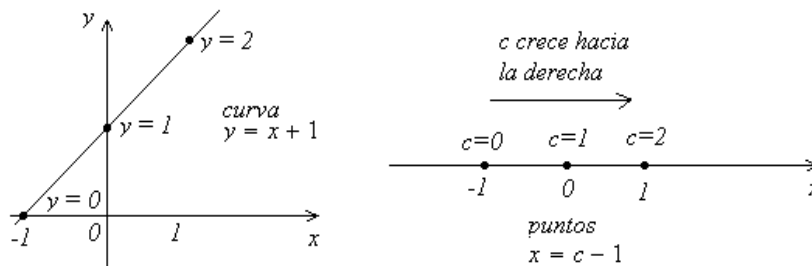


4. Identifica los conjuntos de nivel de la función de utilidad  $u = u(x, y)$ , con  $u : \mathbb{R}_{++}^2 \rightarrow \mathbb{R}$  definida por  $u(x, y) = \frac{1}{2} \ln x + \frac{1}{2} \ln y$ .

Nota que  $u(x, y) = \ln(x^{1/2} y^{1/2})$ , de modo que  $u$  es el logaritmo de una función como la del ejemplo 3. Sus curvas de nivel en  $\mathbb{R}^2$ , o *curvas de indiferencia*, son hipérbolas de la forma  $y = d/x$ , con  $d = e^{2c}$ , que son similares a las del ejemplo anterior, pero están en otra escala.

5. Identifica los conjuntos de nivel de la función  $y = f(x)$ , con  $f : \mathbb{R} \rightarrow \mathbb{R}$  definida por  $f(x) = x + 1$ .

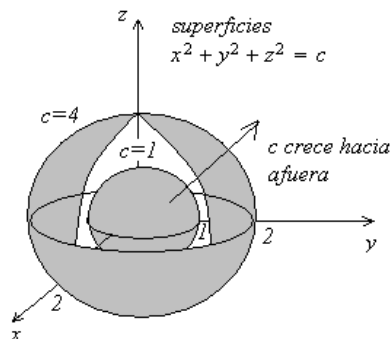
La curva  $y = x + 1$  representa una recta en  $\mathbb{R}^2$ , como se muestra en la figura de la izquierda. En la figura de la derecha se muestran algunos de sus conjuntos de nivel en  $\mathbb{R}$ , dados por los puntos  $x = c - 1$  en  $\mathbb{R}$ .



## 2.3 Superficies cuadráticas y otras superficies

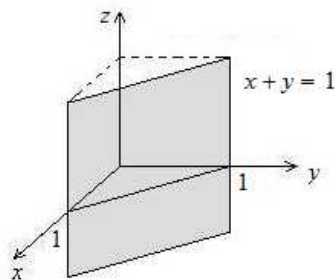
6. Identifica los conjuntos de nivel de la función  $w = f(x, y, z)$ , con  $f : \mathbb{R}^3 \rightarrow \mathbb{R}$  definida por  $f(x, y, z) = x^2 + y^2 + z^2$ .

La función  $w = x^2 + y^2 + z^2$  representa una hipersuperficie en  $\mathbb{R}^4$ , de modo que no podemos representarla gráficamente. Sus conjuntos de nivel son las superficies en  $\mathbb{R}^3$ , dadas por las esferas  $x^2 + y^2 + z^2 = c$  con centro en el origen y radio  $\sqrt{c}$ ,  $c \geq 0$ .



7. Determina la superficie de nivel de la función  $f(x, y, z) = \ln(2 - x - y)$  en  $\mathbb{R}^3$  que contiene al punto  $P(1, 0, -3)$ .

Las superficies de nivel de  $f$  son los planos  $x + y = 2 - e^c$  en  $\mathbb{R}^3$ . En particular, el plano que contiene al punto  $P(1, 0, -3)$  es tal que  $1 + 0 = 2 - e^c$ , es decir,  $c = 0$ . Así, la superficie de nivel buscada es  $x + y = 1$  (con  $z$  libre).



## 2.3 Superficies cuadráticas y otras superficies

**Definición.** Una *superficie* es un conjunto de puntos  $(x, y, z) \in \mathbb{R}^3$  que satisfacen una relación de la forma  $F(x, y, z) = 0$ .

Así, por ejemplo, la ecuación  $x^2 - y^2 + z^2 = 1$  representa una superficie en  $\mathbb{R}^3$ . Cabe señalar que no toda superficie es una función, como veremos a lo largo de esta sección.

## Capítulo 2 Funciones de varias variables

**Definición.** Las *trazas* de una superficie en  $\mathbb{R}^3$  son las curvas formadas por la intersección de la superficie con cada uno de los planos coordenados.

Por ejemplo, para el plano  $2x + y + 3z = 12$  en  $\mathbb{R}^3$ , su traza  $xy$  es la curva

$$2x + y = 12,$$

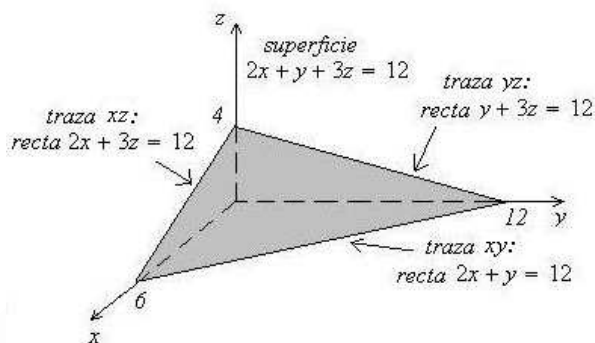
obtenida de la intersección esta superficie con el plano  $z = 0$ . Similarmente, su traza  $yz$  es

$$y + 3z = 12,$$

obtenida a partir de la intersección con el plano  $x = 0$ , y su traza  $xz$  es

$$2x + 3z = 12,$$

obtenida a partir de la intersección con el plano  $y = 0$ .



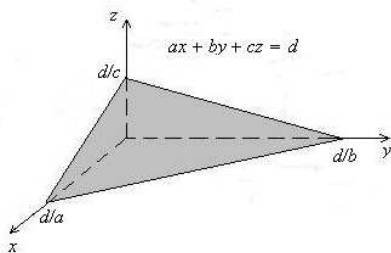
### Algunas superficies famosas en $\mathbb{R}^3$

#### A) Planos

Como ya vimos, un plano  $\pi$  es cualquier conjunto de puntos de la forma

$$\pi = \{ (x, y, z) \in \mathbb{R}^3 \mid ax + by + cz = d \},$$

donde  $a, b, c, d$  son constantes. En la siguiente figura se muestra la gráfica del plano, para  $a, b, c, d \neq 0$ .



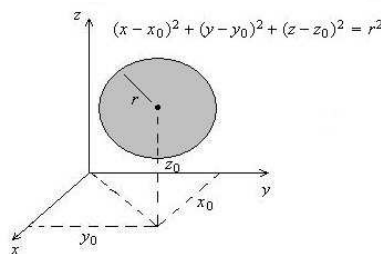
## 2.3 Superficies cuadráticas y otras superficies

### B) Esferas

Una esfera  $S$  de radio  $r$  y centro en  $(x_0, y_0, z_0)$  es un conjunto de puntos de la forma

$$S = \{ (x, y, z) \in \mathbb{R}^3 \mid (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = r^2 \}.$$

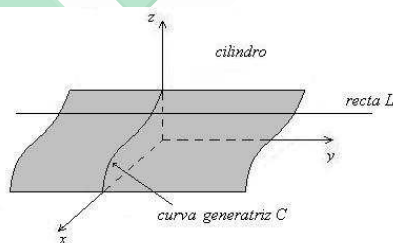
Claramente, si  $r = 0$  el único elemento de  $S$  sería el punto  $(x_0, y_0, z_0)$ .



Por ejemplo, la ecuación  $(x - 1)^2 + y^2 + (z + 3)^2 = 4$  representa una esfera de radio 2 y centro en el punto  $(1, 0, -3)$ .

### C) Cilindros

**Definición.** Sea  $C$  una curva plana y sea  $L$  una recta que no está en el plano de  $C$ . Un *cilindro* es la superficie formada por el conjunto de todas las rectas paralelas a  $L$  que cortan a  $C$ . A  $C$  se le llama la *curva generatriz*, o *directriz*, del cilindro y a las rectas paralelas se les llama *rectas generatrices*. Un *cilindro recto* es aquel tal que  $L$  es perpendicular al plano de  $C$ .



En el caso particular de un cilindro paralelo a alguno de los ejes coordenados, la ecuación correspondiente al cilindro no contiene a la variable de ese eje.

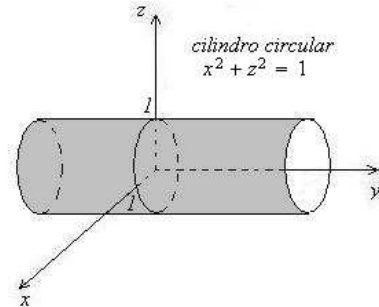
Ejemplos:

1. Esboza la gráfica de  $x^2 + z^2 = 1$  en  $\mathbb{R}^3$ .

Como en esta ecuación no aparece la variable  $y$ , se trata de una superficie en donde esa variable es libre. La ecuación representa un *cilindro circular* que

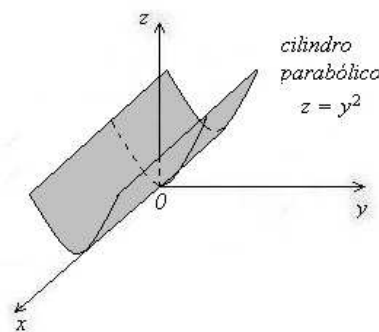
## Capítulo 2 Funciones de varias variables

se extiende a lo largo del eje  $y$ , cuya curva generatriz  $C$  es la circunferencia  $x^2 + z^2 = 1$ .



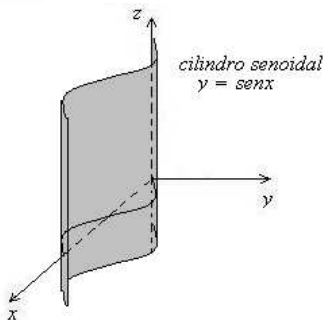
2. Esboza la gráfica de  $z = y^2$  en  $\mathbb{R}^3$ .

Como en esta ecuación no aparece la variable  $x$ , se trata de una superficie en donde esa variable es libre. La ecuación representa un *cilindro parabólico* que se extiende a lo largo del eje  $x$ , cuya curva generatriz  $C$  es la parábola  $z = y^2$ .



3. Esboza la gráfica de  $y = \text{sen} x$  en  $\mathbb{R}^3$ .

Como en esta ecuación no aparece la variable  $z$ , se trata de una superficie en donde esa variable es libre. La ecuación representa un *cilindro senoidal*, cuya curva generatriz  $C$  es la función  $y = \text{sen} x$ , y que se extiende a lo largo del eje  $z$ :



## 2.3 Superficies cuadráticas y otras superficies

### D) Superficies cuadráticas

Para estudiar las superficies cuadráticas se necesita conocer el tema de cónicas. El lector puede encontrar una breve discusión sobre las ecuaciones y gráficas de las cónicas en el Apéndice A.

**Definición.** Una *superficie cuadrática* es la gráfica en  $\mathbb{R}^3$  de una ecuación de segundo grado en las variables  $x, y, z$ , de la forma

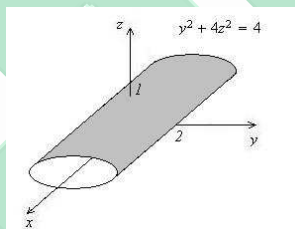
$$Ax^2 + By^2 + Cz^2 + Dxy + Eyz + Fxz + Gx + Hy + Iz + J = 0,$$

con  $A, B, \dots, J$  constantes, y en donde  $A \neq 0$ ,  $B \neq 0$  o  $C \neq 0$ .

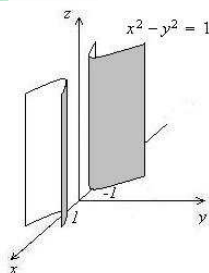
Las esferas y algunos tipos de cilindros son casos particulares de superficies cuadráticas, como se muestra en los siguientes ejemplos.

Ejemplos:

1. La ecuación  $y^2 + 4z^2 = 4$  describe a un cilindro elíptico, que corre a lo largo del eje  $x$ .



2. La ecuación  $x^2 - y^2 = 1$  describe a un cilindro hiperbólico, que corre a lo largo del eje  $z$ .



A continuación presentamos algunas de las superficies cuadráticas más notables, que en general no representan funciones en  $\mathbb{R}^3$ . Se discutirán los casos más simples, en donde las superficies están centradas en el origen, o bien, tendrán a los ejes coordenados como eje de simetría. En todos los casos, supondremos que  $a, b, c \neq 0$ .

**1. Elipsoide:**

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1.$$

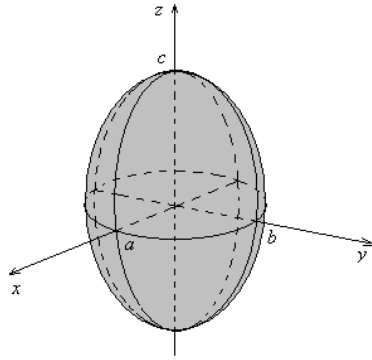
Traza  $xy$ : Elipse  $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$

Traza  $xz$ : Elipse  $\frac{x^2}{a^2} + \frac{z^2}{c^2} = 1$

Traza  $yz$ : Elipse  $\frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$

Curvas de nivel ( $z = K$ ,  $|K| < |c|$ ):

Elipses  $\frac{x^2}{a^2 \left(1 - \frac{K^2}{c^2}\right)} + \frac{y^2}{b^2 \left(1 - \frac{K^2}{c^2}\right)} = 1$



**2. Paraboloide elíptico:**

$$\frac{z}{c} = \frac{x^2}{a^2} + \frac{y^2}{b^2}.$$

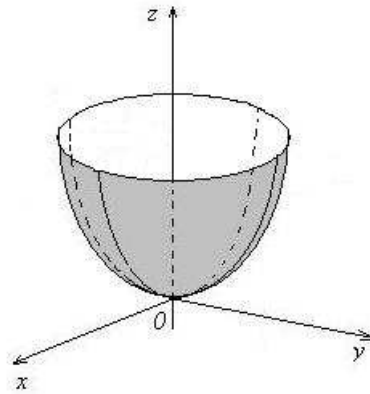
Traza  $xy$ : El origen

Traza  $xz$ : Parábola  $z = \left(\frac{c}{a^2}\right) x^2$

Traza  $yz$ : Parábola  $z = \left(\frac{c}{b^2}\right) y^2$

Curvas de nivel ( $z = K$ ):

Elipses  $\frac{x^2}{a^2 \left(\frac{K}{c}\right)} + \frac{y^2}{b^2 \left(\frac{K}{c}\right)} = 1$



Cuando  $a = b$  se trata de un *paraboloide circular* o *paraboloide de revolución*.

Otras representaciones están dadas por las ecuaciones  $\frac{x}{a} = \frac{y^2}{b^2} + \frac{z^2}{c^2}$  (simetría con respecto al eje  $x$ ) y  $\frac{y}{b} = \frac{x^2}{a^2} + \frac{z^2}{c^2}$  (simetría con respecto al eje  $y$ ).



## 2.3 Superficies cuadráticas y otras superficies

### 3. Hiperboloide de una hoja:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1.$$

Traza  $xy$ : Elipse  $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$

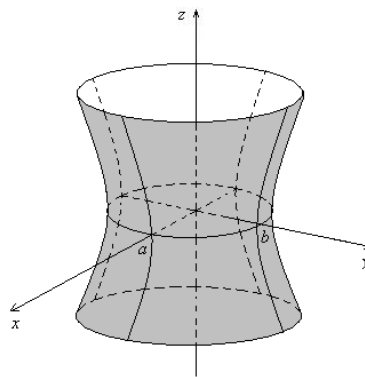
Traza  $xz$ : Hipérbola  $\frac{x^2}{a^2} - \frac{z^2}{c^2} = 1$

Traza  $yz$ : Hipérbola  $\frac{y^2}{b^2} - \frac{z^2}{c^2} = 1$

Curvas de nivel ( $z = K$ ):

$$\text{Elipses } \frac{x^2}{a^2 \left(1 + \frac{K^2}{c^2}\right)} + \frac{y^2}{b^2 \left(1 + \frac{K^2}{c^2}\right)} = 1$$

Otras representaciones son  $\frac{z^2}{c^2} + \frac{y^2}{b^2} - \frac{x^2}{a^2} = 1$  y  $\frac{x^2}{a^2} + \frac{z^2}{c^2} - \frac{y^2}{b^2} = 1$ .



### 4. Hiperboloide de dos hojas:

$$\frac{z^2}{c^2} - \frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$

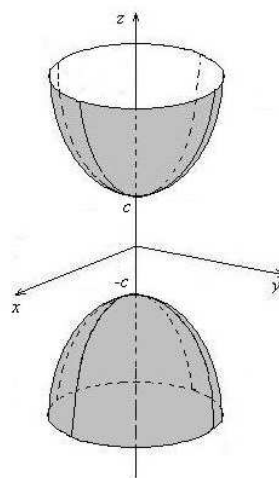
Traza  $xy$ : No existe

Traza  $xz$ : Hipérbola  $\frac{z^2}{c^2} - \frac{x^2}{a^2} = 1$

Traza  $yz$ : Hipérbola  $\frac{z^2}{c^2} - \frac{y^2}{b^2} = 1$

Curvas de nivel ( $z = K$ ,  $|K| \geq |c|$ ):

$$\text{Elipses } \frac{x^2}{a^2 \left(\frac{K^2}{c^2} - 1\right)} + \frac{y^2}{b^2 \left(\frac{K^2}{c^2} - 1\right)} = 1$$



Otras representaciones son  $\frac{x^2}{a^2} - \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1$  y  $\frac{y^2}{b^2} - \frac{x^2}{a^2} - \frac{z^2}{c^2} = 1$ .

### 5. Paraboloide hiperbólico:

$$\frac{z}{c} = \frac{x^2}{a^2} - \frac{y^2}{b^2}$$

Traza  $xy$ : Rectas  $\frac{y}{b} = \pm \frac{x}{a}$

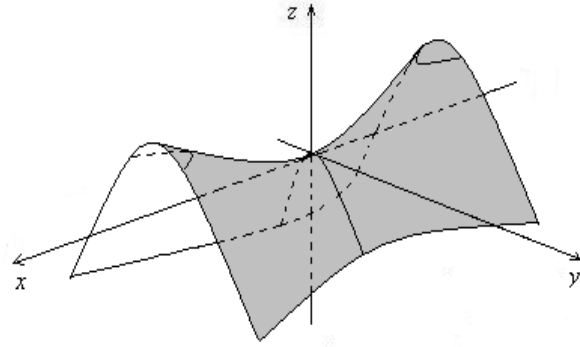
Traza  $xz$ : Parábola  $z = \left(\frac{c}{a^2}\right) x^2$

Traza  $yz$ : Parábola  $z = -\left(\frac{c}{b^2}\right) y^2$

Curvas de nivel ( $z = K$ ):

Hipérbolas  $\frac{x^2}{a^2 \left(\frac{K}{c}\right)} - \frac{y^2}{b^2 \left(\frac{K}{c}\right)} = 1$

Otras representaciones son  $\frac{x}{a} = \frac{y^2}{b^2} - \frac{z^2}{c^2}$  y  $\frac{y}{b} = \frac{x^2}{a^2} - \frac{z^2}{c^2}$ .



### 6. Cono elíptico:

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

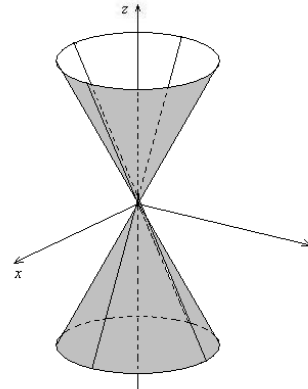
Traza  $xy$ : El origen

Traza  $xz$ : Rectas  $z = \pm \left(\frac{c}{a}\right) x$

Traza  $yz$ : Rectas  $z = \pm \left(\frac{c}{b}\right) y$

Curvas de nivel ( $z = K$ ):

Elipses  $\frac{x^2}{a^2 \left(\frac{K^2}{c^2}\right)} + \frac{y^2}{b^2 \left(\frac{K^2}{c^2}\right)} = 1$



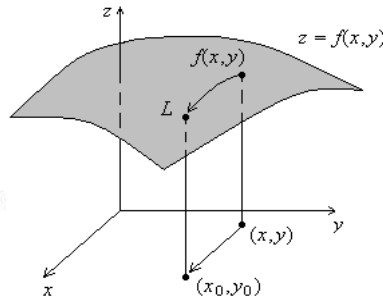
Cuando  $a = b$  se trata de un *cono circular*.

Otras representaciones son  $\frac{y^2}{b^2} + \frac{z^2}{c^2} = \frac{x^2}{a^2}$  y  $\frac{x^2}{a^2} + \frac{z^2}{c^2} = \frac{y^2}{b^2}$ .

## 2.4 Límites y continuidad

En esta sección definiremos los conceptos de límite y continuidad, necesarios para establecer el significado de diferenciabilidad para funciones de varias variables, que se presenta en el capítulo 3. Aunque los resultados que aquí presentamos son válidos en general para funciones cuyo dominio está en  $\mathbb{R}^n$ , los ejemplos y métodos discutidos se centrarán al caso de funciones con dominio en  $\mathbb{R}^2$ .

El concepto de límite para funciones de varias variables es una extensión al de funciones de una variable. La siguiente figura ilustra el significado geométrico de que una función de dos variables,  $z = f(x, y)$ , tenga un límite  $L$  cuando el punto  $(x, y)$  tiende a un punto dado  $(x_0, y_0)$  en el dominio de  $f$ .



Cuando ese límite  $L$  existe, utilizamos la notación

$$\lim_{(x,y) \rightarrow (x_0,y_0)} f(x,y) = L.$$

Desde el punto de vista formal, el límite de  $f$  se define de la siguiente manera.

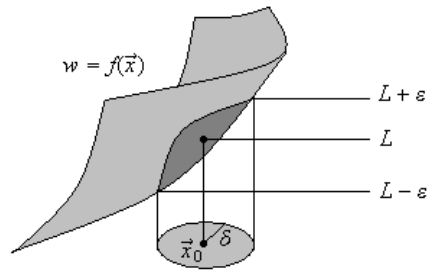
**Definición.** Sea  $f$  una función definida en todo punto  $\vec{x}$  del interior de una vecindad con centro en  $\vec{x}_0$ , excepto quizá en  $\vec{x}_0$ . Se dice que  $f$  tiene límite  $L$  cuando  $\vec{x}$  tiende a  $\vec{x}_0$ , y se escribe

$$\lim_{\vec{x} \rightarrow \vec{x}_0} f(\vec{x}) = L,$$

si para cada número  $\varepsilon > 0$  existe un correspondiente número  $\delta(\varepsilon) > 0$  tal que para todo  $\vec{x}$  en el dominio de  $f$

$$0 < \|\vec{x} - \vec{x}_0\| < \delta \quad \Rightarrow \quad |f(\vec{x}) - L| < \varepsilon.$$

## Capítulo 2 Funciones de varias variables



Cabe señalar que la definición de límite también se aplica al caso de puntos frontera  $\vec{x}_0$  que no estén en el dominio, siempre y cuando los puntos  $\vec{x}$  sí lo estén.

Ejemplos:

1. Demuestra formalmente que  $\lim_{(x,y) \rightarrow (0,0)} x = 0$ .

Para cada número  $\varepsilon > 0$  arbitrario queremos encontrar un correspondiente número  $\delta(\varepsilon)$  tal que

$$0 < \left\| \vec{x} - \vec{0} \right\| < \delta \Rightarrow |x - 0| < \varepsilon,$$

es decir,

$$0 < \sqrt{x^2 + y^2} < \delta \Rightarrow |x| < \varepsilon.$$

Para ello, nota que

$$|x| = \sqrt{x^2} \leq \sqrt{x^2 + y^2} < \delta,$$

por lo que simplemente puedes tomar

$$\delta(\varepsilon) = \varepsilon.$$

2. Demuestra formalmente que  $\lim_{(x,y) \rightarrow (0,0)} \frac{x^2}{\sqrt{x^2 + y^2}} = 0$ .

Para cada  $\varepsilon > 0$  buscamos una  $\delta(\varepsilon)$  tal que

$$0 < \left\| \vec{x} - \vec{0} \right\| < \delta \Rightarrow \left| \frac{x^2}{\sqrt{x^2 + y^2}} - 0 \right| < \varepsilon,$$

es decir,

$$0 < \sqrt{x^2 + y^2} < \delta \Rightarrow \frac{x^2}{\sqrt{x^2 + y^2}} < \varepsilon.$$

Para ello, nota que

$$0 < \frac{x^2}{\sqrt{x^2 + y^2}} \leq \frac{x^2 + y^2}{\sqrt{x^2 + y^2}} = \sqrt{x^2 + y^2} < \delta,$$

## 2.4 Límites y continuidad

por lo que puedes tomar

$$\delta(\varepsilon) = \varepsilon.$$

3. Demuestra formalmente que  $\lim_{(x,y) \rightarrow (0,0)} \frac{2x^2y}{x^2 + y^2} = 0$ .

Para cada  $\varepsilon > 0$  buscamos una  $\delta(\varepsilon)$  tal que

$$0 < \left\| \vec{x} - \vec{0} \right\| < \delta \Rightarrow \left| \frac{2x^2y}{x^2 + y^2} - 0 \right| < \varepsilon,$$

es decir,

$$0 < \sqrt{x^2 + y^2} < \delta \Rightarrow \left| \frac{2x^2y}{x^2 + y^2} \right| < \varepsilon.$$

Como

$$0 \leq |y| = \sqrt{y^2} \leq \sqrt{x^2 + y^2} < \delta,$$

por lo tanto,

$$\left| \frac{2x^2y}{x^2 + y^2} \right| \leq \left| \frac{2x^2y}{x^2} \right| = 2|y| < 2\delta.$$

De esta manera, puedes tomar

$$\delta(\varepsilon) = \frac{\varepsilon}{2}.$$

### Propiedades de los límites

Si  $\lim_{\vec{x} \rightarrow \vec{x}_0} f(\vec{x}) = L_1$  y  $\lim_{\vec{x} \rightarrow \vec{x}_0} g(\vec{x}) = L_2$ , entonces

1. Regla de la suma:

$$\lim_{\vec{x} \rightarrow \vec{x}_0} [f(\vec{x}) + g(\vec{x})] = L_1 + L_2$$

2. Regla del múltiplo constante:

$$\lim_{\vec{x} \rightarrow \vec{x}_0} [k f(\vec{x})] = kL_1, \quad k \in \mathbb{R}$$

3. Regla del producto:

$$\lim_{\vec{x} \rightarrow \vec{x}_0} [f(\vec{x})g(\vec{x})] = L_1L_2$$

4. Regla del cociente:

$$\lim_{\vec{x} \rightarrow \vec{x}_0} \left[ \frac{f(\vec{x})}{g(\vec{x})} \right] = \frac{L_1}{L_2}, \quad L_2 \neq 0$$

5. Regla de la potencia:

$$\lim_{\vec{x} \rightarrow \vec{x}_0} [f(\vec{x})]^{m/n} = [L_1]^{m/n}, \quad m, n \in \mathbb{Z}, \quad n \neq 0, \quad \text{si } [L_1]^{m/n} \in \mathbb{R}$$

Ejemplos:

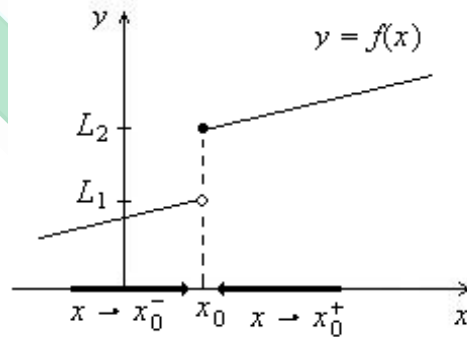
1.  $\lim_{(x,y) \rightarrow (-1,3)} (2x + y) = 2(-1) + 3 = 1.$
2.  $\lim_{(x,y) \rightarrow (3,4)} \sqrt{x^2 + y^2} = \sqrt{3^2 + 4^2} = 5.$
3.  $\lim_{(x,y) \rightarrow (1,0)} \frac{x^2 - 3xy}{xy^2 - 2x + x^3} = \frac{1 - 0}{0 - 2 + 1} = -1.$
4.  $\lim_{(x,y) \rightarrow (0,0)} \frac{x^2 - xy}{\sqrt{x} - \sqrt{y}} = \lim_{(x,y) \rightarrow (0,0)} \left[ \frac{(x^2 - xy)}{(\sqrt{x} - \sqrt{y})} \frac{(\sqrt{x} + \sqrt{y})}{(\sqrt{x} + \sqrt{y})} \right]$   
 $= \lim_{(x,y) \rightarrow (0,0)} \left[ \frac{x(x - y)(\sqrt{x} + \sqrt{y})}{(x - y)} \right] = \lim_{(x,y) \rightarrow (0,0)} [x(\sqrt{x} + \sqrt{y})] = 0.$

### Prueba de las dos trayectorias para demostrar la no existencia de un límite

Si una función  $f(\vec{x})$  tiene límites diferentes a lo largo de dos trayectorias distintas a medida que  $\vec{x}$  tiende a  $\vec{x}_0$ , entonces el límite  $\lim_{\vec{x} \rightarrow \vec{x}_0} f(\vec{x})$  no existe.

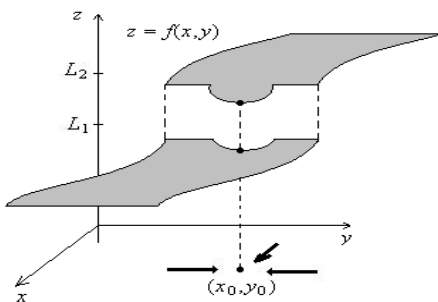
En el caso de funciones  $f$  de una variable, la no existencia del límite se prueba simplemente el límite de  $f$  por las únicas dos trayectorias posibles, a saber,  $x \rightarrow x_0^-$  y  $x \rightarrow x_0^+$ , y mostrando que ambos límites laterales son distintos:

$$\lim_{x \rightarrow x_0^-} f(x) = L_1 \neq L_2 = \lim_{x \rightarrow x_0^+} f(x)$$



## 2.4 Límites y continuidad

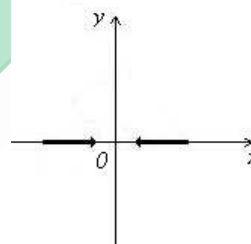
La prueba de las dos trayectorias para el caso de funciones  $f$  de dos variables presenta una mayor dificultad que en el caso anterior, ya que en este caso existe una infinidad de trayectorias posibles en el plano para llegar de  $(x, y)$  a  $(x_0, y_0)$ .



Ejemplos:

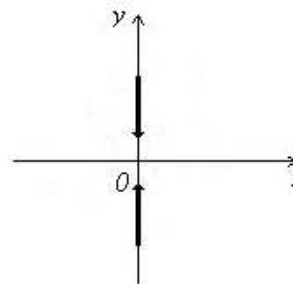
1. Demuestra que no existe el límite de  $f(x, y) = \frac{x^2 - 3y^2}{x^2 + 2y^2}$  en el punto  $(0, 0)$ .
  - i) Tomando el límite a lo largo del eje  $x$  ( $y = 0$ ):

$$\begin{aligned} \lim_{(x,y) \rightarrow (0,0)} f(x, y) &= \lim_{x \rightarrow 0} f(x, 0) \\ &= \lim_{x \rightarrow 0} \frac{x^2 - 0}{x^2 + 0} \\ &= \lim_{x \rightarrow 0} 1 = 1. \end{aligned}$$



- ii) Tomando el límite a lo largo del eje  $y$  ( $x = 0$ ):

$$\begin{aligned} \lim_{(x,y) \rightarrow (0,0)} f(x, y) &= \lim_{y \rightarrow 0} f(0, y) \\ &= \lim_{y \rightarrow 0} \frac{0 - 3y^2}{0 + 2y^2} \\ &= \lim_{y \rightarrow 0} \left( -\frac{3}{2} \right) = -\frac{3}{2}. \end{aligned}$$



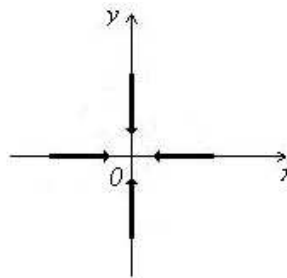
Como los límites son distintos, no existe  $\lim_{(x,y) \rightarrow (0,0)} \frac{x^2 - 3y^2}{x^2 + 2y^2}$ .

## Capítulo 2 Funciones de varias variables

2. Demuestra que no existe el límite de  $f(x, y) = \frac{xy}{x^2 + y^2}$  en el punto  $(0, 0)$ .

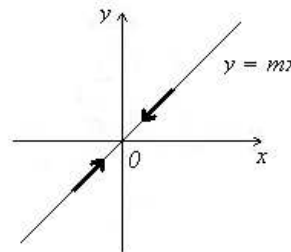
i) Tomando el límite a lo largo de los ejes coordenados:

$$\lim_{(x,y) \rightarrow (0,0)} f(x, 0) = 0 \text{ y } \lim_{(x,y) \rightarrow (0,0)} f(0, y) = 0.$$



ii) Tomando el límite a lo largo de las rectas  $y = mx$ :

$$\begin{aligned} \lim_{(x,y) \rightarrow (0,0)} f(x, y) &= \lim_{x \rightarrow 0} f(x, mx) \\ &= \lim_{x \rightarrow 0} \frac{x(mx)}{x^2 + (mx)^2} \\ &= \lim_{x \rightarrow 0} \frac{m}{1 + m^2} \\ &= \frac{m}{1 + m^2} \neq 0. \end{aligned}$$

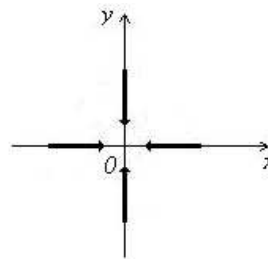


Como hay un límite distinto para cada  $m$ , no existe  $\lim_{(x,y) \rightarrow (0,0)} \frac{xy}{x^2 + y^2}$ .

3. Demuestra que no existe el límite de  $f(x, y) = \frac{x^3 y}{x^6 + y^2}$  en el punto  $(0, 0)$ .

i) Tomando el límite a lo largo de los ejes coordenados:

$$\lim_{(x,y) \rightarrow (0,0)} f(x, 0) = 0 \text{ y } \lim_{(x,y) \rightarrow (0,0)} f(0, y) = 0.$$

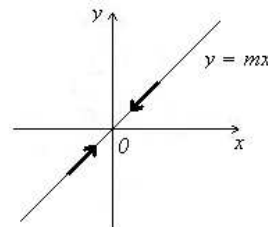




## 2.4 Límites y continuidad

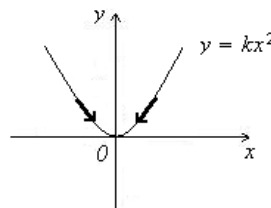
ii) Tomando el límite a lo largo de las rectas  $y = mx$ :

$$\begin{aligned}\lim_{(x,y) \rightarrow (0,0)} f(x,y) &= \lim_{x \rightarrow 0} f(x, mx) \\ &= \lim_{x \rightarrow 0} \frac{x^3(mx)}{x^6 + (mx)^2} \\ &= \lim_{x \rightarrow 0} \frac{mx^2}{x^4 + m^2} = 0.\end{aligned}$$



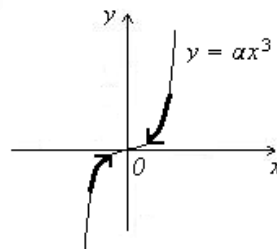
iii) Tomando el límite a lo largo de las parábolas  $y = kx^2$ :

$$\begin{aligned}\lim_{(x,y) \rightarrow (0,0)} f(x,y) &= \lim_{x \rightarrow 0} f(x, kx^2) \\ &= \lim_{x \rightarrow 0} \frac{x^3(kx^2)}{x^6 + (kx^2)^2} \\ &= \lim_{x \rightarrow 0} \frac{kx}{x^2 + k^2} = 0.\end{aligned}$$



iv) Tomando el límite a lo largo de las cúbicas  $y = \alpha x^3$ :

$$\begin{aligned}\lim_{(x,y) \rightarrow (0,0)} f(x,y) &= \lim_{x \rightarrow 0} f(x, \alpha x^3) \\ &= \lim_{x \rightarrow 0} \frac{x^3(\alpha x^3)}{x^6 + (\alpha x^3)^2} \\ &= \lim_{x \rightarrow 0} \frac{\alpha}{1 + \alpha^2} = \frac{\alpha}{1 + \alpha^2} \neq 0.\end{aligned}$$



Como el límite es distinto para cada  $\alpha$ , no existe  $\lim_{(x,y) \rightarrow (0,0)} \frac{x^3 y}{x^6 + y^2}$ .

**Definición.** Una función  $f(\vec{x})$  es *continua en un punto*  $\vec{x}_0$ , si

1.  $f$  está definida en  $\vec{x}_0$ ,
2.  $\lim_{\vec{x} \rightarrow \vec{x}_0} f(\vec{x})$  existe,
3.  $\lim_{\vec{x} \rightarrow \vec{x}_0} f(\vec{x}) = f(\vec{x}_0)$ .

La función es *continua* si lo es en cada punto de su dominio.

Ejemplos:

1. Muestra que  $f(x, y) = 2xy^2 + 3x$  es continua en  $(2, -1)$ .

La función es polinomial, de modo que está definida para todo punto de  $\mathbb{R}^2$ , y en particular en el punto  $(2, -1)$ , con  $f(2, -1) = 10$ . Por otra parte,  $\lim_{(x,y) \rightarrow (2,-1)} (2xy^2 + 3x) = 10$ , de modo que el límite existe. Por último, como  $\lim_{(x,y) \rightarrow (2,-1)} (2xy^2 + 3x) = f(2, -1)$ , por lo tanto  $f(x, y) = 2xy^2 + 3x$  es continua en  $(2, -1)$ .

2. Muestra que la siguiente función es continua en cada punto, excepto en el origen:

$$f(x, y) = \begin{cases} \frac{x^3 y}{x^6 + y^2}, & (x, y) \neq (0, 0) \\ 0, & (x, y) = (0, 0). \end{cases}$$

La función es continua en cada punto  $(x, y) \neq (0, 0)$ , ya que sus valores están dados por una función racional de  $x$  y  $y$ . Sin embargo, como ya mostramos en un ejercicio anterior, la función no tiene límite en el origen. Por lo tanto, la función es continua en cada punto, excepto en el origen.

## Capítulo 3

### Diferenciación

En este capítulo extendemos el concepto de diferenciación para el caso de funciones de varias variables.

#### 3.1 Derivadas parciales. Interpretación geométrica

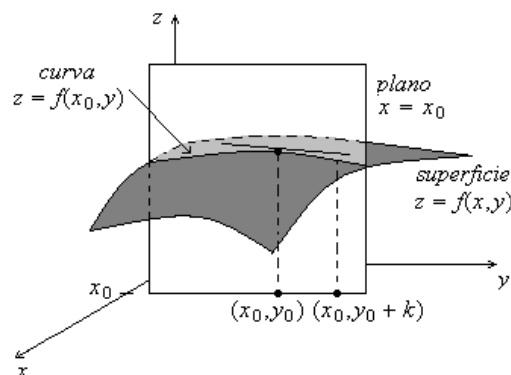
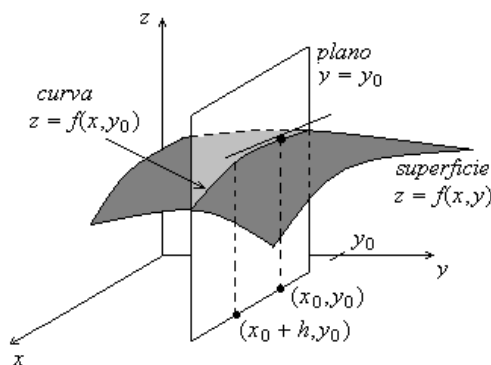
Por simplicidad, aquí nos restringiremos al caso de funciones  $z = f(x, y)$  con dominio en  $\mathbb{R}^2$ , aunque los resultados pueden ser fácilmente generalizados al caso de funciones con dominio en  $\mathbb{R}^n$ .

**Definición.** La *derivada parcial con respecto a  $x$*  de la función continua  $f(x, y)$  en un punto interior  $(x_0, y_0)$  de su dominio está dada por

$$\left. \frac{\partial f}{\partial x} \right|_{(x_0, y_0)} = f_x(x_0, y_0) = \lim_{h \rightarrow 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h},$$

cuando este límite existe. Similarmente, la *derivada parcial con respecto a  $y$*  de  $f$  en  $(x_0, y_0)$  está dada por

$$\left. \frac{\partial f}{\partial y} \right|_{(x_0, y_0)} = f_y(x_0, y_0) = \lim_{k \rightarrow 0} \frac{f(x_0, y_0 + k) - f(x_0, y_0)}{k}.$$



### Capítulo 3 Diferenciación

Las derivadas parciales  $f_x$  y  $f_y$  dan la razón de cambio instantánea de la función  $f(x, y)$  en el punto  $(x_0, y_0)$ , en las direcciones de los vectores base  $\hat{i}$  y  $\hat{j}$ , respectivamente. En otras palabras, la derivada parcial  $f_x$  es la pendiente de la recta tangente a la curva  $z = f(x, y_0)$  en el punto  $P(x_0, y_0, f(x_0, y_0))$  del plano  $y = y_0$ . Asimismo, la derivada parcial  $f_y$  es la pendiente de la recta tangente a la curva  $z = f(x_0, y)$  en el punto  $P(x_0, y_0, f(x_0, y_0))$  del plano  $x = x_0$ .

A partir de la definición, a continuación determinamos la derivada parcial  $f_x$  de la función  $f(x, y) = x^2y^3$ :

$$\begin{aligned}\frac{\partial x^2y^3}{\partial x} &= \lim_{h \rightarrow 0} \frac{(x+h)^2y^3 - x^2y^3}{h} \\ &= \lim_{h \rightarrow 0} \frac{2xhy^3 + h^2y^3}{h} \\ &= \lim_{h \rightarrow 0} (2xy^3 + hy^3) \\ &= 2xy^3.\end{aligned}$$

Nota que este resultado es equivalente a obtener directamente la derivada de  $f$  con respecto a  $x$ , como si  $y$  estuviera fija:

$$\frac{\partial x^2y^3}{\partial x} = y^3 \frac{\partial x^2}{\partial x} = y^3(2x) = 2xy^3.$$

Es posible demostrar que esto es válido en general, es decir, para determinar la derivada parcial  $f_x$  de una función  $z = f(x, y)$  simplemente se toma la derivada de  $f$  con respecto a  $x$ , manteniendo fijo el valor de  $y$ . Similarmente, para obtener la derivada parcial  $f_y$  de la función, se toma la derivada de  $f$  con respecto a  $y$ , manteniendo fijo el valor de  $x$ .

Ejemplos:

1. Sea  $f(x, y) = x^3y + \frac{2y^2}{x}$ . Determina las derivadas parciales  $f_x$  y  $f_y$ .

Las derivadas parciales  $f_x$  y  $f_y$  están dadas por

$$f_x = \frac{\partial}{\partial x} \left( x^3y + \frac{2y^2}{x} \right) = y \frac{\partial}{\partial x} (x^3) + 2y^2 \frac{\partial}{\partial x} \left( \frac{1}{x} \right) = 3x^2y - \frac{2y^2}{x^2},$$

$$f_y = \frac{\partial}{\partial y} \left( x^3y + \frac{2y^2}{x} \right) = x^3 \frac{\partial}{\partial y} (y) + \left( \frac{1}{x} \right) \frac{\partial}{\partial y} (2y^2) = x^3 + \frac{4y}{x}.$$

### 3.1 Derivadas parciales. Interpretación geométrica

2. Sea  $z = x \operatorname{sen}(xy)$ . Determina las derivadas parciales  $z_x$  y  $z_y$ .

Las derivadas parciales  $z_x$  y  $z_y$  están dadas por

$$\begin{aligned} z_x &= \frac{\partial [x \operatorname{sen}(xy)]}{\partial x} \\ &= x \left( \frac{\partial \operatorname{sen}(xy)}{\partial x} \right) + \left( \frac{\partial x}{\partial x} \right) \operatorname{sen}(xy) \\ &= x (\cos(xy) \cdot y) + 1 \cdot \operatorname{sen}(xy) \\ &= xy \cos(xy) + \operatorname{sen}(xy), \\ z_y &= \frac{\partial [x \operatorname{sen}(xy)]}{\partial y} \\ &= x \left( \frac{\partial \operatorname{sen}(xy)}{\partial y} \right) + \left( \frac{\partial x}{\partial y} \right) \operatorname{sen}(xy) \\ &= x (\cos(xy) \cdot x) + 0 \cdot \operatorname{sen}(xy) \\ &= x^2 \cos(xy). \end{aligned}$$

3. Sea  $P(L, K) = L^{1/3} K^{2/3}$  una función de producción Cobb-Douglas. Encuentra los productos marginales  $P_L$  y  $P_K$ .

En este caso, se tiene simplemente

$$\begin{aligned} P_L &= \frac{\partial (L^{1/3} K^{2/3})}{\partial L} = \frac{1}{3} L^{-2/3} K^{2/3} = \frac{1}{2} \left( \frac{K}{L} \right)^{2/3}, \\ P_K &= \frac{\partial (L^{1/3} K^{2/3})}{\partial K} = \frac{2}{3} L^{1/3} K^{-1/3} = \frac{1}{2} \left( \frac{L}{K} \right)^{1/3}. \end{aligned}$$

4. Sea  $u(x_1, x_2) = \ln(x_1^{\alpha_1} x_2^{\alpha_2})$  la función de utilidad para una canasta de dos bienes, con  $x_1, x_2, \alpha_1, \alpha_2 > 0$ . Encuentra las utilidades marginales  $u_{x_1}$  y  $u_{x_2}$ .

Las utilidades marginales están dadas por

$$\begin{aligned} u_{x_1} &= \frac{\partial \ln(x_1^{\alpha_1} x_2^{\alpha_2})}{\partial x_1} = \frac{\partial (\alpha_1 \ln x_1 + \alpha_2 \ln x_2)}{\partial x_1} = \frac{\alpha_1}{x_1}, \\ u_{x_2} &= \frac{\partial \ln(x_1^{\alpha_1} x_2^{\alpha_2})}{\partial x_2} = \frac{\partial (\alpha_1 \ln x_1 + \alpha_2 \ln x_2)}{\partial x_2} = \frac{\alpha_2}{x_2}. \end{aligned}$$

5. Sea  $u(c, \sigma) = \frac{c^{1-\sigma} - 1}{1-\sigma}$ . Determina  $u_c$  y  $u_\sigma$ .

$$\begin{aligned} u_c &= \frac{\partial}{\partial c} \left( \frac{c^{1-\sigma} - 1}{1-\sigma} \right) = c^{-\sigma}, \\ u_\sigma &= \frac{\partial}{\partial \sigma} \left( \frac{c^{1-\sigma} - 1}{1-\sigma} \right) = \frac{-(1-\sigma) c^{1-\sigma} \ln c + (c^{1-\sigma} - 1)}{(1-\sigma)^2} \end{aligned}$$

### Capítulo 3 Diferenciación

6. La función de producción CES (constant elasticity of substitution) está dada por  $P(L, K, \alpha_1, \alpha_2, \rho) = (\alpha_1 L^\rho + \alpha_2 K^\rho)^{1/\rho}$ , con  $-\infty < \rho \leq 1$ ,  $\rho \neq 0$ . Para esta función determina: a)  $\partial P / \partial K$ , b)  $\partial P / \partial \alpha_1$  y c)  $\partial P / \partial \rho$ .

a) La parcial  $\partial P / \partial K$  está dada directamente por

$$\begin{aligned} \frac{\partial P}{\partial K} &= \frac{\partial(\alpha_1 L^\rho + \alpha_2 K^\rho)^{1/\rho}}{\partial K} \\ &= \frac{1}{\rho} (\alpha_1 L^\rho + \alpha_2 K^\rho)^{(1/\rho)-1} \frac{\partial(\alpha_1 L^\rho + \alpha_2 K^\rho)}{\partial K} \\ &= \frac{1}{\rho} (\alpha_1 L^\rho + \alpha_2 K^\rho)^{(1/\rho)-1} (\alpha_2 \rho K^{\rho-1}) \\ &= \alpha_2 K^{\rho-1} (\alpha_1 L^\rho + \alpha_2 K^\rho)^{\frac{1-\rho}{\rho}} \end{aligned}$$

b) Similarmente, la parcial  $\partial P / \partial \alpha_1$  es

$$\begin{aligned} \frac{\partial P}{\partial \alpha_1} &= \frac{\partial(\alpha_1 L^\rho + \alpha_2 K^\rho)^{1/\rho}}{\partial \alpha_1} \\ &= \frac{1}{\rho} (\alpha_1 L^\rho + \alpha_2 K^\rho)^{(1/\rho)-1} \frac{\partial(\alpha_1 L^\rho + \alpha_2 K^\rho)}{\partial \alpha_1} \\ &= \frac{1}{\rho} (\alpha_1 L^\rho + \alpha_2 K^\rho)^{(1/\rho)-1} (L^\rho) \\ &= \frac{L^\rho}{\rho} (\alpha_1 L^\rho + \alpha_2 K^\rho)^{\frac{1-\rho}{\rho}} \end{aligned}$$

c) Para hallar  $\partial P / \partial \rho$  conviene usar derivación logarítmica, ya que la variable  $\rho$  aparece tanto en la base como en la potencia. Para ello, partimos de

$$\ln P(L, K) = \ln(\alpha_1 L^\rho + \alpha_2 K^\rho)^{1/\rho} = \frac{1}{\rho} \ln(\alpha_1 L^\rho + \alpha_2 K^\rho),$$

de modo que

$$\begin{aligned} \frac{1}{P} \frac{\partial P}{\partial \rho} &= \frac{1}{\rho} \frac{\partial(\alpha_1 L^\rho + \alpha_2 K^\rho) / \partial \rho}{(\alpha_1 L^\rho + \alpha_2 K^\rho)} - \frac{1}{\rho^2} \ln(\alpha_1 L^\rho + \alpha_2 K^\rho) \\ &= \frac{1}{\rho} \frac{\alpha_1 L^\rho \ln L + \alpha_2 K^\rho \ln K}{(\alpha_1 L^\rho + \alpha_2 K^\rho)} - \frac{1}{\rho^2} \ln(\alpha_1 L^\rho + \alpha_2 K^\rho), \end{aligned}$$

en donde se utilizó la derivada de un producto y la fórmula de la derivada de  $a^x$  (para obtener  $\partial L^\rho / \partial \rho$  y  $\partial K^\rho / \partial \rho$ ). Finalmente,

$$\begin{aligned} \frac{\partial P}{\partial \rho} &= P \left[ \frac{1}{\rho} \frac{\alpha_1 L^\rho \ln L + \alpha_2 K^\rho \ln K}{(\alpha_1 L^\rho + \alpha_2 K^\rho)} - \frac{1}{\rho^2} \ln(\alpha_1 L^\rho + \alpha_2 K^\rho) \right] \\ &= (\alpha_1 L^\rho + \alpha_2 K^\rho)^{1/\rho} \left[ \frac{1}{\rho} \frac{\alpha_1 L^\rho \ln L + \alpha_2 K^\rho \ln K}{(\alpha_1 L^\rho + \alpha_2 K^\rho)} - \frac{1}{\rho^2} \ln(\alpha_1 L^\rho + \alpha_2 K^\rho) \right]. \end{aligned}$$

### 3.1 Derivadas parciales. Interpretación geométrica

7. Sea  $u(c_1, c_2, \dots, c_n) = \sum_{j=2}^n (\ln c_j + \ln c_{j-1})$  la función de utilidad para la trayectoria de consumo  $c_1, c_2, \dots, c_n$ . Determina  $u_{c_1}, u_{c_2}, \dots, u_{c_n}$ . La función de utilidad está dada por

$$\begin{aligned} u(c_1, c_2, \dots, c_n) &= \sum_{j=2}^n (\ln c_j + \ln c_{j-1}) \\ &= (\ln c_2 + \ln c_1) + (\ln c_3 + \ln c_2) + \dots \\ &\quad + (\ln c_{n-1} + \ln c_{n-2}) + (\ln c_n + \ln c_{n-1}). \end{aligned}$$

De esta manera, las utilidades marginales son

$$\begin{aligned} u_{c_1} &= \frac{\partial u(c_1, c_2, \dots, c_n)}{\partial c_1} = \frac{1}{c_1} \\ u_{c_j} &= \frac{\partial u(c_1, c_2, \dots, c_n)}{\partial c_j} = \frac{2}{c_j}, \quad j = 2, \dots, n-1 \\ u_{c_n} &= \frac{\partial u(c_1, c_2, \dots, c_n)}{\partial c_n} = \frac{1}{c_n} \end{aligned}$$

Similarmente al caso de funciones de una variable, es posible definir derivadas parciales de orden superior y mixtas para funciones de varias variables. En particular, para una función  $f(x, y)$  de dos variables hay 4 posibles derivadas parciales de orden 2, a saber

$$\begin{aligned} f_{xx} &= \frac{\partial}{\partial x} \left( \frac{\partial f}{\partial x} \right) = \frac{\partial^2 f}{\partial x^2}, & f_{xy} &= \frac{\partial}{\partial y} \left( \frac{\partial f}{\partial x} \right) = \frac{\partial^2 f}{\partial y \partial x}, \\ f_{yx} &= \frac{\partial}{\partial x} \left( \frac{\partial f}{\partial y} \right) = \frac{\partial^2 f}{\partial x \partial y}, & f_{yy} &= \frac{\partial}{\partial y} \left( \frac{\partial f}{\partial y} \right) = \frac{\partial^2 f}{\partial y^2}. \end{aligned}$$

Asimismo, hay 8 posibles derivadas parciales de orden 3, del tipo

$$f_{xxx} = \frac{\partial}{\partial x} \left( \frac{\partial}{\partial x} \left( \frac{\partial f}{\partial x} \right) \right) = \frac{\partial^3 f}{\partial x^3}, \quad f_{xxy} = \frac{\partial}{\partial y} \left( \frac{\partial}{\partial x} \left( \frac{\partial f}{\partial x} \right) \right) = \frac{\partial^3 f}{\partial y \partial x^2}, \text{ etc...}$$

y habrían  $2^n$  posibles derivadas parciales de orden  $n$ .

Ejemplos:

1. Verifica que  $z = \ln(x^2 + y^2)$  satisface la ecuación de Laplace,  $\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2} = 0$ .

Como

$$\frac{\partial z}{\partial x} = \frac{2x}{x^2 + y^2}, \quad \frac{\partial z}{\partial y} = \frac{2y}{x^2 + y^2},$$

### Capítulo 3 Diferenciación

por lo tanto

$$\begin{aligned}\frac{\partial^2 z}{\partial x^2} &= \frac{\partial}{\partial x} \left( \frac{\partial z}{\partial x} \right) = \frac{\partial}{\partial x} \left( \frac{2x}{x^2 + y^2} \right) = \frac{2(y^2 - x^2)}{(x^2 + y^2)^2} \\ \frac{\partial^2 z}{\partial y^2} &= \frac{\partial}{\partial y} \left( \frac{\partial z}{\partial y} \right) = \frac{\partial}{\partial y} \left( \frac{2y}{x^2 + y^2} \right) = \frac{2(x^2 - y^2)}{(x^2 + y^2)^2}.\end{aligned}$$

Así,

$$\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2} = \frac{2(y^2 - x^2)}{(x^2 + y^2)^2} + \frac{2(x^2 - y^2)}{(x^2 + y^2)^2} = 0.$$

2. Encuentra todas las derivadas parciales de segundo orden de  $h(r, \theta) = r^3 e^{-\theta/2}$ .

Como

$$h_r(r, \theta) = 3r^2 e^{-\theta/2}, \quad h_\theta(r, \theta) = -\frac{1}{2} r^3 e^{-\theta/2},$$

por lo tanto,

$$\begin{aligned}h_{rr}(r, \theta) &= \frac{\partial}{\partial r} (h_r) = 6r e^{-\theta/2} \\ h_{r\theta}(r, \theta) &= \frac{\partial}{\partial \theta} (h_r) = -\frac{3}{2} r^2 e^{-\theta/2} \\ h_{\theta r}(r, \theta) &= \frac{\partial}{\partial r} (h_\theta) = -\frac{3}{2} r^2 e^{-\theta/2} \\ h_{\theta\theta}(r, \theta) &= \frac{\partial}{\partial \theta} (h_\theta) = \frac{1}{4} r^3 e^{-\theta/2}.\end{aligned}$$

3. Encuentra todas las derivadas parciales de segundo orden de  $f(x, y) = y \ln \left( \frac{x}{y} \right)$ .

Conviene reescribir  $f$  como

$$f(x, y) = y (\ln x - \ln y) = y \ln x - y \ln y$$

Como

$$f_x(x, y) = \frac{y}{x}, \quad f_y(x, y) = \ln x - \ln y - 1,$$

por lo tanto,

$$\begin{aligned}f_{xx}(x, y) &= \frac{\partial}{\partial x} (f_x) = -\frac{y}{x^2} \\ f_{xy}(x, y) &= \frac{\partial}{\partial y} (f_x) = \frac{1}{x} \\ f_{yx}(x, y) &= \frac{\partial}{\partial x} (f_y) = \frac{1}{x} \\ f_{yy}(x, y) &= \frac{\partial}{\partial y} (f_y) = -\frac{1}{y}.\end{aligned}$$



### 3.1 Derivadas parciales. Interpretación geométrica

Observamos que en los ejemplos 2 y 3 las derivadas mixtas son iguales ( $h_{r\theta} = h_{\theta r}$  y  $f_{xy} = f_{yx}$ ). ¿Es éste un resultado general? El siguiente teorema establece bajo qué condiciones se cumple esto.

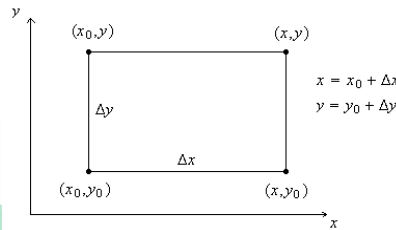
**Teorema sobre derivadas parciales mixtas.**<sup>1</sup> Si  $f(x, y)$  es una función de clase  $C^2$  (dos veces diferenciable con continuidad) en una región abierta que contiene al punto  $(x_0, y_0)$ , entonces las derivadas parciales mixtas son iguales en ese punto, esto es,

$$f_{xy}(x_0, y_0) = f_{yx}(x_0, y_0).$$

Demostración:

Sean  $x = x_0 + \Delta x$  y  $y = y_0 + \Delta y$  y considera la expresión

$$S(\Delta x, \Delta y) = f(x, y) - f(x, y_0) - f(x_0, y) + f(x_0, y_0).$$



Manteniendo  $y_0$  y  $y$  fijos, definimos la función

$$g(x) = f(x, y) - f(x, y_0),$$

de donde

$$S(\Delta x, \Delta y) = g(x) - g(x_0).$$

Por el teorema del valor medio para funciones de una variable,  $g(x) - g(x_0) = g'(c) \Delta x$ , para algún  $c$  entre  $x_0$  y  $x$ . Como  $g'(x) = f_x(x, y) - f_x(x, y_0)$ , por lo tanto

$$S(\Delta x, \Delta y) = [f_x(c, y) - f_x(c, y_0)] \Delta x.$$

Ahora definimos la función

$$h(y) = f_x(c, y) - f_x(c, y_0),$$

con  $h(y_0) = 0$ . Aplicando otra vez el teorema del valor medio,  $h(y) - h(y_0) = h'(d) \Delta y$ , para algún  $d$  entre  $y_0$  y  $y$ . Así,

$$S(\Delta x, \Delta y) = [h'(d) \Delta y] \Delta x.$$

De la definición de  $h(y)$  se sigue que  $h'(y) = f_{xy}(c, y)$ , de donde

$$S(\Delta x, \Delta y) = f_{xy}(c, d) \Delta y \Delta x.$$

<sup>1</sup> Este teorema se atribuye a diversos autores, tales como Euler, Young, Clairaut, Schwarz.

### Capítulo 3 Diferenciación

Como  $f_{xy}$  es continua, se sigue que

$$f_{xy}(x_0, y_0) = \lim_{(\Delta x, \Delta y) \rightarrow (0,0)} \frac{S(\Delta x, \Delta y)}{\Delta x \Delta y}.$$

Como  $S$  es simétrica en  $\Delta x$  y  $\Delta y$ , de forma similar se demuestra que  $f_{yx}(x_0, y_0)$  está dada por la misma fórmula límite, lo que prueba el resultado.

Ejemplo:

Determina  $f_{yyyyxx}$  para la función  $f(x, y) = xe^{y^2}$ .

Para determinar  $f_{yyyyxx}$  hay que encontrar  $f_y$ , luego  $f_{yy}$ , etc... Sin embargo, aquí resulta menos laborioso usar la igualdad  $f_{yyyyxx} = f_{xxyyyy}$ . Como  $f_x = e^{y^2}$ , por lo tanto  $f_{xx} = 0$ , de modo que  $f_{xxyyyy} = 0$ . Concluimos que  $f_{yyyyxx} = 0$ .

La mayoría de las funciones de interés en economía satisfacen las hipótesis del teorema de la igualdad de las derivadas parciales mixtas. En un problema matemático general, esto no necesariamente sucede. Un ejemplo de ello es la función

$$f(x, y) = \begin{cases} \frac{xy(x^2 - y^2)}{x^2 + y^2}, & (x, y) \neq (0, 0) \\ 0, & (x, y) = (0, 0). \end{cases}$$

Es fácil demostrar que las segundas derivadas mixtas  $f_{xy}$  y  $f_{yx}$  son iguales en todos los puntos del dominio, excepto en el origen. En este último punto, se tiene  $f_{xy}(0, 0) = -1$ , mientras que  $f_{yx}(0, 0) = 1$ .

### 3.2 Diferenciabilidad. Linealización y diferenciales

Para comprender el concepto de diferenciabilidad para funciones de varias variables, recordemos primero el caso de funciones de una variable.

Sea  $f : D \subset \mathbb{R} \rightarrow \mathbb{R}$ , con  $y = f(x)$ . Sea  $x_0 \in D$  y considera el cambio

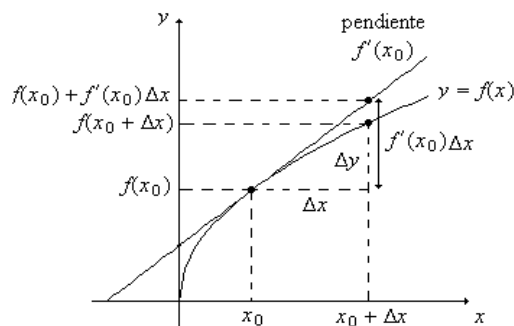
$$\Delta y = f(x_0 + \Delta x) - f(x_0)$$

de  $f$ , al incrementarse  $x_0$  en un valor  $\Delta x = x - x_0$ . Se dice que  $f$  es *diferenciable* en  $x_0$  si  $\Delta y$  está dado por

$$\Delta y = f'(x_0) \Delta x + \varepsilon \Delta x,$$

donde  $\varepsilon \rightarrow 0$  a medida que  $\Delta x \rightarrow 0$ . Geométricamente, esto significa que podemos aproximar el cambio  $\Delta y$  en la altura de la curva  $y = f(x)$  por el cambio  $f'(x_0) \Delta x$  obtenido a partir de la pendiente  $f'(x_0)$  de la curva en  $x_0$ , con un error  $\varepsilon \Delta x$  que decrece a medida que  $x$  se acerca a  $x_0$ .

### 3.2 Diferenciabilidad. Linealización y diferenciales



En ese caso, podemos aproximar

$$\Delta y \cong f'(x_0) \Delta x,$$

de modo que

$$f(x_0 + \Delta x) \cong f(x_0) + f'(x_0) \Delta x.$$

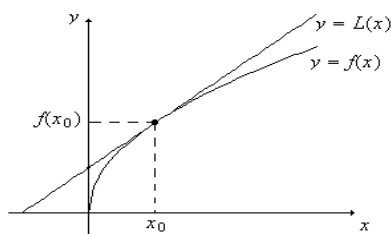
El término del lado derecho de esta expresión se conoce como la *linealización*  $L(x)$  de  $f$  en  $x_0$ ,

$$L(x) = f(x_0) + f'(x_0) \Delta x.$$

La ecuación

$$y = L(x) = f(x_0) + f'(x_0) \Delta x$$

es una ecuación lineal de la forma  $y = y_0 + m(x - x_0)$ , con  $y_0 = f(x_0)$  y  $m = f'(x_0)$ , y representa la ecuación de la *recta tangente* a la curva  $y = f(x)$  en el punto  $(x_0, f(x_0))$  de esa curva.



Concluimos que una función  $f(x)$  es diferenciable en un punto si existe una recta tangente a la curva  $y = f(x)$  en ese punto.

Ejemplo:

Analiza la diferenciabilidad de la función  $f(x) = \ln(1 + x)$  en el punto  $x = 0$ .

La linealización  $L(x)$  de la función  $f(x) = \ln(1 + x)$  en  $x = 0$  está dada por

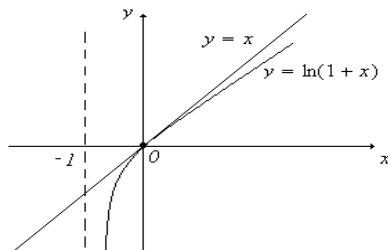
$$L(x) = f(0) + f'(0)(x - 0) = \ln(1 + 0) + \frac{1}{1 + 0}(x - 0) = x,$$

### Capítulo 3 Diferenciación

de donde

$$\ln(1+x) \simeq x, \quad \text{cuando } x \simeq 0.$$

Así,  $f$  es diferenciable en  $x = 0$ , ya que posee una recta tangente en el punto  $(0, 0)$  dada por  $y = x$ .



Por último, para valores muy pequeños de  $\Delta x$  los incrementos se convierten en diferenciales,  $\Delta x \approx dx$ ,  $\Delta y \approx dy$ , de modo que el resultado  $\Delta y \cong f'(x_0) \Delta x$  conduce a la expresión familiar para la diferencial de  $y$ , dada por

$$dy = f'(x_0) dx.$$

A continuación generalizamos los resultados anteriores al caso de funciones de dos variables.

**Definición.** Sea  $f : D \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ , con  $z = f(x, y)$ . Sea  $(x_0, y_0)$  un punto interior de  $D$  y considere el cambio

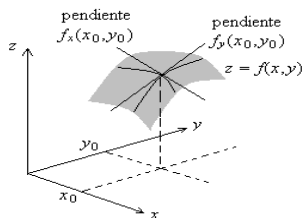
$$\Delta z = f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0, y_0)$$

de  $f$ , al incrementarse  $x_0$  en un valor  $\Delta x = x - x_0$  y  $y_0$  en un valor  $\Delta y = y - y_0$ . Se dice que  $f$  es *diferenciable* en  $(x_0, y_0)$  si  $f_x(x_0, y_0)$  y  $f_y(x_0, y_0)$  existen y si el cambio  $\Delta z$  satisface una ecuación de la forma

$$\Delta z = f_x(x_0, y_0)\Delta x + f_y(x_0, y_0)\Delta y + \varepsilon_1\Delta x + \varepsilon_2\Delta y,$$

en donde  $\varepsilon_1, \varepsilon_2 \rightarrow 0$  cuando  $\Delta x, \Delta y \rightarrow 0$ .

Esto significa que podemos aproximar el cambio  $\Delta z$  en la altura de la superficie  $z = f(x, y)$  por la suma de los cambios  $f_x(x_0, y_0)\Delta x + f_y(x_0, y_0)\Delta y$  obtenidos a partir de las derivadas parciales en  $(x_0, y_0)$ , con un error  $\varepsilon_1\Delta x + \varepsilon_2\Delta y$  que decrece a medida que  $(x, y)$  se acerca a  $(x_0, y_0)$ .



### 3.2 Diferenciabilidad. Linealización y diferenciales

En ese caso, podemos aproximar

$$\Delta z \cong f_x(x_0, y_0)\Delta x + f_y(x_0, y_0)\Delta y,$$

de modo que

$$f(x_0 + \Delta x, y_0 + \Delta y) \cong f(x_0, y_0) + f_x(x_0, y_0)\Delta x + f_y(x_0, y_0)\Delta y.$$

El término de la derecha se conoce como la linealización  $L(x, y)$  de  $f$  en  $(x_0, y_0)$ .

**Definición.** La *linealización*  $L(x, y)$  de una función diferenciable  $f(x, y)$  en un punto  $(x_0, y_0)$  de su dominio es la función

$$L(x, y) = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0).$$

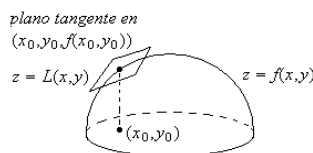
La ecuación

$$z = L(x, y) = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$$

es una ecuación lineal de la forma  $z = z_0 + a(x - x_0) + b(y - y_0)$ , con  $z_0 = f(x_0, y_0)$ ,  $a = f_x(x_0, y_0)$  y  $b = f_y(x_0, y_0)$ , de modo que representa un plano. Nota que la intersección de este plano con el plano  $y = y_0$  es la recta  $z = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0)$ , que es tangente a la superficie  $z = f(x, y)$  en el punto  $(x_0, y_0, f(x_0, y_0))$ . Asimismo, la intersección de este plano con el plano  $x = x_0$  es la recta  $z = f(x_0, y_0) + f_y(x_0, y_0)(y - y_0)$ , que también es tangente a la superficie  $z = f(x, y)$  en el punto  $(x_0, y_0, f(x_0, y_0))$ . De esta manera, la ecuación

$$z = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$$

es la ecuación del *plano tangente* a la superficie  $z = f(x, y)$  en el punto  $(x_0, y_0, f(x_0, y_0))$  de esa superficie.



Concluimos que una función  $f(x, y)$  es diferenciable en un punto si existe un plano tangente a la superficie  $z = f(x, y)$  en ese punto.

Ejemplo:

Analiza la diferenciabilidad de la función  $f(x, y) = x^2 - xy + \frac{1}{2}y^2 + 3$  en el punto  $(x_0, y_0) = (3, 2)$ .

### Capítulo 3 Diferenciación

Encontremos la linealización  $L(x, y)$  de la función  $f$  en  $(x, y) = (3, 2)$ . Como

$$f(x, y) = x^2 - xy + \frac{1}{2}y^2 + 3, \quad f_x(x, y) = 2x - y, \quad f_y(x, y) = -x + y,$$

por lo tanto  $f(3, 2) = 8$ ,  $f_x(3, 2) = 4$ ,  $f_y(3, 2) = -1$ , de modo que,

$$\begin{aligned} z = L(x, y) &= f(3, 2) + f_x(3, 2)(x - 3) + f_y(3, 2)(y - 2) \\ &= 8 + 4(x - 3) - 1(y - 2) = 4x - y - 2. \end{aligned}$$

Así,  $f$  es diferenciable en  $(3, 2)$ , ya que posee un plano tangente en el punto  $(3, 2, 8)$ , dado por su linealización,  $z = 4x - y - 2$ .

Por último, para valores muy pequeños de  $\Delta x$  y  $\Delta y$  los incrementos se convierten en diferenciales,  $\Delta x \approx dx$ ,  $\Delta y \approx dy$ ,  $\Delta z \approx dz$ , de modo que el resultado  $\Delta z \cong f_x(x_0, y_0)\Delta x + f_y(x_0, y_0)\Delta y$  conduce a la expresión  $dz = f_x(x_0, y_0)dx + f_y(x_0, y_0)dy$ , conocida como la diferencial total.

**Definición.** La *diferencial total*  $dz$  de una función diferenciable  $z = f(x, y)$  en un punto interior  $(x_0, y_0)$  de su dominio está dada por

$$dz = f_x(x_0, y_0)dx + f_y(x_0, y_0)dy.$$

La diferencial total puede utilizarse para *aproximar* el cambio en el valor de  $f$  cuando el punto  $(x_0, y_0)$  cambia a un valor cercano  $(x_0 + \Delta x, y_0 + \Delta y)$ , mediante

$$\Delta z \approx f_x(x_0, y_0)\Delta x + f_y(x_0, y_0)\Delta y.$$

Ejemplos:

1. Sea  $z = f(x, y)$ , con  $f(x, y) = x^2 e^{3y}$ . Encuentra la diferencial total  $dz$  en el punto  $(1, 0)$ . Utiliza ésta para estimar el cambio en  $z$  cuando  $x$  disminuye, de  $x = 1$  a  $x = 0.99$ , y  $y$  se incrementa, de  $y = 0$  a  $y = 0.02$ .

Como

$$f_x(x, y) = 2xe^{3y} \text{ y } f_y(x, y) = 3x^2 e^{3y},$$

por lo tanto,  $f_x(1, 0) = 2$  y  $f_y(1, 0) = 3$ . De este modo, la diferencial total de  $f$  está dada por

$$dz = 2dx + 3dy.$$

En el punto inicial la función tiene un valor  $z = f(1, 0) = 1$ . Al disminuir  $x$  en  $\Delta x = 0.99 - 1 = -0.01$  y al incrementarse  $y$  en  $\Delta y = 0.02 - 0 = 0.02$ , la función cambia aproximadamente en

$$\Delta z \approx 2 \Delta x + 3 \Delta y = 2(-0.01) + 3(0.02) = 0.04.$$

### 3.3 Regla de la cadena

En otras palabras,  $z$  cambia de 1 a 1.04, aproximadamente. Nota que el cambio exacto de  $z$  es

$$\Delta z_{exacto} = (0.99)^2 e^{3(0.02)} - (1)^2 e^{3(0)} = 0.0407.$$

2. Sea  $Q = P(L, K)$  la producción, con  $P(L, K) = 4L^{1/4}K^{3/4}$ . Aproxima el efecto que tendría sobre la producción que el trabajo disminuya, de  $L_0 = 625$  a  $L = 623$ , y el capital se incremente, de  $K_0 = 10\,000$  a  $K = 10\,010$ .

La diferencial total  $dQ$  en los niveles iniciales  $(L_0, K_0) = (625, 10\,000)$  es

$$dQ = P_L(625, 10\,000) dL + P_K(625, 10\,000) dK.$$

Los productos marginales están dados por

$$P_L(L, K) = \left(\frac{K}{L}\right)^{3/4} \text{ y } P_K(L, K) = 3 \left(\frac{L}{K}\right)^{1/4},$$

de modo que  $P_L(625, 10\,000) = 8$  y  $P_K(625, 10\,000) = 1.5$ . Así, la diferencial total de  $Q$  en  $(625, 10\,000)$  es

$$dQ = 8 dL + 1.5 dK.$$

Esta expresión permite aproximar el cambio en  $Q$  ante un pequeño cambio en los insumos  $(L, K)$  alrededor de  $(625, 10\,000)$ , mediante

$$\Delta Q \approx 8 \Delta L + 1.5 \Delta K.$$

Tomando en cuenta que  $\Delta L = -2$  y  $\Delta K = 10$ , por lo tanto

$$\Delta Q \approx 8(-2) + 1.5(10) = -1.$$

Así,  $Q$  decrece aproximadamente en 1. Nota que el cambio exacto es

$$\begin{aligned} \Delta Q_{exacto} &= P(623, 10\,010) - P(625, 10\,000) \\ &= 19\,998.967 - 20\,000 = -1.033. \end{aligned}$$

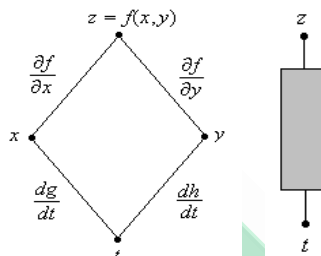
### 3.3 Regla de la cadena

En esta sección mostramos cómo se generaliza la regla de la cadena para la derivada de la composición de funciones de varias variables. Para tal fin, nos ayudamos con los llamados *diagramas de árbol*, que son esquemas en los que se especifica la dependencia que guardan entre sí las variables involucradas. A continuación, ilustramos la regla de la cadena a través de varios ejemplos. En cada caso, la figura de la izquierda representa el diagrama de árbol correspondiente y la figura de la derecha muestra la dependencia final de  $z$ , sin tomar en cuenta a las variables intermedias.

### Capítulo 3 Diferenciación

A) Sea  $z = f(x, y)$ , con  $x = g(t)$ ,  $y = h(t)$ .

$$\therefore \frac{dz}{dt} = \frac{\partial f}{\partial x} \frac{dg}{dt} + \frac{\partial f}{\partial y} \frac{dh}{dt}$$



Esta expresión, o equivalentemente  $\frac{dz}{dt} = \frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt}$ , se denomina la *derivada total* de  $z$  con respecto a  $t$ . Compara ésta con la diferencial total  $dz = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy$  de la sección 3.2. Nota que la derivada total  $\frac{dz}{dt}$  se puede expresar como un producto de matrices, de la forma

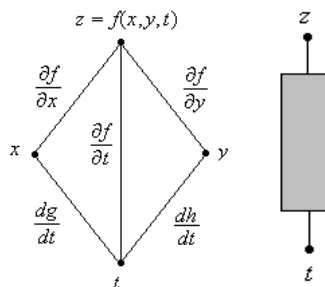
$$\frac{dz}{dt} = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{pmatrix} \begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{pmatrix}.$$

B) Sea  $z = f(x, y, t)$ , con  $x = g(t)$ ,  $y = h(t)$ .

$$\therefore \frac{dz}{dt} = \frac{\partial f}{\partial x} \frac{dg}{dt} + \frac{\partial f}{\partial y} \frac{dh}{dt} + \frac{\partial f}{\partial t}$$

o bien

$$\frac{dz}{dt} = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial t} \end{pmatrix} \begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \\ 1 \end{pmatrix}$$





### 3.3 Regla de la cadena

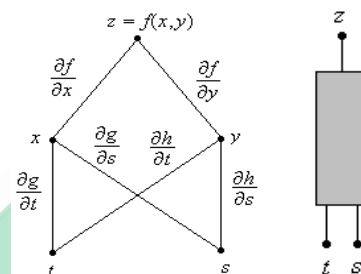
C) Sea  $z = f(x, y)$ , con  $x = g(t, s)$ ,  $y = h(t, s)$ .

$$\therefore \frac{\partial z}{\partial t} = \frac{\partial f}{\partial x} \frac{\partial g}{\partial t} + \frac{\partial f}{\partial y} \frac{\partial h}{\partial t}$$

$$\frac{\partial z}{\partial s} = \frac{\partial f}{\partial x} \frac{\partial g}{\partial s} + \frac{\partial f}{\partial y} \frac{\partial h}{\partial s}$$

o bien

$$\begin{pmatrix} \frac{\partial z}{\partial t} & \frac{\partial z}{\partial s} \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{pmatrix} \begin{pmatrix} \frac{\partial g}{\partial t} & \frac{\partial g}{\partial s} \\ \frac{\partial h}{\partial t} & \frac{\partial h}{\partial s} \end{pmatrix}$$



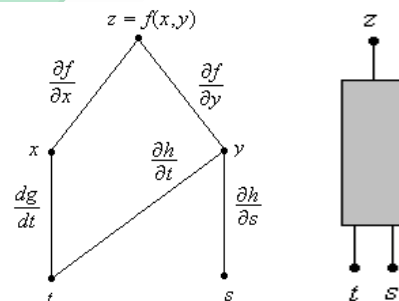
D) Sea  $z = f(x, y)$ , con  $x = g(t)$ ,  $y = h(t, s)$ .

$$\therefore \frac{\partial z}{\partial t} = \frac{\partial f}{\partial x} \frac{dg}{dt} + \frac{\partial f}{\partial y} \frac{\partial h}{\partial t}$$

$$\frac{\partial z}{\partial s} = \frac{\partial f}{\partial y} \frac{\partial h}{\partial s}$$

o bien

$$\begin{pmatrix} \frac{\partial z}{\partial t} & \frac{\partial z}{\partial s} \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{pmatrix} \begin{pmatrix} \frac{dg}{dt} & 0 \\ \frac{\partial h}{\partial t} & \frac{\partial h}{\partial s} \end{pmatrix}$$



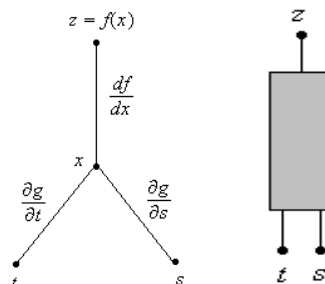
E) Sea  $z = f(x)$ , con  $x = g(t, s)$ .

$$\therefore \frac{\partial z}{\partial t} = \frac{df}{dx} \frac{\partial g}{\partial t}$$

$$\frac{\partial z}{\partial s} = \frac{df}{dx} \frac{\partial g}{\partial s}$$

o bien

$$\begin{pmatrix} \frac{\partial z}{\partial t} & \frac{\partial z}{\partial s} \end{pmatrix} = \begin{pmatrix} \frac{df}{dx} \end{pmatrix} \begin{pmatrix} \frac{\partial g}{\partial t} & \frac{\partial g}{\partial s} \end{pmatrix}$$



### Capítulo 3 Diferenciación

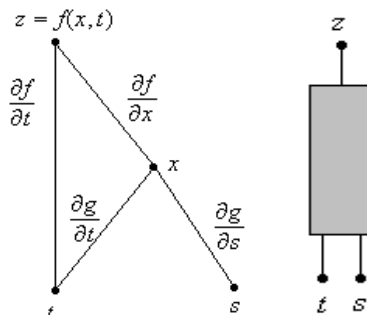
F) Sea  $z = f(x, t)$ , con  $x = g(t, s)$ .

$$\therefore \frac{\partial z}{\partial t} = \frac{\partial f}{\partial x} \frac{\partial g}{\partial t} + \frac{\partial f}{\partial t}$$

$$\frac{\partial z}{\partial s} = \frac{\partial f}{\partial x} \frac{\partial g}{\partial s}$$

o bien

$$\begin{pmatrix} \frac{\partial z}{\partial t} & \frac{\partial z}{\partial s} \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial t} \end{pmatrix} \begin{pmatrix} \frac{\partial g}{\partial t} & \frac{\partial g}{\partial s} \\ 1 & 0 \end{pmatrix}$$



Con este procedimiento podemos obtener la regla de la cadena para cualquier otro caso. Sólo debes tener cuidado en el uso correcto de derivadas ordinarias ( $d/dt$ ) o de derivadas parciales ( $\partial/\partial t$ ) en cada etapa del proceso de derivación.

Ejemplos:

1. Sea  $z = f(x, y) = xy$ , donde  $x = g(t) = \cos t$  y  $y = h(t) = \sin t$ . Encuentra  $dz/dt$ .

Primero notamos que  $\partial f/\partial x = y$  y  $\partial f/\partial y = x$ , de modo que

$$\frac{dz}{dt} = \frac{\partial f}{\partial x} \frac{dg}{dt} + \frac{\partial f}{\partial y} \frac{dh}{dt} = y \frac{dg}{dt} + x \frac{dh}{dt}.$$

Por otra parte, como  $dg/dt = -\sin t$  y  $dh/dt = \cos t$ , por lo tanto,

$$\frac{dz}{dt} = (\sin t)(-\sin t) + (\cos t)(\cos t) = -\sin^2 t + \cos^2 t.$$

Nota que hubieras obtenido el mismo resultado al sustituir  $x = \cos t$  y  $y = \sin t$  en  $z = xy$  y luego derivar el producto respecto a  $t$ .

2. Sea  $Q = P(L, K)$  una función de producción, donde el trabajo  $L(t)$  y el capital  $K(t)$  son funciones del tiempo  $t$ . Encuentra una expresión que establezca cómo cambia la producción en el tiempo,  $dQ/dt$ .

Este caso es similar al anterior. Sin embargo, aquí desconocemos la dependencia explícita de las variables, de modo que sólo se tiene el resultado general

$$\begin{aligned} \frac{dQ}{dt} &= \frac{\partial P}{\partial L} \frac{dL}{dt} + \frac{\partial P}{\partial K} \frac{dK}{dt} \\ &= P_L \frac{dL}{dt} + P_K \frac{dK}{dt}. \end{aligned}$$

### 3.4 Diferenciación implícita

3. Sea  $Q = P(L, K, t)$  una función de producción que depende del tiempo, no sólo indirectamente, a través del trabajo y el capital,  $L(t)$  y  $K(t)$ , sino también directamente (un ejemplo de esto sería  $Q = A(t)L^{1/2}K^{1/2}$ ). Encuentra una expresión para  $dQ/dt$ .

Aquí sólo hay que agregar el término  $\partial P/\partial t$  al resultado anterior, quedando

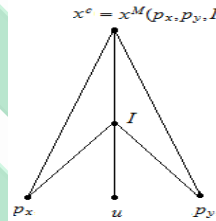
$$\begin{aligned}\frac{dQ}{dt} &= \frac{\partial P}{\partial L} \frac{dL}{dt} + \frac{\partial P}{\partial K} \frac{dK}{dt} + \frac{\partial P}{\partial t} \\ &= P_L \frac{dL}{dt} + P_K \frac{dK}{dt} + \frac{\partial P}{\partial t}.\end{aligned}$$

4. Determina  $\partial x^c / \partial p_x$ , si

$$x^c(p_x, p_y, u) = x^M(p_x, p_y, I), \quad \text{con } I = E(p_x, p_y, u).$$

En este caso, se tiene

$$\frac{\partial x^c}{\partial p_x} = \frac{\partial x^M}{\partial p_x} + \frac{\partial x^M}{\partial I} \frac{\partial E}{\partial p_x},$$



que es conocida como la ecuación de Slutsky.

### 3.4 Diferenciación implícita

Seguramente recuerdas cómo obtener la derivada  $dy/dx$  cuando  $y$  y  $x$  están relacionadas a través de una función implícita  $F(x, y) = 0$ . Por ejemplo, determinemos  $dy/dx$  en la ecuación

$$xe^{xy} + y - 1 = 0,$$

que define implícitamente a  $y$  como función de  $x$ . Para ello, derivas ambos lados respecto a  $x$ , obteniendo

$$\begin{aligned}x \cdot e^{xy} \left( x \cdot \frac{dy}{dx} + y \cdot 1 \right) + e^{xy} \cdot 1 + \frac{dy}{dx} &= 0 \\ \therefore \frac{dy}{dx} \cdot (x^2 e^{xy} + 1) &= -xye^{xy} - e^{xy} \\ \therefore \frac{dy}{dx} &= -\frac{xye^{xy} + e^{xy}}{x^2 e^{xy} + 1}.\end{aligned}$$

### Capítulo 3 Diferenciación

En el caso de una ecuación implícita de la forma  $F(x, y, z) = 0$ , para obtener las derivadas parciales  $\partial z/\partial x$  y  $\partial z/\partial y$  puedes seguir un procedimiento similar al anterior. Por ejemplo, consideremos la ecuación

$$yz - \ln z = x + y,$$

que define a  $z$  como una función implícita, diferenciable, de  $x$  y  $y$ . Para encontrar  $\partial z/\partial x$  derivamos la ecuación respecto a  $x$ , tomando a  $y$  fija, con lo cual se obtiene

$$\begin{aligned} y \cdot \frac{\partial z}{\partial x} - \frac{1}{z} \cdot \frac{\partial z}{\partial x} &= 1 \\ \frac{\partial z}{\partial x} \left( y - \frac{1}{z} \right) &= 1 \\ \frac{\partial z}{\partial x} &= \frac{z}{yz - 1}. \end{aligned}$$

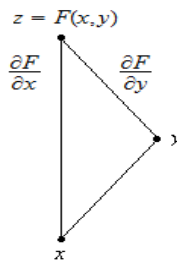
La derivada  $\partial z/\partial y$  se obtendría de manera análoga, obteniendo

$$\frac{\partial z}{\partial y} = \frac{z(1 - z)}{yz - 1}.$$

A continuación presentamos una técnica alternativa para obtener estas derivadas parciales, de una manera más simple, utilizando la regla de la cadena.

**Caso 1.** Queremos encontrar la derivada  $dy/dx$ , suponiendo que la ecuación  $F(x, y) = 0$  define a  $y$  como una función implícita, diferenciable, de  $x$ . Para ello, vamos a suponer que  $F(x, y) = 0$  es la curva de nivel  $z = 0$  de una función  $z = F(x, y)$  en  $\mathbb{R}^3$ . Así,

$$\begin{aligned} z &= F(x, y) = 0 \\ \therefore \frac{dz}{dx} &= \frac{\partial F}{\partial x} + \frac{\partial F}{\partial y} \frac{dy}{dx} = 0 \\ \therefore F_x + F_y \frac{dy}{dx} &= 0 \\ \therefore \frac{dy}{dx} &= -\frac{F_x}{F_y}. \end{aligned}$$



**Teorema.** Si  $F(x, y)$  es diferenciable, y la ecuación  $F(x, y) = 0$  define a  $y$  como una función implícita, diferenciable, de  $x$ , entonces

$$\frac{dy}{dx} = -\frac{F_x}{F_y},$$

en todos los puntos de la curva  $F(x, y) = 0$  en donde  $F_y \neq 0$ .

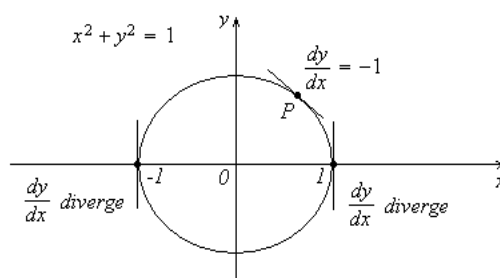
### 3.4 Diferenciación implícita

Ejemplos:

1. Determina en qué puntos la relación  $x^2 + y^2 = 1$  define a  $y$  como una función diferenciable de  $x$ , y en ese caso encuentra  $dy/dx$ .

Sea  $F(x, y) = x^2 + y^2 - 1$ . La relación  $F(x, y) = x^2 + y^2 - 1 = 0$  define a  $y$  como función implícita, diferenciable, de  $x$  cuando  $F_y = 2y \neq 0$ , es decir, en todos los puntos de la circunferencia en donde  $y \neq 0$ . En ese caso,

$$\frac{dy}{dx} = -\frac{F_x}{F_y} = -\frac{2x}{2y} = -\frac{x}{y}.$$



Así, por ejemplo, la derivada  $\frac{dy}{dx}$  en el punto  $P(1/\sqrt{2}, 1/\sqrt{2})$  es

$$\left. \frac{dy}{dx} \right|_P = -\frac{1/\sqrt{2}}{1/\sqrt{2}} = -1.$$

Notamos que en los puntos con  $y = 0$  la derivada se indetermina, de modo que ahí  $y$  no es función diferenciable de  $x$ .

2. Determina en qué puntos la relación  $xe^{xy} + y - 1 = 0$  define a  $y$  como una función diferenciable de  $x$ , y en ese caso encuentra  $dy/dx$ .

Definimos  $F(x, y) = xe^{xy} + y - 1$ , de modo que  $F_x = xye^{xy} + e^{xy}$  y  $F_y = x^2e^{xy} + 1$ . La relación  $F(x, y) = xe^{xy} + y - 1 = 0$  define a  $y$  como una función diferenciable de  $x$  en aquellos puntos sobre la curva tales que  $F_y = x^2e^{xy} + 1 \neq 0$  (que en este caso siempre se cumple). En ese caso,

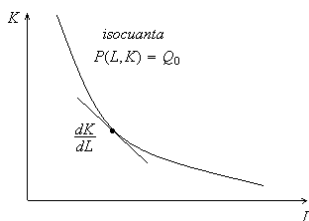
$$\frac{dy}{dx} = -\frac{F_x}{F_y} = -\frac{xye^{xy} + e^{xy}}{x^2e^{xy} + 1}.$$

3. Demuestra que, a lo largo de una isocuanta  $P(L, K) = Q_0$  ( $Q_0 = \text{const}$ ) de una función de producción  $Q = P(L, K)$  se cumple la relación  $-\frac{dK}{dL} = \frac{P_L}{P_K}$  (tasa marginal de sustitución técnica).

### Capítulo 3 Diferenciación

Sea  $F(L, K) = P(L, K) - Q_0$ . La relación  $F(L, K) = P(L, K) - Q_0 = 0$  define a  $K$  como una función diferenciable de  $L$ , si  $F_K = P_K \neq 0$ . En ese caso,

$$\frac{dK}{dL} = -\frac{F_L}{F_K} = -\frac{P_L}{P_K}.$$



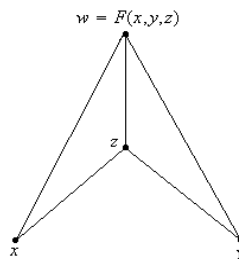
**Caso 2.** Queremos encontrar las derivadas parciales  $\partial z/\partial x$  y  $\partial z/\partial y$ , si la ecuación  $F(x, y, z) = 0$  define a  $z$  como una función implícita, diferenciable, de  $x$  y  $y$ . Para ello, vamos a suponer que  $F(x, y, z) = 0$  es la superficie de nivel  $w = 0$  de una función  $w = F(x, y, z)$  en  $\mathbb{R}^4$ . Así,

$$w = F(x, y, z) = 0$$

$$\therefore \frac{\partial w}{\partial x} = \frac{\partial F}{\partial x} + \frac{\partial F}{\partial z} \frac{\partial z}{\partial x} = F_x + F_z \frac{\partial z}{\partial x} = 0$$

$$\frac{\partial w}{\partial y} = \frac{\partial F}{\partial y} + \frac{\partial F}{\partial z} \frac{\partial z}{\partial y} = F_y + F_z \frac{\partial z}{\partial y} = 0$$

$$\therefore \frac{\partial z}{\partial x} = -\frac{F_x}{F_z}, \quad \frac{\partial z}{\partial y} = -\frac{F_y}{F_z}.$$



**Teorema.** Si  $F(x, y, z)$  es diferenciable, y la ecuación  $F(x, y, z) = 0$  define a  $z$  como una función implícita, diferenciable, de  $x$  y  $y$ , entonces

$$\frac{\partial z}{\partial x} = -\frac{F_x}{F_z}, \quad \frac{\partial z}{\partial y} = -\frac{F_y}{F_z},$$

en todos los puntos de la superficie  $F(x, y, z) = 0$  en donde  $F_z \neq 0$ .

Este teorema permite determinar fácilmente las derivadas parciales  $\partial z/\partial x$  y  $\partial z/\partial y$  en una relación implícita de la forma  $F(x, y, z) = 0$ , siempre y cuando  $F$  sea diferenciable y su derivada parcial  $F_z$  no se anule en el punto  $(x, y, z)$ . Este

### 3.4 Diferenciación implícita

resultado es muy útil, ya que podemos encontrar las derivadas  $\partial z/\partial x$  y  $\partial z/\partial y$  sin necesidad de conocer la función  $z(x, y)$ .

Ejemplos:

1. Determina en qué puntos la relación  $yz - \ln z = x + y$  define a  $z$  como una función diferenciable de  $x$  y  $y$ . En ese caso, encuentra  $\partial z/\partial x$  y  $\partial z/\partial y$ .

Definimos  $F(x, y, z) = yz - \ln z - x - y$ . Así,  $F(x, y, z) = yz - \ln z - x - y$  define a  $z$  como una función diferenciable de  $x$  y  $y$  en todos los puntos  $(x, y, z)$  tales que  $F_z = y - 1/z \neq 0$ , es decir, en todos los puntos de la superficie en donde  $yz - 1 \neq 0$ . En ese caso,

$$\frac{\partial z}{\partial x} = -\frac{F_x}{F_z} = -\left(\frac{-1}{y - 1/z}\right) = \frac{z}{yz - 1},$$

$$\frac{\partial z}{\partial y} = -\frac{F_y}{F_z} = -\left(\frac{z - 1}{y - 1/z}\right) = \frac{z(1 - z)}{yz - 1}.$$

2. Determina si  $3xe^{yz} - ye^{xz} - 1 = 0$  define a  $z$  como una función diferenciable de  $x$  y  $y$  en el punto  $P(1, 2, 0)$ . De ser así, calcula  $\partial z/\partial x$  y  $\partial z/\partial y$  en  $P$ .

Definimos  $F(x, y, z) = 3xe^{yz} - ye^{xz} - 1$ . Como  $F_z(x, y, z) = 3xye^{yz} - xye^{xz}$ , por lo tanto  $F_z(1, 2, 0) = 4 \neq 0$ , de modo que la relación  $3xe^{yz} - ye^{xz} - 1 = 0$  sí define a  $z$  como una función diferenciable de  $x$  y  $y$  cerca del punto  $P(1, 2, 0)$ . Finalmente, como

$$\frac{\partial z}{\partial x} = -\frac{F_x}{F_z} = -\frac{3e^{yz} - ye^{xz}}{xy(3e^{yz} - e^{xz})},$$

$$\frac{\partial z}{\partial y} = -\frac{F_y}{F_z} = -\frac{3xe^{yz} - e^{xz}}{xy(3e^{yz} - e^{xz})},$$

por lo tanto

$$\left.\frac{\partial z}{\partial x}\right|_P = -\frac{3}{4} \quad \text{y} \quad \left.\frac{\partial z}{\partial y}\right|_P = \frac{1}{4}.$$

3. Sean  $D(p, w)$  y  $S(p, t)$  las funciones de demanda ( $D$ ) y de oferta ( $S$ ) de un bien, en términos del precio  $p$  de éste en el mercado, el salario  $w$  y el impuesto  $t$  sobre el producto, y en donde sus derivadas parciales satisfacen  $D_p < 0$ ,  $D_w > 0$ ,  $S_p > 0$  y  $S_t < 0$ . Si en el equilibrio se cumple  $D(p, w) = S(p, t)$ , determina bajo qué condiciones la relación de equilibrio define a  $p$  como una función diferenciable de las variables  $w$  y  $t$ , y en ese caso encuentra expresiones para  $\partial p/\partial w$  y  $\partial p/\partial t$ .

### Capítulo 3 Diferenciación

Podemos escribir la condición de equilibrio como  $D(p, w) - S(p, t) = 0$ , de modo que definimos  $F(p, w, t) = D(p, w) - S(p, t)$ . Así, la relación  $D(p, w) - S(p, t) = 0$  define a  $p$  como una función diferenciable de las variables  $w$  y  $t$ , en los puntos tales que  $F_p(p, w, t) = D_p(p, w) - S_p(p, t) \neq 0$ , es decir, en todos los puntos en donde  $D_p(p, w) \neq S_p(p, t)$ . En ese caso,

$$\frac{\partial p}{\partial w} = -\frac{F_w}{F_p} = -\frac{D_w}{D_p - S_p} = \frac{D_w}{S_p - D_p} > 0,$$

$$\frac{\partial p}{\partial t} = -\frac{F_t}{F_p} = -\frac{(-S_t)}{D_p - S_p} = \frac{S_t}{D_p - S_p} > 0.$$

Es interesante notar cómo hemos podido deducir este resultado general, aun sin conocer la forma explícita de las funciones de demanda y de oferta.

En el caso general de una ecuación que relaciona a más de tres variables el teorema de la función implícita se generaliza de la siguiente manera.

**Teorema.** Si  $F(x_1, x_2, \dots, x_n, w)$  es diferenciable, y la ecuación  $F(x_1, x_2, \dots, x_n, w) = 0$  define a  $w$  como una función implícita, diferenciable, de  $x_1, x_2, \dots, x_n$ , entonces

$$\frac{\partial w}{\partial x_i} = -\frac{F_{x_i}}{F_w}, \quad i = 1, 2, \dots, n,$$

en todos los puntos de la hipersuperficie  $F(x_1, x_2, \dots, x_n, w) = 0$  para los cuales

$$F_w \neq 0.$$

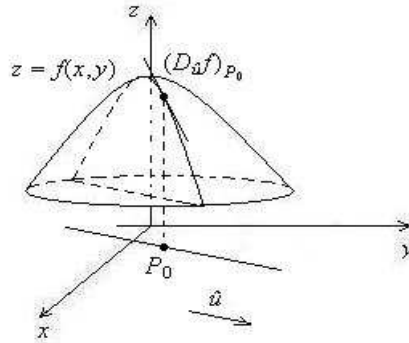
### 3.5 Derivada direccional y vector gradiente. Recta normal y plano tangente

En la sección 3.1 definimos el concepto de derivada parcial de una función  $f$  como la razón de cambio instantánea de  $f$  con respecto a cada una de sus variables independientes, manteniendo las otras fijas. Para una función  $f(x, y)$  de dos variables, la derivada parcial  $\partial f / \partial x$  representa la derivada de  $f$  en la dirección  $\hat{i}$ , mientras que  $\partial f / \partial y$  es la derivada de  $f$  en dirección  $\hat{j}$ . A continuación generalizamos el concepto de derivada de  $f$ , tomando en cuenta cambios simultáneos entre sus variables independientes, esto es, en cualquier dirección arbitraria del plano  $xy$ . A esto se le conoce como la derivada direccional de  $f$ .



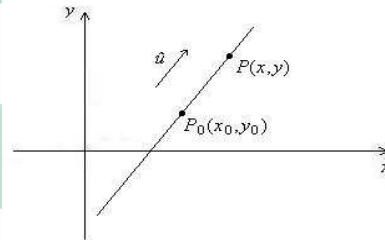
### 3.5 Derivada direccional y vector gradiente. Recta normal y plano tangente

La derivada direccional  $(D_{\hat{u}}f)_{P_0}$  de una función  $z = f(x, y)$  en un punto  $P_0(x_0, y_0)$  de su dominio representa la razón de cambio de  $f$  a lo largo de una dirección arbitraria  $\hat{u}$  en el plano  $xy$ , a partir de  $P_0$ .



Para calcular  $(D_{\hat{u}}f)_{P_0}$  es necesario determinar el cambio que experimenta la función  $z = f(x, y)$  cuando el punto  $P_0(x_0, y_0)$  se mueve en línea recta en la dirección  $\hat{u} = u_1\hat{i} + u_2\hat{j}$  hacia otro punto cercano  $P(x, y)$ , con

$$\begin{aligned} x &= x_0 + u_1 s \\ y &= y_0 + u_2 s, \quad s \in \mathbb{R}. \end{aligned}$$



**Definición.** La *derivada direccional*  $(D_{\hat{u}}f)_{P_0}$  de una función diferenciable  $f$  en un punto  $P_0(x_0, y_0)$  de su dominio, en la dirección del vector unitario  $\hat{u} = u_1\hat{i} + u_2\hat{j}$ , es el número

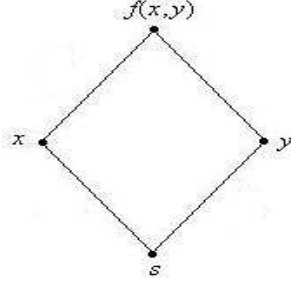
$$(D_{\hat{u}}f)_{P_0} = \left( \frac{df}{ds} \right)_{\hat{u}, P_0} = \lim_{s \rightarrow 0} \frac{f(x_0 + su_1, y_0 + su_2) - f(x_0, y_0)}{s},$$

siempre que este límite exista.

El cálculo de la derivada direccional a partir de la definición anterior no resulta práctico en general (puedes ver un ejemplo en el Thomas-Finney). Para reescribir la derivada  $(df/ds)_{\hat{u}, P_0}$  en términos de una expresión más fácil de calcular utilizaremos la regla de la cadena. Tomando en cuenta que  $z = f(x, y)$ ,

### Capítulo 3 Diferenciación

con  $x = x(s) = x_0 + u_1 s$  y  $y = y(s) = y_0 + u_2 s$ , se tiene



$$\begin{aligned} (D_{\hat{u}}f)_{P_0} &= \left( \frac{df}{ds} \right)_{\hat{u}, P_0} = \left( \frac{\partial f}{\partial x} \right)_{P_0} \frac{dx}{ds} + \left( \frac{\partial f}{\partial y} \right)_{P_0} \frac{dy}{ds}, \\ &= \left( \frac{\partial f}{\partial x} \right)_{P_0} u_1 + \left( \frac{\partial f}{\partial y} \right)_{P_0} u_2 \\ &= \left[ \left( \frac{\partial f}{\partial x} \right)_{P_0} \hat{i} + \left( \frac{\partial f}{\partial y} \right)_{P_0} \hat{j} \right] \cdot (u_1 \hat{i} + u_2 \hat{j}). \end{aligned}$$

Esta última expresión puede simplificarse, introduciendo la definición de vector gradiente de  $f$ .

**Definición.** El *gradiente* de una función diferenciable  $f(x, y)$  en cada punto interior de su dominio es el vector

$$\nabla f(x, y) = \frac{\partial f(x, y)}{\partial x} \hat{i} + \frac{\partial f(x, y)}{\partial y} \hat{j}.$$

El símbolo  $\nabla$  se conoce como “nabla” y la notación  $\nabla f$  se lee “gradiente de  $f$ ” o “nabla de  $f$ ”. En términos de este vector, la derivada direccional se simplifica como lo establece el siguiente teorema.

**Teorema.** Si las derivadas parciales de  $f(x, y)$  están definidas en el punto  $P_0(x_0, y_0)$ , entonces

$$(D_{\hat{u}}f)_{P_0} = \nabla f(x_0, y_0) \cdot \hat{u}.$$

Nota que en los casos particulares  $\hat{u} = \hat{i}$  o  $\hat{u} = \hat{j}$  la derivada direccional se convierte en las derivadas parciales, es decir,

$$(D_{\hat{i}}f)_{P_0} = \left( \frac{\partial f}{\partial x} \right)_{P_0}, \quad (D_{\hat{j}}f)_{P_0} = \left( \frac{\partial f}{\partial y} \right)_{P_0}.$$

### 3.5 Derivada direccional y vector gradiente. Recta normal y plano tangente

Ejemplos:

1. Calcula el vector gradiente de la función  $f(x, y) = xe^y$  en el punto  $P_0(3, 0)$ .  
Como  $f_x(x, y) = e^y$  y  $f_y(x, y) = xe^y$ , por lo tanto el vector gradiente  $\nabla f$  en cada punto  $(x, y)$  está dado por

$$\nabla f(x, y) = e^y \hat{i} + xe^y \hat{j}.$$

De esta manera, el vector gradiente de  $f$  en el punto  $P_0(3, 0)$  es

$$\nabla f(3, 0) = \hat{i} + 3\hat{j}.$$

2. Calcula la derivada direccional de  $f(x, y) = xe^y$  en el punto  $P_0(3, 0)$ , en la dirección del vector  $\vec{A} = 4\hat{i} - 3\hat{j}$ .

De acuerdo con el ejercicio anterior,  $\nabla f(3, 0) = \hat{i} + 3\hat{j}$ . Por otra parte, el vector unitario de  $\vec{A}$  es

$$\hat{A} = \frac{\vec{A}}{\|\vec{A}\|} = \frac{4}{5} \hat{i} - \frac{3}{5} \hat{j}.$$

De esta manera, la derivada direccional de  $f$  en el punto  $P_0$  en la dirección del vector  $\vec{A}$  está dada por

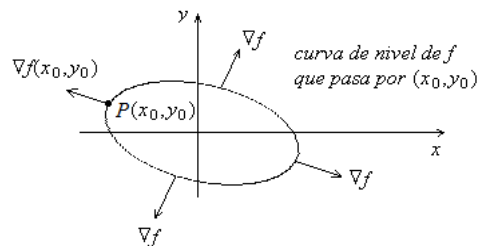
$$(D_{\hat{A}}f)_{P_0} = \nabla f(3, 0) \cdot \hat{A} = (\hat{i} + 3\hat{j}) \cdot \left( \frac{4}{5} \hat{i} - \frac{3}{5} \hat{j} \right) = \frac{4}{5} - \frac{9}{5} = -1.$$

Esto significa que, al cambiar el punto  $P_0$  hacia otro punto muy cercano en la dirección de  $\hat{A}$ , la función  $f$  decrece aproximadamente en 1 unidad.

#### Significado geométrico del gradiente

De acuerdo con su definición, el gradiente de una función en  $\mathbb{R}^3$  es un vector en  $\mathbb{R}^2$ , esto es, el gradiente de  $f$  es un vector que habita en el dominio de  $f$ . Este vector tiene un significado geométrico muy interesante, como se describe en los dos teoremas enunciados a continuación.

**Teorema 1.** Sea  $z = f(x, y)$  una superficie en  $\mathbb{R}^3$ . En cada punto interior  $P(x_0, y_0)$  del dominio de  $f$ , el gradiente  $\nabla f(x_0, y_0)$  es un vector perpendicular a la curva de nivel de  $f$  que contiene a  $P$ .



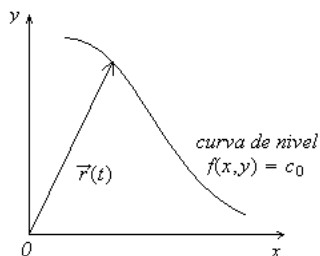
### Capítulo 3 Diferenciación

Demostración:

Supongamos que cada curva de nivel  $f(x, y) = c_0$  de la función  $f$  puede escribirse en forma paramétrica como

$$\vec{r}(t) = g(t) \hat{i} + h(t) \hat{j},$$

en donde  $x = g(t)$  y  $y = h(t)$ .



Así, la curva de nivel  $f(x, y) = c_0$  obedece la ecuación

$$f(g(t), h(t)) = c_0.$$

Derivando ambos lados de esta ecuación con respecto al parámetro  $t$ , se tiene

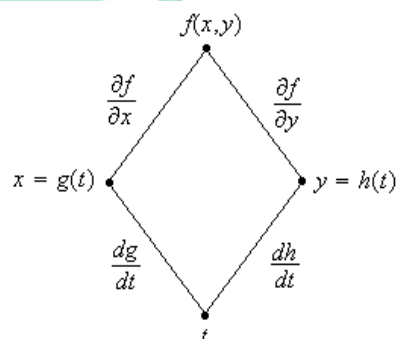
$$\frac{d f(g(t), h(t))}{dt} = \frac{d c_0}{dt} = 0$$

$$\therefore \frac{\partial f}{\partial x} \frac{dg}{dt} + \frac{\partial f}{\partial y} \frac{dh}{dt} = 0$$

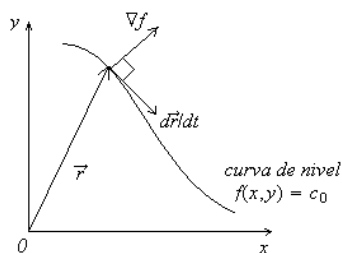
$$\therefore \left( \frac{\partial f}{\partial x} \hat{i} + \frac{\partial f}{\partial y} \hat{j} \right) \cdot \left( \frac{dg}{dt} \hat{i} + \frac{dh}{dt} \hat{j} \right) = 0$$

es decir,

$$\nabla f \cdot \frac{d\vec{r}}{dt} = 0.$$



De esta manera, el gradiente  $\nabla f$  en cada punto  $\vec{r}$  es perpendicular al vector  $d\vec{r}/dt$ , que es tangente a la curva de nivel en ese punto. En otras palabras, el gradiente  $\nabla f(x_0, y_0)$  es un vector perpendicular a la curva de nivel de  $f$  que contiene a  $P$ .



### 3.5 Derivada direccional y vector gradiente. Recta normal y plano tangente

Ejemplo:

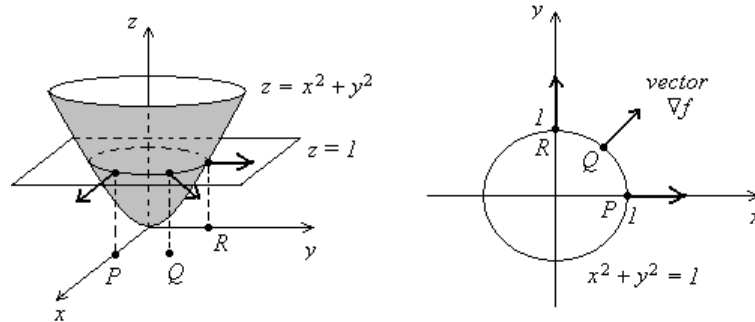
Encuentra un vector perpendicular a la curva de nivel  $z = 1$  de la función  $f(x, y) = x^2 + y^2$  en los puntos  $P(1, 0)$ ,  $Q(1/\sqrt{2}, 1/\sqrt{2})$  y  $R(0, 1)$ .

El gradiente de la función en cada punto de su dominio es el vector

$$\nabla f(x, y) = 2x \hat{i} + 2y \hat{j}.$$

Por lo tanto, un vector perpendicular a la curva de nivel  $x^2 + y^2 = 1$  de  $f$  en los puntos  $P$ ,  $Q$  y  $R$  es, respectivamente,

$$\nabla f|_P = 2 \hat{i}, \quad \nabla f|_Q = \frac{2}{\sqrt{2}} \hat{i} + \frac{2}{\sqrt{2}} \hat{j}, \quad \nabla f|_R = 2 \hat{j}.$$



**Teorema 2.** Sea  $z = f(x, y)$  una superficie en  $\mathbb{R}^3$ , y sea  $f(x, y) = c_0$  la curva de nivel de  $f$  que contiene al punto interior  $P(x_0, y_0)$  del dominio de  $f$ . Entonces el vector  $\nabla f|_{(x_0, y_0)}$  apunta en la dirección en la que  $f$  crece más rápidamente a partir del punto  $P(x_0, y_0)$ .

Demostración:

Reescribimos la derivada direccional  $D_{\hat{u}}f$  como

$$D_{\hat{u}}f = \nabla f \cdot \hat{u} = \|\nabla f\| \|\hat{u}\| \cos \theta = \|\nabla f\| \cos \theta,$$

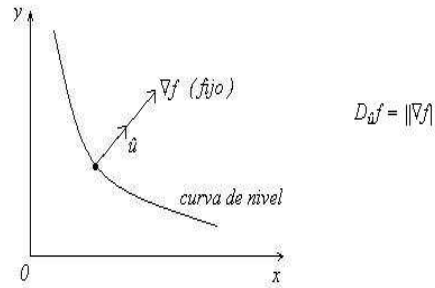
en donde  $\theta$  es el ángulo entre el vector  $\nabla f$ , que en cada punto es un vector fijo, y el vector de dirección  $\hat{u}$ , que es arbitrario. Como  $D_{\hat{u}}f$  está dado por un producto escalar, su valor y su signo dependen del ángulo entre estos dos vectores. Así, para ángulos entre  $0$  y  $\pi/2$  la derivada direccional es positiva y para ángulos entre  $\pi/2$  y  $\pi$  ésta es negativa. Si el ángulo relativo es  $\pi/2$ , la derivada direccional es cero; esto significa que la función  $f$  no cambia en la dirección perpendicular al gradiente  $\nabla f$ , es decir, a lo largo de una curva de nivel de  $f$ .

### Capítulo 3 Diferenciación

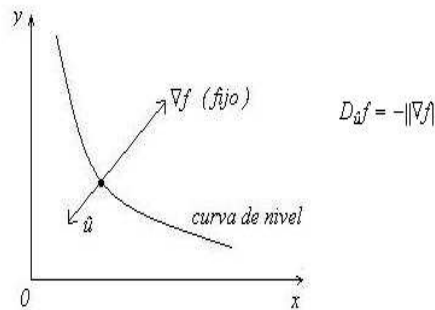
De lo anterior se desprende que

$$-||\nabla f|| \leq D_{\hat{u}}f \leq ||\nabla f||.$$

De esta manera, el valor máximo de  $D_{\hat{u}}f$  está dado por  $||\nabla f||$ , y éste se alcanza cuando  $\hat{u}$  apunta en la misma dirección y sentido ( $\theta = 0$ ) que el gradiente  $\nabla f$ .



Correspondientemente, el valor mínimo de  $D_{\hat{u}}f$  está dado por  $-||\nabla f||$ , y éste se alcanza cuando  $\hat{u}$  apunta en el sentido opuesto ( $\theta = \pi$ ) que el gradiente  $\nabla f$ .



Así, el gradiente de una función  $f$  en un punto dado apunta hacia la dirección de máximo crecimiento de  $f$  en ese punto.

Ejemplos:

1. Encuentra la dirección en la que la función  $f(x, y) = xe^y$ , en el punto  $P(3, 0)$ :
  - i) crece más rápidamente, ii) decrece más rápidamente.
  - i) En el punto  $P(3, 0)$  la función crece más rápidamente en la dirección del vector  $\nabla f(3, 0) = \hat{i} + 3\hat{j}$ , que está dada por  $\hat{u} = \frac{1}{\sqrt{10}} \hat{i} + \frac{3}{\sqrt{10}} \hat{j}$ .
  - ii) En el punto  $P(3, 0)$  la función decrece más rápidamente en la dirección del vector  $-\nabla f(3, 0) = -\hat{i} - 3\hat{j}$ , que está dada por  $-\hat{u} = -\frac{1}{\sqrt{10}} \hat{i} - \frac{3}{\sqrt{10}} \hat{j}$ .

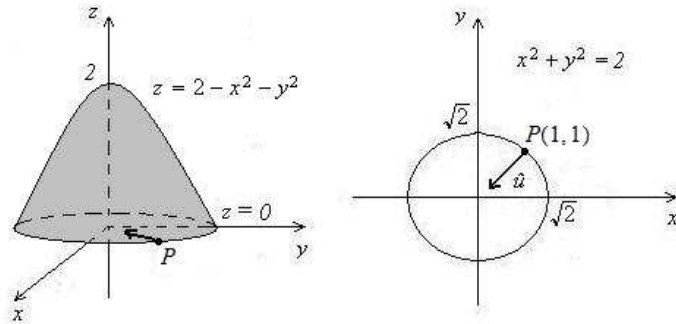
### 3.5 Derivada direccional y vector gradiente. Recta normal y plano tangente

2. Encuentra la dirección en la cual la función  $f(x, y) = 2 - x^2 - y^2$  crece más rápidamente en el punto  $P(1, 1)$ . Ilustra tu resultado gráficamente.

En el punto  $P(1, 1)$  la función crece más rápidamente en la dirección del vector  $\nabla f(x, y)|_{(1,1)} = -2x \hat{i} - 2y \hat{j}|_{(1,1)} = -2 \hat{i} - 2 \hat{j}$ , dada por

$$\hat{u} = -\frac{2}{\sqrt{8}} \hat{i} - \frac{2}{\sqrt{8}} \hat{j} = -\frac{1}{\sqrt{2}} \hat{i} - \frac{1}{\sqrt{2}} \hat{j}.$$

Como  $P(1, 1)$  está en la curva de nivel  $x^2 + y^2 = 2$  de la función (correspondiente a  $z = 0$ ), el resultado anterior muestra que en ese punto la función crece más rápidamente en la dirección perpendicular hacia adentro.

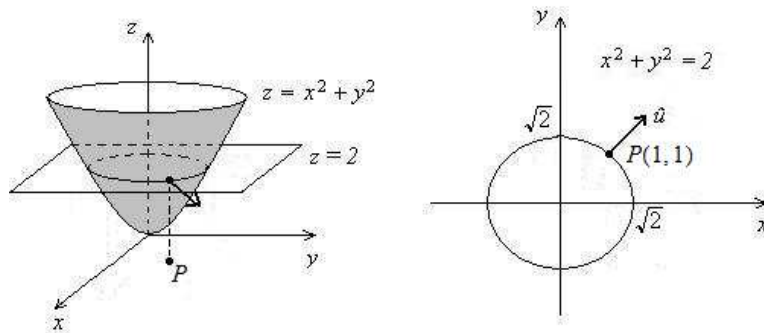


3. Encuentra la dirección en la cual la función  $f(x, y) = x^2 + y^2$  crece más rápidamente en el punto  $P(1, 1)$ . Ilustra tu resultado gráficamente.

En el punto  $P(1, 1)$  la función crece más rápidamente en la dirección del vector  $\nabla f(x, y)|_{(1,1)} = 2x \hat{i} + 2y \hat{j}|_{(1,1)} = 2 \hat{i} + 2 \hat{j}$ , dada por

$$\hat{u} = \frac{2}{\sqrt{8}} \hat{i} + \frac{2}{\sqrt{8}} \hat{j} = \frac{1}{\sqrt{2}} \hat{i} + \frac{1}{\sqrt{2}} \hat{j}.$$

Como  $P(1, 1)$  está en la curva de nivel  $x^2 + y^2 = 2$  de la función (correspondiente a  $z = 2$ ), el resultado anterior muestra que en ese punto la función crece más rápidamente en la dirección perpendicular hacia afuera.



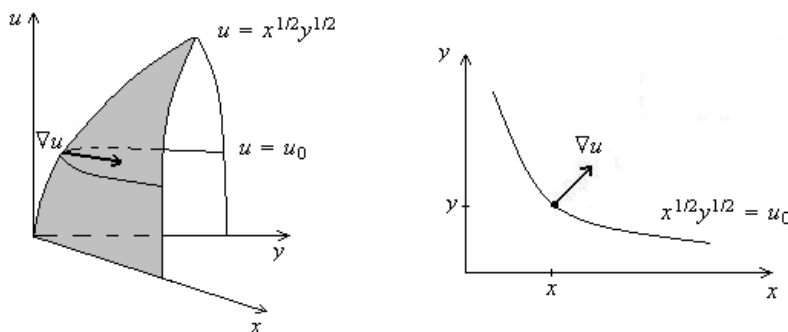
## Capítulo 3 Diferenciación

4. Encuentra la dirección de máximo crecimiento de una función de utilidad de la forma  $u(x, y) = x^{1/2}y^{1/2}$ ,  $x, y > 0$ , en cada punto de sus curvas de indiferencia. Ilustra tu resultado.

Para cada canasta  $(x, y)$ , el gradiente de la función  $u$  está dado por el vector

$$\nabla u(x, y) = \frac{1}{2} \left( \frac{y}{x} \right)^{1/2} \hat{i} + \frac{1}{2} \left( \frac{x}{y} \right)^{1/2} \hat{j},$$

cuyas componentes son ambas positivas. Así, para cada curva de indiferencia  $x^{1/2}y^{1/2} = u_0$  el vector gradiente  $\nabla u$  en cada punto apunta en la dirección perpendicular mostrada en la figura.



### Recta normal y plano tangente

Generalizando los resultados anteriores, en el caso de una función diferenciable  $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$  en  $\mathbb{R}^{n+1}$ , dada por  $w = f(x_1, \dots, x_n)$ , el gradiente  $\nabla f$  en cada punto  $(x_1, \dots, x_n) \in S$  es el vector

$$\nabla f = \left( \frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right).$$

Este vector en  $\mathbb{R}^n$  es perpendicular a los conjuntos de nivel de  $f$ , dados por  $f(x_1, \dots, x_n) = c$ , que también habitan en  $\mathbb{R}^n$ . En particular, en el tema anterior vimos que para una superficie  $z = f(x, y)$  en  $\mathbb{R}^3$  el vector  $\nabla f$  es perpendicular a sus curvas de nivel,  $f(x, y) = c$ , en  $\mathbb{R}^2$ .

Siguiendo con este razonamiento, para una hipersuperficie  $w = f(x, y, z)$  en  $\mathbb{R}^4$  el vector  $\nabla f$  será perpendicular a sus superficies de nivel,  $f(x, y, z) = c$ , en  $\mathbb{R}^3$ . Este hecho puede utilizarse para encontrar un vector normal a una superficie dada, invirtiendo el argumento, de la siguiente manera. Nos interesa encontrar un vector que sea perpendicular a una cierta superficie

$$z = f(x, y).$$



### 3.5 Derivada direccional y vector gradiente. Recta normal y plano tangente

Para ello, basta con suponer que la superficie  $z = f(x, y)$  en  $\mathbb{R}^3$  es el conjunto de nivel  $w = 0$  de la hipersuperficie

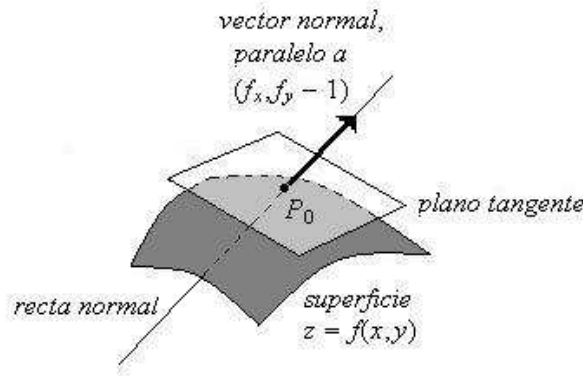
$$w = f(x, y) - z$$

en  $\mathbb{R}^4$ , de modo que un vector normal a la superficie original sería, precisamente,

$$\nabla w = (w_x, w_y, w_z) = (f_x, f_y, -1).$$

**Teorema.** Un vector normal a la superficie  $z = f(x, y)$  en  $\mathbb{R}^3$  es el vector  $(f_x, f_y, -1)$ .

A partir de este teorema podemos encontrar fácilmente las ecuaciones del plano tangente y de la recta normal a cualquier superficie  $z = f(x, y)$  en  $\mathbb{R}^3$ , en cualquier punto  $P_0(x_0, y_0, z_0)$  de la superficie.



Así, si  $z = f(x, y)$  representa una superficie en  $\mathbb{R}^3$ . Entonces:

i) La ecuación del plano tangente a la superficie en el punto  $P_0(x_0, y_0, z_0)$  es

$$f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) - (z - z_0) = 0.$$

ii) La ecuación de la recta normal a la superficie en el punto  $P_0(x_0, y_0, z_0)$  es

$$\begin{aligned} x &= x_0 + f_x(x_0, y_0) t, \\ y &= y_0 + f_y(x_0, y_0) t, \\ z &= z_0 - t, \quad t \in \mathbb{R}. \end{aligned}$$

### Capítulo 3 Diferenciación

Nota que la ecuación del plano tangente en el inciso i) puede también escribirse como

$$z = z_0 + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$$

o, equivalentemente, como

$$z = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0),$$

que es precisamente la linealización  $L(x, y)$  de la función  $z = f(x, y)$  en el punto  $(x_0, y_0)$ , que estudiamos con anterioridad.

Ejemplo:

Encuentra las ecuaciones del plano tangente y la recta normal a la superficie  $z = f(x, y)$ , con  $f(x, y) = 9 - x^2 - y^2$ , en el punto  $(1, 2, 4)$ .

En este caso,  $f_x(x, y) = -2x$  y  $f_y(x, y) = -2y$ , de modo que  $f_x(1, 2) = -2$  y  $f_y(1, 2) = -4$ . Por lo tanto, el vector normal a la superficie en el punto dado es  $(-2, -4, -1)$ . De esta manera, la ecuación del plano tangente es  $-2(x - 1) - 4(y - 2) - (z - 4) = 0$ , o bien

$$2x + 4y + z = 14.$$

Por otra parte, la ecuación de la recta normal es

$$x = 1 - 2t, \quad y = 2 - 4t, \quad z = 4 - t, \quad t \in \mathbb{R}.$$

Para finalizar esta sección, cabe señalar que un razonamiento análogo a este último puede aplicarse para encontrar las ecuaciones de la recta tangente y la recta normal a una curva en  $\mathbb{R}^2$ , así como las del hiperplano tangente y la recta normal a una hipersuperficie en  $\mathbb{R}^{n+1}$ , con  $n \geq 3$ .

### 3.6 Funciones homogéneas. Teorema de Euler

**Definición.** Se dice que una función  $f(x_1, x_2, \dots, x_n)$  es *homogénea de grado  $k$*  si satisface

$$f(\lambda x_1, \lambda x_2, \dots, \lambda x_n) = \lambda^k f(x_1, x_2, \dots, x_n), \quad \lambda \in \mathbb{R}^+.$$

### 3.6 Funciones homogéneas. Teorema de Euler

Ejemplos:

1. La función  $f(x) = x^2$  es homogénea de grado 2, ya que

$$f(\lambda x) = (\lambda x)^2 = \lambda^2 x^2 = \lambda^2 f(x).$$

2. La función  $f(x) = x^{-1}$  es homogénea de grado  $-1$ , ya que

$$f(\lambda x) = (\lambda x)^{-1} = \lambda^{-1} x^{-1} = \lambda^{-1} f(x).$$

3. La función  $f(x) = 2$  es homogénea de grado 0, ya que

$$f(\lambda x) = 2 = \lambda^0 \cdot 2 = \lambda^0 f(x).$$

4. La función  $f(x) = x^2 + 2x$  no es homogénea, ya que

$$f(\lambda x) = (\lambda x)^2 + 2(\lambda x) = \lambda^2 x^2 + 2\lambda x \neq \lambda^k f(x).$$

5. La función  $f(x, y) = x^2 y$  es homogénea de grado 3, ya que

$$f(\lambda x, \lambda y) = (\lambda x)^2 (\lambda y) = \lambda^3 (x^2 y) = \lambda^3 f(x, y).$$

6. La función  $f(x, y) = \frac{x^3}{xy + y^2}$  es homogénea de grado 1, ya que

$$f(\lambda x, \lambda y) = \frac{(\lambda x)^3}{(\lambda x)(\lambda y) + (\lambda y)^2} = \frac{\lambda^3 x^3}{\lambda^2 xy + \lambda^2 y^2} = \lambda \frac{x^3}{xy + y^2} = \lambda f(x, y).$$

7. La función  $f(x, y) = e^{x/y}$  es homogénea de grado 0, ya que

$$f(\lambda x, \lambda y) = e^{(\lambda x)/(\lambda y)} = e^{x/y} = \lambda^0 e^{x/y} = \lambda^0 f(x, y).$$

8. La función  $f(x, y) = \frac{x^3}{xy + y}$  no es homogénea, ya que

$$f(\lambda x, \lambda y) = \frac{(\lambda x)^3}{(\lambda x)(\lambda y) + (\lambda y)} = \frac{\lambda^3 x^3}{\lambda^2 xy + \lambda y} \neq \lambda^k f(x, y).$$

9. Las funciones de producción (utilidad) tipo Cobb-Douglas  $f(x, y) = x^\alpha y^\beta$  son homogéneas de grado  $\alpha + \beta$ , ya que

$$f(\lambda x, \lambda y) = (\lambda x)^\alpha (\lambda y)^\beta = \lambda^{\alpha+\beta} x^\alpha y^\beta = \lambda^{\alpha+\beta} f(x, y).$$

Si  $\alpha + \beta > 1$  se tiene rendimientos crecientes a escala, si  $\alpha + \beta = 1$  se tiene rendimientos constantes a escala, y si  $\alpha + \beta < 1$  los rendimientos a escala son decrecientes.

### Capítulo 3 Diferenciación

10. El logaritmo de una función de producción (utilidad) tipo Cobb-Douglas, a saber,  $f(x, y) = \ln(x^\alpha y^\beta) = \alpha \ln x + \beta \ln y$ , no es una función homogénea, ya que

$$f(\lambda x, \lambda y) = \ln[(\lambda x)^\alpha (\lambda y)^\beta] = (\alpha + \beta) \ln \lambda + (\alpha \ln x + \beta \ln y) \neq \lambda^k f(x, y).$$

Para entender el significado de la homogeneidad de una función, supongamos que  $f(x, y)$  es la función de producción correspondiente a los insumos  $(x, y)$ , y preguntémonos cuál sería la nueva producción si ambos insumos se duplicaran, es decir si fueran  $(2x, 2y)$ . En ese caso:

i) Si  $f$  fuera homogénea de grado 1, entonces  $f(2x, 2y) = 2^1 f(x, y) = 2f(x, y)$ , es decir, la nueva producción sería el doble de la original.

ii) Si  $f$  fuera homogénea de grado 2, entonces  $f(2x, 2y) = 2^2 f(x, y) = 4f(x, y)$ , es decir, la nueva producción sería cuatro veces la original.

iii) Si  $f$  fuera homogénea de grado 0, entonces  $f(2x, 2y) = 2^0 f(x, y) = f(x, y)$ , es decir, la nueva producción sería igual a la original.

iv) Si  $f$  fuera homogénea de grado  $-1$ , entonces  $f(2x, 2y) = 2^{-1} f(x, y) = \frac{1}{2} f(x, y)$ , es decir, la nueva producción sería la mitad de la original.

v) Si  $f$  no fuera homogénea, la nueva producción no sería un múltiplo de la original.

**Teorema.** Si  $f(x, y)$  es una función homogénea de grado  $k$ , entonces

$$f(x, y) = x^k f(1, y/x),$$

$$f(x, y) = y^k f(x/y, 1).$$

Demostración:

En el primer caso, considera que la variable  $x$  juega el papel del factor  $\lambda$ , de modo que  $f(x, y) = f(x(1), x(y/x)) = x^k f(1, y/x)$ . El segundo caso se demuestra de manera similar, tomando a la variable  $y$  como el factor  $\lambda$ .

Este resultado es de particular interés en economía. Así, por ejemplo, si  $Q = P(L, K)$  representa la producción en función del trabajo  $L$  y el capital  $K$ , y

### 3.6 Funciones homogéneas. Teorema de Euler

si ésta es una función homogénea de grado 1, entonces la producción per cápita  $q = \frac{P(L,K)}{L}$  puede expresarse simplemente en términos del capital per cápita  $k = K/L$  de acuerdo con

$$q = \frac{P(L, K)}{L} = \frac{L P(1, K/L)}{L} = P(1, K/L) = p(k).$$

Como caso particular, si  $P(L, K) = L^{1/4} K^{3/4}$ , entonces

$$q = \frac{L^{1/4} K^{3/4}}{L} = \left( \frac{K}{L} \right)^{3/4} = k^{3/4}.$$

Es decir, la producción per cápita es la función  $p(k) = k^{3/4}$ .

**Definición.** Sea  $f(\vec{x})$  una función definida en un dominio  $D \subset \mathbb{R}^n$ . Se dice que  $f$  es *homotética* si para todos  $\vec{x}_1, \vec{x}_2 \in D$  y para todo  $\lambda > 0$  se cumple

$$f(\vec{x}_1) = f(\vec{x}_2) \Rightarrow f(\lambda \vec{x}_1) = f(\lambda \vec{x}_2).$$

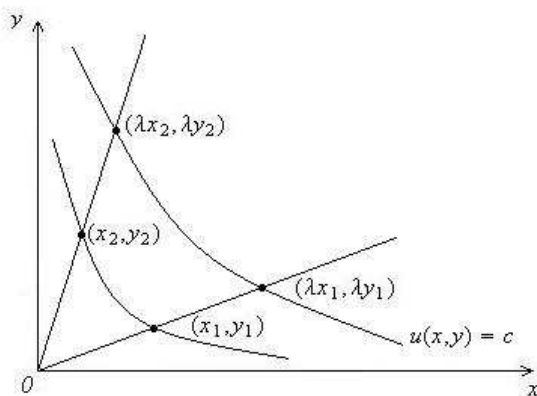
Por ejemplo, si la función de utilidad  $u(x, y)$  de un consumidor es una función homotética, entonces si a un consumidor le es indiferente elegir entre dos canastas  $\vec{x}_1 = (x_1, y_1)$  y  $\vec{x}_2 = (x_2, y_2)$ , es decir,

$$u(x_1, y_1) = u(x_2, y_2),$$

entonces le seguirá siendo indiferente si en cualquiera de estas canastas los dos bienes se aumentan en la misma proporción, es decir,

$$u(\lambda x_1, \lambda y_1) = u(\lambda x_2, \lambda y_2).$$

En otras palabras, una función de utilidad homotética preserva las preferencias de un consumidor.



### Capítulo 3 Diferenciación

**Teorema.** Toda función homogénea es homotética.

Demostración:

Sea  $f$  homogénea de grado  $k$ , y sean  $\vec{x}$  y  $\vec{y}$  tales que  $f(\vec{x}) = f(\vec{y})$ . Entonces

$$f(\lambda\vec{x}) = \lambda^k f(\vec{x}) = \lambda^k f(\vec{y}) = f(\lambda\vec{y}),$$

y por lo tanto  $f$  es homotética.

Es importante señalar que no toda función homotética es homogénea. Por ejemplo, dada una función homogénea  $u$  (por ejemplo, la función de utilidad) su logaritmo natural  $\ln u$  no es una función homogénea, pero es fácil demostrar que sí es una función homotética. La razón de ello será evidente a la luz de las siguientes consideraciones.

**Definición.** Se dice que  $F(\vec{x})$  es una *transformación monotónica creciente* de una función homogénea  $f(\vec{x})$ , si existe una función creciente,  $H$ , tal que

$$F(\vec{x}) = H(f(\vec{x})).$$

En otras palabras, una transformación monotónica creciente es una composición de funciones de la forma  $H(f(\vec{x}))$ , en donde  $H$  es una función creciente y  $f$  es una función homogénea. Así, por ejemplo, dada una función de utilidad homogénea  $u(x, y)$ , la función  $\ln u$  es una transformación monotónica creciente de  $u$ .

**Teorema.** Toda transformación monotónica es una función homotética.

Demostración:

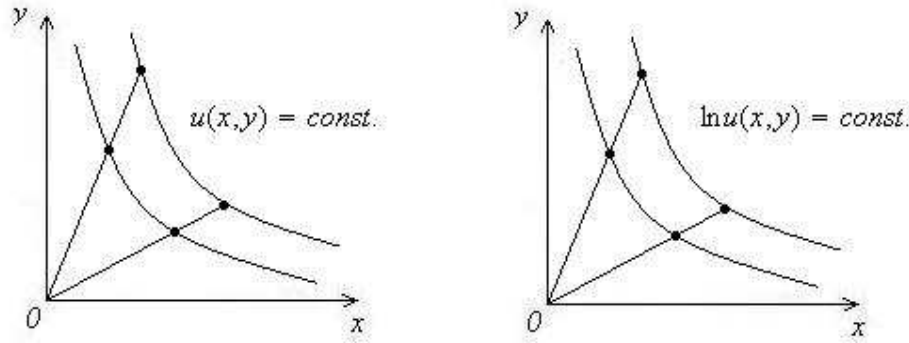
Sea  $F(\vec{x}) = H(f(\vec{x}))$  una transformación monotónica creciente de una función  $f$  homogénea de grado  $k$ , y sean  $\vec{x}$  y  $\vec{y}$  tales que  $f(\vec{x}) = f(\vec{y})$ . De esta manera,  $F(\vec{x}) = F(\vec{y})$ . Como

$$F(\lambda\vec{x}) = H(f(\lambda\vec{x})) = H(\lambda^k f(\vec{x})) = H(\lambda^k f(\vec{y})) = H(f(\lambda\vec{y})) = F(\lambda\vec{y}),$$

por lo tanto,  $F$  es homotética.

### 3.6 Funciones homogéneas. Teorema de Euler

Así, aunque  $\ln u$  no es una función homogénea, sino más bien una transformación monotónica de una función homogénea  $u$ , es claro que  $\ln u$  será una función homotética, preservando la característica de llevar canastas en una misma curva de indiferencia hacia canastas en otra misma curva de indiferencia, al aumentar los bienes en una misma proporción.



Los resultados anteriores son válidos para funciones en general, aunque éstas no sean diferenciables. En el caso particular de funciones homogéneas diferenciables, existen teoremas adicionales, que son de gran interés y utilidad en economía, como se muestra a continuación.

**Teorema.** Las primeras derivadas parciales  $\partial f / \partial x_1, \dots, \partial f / \partial x_n$  de una función homogénea, diferenciable,  $f(x_1, \dots, x_n)$  de grado  $k$ , son homogéneas de grado  $k - 1$ .

Demostración:

Sea  $f(x_1, \dots, x_n)$  una función homogénea de grado  $k$ . Por lo tanto, para todo  $\lambda > 0$ ,

$$f(\lambda x_1, \lambda x_2, \dots, \lambda x_n) = \lambda^k f(x_1, x_2, \dots, x_n).$$

Derivando con respecto a  $x_i$  ambos lados de la igualdad,  $i = 1, \dots, n$ , se tiene

$$\begin{aligned} \frac{\partial}{\partial x_i} f(\lambda x_1, \lambda x_2, \dots, \lambda x_n) &= \lambda \frac{\partial}{\partial (\lambda x_i)} f(\lambda x_1, \lambda x_2, \dots, \lambda x_n) \\ &= \lambda^k \frac{\partial}{\partial x_i} f(x_1, x_2, \dots, x_n), \end{aligned}$$

de modo que

$$\frac{\partial}{\partial (\lambda x_i)} f(\lambda x_1, \lambda x_2, \dots, \lambda x_n) = \lambda^{k-1} \frac{\partial}{\partial x_i} f(x_1, x_2, \dots, x_n).$$

### Capítulo 3 Diferenciación

Por lo tanto, cada derivada parcial  $\partial f / \partial x_i$  es homogénea de grado  $k - 1$ .

Ejemplo:

Sea  $P(L, K)$  una función de producción homogénea de grado 1. Si se sabe que  $P(150, 50) = 550$ ,  $P_L(150, 50) = 3$  y  $P_K(150, 50) = 2$ , calcula  $P(30, 10)$ ,  $P_L(30, 10)$  y  $P_K(30, 10)$ .

Como la producción  $P$  es una función homogénea de grado 1, por lo tanto

$$P(30, 10) = P\left(\frac{1}{5}(150), \frac{1}{5}(50)\right) = \left(\frac{1}{5}\right)^1 P(150, 50) = \frac{1}{5}(550) = 110.$$

Por otra parte, los productos marginales  $P_L$  y  $P_K$  son funciones homogéneas de grado 0, de modo que

$$\begin{aligned} P_L(30, 10) &= P_L\left(\frac{1}{5}(150), \frac{1}{5}(50)\right) \\ &= \left(\frac{1}{5}\right)^0 P_L(150, 50) = P_L(150, 50) = 3, \end{aligned}$$

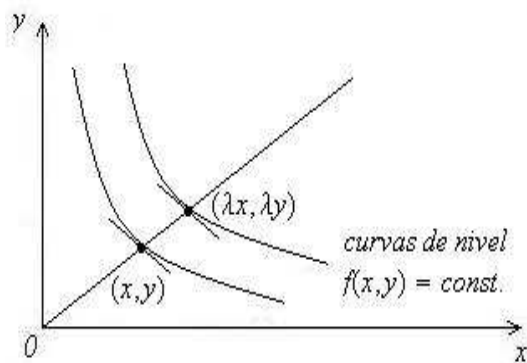
$$\begin{aligned} P_K(30, 10) &= P_K\left(\frac{1}{5}(150), \frac{1}{5}(50)\right) \\ &= \left(\frac{1}{5}\right)^0 P_K(150, 50) = P_K(150, 50) = 2. \end{aligned}$$

Una consecuencia del teorema anterior es que las curvas de nivel de una función homogénea tienen la misma pendiente a lo largo de puntos que se encuentran en rectas que pasan por el origen. Es decir, si  $f(x, y) = c$  representa una curva de nivel de una función homogénea  $z = f(x, y)$ , entonces

$$\left. \frac{dy}{dx} \right|_{(\lambda x, \lambda y)} = \left. \frac{dy}{dx} \right|_{(x, y)}.$$



### 3.6 Funciones homogéneas. Teorema de Euler



Para demostrar este resultado, supongamos que  $z = f(x, y)$  es una función homogénea de grado  $k$  y consideremos la curva de nivel  $f(x, y) = c$  que contiene al punto  $(\lambda x, \lambda y)$ . Sea  $F(x, y) = f(x, y) - c = 0$ , de modo que, de acuerdo con el teorema de la función implícita, se tiene

$$\left. \frac{dy}{dx} \right|_{(\lambda x, \lambda y)} = -\frac{f_x(\lambda x, \lambda y)}{f_y(\lambda x, \lambda y)} = -\frac{\lambda^{k-1} f_x(x, y)}{\lambda^{k-1} f_y(x, y)} = -\frac{f_x(x, y)}{f_y(x, y)} = \left. \frac{dy}{dx} \right|_{(x, y)}.$$

Así, si una función de producción  $P(L, K)$  es homogénea, entonces la tasa marginal de sustitución técnica,  $P_L/P_K = -dK/dL$ , es constante a lo largo de rayos que salen del origen. Este resultado es válido para transformaciones monótonicas en general, y no sólo para funciones homogéneas.

**Teorema de Euler.** Sea  $f(x_1, x_2, \dots, x_n)$  una función diferenciable. Si  $f$  es homogénea de grado  $k$ , entonces se cumple  $\vec{x} \cdot \nabla f(\vec{x}) = k f(\vec{x})$ , es decir,

$$x_1 f_1(\vec{x}) + x_2 f_2(\vec{x}) + \dots + x_n f_n(\vec{x}) = k f(\vec{x}),$$

en donde  $f_i = \partial f / \partial x_i$ , para todo  $1 \leq i \leq n$ .

**Demostración:**

Sea  $f(x_1, x_2, \dots, x_n)$  una función diferenciable. Si  $f$  es homogénea de grado  $k$ , entonces, para todo  $\lambda > 0$ ,

$$f(\lambda x_1, \lambda x_2, \dots, \lambda x_n) = \lambda^k f(x_1, x_2, \dots, x_n).$$

Derivando con respecto a  $\lambda$  ambos lados de la igualdad, se tiene

$$\frac{\partial f}{\partial(\lambda x_1)} \frac{\partial(\lambda x_1)}{\partial \lambda} + \dots + \frac{\partial f}{\partial(\lambda x_n)} \frac{\partial(\lambda x_n)}{\partial \lambda} = k \lambda^{k-1} f(x_1, x_2, \dots, x_n).$$

### Capítulo 3 Diferenciación

Como  $\partial(\lambda x_i)/\partial\lambda = x_i$ , para todo  $i = 1, \dots, n$ , por lo tanto

$$x_1 \frac{\partial f}{\partial(\lambda x_1)} + \dots + x_n \frac{\partial f}{\partial(\lambda x_n)} = k\lambda^{k-1}f.$$

Evaluando esta expresión en  $\lambda = 1$  se obtiene el resultado deseado.

Ejemplo:

Como  $f(x, y) = x^\alpha y^\beta$  es homogénea de grado  $\alpha + \beta$ , por lo tanto

$$x \frac{\partial f}{\partial x} + y \frac{\partial f}{\partial y} = x (\alpha x^{\alpha-1} y^\beta) + y (\beta x^\alpha y^{\beta-1}) = (\alpha + \beta) x^\alpha y^\beta = (\alpha + \beta) f.$$

## Capítulo 4

# Funciones cóncavas y cuasicóncavas

### 4.1 Polinomio de Taylor de orden 2. Matriz hessiana

El objetivo de esta sección es introducir el concepto de matriz hessiana, que es una matriz de segundas derivadas parciales que se utiliza para establecer la concavidad o convexidad de funciones diferenciables en varias variables. A continuación se motiva su definición a partir de lo que se conoce como aproximación cuadrática, o polinomio de Taylor de orden 2.

Comencemos con el caso de una función diferenciable en una variable,  $f : S \rightarrow \mathbb{R}$ , con  $S \subset \mathbb{R}$  un intervalo abierto. Como  $f$  es diferenciable en  $S$ , existe una recta tangente a la curva  $y = f(x)$  en cada punto  $a \in S$ , dada por

$$y = f(a) + f'(a)(x - a).$$

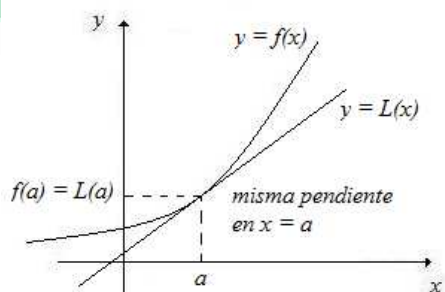
Esto significa que para valores de  $x$  cercanos a  $a$  podemos aproximar la función  $f$  por su *linealización*  $L$ , dada por

$$L(x) = f(a) + f'(a)(x - a),$$

que estudiamos en la sección 3.2. Esta es una función polinomial de grado 1 en  $x$ , y es tal que satisface

$$L(a) = f(a) \quad \text{y} \quad L'(a) = f'(a),$$

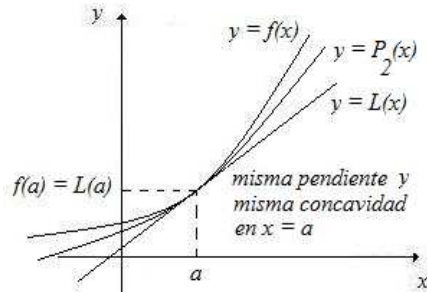
es decir, en  $x = a$  las funciones  $L$  y  $f$  toman el mismo valor y además tienen la misma pendiente.



## Capítulo 4 Funciones cóncavas y cuasicóncavas

Cuando  $f$  es doblemente diferenciable en  $S$  se puede construir una mejor aproximación para  $f$  que, además de contar con las características de la aproximación lineal, tenga la misma concavidad que  $f$  en  $x = a$ . En otras palabras, se busca una función  $P_2(x)$  que satisfaga

$$P_2(a) = f(a), \quad P_2'(a) = f'(a) \quad \text{y} \quad P_2''(a) = f''(a).$$



Es fácil verificar que una función que cumple estas condiciones es

$$P_2(x) = f(a) + f'(a)(x - a) + \frac{1}{2}f''(a)(x - a)^2,$$

conocida como el *polinomio de Taylor de orden 2 generado por  $f(x)$  alrededor de  $x = a$* . La función  $P_2$  representa una *aproximación cuadrática* de  $f$  cerca de  $x = a$ . El factor  $\frac{1}{2}$  en el término cuadrático es necesario para que se verifique la igualdad de las segundas derivadas,  $P_2''(a) = f''(a)$ . Nota que

$$P_2(x) = L(x) + \frac{1}{2}f''(a)(x - a)^2.$$

Se puede demostrar que, por lo general, la función cuadrática  $P_2$  es una mejor aproximación de  $f$  que la función lineal  $L$ .

Ejemplo:

Encuentra la linealización  $L(x)$  y el polinomio de Taylor  $P_2(x)$  generados por la función  $f(x) = \sqrt{x}$  alrededor de  $x = 1$ . Utiliza estas funciones para aproximar el valor de  $\sqrt{1.1}$  y compara los resultados obtenidos con el valor exacto.

La linealización  $L(x)$  y el polinomio de Taylor  $P_2(x)$  generados por una función doblemente diferenciable  $f(x)$  en  $x = 1$  están dados por

$$\begin{aligned} L(x) &= f(1) + f'(1)(x - 1), \\ P_2(x) &= f(1) + f'(1)(x - 1) + \frac{1}{2}f''(1)(x - 1)^2. \end{aligned}$$

#### 4.1 Polinomio de Taylor de orden 2. Matriz hessiana

En particular, para la función  $f(x) = \sqrt{x}$  se tiene

$$f(1) = \sqrt{1}, \quad f'(1) = \frac{1}{2\sqrt{1}} \quad \text{y} \quad f''(1) = -\frac{1}{4(1)^{3/2}},$$

de modo que

$$L(x) = 1 + \frac{1}{2}(x - 1),$$

$$P_2(x) = 1 + \frac{1}{2}(x - 1) + \frac{1}{2} \left( -\frac{1}{4} \right) (x - 1)^2.$$

De esta manera, las aproximaciones lineal y cuadrática de la función  $f(x) = \sqrt{x}$  alrededor de  $x = 1$  son, respectivamente,

$$(\sqrt{x})_{\text{lineal}} \approx 1 + \frac{1}{2}(x - 1),$$

$$(\sqrt{x})_{\text{cuad}} \approx 1 + \frac{1}{2}(x - 1) - \frac{1}{8}(x - 1)^2.$$

Se tiene, entonces,

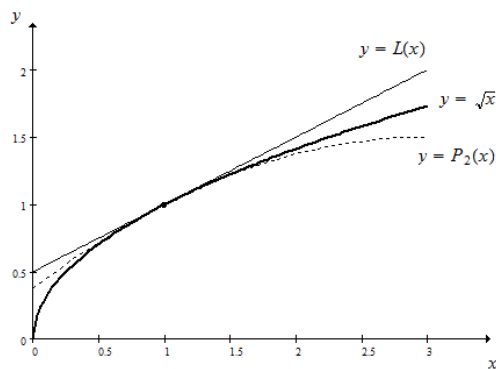
$$(\sqrt{1.1})_{\text{lineal}} \approx 1 + \frac{1}{2}(1.1 - 1) = 1.05,$$

$$(\sqrt{1.1})_{\text{cuad}} \approx 1 + \frac{1}{2}(1.1 - 1) - \frac{1}{8}(1.1 - 1)^2 = 1.04875.$$

Compara estos resultados con el valor exacto de  $\sqrt{1.1}$ , dado por

$$\sqrt{1.1} = 1.048808848 \dots$$

Es claro que la aproximación cuadrática  $(\sqrt{1.1})_{\text{cuad}}$  obtenida de  $P_2(x)$  da un resultado más exacto que la aproximación lineal  $(\sqrt{1.1})_{\text{lineal}}$  obtenida de  $L(x)$ . La razón geométrica de este hecho se ilustra en la siguiente figura, en donde están graficadas las funciones  $y = \sqrt{x}$ ,  $y = L(x)$  y  $y = P_2(x)$ . Aunque las tres funciones son tangentes en el punto  $(1, 1)$ , observa que sólo  $P_2(x)$  preserva la concavidad de la función  $\sqrt{x}$  en ese punto.



## Capítulo 4 Funciones cóncavas y cuasicóncavas

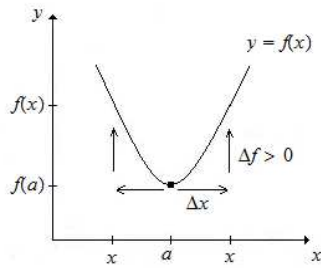
La aproximación cuadrática  $P_2(x)$  para funciones  $f$  doblemente diferenciables es muy importante en el tema de optimización, puesto que el signo de la segunda derivada  $f''$  en  $x = a$  determina la concavidad de  $f$  en ese punto. Para precisar esta idea, supón que la función  $f$  posee un valor extremo local (máximo o mínimo) en  $x = a$ , de modo que  $f'(a) = 0$ . En ese caso, la aproximación cuadrática de  $f(x)$  se reduce a

$$f(x) \approx P_2(x) = f(a) + \frac{1}{2}f''(a)(x - a)^2.$$

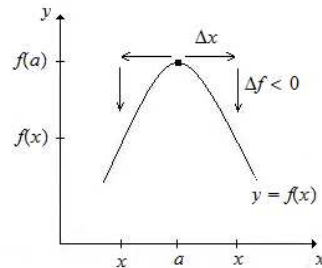
Introduciendo la notación  $\Delta x = x - a$  y  $\Delta f = f(x) - f(a)$ , se tiene

$$\Delta f \approx \frac{1}{2}f''(a)(\Delta x)^2.$$

Si  $f''(a) > 0$  entonces  $\Delta f > 0$ , por lo que  $f$  se incrementa a partir de  $f(a)$ . Como se observa en la figura de la izquierda, esto implica que  $f$  tiene un mínimo en  $x = a$ , de modo que  $f$  es convexa en ese punto. Por el contrario, si  $f''(a) < 0$  entonces  $\Delta f < 0$ , por lo que  $f$  disminuye a partir de  $f(a)$ . De acuerdo con la figura de la derecha, esto implica que  $f$  tiene un máximo en  $x = a$ , de modo que  $f$  es cóncava en ese punto.



$f''(a) > 0$   
 $\Rightarrow f$  convexa en  $x = a$



$f''(a) < 0$   
 $\Rightarrow f$  cóncava en  $x = a$

El criterio de la segunda derivada para funciones  $f$  de una variable se resume, entonces, de la siguiente manera:

- i)  $f''(a) > 0 \Rightarrow f$  es convexa en  $x = a$ ,
- ii)  $f''(a) < 0 \Rightarrow f$  es cóncava en  $x = a$ .

Los resultados anteriores pueden extenderse al caso de una función diferenciable de dos variables,  $f : S \rightarrow \mathbb{R}$ , con  $S \subset \mathbb{R}^2$  abierto y convexo. Como  $f$  es diferenciable en  $S$ , existe un plano tangente a la superficie  $z = f(x, y)$  en cada punto  $(a, b) \in S$ , dado por

$$z = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b).$$

#### 4.1 Polinomio de Taylor de orden 2. Matriz hessiana

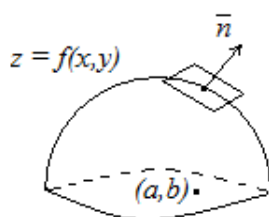
Esto significa que para aquellos puntos  $(x, y)$  cercanos a  $(a, b)$  podemos aproximar la función  $f$  por su *linealización*  $L$ , dada por

$$L(x, y) = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b),$$

que estudiamos en la sección 3.2. Esta es una función polinomial de grado 1, y es tal que satisface

$$L(a, b) = f(a, b), \quad L_x(a, b) = f_x(a, b) \quad \text{y} \quad L_y(a, b) = f_y(a, b),$$

es decir, en  $(x, y) = (a, b)$  las funciones  $L$  y  $f$  toman el mismo valor y además tienen el mismo vector normal,  $\vec{n} = (f_x, f_y, -1)$ .



Cuando  $f$  es doblemente diferenciable en  $S$  es posible construir una aproximación cuadrática,  $P_2(x, y)$ , que presente la misma concavidad que  $f$  en  $(x, y) = (a, b)$ , estableciendo la igualdad de todas sus derivadas parciales de orden 2. Es fácil verificar que una función que satisface estas condiciones es

$$P_2(x, y) = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b) + \frac{1}{2} [f_{xx}(a, b)(x - a)^2 + 2f_{xy}(a, b)(x - a)(y - b) + f_{yy}(a, b)(y - b)^2],$$

conocida como el *polinomio de Taylor de orden 2* generado por  $f(x, y)$  alrededor de  $(x, y) = (a, b)$ . Tomando como base esta aproximación, a continuación desarrollaremos un criterio para establecer la concavidad local de  $f$ , dependiendo del signo de las derivadas  $f_{xx}$ ,  $f_{yy}$  y  $f_{xy}$  en el punto  $(x, y) = (a, b)$ .

Para ello, supongamos que la función  $f$  posee un valor extremo local en  $(x, y) = (a, b)$ , de modo que  $f_x(a, b) = f_y(a, b) = 0$ . En ese caso,

$$\begin{aligned} f(x, y) &\approx P_2(x, y) \\ &= f(a, b) + \frac{1}{2} [f_{xx}(a, b)(x - a)^2 + 2f_{xy}(a, b)(x - a)(y - b) + f_{yy}(a, b)(y - b)^2]. \end{aligned}$$

Introduciendo la notación  $\Delta x = x - a$ ,  $\Delta y = y - b$  y  $\Delta f = f(x, y) - f(a, b)$ , se tiene

$$\Delta f \approx \frac{1}{2} [f_{xx}(a, b)(\Delta x)^2 + 2f_{xy}(a, b)(\Delta x)(\Delta y) + f_{yy}(a, b)(\Delta y)^2].$$

## Capítulo 4 Funciones cóncavas y cuasicóncavas

Claramente, el signo de  $\Delta f$  está determinado por el signo de las segundas derivadas parciales, pero no de una manera directa como en el caso de funciones de una variable.

Para encontrar condiciones suficientes sobre el signo de  $\Delta f$  partimos de

$$2\Delta f \approx f_{xx}(\Delta x)^2 + 2f_{xy}(\Delta x)(\Delta y) + f_{yy}(\Delta y)^2,$$

en donde, para simplificar la notación, hemos omitido que las derivadas parciales están evaluadas en  $(a, b)$ . Ahora multiplicamos ambos lados de esta ecuación por  $f_{xx}$  (o bien, por  $f_{yy}$ , si  $f_{xx} = 0$ ), obteniendo

$$2f_{xx}\Delta f \approx f_{xx}^2(\Delta x)^2 + 2f_{xx}f_{xy}(\Delta x)(\Delta y) + f_{xx}f_{yy}(\Delta y)^2.$$

Luego completamos cuadrados en el lado derecho de la ecuación,

$$\begin{aligned} 2f_{xx}\Delta f &\approx f_{xx}^2(\Delta x)^2 + 2f_{xx}f_{xy}(\Delta x)(\Delta y) + f_{xy}^2(\Delta y)^2 + [f_{xx}f_{yy}(\Delta y)^2 - f_{xy}^2(\Delta y)^2] \\ &= [f_{xx}(\Delta x) + f_{xy}(\Delta y)]^2 + [f_{xx}f_{yy} - f_{xy}^2](\Delta y)^2. \end{aligned}$$

Por último, dividimos ambos lados de la ecuación por  $2f_{xx}$ , obteniendo

$$\Delta f \approx \frac{[f_{xx}(\Delta x) + f_{xy}(\Delta y)]^2}{2f_{xx}} + \frac{[f_{xx}f_{yy} - f_{xy}^2](\Delta y)^2}{2f_{xx}}.$$

Observamos que la condición  $f_{xx} > 0$  y  $f_{xx}f_{yy} - f_{xy}^2 > 0$  en  $(a, b)$  es suficiente para garantizar que  $\Delta f > 0$  a partir de ese punto; siguiendo un razonamiento análogo al caso de una función de una variable, se tiene entonces que  $f$  es convexa en  $(a, b)$ . Similarmente, la condición  $f_{xx} < 0$  y  $f_{xx}f_{yy} - f_{xy}^2 > 0$  en  $(a, b)$  es suficiente para garantizar que  $\Delta f < 0$  a partir de ese punto; de este modo,  $f$  es cóncava en  $(a, b)$ . Se llega entonces a las siguientes condiciones de suficiencia:

- i)  $f_{xx} > 0$  y  $f_{xx}f_{yy} - f_{xy}^2 > 0$  en  $(a, b) \Rightarrow f$  es convexa en  $(a, b)$ ,
- ii)  $f_{xx} < 0$  y  $f_{xx}f_{yy} - f_{xy}^2 > 0$  en  $(a, b) \Rightarrow f$  es cóncava en  $(a, b)$ .

Los resultados anteriores pueden expresarse de una manera más simple, utilizando un lenguaje matricial. Para ello, primero reescribimos la expresión

$$\Delta f \approx \frac{1}{2} [f_{xx}(a, b)(\Delta x)^2 + 2f_{xy}(a, b)(\Delta x)(\Delta y) + f_{yy}(a, b)(\Delta y)^2]$$

como un producto de matrices, de la forma

$$\begin{aligned} \Delta f &\approx \frac{1}{2} \begin{pmatrix} \Delta x & \Delta y \end{pmatrix} \begin{pmatrix} f_{xx}(a, b) & f_{xy}(a, b) \\ f_{xy}(a, b) & f_{yy}(a, b) \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} \\ &= \frac{1}{2} (\overrightarrow{\Delta x})^T H (\overrightarrow{\Delta x}), \end{aligned}$$



## 4.2 Funciones cóncavas y funciones convexas

en donde  $\vec{\Delta x} = \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}$  es el vector de incrementos,  $(\vec{\Delta x})^T$  es su transpuesta y

$$H = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix},$$

es la matriz de segundas derivadas de  $f$  evaluadas en el punto  $(a, b)$ , conocida como *matriz Hessiana* de  $f$ . Nota que el determinante  $|H|$  de  $H$  está dado por

$$|H| = f_{xx}f_{yy} - f_{xy}^2,$$

que es precisamente uno de los términos en las condiciones suficientes que desarrollamos con anterioridad. De esta manera, dichas condiciones suficientes pueden expresarse en términos de la matriz hessiana  $H$  de  $f$ , como:

- i)  $f_{xx} > 0$  y  $|H| > 0$  en  $(a, b) \Rightarrow f$  es convexa en  $(a, b)$ ,
- ii)  $f_{xx} < 0$  y  $|H| > 0$  en  $(a, b) \Rightarrow f$  es cóncava en  $(a, b)$ .

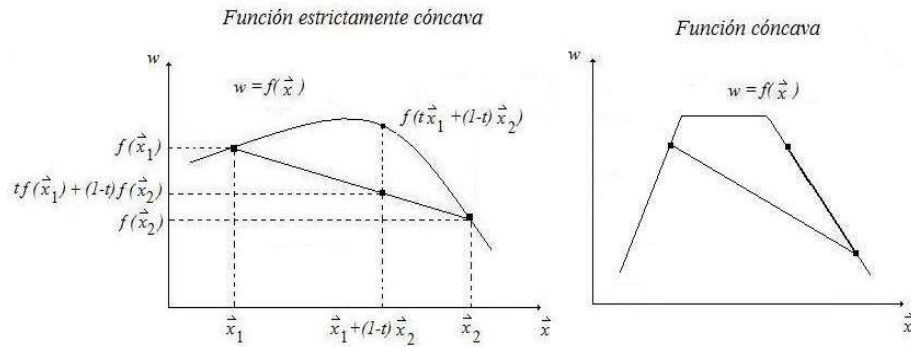
## 4.2 Funciones cóncavas y funciones convexas

En esta sección formalizamos el concepto de concavidad o convexidad de una función, que reviste de gran interés en el tema de optimización de funciones. Iniciamos el tema presentando las definiciones de función cóncava y de función convexa, que son válidas aun para el caso de funciones no diferenciables.

**Definición.** Sea  $S \subset \mathbb{R}^n$  un conjunto convexo, y sea  $f : S \rightarrow \mathbb{R}$ . Se dice que  $f$  es una *función cóncava* si para todos  $\vec{x}_1, \vec{x}_2 \in S$  se satisface

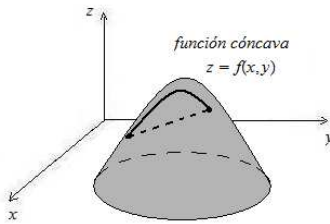
$$f(t\vec{x}_1 + (1-t)\vec{x}_2) \geq tf(\vec{x}_1) + (1-t)f(\vec{x}_2),$$

para todo número  $t \in [0, 1]$ . La función es *estrictamente cóncava* si la desigualdad es estricta, para  $t \in (0, 1)$  y  $\vec{x}_1 \neq \vec{x}_2$ .



## Capítulo 4 Funciones cóncavas y cuasicóncavas

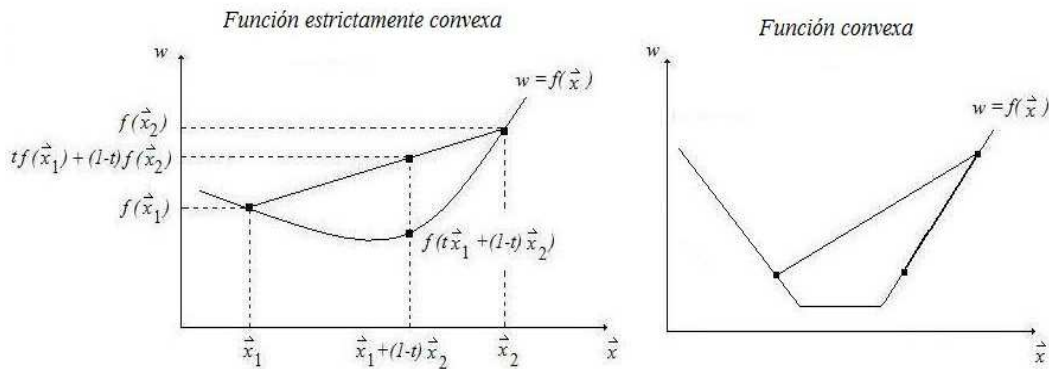
En otras palabras, una función es cóncava si la recta que une cualesquiera dos puntos de su gráfica queda por debajo de la gráfica, o en la gráfica, pero nunca por encima de ella. Si la recta queda siempre por debajo de la gráfica, la función es estrictamente cóncava. Nota que este argumento es válido en general para superficies o hipersuperficies, como se ilustra en la siguiente figura.



**Definición.** Sea  $S \subset \mathbb{R}^n$  un conjunto convexo, y sea  $f : S \rightarrow \mathbb{R}$ . Se dice que  $f$  es una *función convexa en  $S$*  si para todos  $\vec{x}_1, \vec{x}_2 \in S$  se satisface

$$f(t\vec{x}_1 + (1-t)\vec{x}_2) \leq tf(\vec{x}_1) + (1-t)f(\vec{x}_2),$$

para todo número  $t \in [0, 1]$ . La función es *estrictamente convexa* si la desigualdad es estricta, para  $t \in (0, 1)$  y  $\vec{x}_1 \neq \vec{x}_2$ .



Ejemplo:

Demuestra que la función  $f(x) = |x|$  es convexa no estricta.

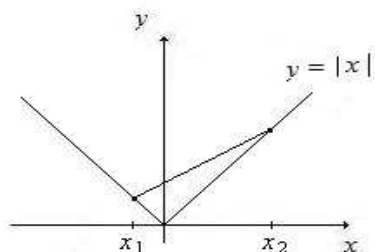
Sean  $x_1, x_2 \in \mathbb{R}$  y sea  $t \in [0, 1]$ . Entonces

$$\begin{aligned} f(tx_1 + (1-t)x_2) &= |tx_1 + (1-t)x_2| \\ &\leq |tx_1| + |(1-t)x_2| \quad (\text{desigualdad del triángulo}) \\ &= |t||x_1| + |1-t||x_2| \quad (\text{propiedades del valor absoluto}) \\ &= t|x_1| + (1-t)|x_2| \quad (t \geq 0 \text{ y } 1-t \geq 0) \\ &= tf(x_1) + (1-t)f(x_2). \end{aligned}$$

## 4.2 Funciones cóncavas y funciones convexas

De esta manera,

$$f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2).$$



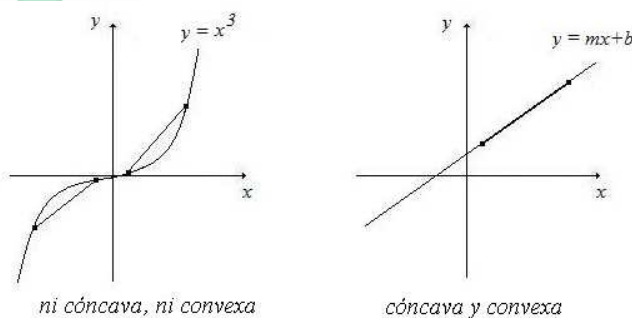
**Teorema.** Sea  $S \subset \mathbb{R}^n$  un conjunto convexo. Sean  $f, g : S \rightarrow \mathbb{R}$  funciones cóncavas en  $S$  y sea  $\alpha \in \mathbb{R}$ .

- a) Si  $\alpha > 0$ , entonces  $\alpha f$  es cóncava.
- b) Si  $\alpha < 0$ , entonces  $\alpha f$  es convexa.
- c)  $f + g$  es cóncava.
- d) Si  $h : S \rightarrow \mathbb{R}$  es una función lineal, entonces  $f + h$  es cóncava.
- e) Sea  $h : Y \rightarrow \mathbb{R}$  una función cóncava y creciente, tal que  $f(S) \subset Y \subset \mathbb{R}$ .

Entonces  $h \circ f$  es cóncava.

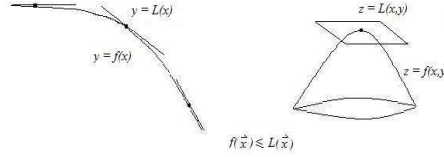
Por ejemplo, si  $P(L, K)$  es una función de producción cóncava y  $p > 0$  es el precio del bien, entonces el ingreso  $pP(L, K)$  es una función cóncava (inciso a). A su vez, si  $C(L, K) = wL + rK$  es una función de costos lineal, también  $-C$  es lineal; así, la función de beneficio  $\Pi = pP(L, K) - C(L, K)$  es cóncava (inciso d). Por otra parte, si  $u$  es una función de utilidad cóncava, la composición  $\ln u$  también lo es, ya que  $\ln x$  es una función cóncava y creciente (inciso e).

Claramente, hay funciones que no son cóncavas ni son convexas, tales como  $f(x) = x^3$ , mientras que hay funciones que son tanto cóncavas como convexas (no estrictas), como ocurre con las funciones lineales (rectas o planos):

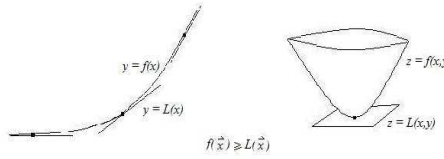


## Capítulo 4 Funciones cóncavas y cuasicóncavas

En el caso particular de que la función  $f$  sea diferenciable, o de clase  $C^1$ , se puede demostrar que  $f(\vec{x})$  es una función cóncava en su dominio si y sólo si su gráfica nunca está por encima de su linealización  $L(\vec{x})$  (recta o plano tangente).



Análogamente,  $f(\vec{x})$  es una función convexa si y sólo si su gráfica nunca está por debajo de su linealización  $L(\vec{x})$ .



Para enunciar estos resultados de una manera formal, reescribimos la linealización  $L(x, y)$  de una función de dos variables,  $f(x, y)$ , alrededor del punto  $(x_0, y_0)$ , como

$$\begin{aligned} L(x, y) &= f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) \\ &= f(x_0, y_0) + (f_x(x_0, y_0), f_y(x_0, y_0)) \cdot (x - x_0, y - y_0) \\ &= f(x_0, y_0) + \nabla f(x_0, y_0) \cdot (x - x_0, y - y_0), \end{aligned}$$

en donde  $\nabla f(x_0, y_0)$  denota el gradiente de  $f$  evaluado en el punto  $(x_0, y_0)$ . Por último, introduciendo la notación  $\vec{x} = (x, y)$  y  $\vec{x}_0 = (x_0, y_0)$ , se obtiene

$$L(\vec{x}) = f(\vec{x}_0) + \nabla f(\vec{x}_0) \cdot (\vec{x} - \vec{x}_0).$$

La ventaja de esta última expresión es que no está limitada a vectores en  $\mathbb{R}^2$ , sino que es válida para vectores en el espacio general  $\mathbb{R}^n$ .

**Teorema.** Sea  $S \subset \mathbb{R}^n$  un conjunto abierto y convexo, y sea  $f : S \rightarrow \mathbb{R}$ , con  $f \in C^1(S)$ . Entonces,

a)  $f$  es cóncava en  $S \Leftrightarrow$  para todos  $\vec{x}, \vec{x}_0 \in S$

$$f(\vec{x}) \leq f(\vec{x}_0) + \nabla f(\vec{x}_0) \cdot (\vec{x} - \vec{x}_0).$$

b)  $f$  es convexa en  $S \Leftrightarrow$  para todos  $\vec{x}, \vec{x}_0 \in S$

$$f(\vec{x}) \geq f(\vec{x}_0) + \nabla f(\vec{x}_0) \cdot (\vec{x} - \vec{x}_0).$$

Si las desigualdades son estrictas para  $\vec{x} \neq \vec{x}_0$ , entonces  $f$  es estrictamente cóncava o estrictamente convexa en  $S$ .

## 4.2 Funciones cóncavas y funciones convexas

Más particularmente, en el caso de una función  $f$  doblemente diferenciable, o de clase  $C^2$ , existe un criterio aun más simple para establecer su concavidad o convexidad. Este criterio se enuncia en el siguiente teorema, y está basado en el signo de los elementos de su matriz hessiana  $H$ , o matriz de las segundas derivadas de  $f$ , como se justificó en la sección 4.1.

**Teorema.** Sea  $S \subset \mathbb{R}^2$  un conjunto abierto y convexo, y sea  $f : S \rightarrow \mathbb{R}$ , con  $f \in C^2(S)$ . Entonces

- a)  $f_{xx} < 0$  y  $f_{xx}f_{yy} - f_{xy}^2 > 0$  en  $S \Rightarrow f$  es estrictamente cóncava en  $S$ ,
- b)  $f_{xx} > 0$  y  $f_{xx}f_{yy} - f_{xy}^2 > 0$  en  $S \Rightarrow f$  es estrictamente convexa en  $S$ ,
- c)  $f_{xx}f_{yy} - f_{xy}^2 < 0$  en  $S \Rightarrow f$  no es cóncava ni es convexa en  $S$ .

Ejemplos:

1. Analiza la concavidad de  $f(x, y) = -2x^2 - 3y^2$  en cada punto  $(x, y) \in \mathbb{R}^2$ .

La matriz Hessiana de  $f$  en cada punto  $(x, y)$  es

$$H(x, y) = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix} = \begin{pmatrix} -4 & 0 \\ 0 & -6 \end{pmatrix}.$$

Como  $f_{xx} = -4 < 0$  y  $|H| = f_{xx}f_{yy} - f_{xy}^2 = 24 > 0$ , por lo tanto  $f$  es estrictamente cóncava, para todos  $(x, y) \in \mathbb{R}^2$ .

2. Analiza la concavidad de  $f(x, y) = x^2 + y^2 + xy$  en cada punto  $(x, y) \in \mathbb{R}^2$ .

La matriz Hessiana de  $f$  en cada punto  $(x, y)$  es

$$H(x, y) = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

Como  $f_{xx} = 2 > 0$  y  $|H| = f_{xx}f_{yy} - f_{xy}^2 = 3 > 0$ , por lo tanto  $f$  es estrictamente convexa, para todos  $(x, y) \in \mathbb{R}^2$ .

3. Analiza la concavidad de  $f(x, y) = x^2 - y^2$  en cada punto  $(x, y) \in \mathbb{R}^2$ .

La matriz Hessiana de  $f$  en cada punto  $(x, y)$  es

$$H(x, y) = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}.$$

Como  $f_{xx} = 2 > 0$  y  $|H| = f_{xx}f_{yy} - f_{xy}^2 = -4 < 0$ , por lo tanto  $f$  no es ni cóncava ni convexa en  $\mathbb{R}^2$ .

## Capítulo 4 Funciones cóncavas y cuasicóncavas

4. Analiza la concavidad de  $f(x, y) = x^3 + y^3$  en cada punto  $(x, y) \in \mathbb{R}^2$ .

La matriz Hessiana de  $f$  en cada punto  $(x, y)$  es

$$H(x, y) = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix} = \begin{pmatrix} 6x & 0 \\ 0 & 6y \end{pmatrix}.$$

Como  $f_{xx} = 6x$  y  $|H| = f_{xx}f_{yy} - f_{xy}^2 = 36xy$ , se tiene que  $f$  es estrictamente convexa en el cuadrante I ( $x > 0, xy > 0$ ), estrictamente cóncava en el cuadrante III ( $x < 0, xy > 0$ ), y ni cóncava ni convexa en los cuadrantes II y IV. De manera global, decimos que  $f$  no es cóncava ni convexa en  $\mathbb{R}^2$ .

Un hecho interesante sobre el teorema anterior es que es de la forma  $\Rightarrow$  y no de la forma  $\Leftrightarrow$ . Esto es, hay funciones estrictamente cóncavas o convexas que no satisfacen las desigualdades estrictas del teorema. Por ejemplo,  $f(x, y) = x^4 + y^4$  es estrictamente convexa en  $\mathbb{R}^2$ , como puede demostrarse a partir de la definición. Sin embargo,  $f_{xx} = 0$  y  $f_{xx}f_{yy} - f_{xy}^2 = 0$  evaluadas en  $(0, 0)$ . El siguiente teorema establece condiciones necesarias y suficientes en relación con la concavidad/convexidad y los signos de las segundas derivadas, aun en el caso con  $f_{xx} = 0$  o  $f_{xx}f_{yy} - f_{xy}^2 = 0$ .

**Teorema.** Sea  $S \subset \mathbb{R}^2$  un conjunto abierto y convexo, y sea  $f : S \rightarrow \mathbb{R}$ , con  $f \in C^2(S)$ . Entonces

- a)  $f_{xx} \leq 0, f_{yy} \leq 0$  y  $f_{xx}f_{yy} - f_{xy}^2 \geq 0$  en  $S \Leftrightarrow f$  es cóncava en  $S$ ,
- b)  $f_{xx} \geq 0, f_{yy} \geq 0$  y  $f_{xx}f_{yy} - f_{xy}^2 \geq 0$  en  $S \Leftrightarrow f$  es convexa en  $S$ .

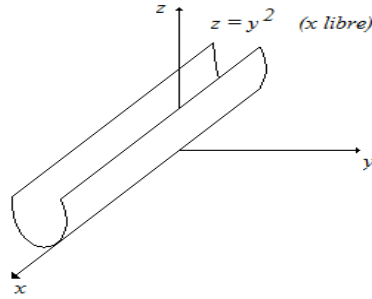
Compara cuidadosamente este teorema con el anterior (nota que aquí entra en escena el signo de  $f_{yy}$ , la implicación es del tipo  $\Leftrightarrow$  en lugar de  $\Rightarrow$ , y además ya no se trata de concavidad/convexidad estricta).

Ejemplos:

1. Analiza la concavidad de la función  $f(x, y) = y^2$ .

Nota que la función  $f(x, y) = y^2$  describe un cilindro parabólico que corre a lo largo del eje  $x$ , de modo que la función es convexa, pero no estricta, en  $\mathbb{R}^2$ .

### 4.3 Funciones cuasicóncavas y funciones cuasiconvexas



Este resultado es consistente con las condiciones necesarias de segundo orden para la matriz hessiana correspondiente,

$$H(x, y) = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix},$$

en donde  $f_{xx} = 0 \geq 0$ ,  $f_{yy} = 2 \geq 0$  y  $f_{xx}f_{yy} - f_{xy}^2 = 0 \geq 0$ .

2. Analiza la concavidad de la función  $f(x, y) = 2x + y$ .

Nota que la función  $f(x, y) = 2x + y$  describe un plano, de modo que la función es cóncava y convexa a la vez (obviamente, no estricta) en  $\mathbb{R}^2$ . Este resultado es consistente con las condiciones necesarias de segundo orden para la matriz hessiana correspondiente,

$$H(x, y) = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

ya que al ser  $f_{xx} = f_{yy} = f_{xx}f_{yy} - f_{xy}^2 = 0$  entonces se cumplen las dos condiciones del teorema anterior.

La generalización de los dos teoremas anteriores para el caso de funciones de tres o más variables puede consultarse en el Apéndice B.

### 4.3 Funciones cuasicóncavas y funciones cuasiconvexas

Además de las funciones cóncavas y convexas, existe otro tipo de funciones denominadas cuasicóncavas y cuasiconvexas, que también revisten de gran importancia en el tema de optimización. Antes de presentar estas últimas, es útil introducir primero la siguiente definición.

## Capítulo 4 Funciones cóncavas y cuasicóncavas

**Definición.** Sea  $S \subset \mathbb{R}^n$  y sean  $f : S \rightarrow \mathbb{R}$  y  $k \in \mathbb{R}$ .

a) El *contorno* de  $f$  en  $k$  es el conjunto

$$C_f(k) = \{ \vec{x} \in S \mid f(\vec{x}) = k \}.$$

b) El *contorno superior* de  $f$  en  $k$  es el conjunto

$$CS_f(k) = \{ \vec{x} \in S \mid f(\vec{x}) \geq k \}.$$

c) El *contorno inferior* de  $f$  en  $k$  es el conjunto

$$CI_f(k) = \{ \vec{x} \in S \mid f(\vec{x}) \leq k \}.$$

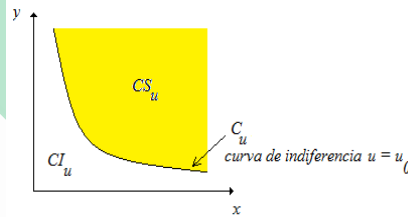
El contorno  $C_f(k)$  es lo que denominamos en la sección 2.2 como conjunto de nivel, o curva de nivel en el caso de funciones  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ . Así, por ejemplo, en el caso de una función de utilidad  $u(x, y)$  correspondiente a la canasta de dos bienes  $(x, y)$ , el contorno  $C_u(u_0)$  es la curva de indiferencia

$$C_u(u_0) = \{ (x, y) \in \mathbb{R}_+^2 \mid u(x, y) = u_0 \},$$

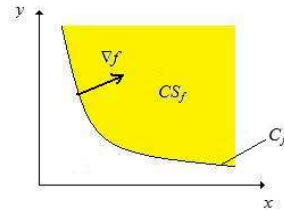
mientras que el contorno superior  $CS_u(u_0)$  representa las preferencias del consumidor,

$$CS_u(u_0) = \{ (x, y) \in \mathbb{R}_+^2 \mid u(x, y) \geq u_0 \},$$

dadas por las canastas que le generan una utilidad mayor o igual que  $u_0$ .



Nota que los contornos  $CS_f(k)$  y  $CI_f(k)$  son subconjuntos del dominio de  $f$  y ambas regiones contienen al contorno  $C_f(k)$ . Para determinar las regiones  $CS_f(k)$  y  $CI_f(k)$  basta con resolver la desigualdad correspondiente a su definición. Existe una manera alternativa, que consiste en identificar solamente el contorno  $C_f(k)$  y graficar en él el vector gradiente  $\nabla f$ , que necesariamente apuntará hacia el contorno superior  $CS_f(k)$ .





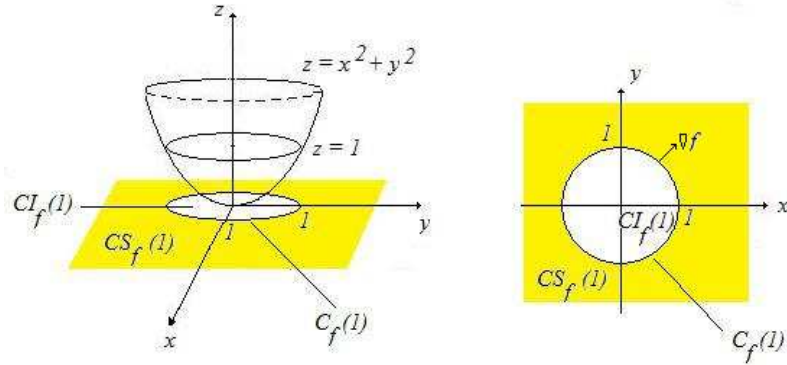
### 4.3 Funciones cuasicóncavas y funciones cuasiconvexas

Ejemplos:

1. Sea  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ , definida por  $f(x, y) = x^2 + y^2$ . Encuentra los contornos  $C_f$ ,  $CS_f$  y  $CI_f$  correspondientes a  $k = 1$ .

En este caso, se tiene directamente

$$\begin{aligned} C_f(1) &= \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}, \\ CS_f(1) &= \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \geq 1\}, \\ CI_f(1) &= \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}. \end{aligned}$$



2. Sea  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ , definida por  $f(x, y) = 2 - x^2 - y^2$ . Encuentra los contornos  $C_f$ ,  $CS_f$  y  $CI_f$  correspondientes a  $k = 1$ .

En este caso se tiene  $C_f(1) = \{(x, y) \in \mathbb{R}^2 \mid 2 - x^2 - y^2 = 1\}$ , es decir,

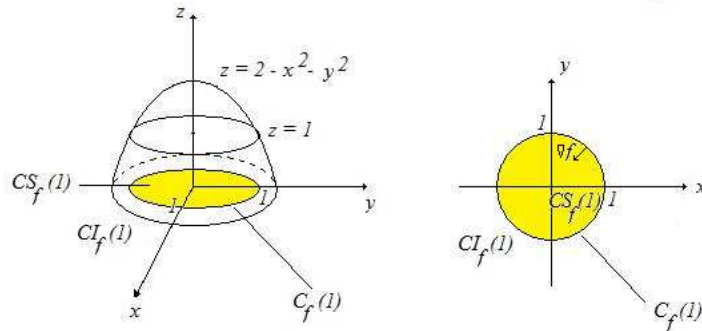
$$C_f(1) = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}.$$

Asimismo,  $CS_f(1) = \{(x, y) \in \mathbb{R}^2 \mid 2 - x^2 - y^2 \geq 1\}$ , es decir,

$$CS_f(1) = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$$

y  $CI_f(1) = \{(x, y) \in \mathbb{R}^2 \mid 2 - x^2 - y^2 \leq 1\}$ , es decir,

$$CI_f(1) = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \geq 1\}.$$



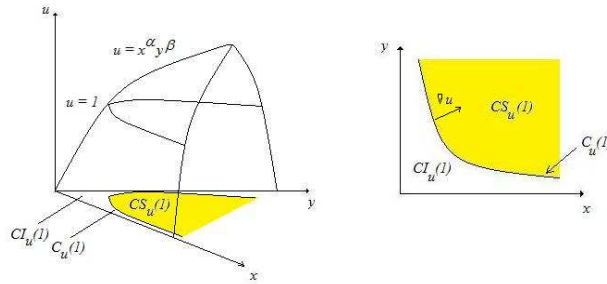
## Capítulo 4 Funciones cóncavas y cuasicóncavas

3. Sea  $u : \mathbb{R}_+^2 \rightarrow \mathbb{R}$  una función de utilidad tipo Cobb-Douglas, definida por  $u(x, y) = x^\alpha y^\beta$ ,  $\alpha, \beta > 0$ . Encuentra los contornos  $C_u$ ,  $CS_u$  y  $CI_u$  correspondientes a  $k = 1$ .

Procediendo de manera análoga a los dos ejemplos anteriores, se tiene

$$\begin{aligned} C_u(1) &= \{(x, y) \in \mathbb{R}_+^2 \mid x^\alpha y^\beta = 1\}, \\ CS_u(1) &= \{(x, y) \in \mathbb{R}_+^2 \mid x^\alpha y^\beta \geq 1\}, \\ CI_u(1) &= \{(x, y) \in \mathbb{R}_+^2 \mid x^\alpha y^\beta \leq 1\}, \end{aligned}$$

que representan las canastas con una utilidad  $u = 1$ ,  $u \geq 1$  y  $u \leq 1$ , respectivamente.



4. Sea  $f : \mathbb{R} \rightarrow \mathbb{R}$ , definida por  $f(x) = 2 - x^2$ . Encuentra los contornos  $C_f$ ,  $CS_f$  y  $CI_f$  correspondientes a  $k = 1$ .

En este caso se tiene  $C_f(1) = \{x \in \mathbb{R} \mid 2 - x^2 = 1\}$ , es decir,

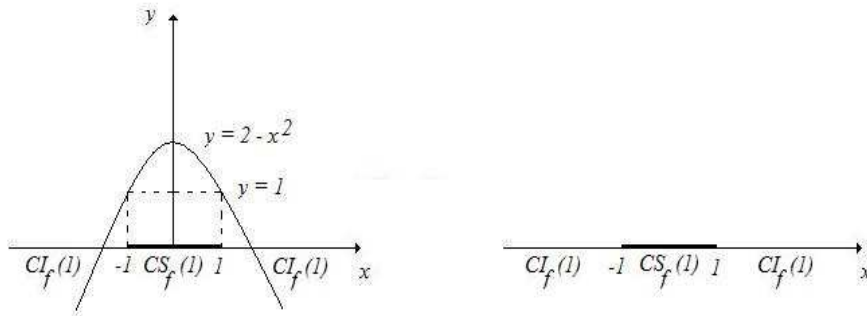
$$C_f(1) = \{x \in \mathbb{R} \mid |x| = 1\}.$$

Asimismo,  $CS_f(1) = \{x \in \mathbb{R} \mid 2 - x^2 \geq 1\}$ , es decir,

$$CS_f(1) = \{x \in \mathbb{R} \mid |x| \leq 1\}$$

y  $CI_f(1) = \{x \in \mathbb{R} \mid 2 - x^2 \leq 1\}$ , es decir,

$$CI_f(1) = \{x \in \mathbb{R} \mid |x| \geq 1\}.$$



### 4.3 Funciones cuasicóncavas y funciones cuasiconvexas

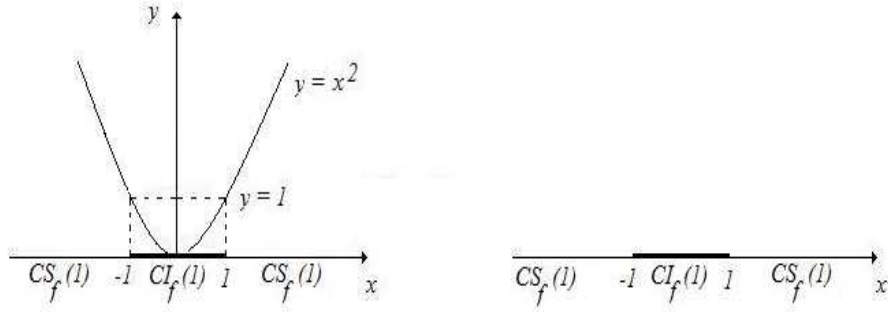
5. Sea  $f : \mathbb{R} \rightarrow \mathbb{R}$ , definida por  $f(x) = x^2$ . Encuentra los contornos  $C_f$ ,  $CS_f$  y  $CI_f$  correspondientes a  $k = 1$ .

En este caso, se tiene

$$C_f(1) = \{x \in \mathbb{R} \mid |x| = 1\},$$

$$CS_f(1) = \{x \in \mathbb{R} \mid |x| \geq 1\},$$

$$CI_f(1) = \{x \in \mathbb{R} \mid |x| \leq 1\}.$$



6. Para la función  $f(x) = \ln(2 - x)$ , encuentra los contornos  $C_f$ ,  $CS_f$  y  $CI_f$  correspondientes a  $k = 0$ .

Primero notamos que el dominio  $D_f$  de la función  $f$  es el conjunto

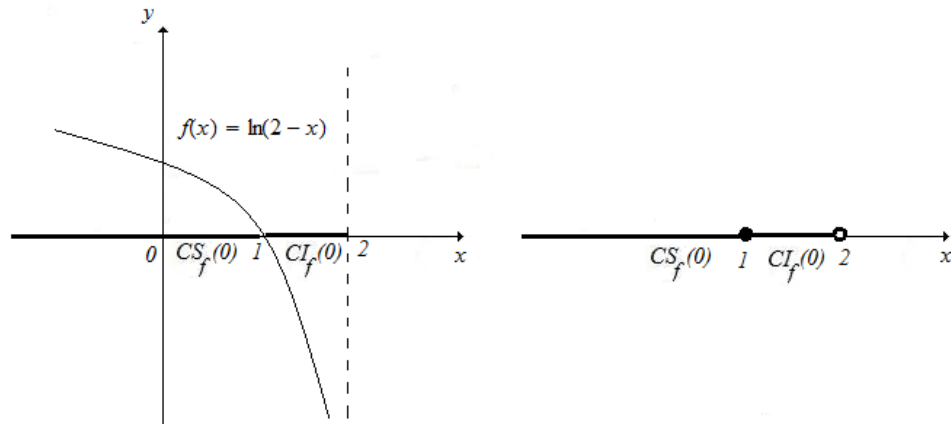
$$D_f = \{x \in \mathbb{R} \mid -\infty < x < 2\}.$$

En este caso, se tiene

$$C_f(0) = \{x \in D_f \mid x = 1\} = \{1\},$$

$$CS_f(0) = \{x \in D_f \mid x \leq 1\} = \{x \in \mathbb{R} \mid -\infty < x \leq 1\},$$

$$CI_f(0) = \{x \in D_f \mid x \geq 1\} = \{x \in \mathbb{R} \mid 1 \leq x < 2\}.$$



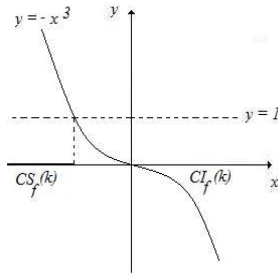
## Capítulo 4 Funciones cóncavas y cuasicóncavas

Los ejemplos anteriores ilustran cómo en el caso de funciones  $f$  en  $\mathbb{R}^3$  los contornos  $C_f$ ,  $CS_f$  y  $CI_f$  están en  $\mathbb{R}^2$ , mientras que para funciones  $f$  en  $\mathbb{R}^2$  los contornos están en  $\mathbb{R}$ . En todos los casos, los contornos son subconjuntos del dominio, como se muestra muy especialmente en el ejemplo 6. A la luz de esos ejemplos debe resultar claro el contenido del siguiente teorema.

### Teorema.

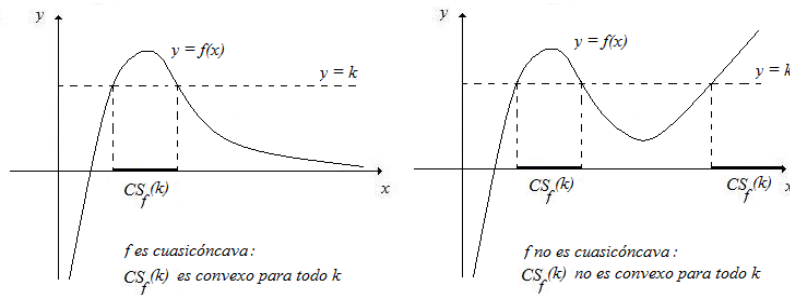
- a)  $f$  es cóncava  $\Rightarrow CS_f(k)$  es convexo, para todo  $k$  en la imagen de  $f$ .
- b)  $f$  es convexa  $\Rightarrow CI_f(k)$  es convexo, para todo  $k$  en la imagen de  $f$ .

Uno podría preguntarse ahora si las implicaciones anteriores pueden ser invertidas. En otras palabras, ¿será cierto que  $CS_f$  convexo implica que  $f$  es cóncava? La respuesta es negativa, como lo ilustra la siguiente gráfica, correspondiente a la función  $f(x) = -x^3$ .



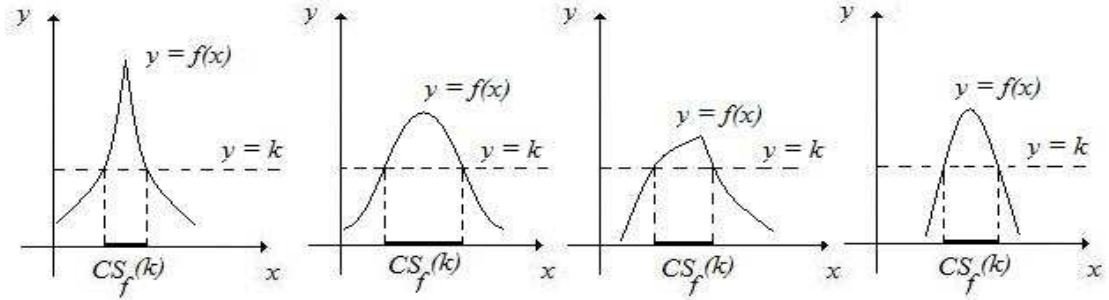
En este ejemplo es claro que tanto  $CS_f(k)$  como  $CI_f(k)$  son convexas para todo  $k \in \mathbb{R}$ , pero la función  $f(x) = -x^3$  no es cóncava ni es convexa en su dominio. Una función como ésta es un ejemplo de función cuasicóncava y cuasiconvexa, a la vez, como se define a continuación.

**Definición.** Sea  $S \subset \mathbb{R}^n$  un conjunto convexo. Se dice que una función  $f : S \rightarrow \mathbb{R}$  es *cuasicóncava*, si para todo  $k$  en la imagen de  $f$  el conjunto  $CS_f(k)$  es convexo. La función es *cuasiconvexa*, si para todo  $k$  en la imagen de  $f$  el conjunto  $CI_f(k)$  es convexo.

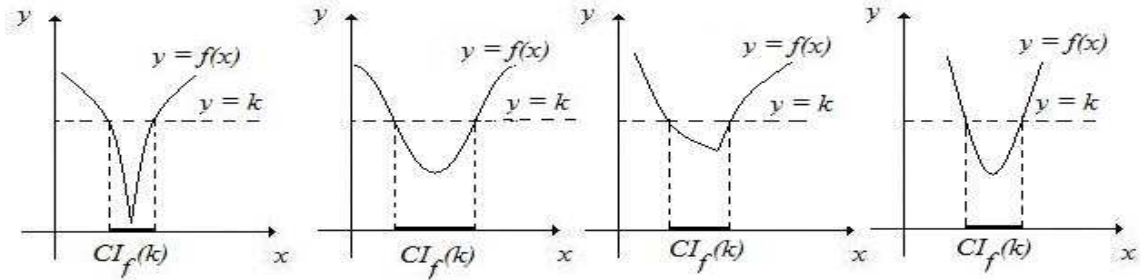


### 4.3 Funciones cuasicóncavas y funciones cuasiconvexas

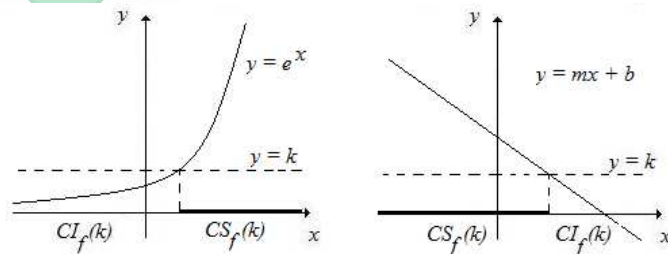
La figura de la izquierda representa una función cuasicóncava, ya que para todo  $k$  el contorno superior  $CS_f(k)$  es convexo. La figura de la derecha muestra una función que no es cuasicóncava, ya que su contorno superior  $CS_f(k)$  no siempre es convexo (depende del valor de  $k$ ). Nota que esta última tampoco es cuasiconvexa, puesto que su contorno inferior  $CI_f(k)$  tampoco es convexo en general. Las siguientes figuras muestran otras funciones cuasicóncavas en  $\mathbb{R}^2$ .



Análogamente, las siguientes figuras muestran ejemplos de funciones cuasiconvexas en  $\mathbb{R}^2$ .

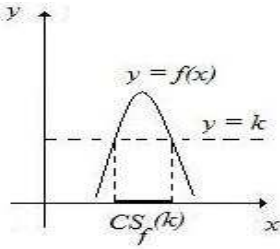


También observa que una función puede ser convexa y cuasicóncava a la vez, como es el caso de la función  $e^x$ , que también es cuasiconvexa, y que todas las funciones lineales (rectas, planos e hiperplanos) son cuasicóncavas y cuasiconvexas, además de ser cóncavas y convexas (no estrictas).



Por último, nota que toda función cóncava (convexa) es también cuasicóncava (cuasiconvexa), pero no viceversa.

## Capítulo 4 Funciones cóncavas y cuasicóncavas



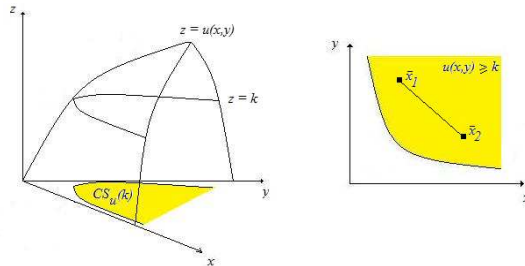
### Teorema

- a)  $f$  cóncava  $\Rightarrow f$  cuasicóncava.
- b)  $f$  convexa  $\Rightarrow f$  cuasiconvexa.

También es posible demostrar que las implicaciones  $CS_f$  convexo  $\Rightarrow f$  cuasicóncava y  $CI_f$  convexo  $\Rightarrow f$  cuasiconvexa de la definición de cuasiconcavidad/cuasiconvexidad son más bien del tipo  $\Longleftrightarrow$ . En resumen, podemos concluir que

$$\begin{aligned} f \text{ cóncava} &\Rightarrow f \text{ cuasicóncava} \Longleftrightarrow CS_f \text{ convexo,} \\ f \text{ convexa} &\Rightarrow f \text{ cuasiconvexa} \Longleftrightarrow CI_f \text{ convexo.} \end{aligned}$$

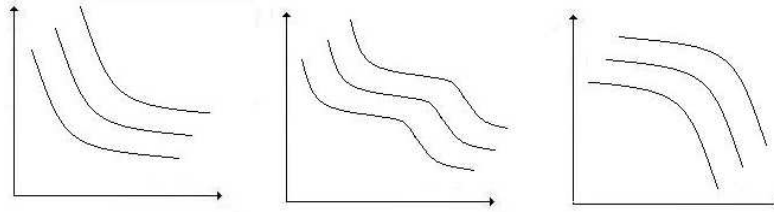
Este resultado es muy importante en economía. Así, por ejemplo, si suponemos que una función de utilidad  $u$  es cuasicóncava, entonces el conjunto de canastas  $\vec{x}$  del espacio de bienes que producen al consumidor una utilidad  $u(\vec{x}) \geq k$  (o sea,  $CS_u(k)$ ) es convexo. De esta manera, cualquier combinación convexa,  $\lambda \vec{x}_1 + (1 - \lambda) \vec{x}_2$ ,  $0 < \lambda < 1$ , de dos canastas  $\vec{x}_1, \vec{x}_2$  con  $u \geq k$  también genera una utilidad mayor o igual a  $k$ .



Se puede demostrar que la convexidad del contorno superior  $CS_u(k)$  garantiza que se preserve el orden en las relaciones de preferencia del consumidor. Si  $CS_u$  es convexo, las curvas de nivel de  $u$  son funciones convexas, como en la figura de la

### 4.3 Funciones cuasicóncavas y funciones cuasiconvexas

izquierda, y no como en las otras dos figuras.

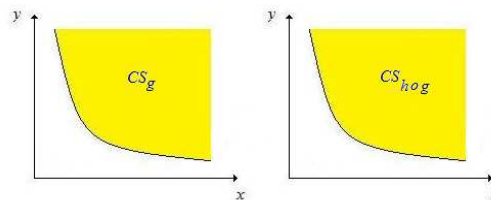


Esto equivale a la condición de que la función  $u$ , que representa el orden de las preferencias del consumidor, sea cuasicóncava. Observa que no tiene sentido imponer la condición más restrictiva de que  $u$  sea una función cóncava, ya que la única propiedad significativa de esta función es el carácter de sus curvas de nivel.

En economía es muy frecuente el uso de una transformación del tipo  $\ln u$ , en lugar de utilizar la función de utilidad  $u$ . El siguiente teorema establece que una transformación monotónica creciente de este tipo no afecta el argumento anterior en relación con las canastas correspondientes a  $\ln u \geq k$ .

**Teorema.** Sea  $S \subset \mathbb{R}^n$  un conjunto convexo y sea  $g : S \rightarrow \mathbb{R}$  una función cuasicóncava. Sea  $h : Y \subset \mathbb{R} \rightarrow \mathbb{R}$  una función creciente, donde  $g(S) \subset Y$ . Entonces la composición  $h \circ g$  es una función cuasicóncava, y además

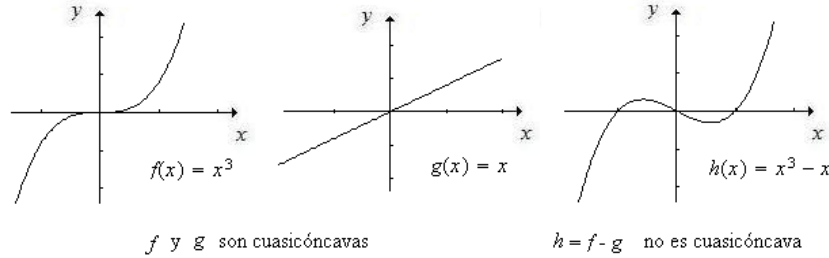
$$CS_g(y) \subset CS_{h \circ g}(h(y)).$$



Es importante señalar que la suma de funciones cuasicóncavas (cuasiconvexas) no necesariamente es una función cuasicóncava (cuasiconvexa), a diferencia de lo que sucede con la suma de funciones cóncavas (convexas) que sí te da una función cóncava (convexa). Por ejemplo, sabemos que  $f(x) = x^3$  y  $g(x) = x$  son ambas funciones cuasicóncavas (y cuasiconvexas), pero  $h(x) = x^3 - x$  no es cuasicóncava

## Capítulo 4 Funciones cóncavas y cuasicóncavas

ni cuasiconvexa.

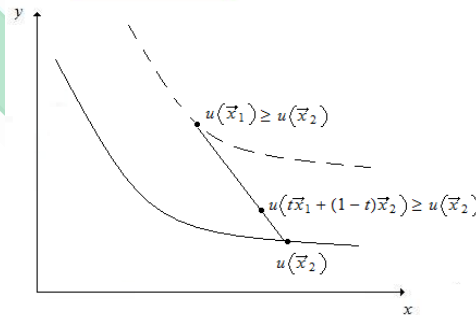


Por último, existen otras maneras de caracterizar la cuasiconcavidad o cuasiconvexidad de funciones sin utilizar el concepto de contorno. La primera de ellas representa una definición alternativa, utilizada frecuentemente en textos de microeconomía. Aquí la presentamos como un teorema, ya que puede deducirse a partir de nuestra previa definición. Observa que no involucra el concepto de diferenciabilidad, por lo que es válida para funciones continuas en general.

**Teorema.** Sea  $f : S \rightarrow \mathbb{R}$ , con  $S \subset \mathbb{R}^n$  convexo. Entonces,  $f$  es cuasicóncava en  $S$  si y sólo si, para todos  $\vec{x}_1, \vec{x}_2 \in S$  y para todo  $t \in [0, 1]$ , se cumple:

- a)  $f(\vec{x}_1) \geq f(\vec{x}_2) \implies f(t\vec{x}_1 + (1-t)\vec{x}_2) \geq f(\vec{x}_2)$ .
- b)  $f(t\vec{x}_1 + (1-t)\vec{x}_2) \geq \min\{f(\vec{x}_1), f(\vec{x}_2)\}$ .

Para comprender el significado geométrico de este teorema, considera las preferencias de un consumidor. Si  $u(\vec{x})$  es la utilidad correspondiente a la canasta  $\vec{x}$  de dos bienes ( $u$  crece al alejarse del origen), y si para cualesquiera dos canastas  $\vec{x}_1, \vec{x}_2$  se cumple  $u(\vec{x}_1) \geq u(\vec{x}_2)$ , entonces para cualquier canasta intermedia  $t\vec{x}_1 + (1-t)\vec{x}_2$ ,  $t \in [0, 1]$  se tendrá  $u(t\vec{x}_1 + (1-t)\vec{x}_2) \geq u(\vec{x}_2)$ .



Por otra parte, para funciones  $f$  doblemente diferenciables en su dominio  $S$  existe un criterio simple para determinar su cuasiconcavidad o cuasiconvexidad. Éste se refiere al signo de la matriz que se obtiene al añadir al hessiano  $H$  una columna y un renglón conteniendo el gradiente de la función  $f$ . Esa matriz



### 4.3 Funciones cuasicóncavas y funciones cuasiconvexas

ampliada se conoce como *hessiano orlado*,  $\overline{H}$ , que en el caso de funciones  $f : S \subset \mathbb{R}^2, f \in C^2(S)$  se define como la matriz

$$\overline{H} = \begin{pmatrix} 0 & f_x & f_y \\ f_x & f_{xx} & f_{xy} \\ f_y & f_{xy} & f_{yy} \end{pmatrix}.$$

En ese caso, es posible demostrar que el criterio de signos correspondiente es

$$\begin{aligned} |\overline{H}| &> 0 \Rightarrow f \text{ es cuasicóncava,} \\ |\overline{H}| &< 0 \Rightarrow f \text{ es cuasiconvexa.} \end{aligned}$$

Utilizando este resultado, es fácil verificar que las funciones tipo Cobb-Douglas,  $f(x, y) = x^\alpha y^\beta$ ,  $\alpha, \beta > 0$ , son cuasicóncavas en general; de éstas, sólo son cóncavas las que satisfacen  $\alpha + \beta \leq 1$ .

# Capítulo 5

## Optimización

En este capítulo aplicaremos los resultados sobre concavidad del capítulo 4 para encontrar los máximos y mínimos de una función  $f$  definida en un dominio convexo  $S$ . Este dominio puede ser simplemente el dominio natural de la función, o bien, la región que resulte al imponer una colección de restricciones. En el primer caso, hablaremos de problemas de optimización libre, que estudiaremos en la sección 5.1. En el segundo caso, hablaremos de problemas de optimización restringida, que presentaremos en las secciones 5.2 y 5.3. Por simplicidad, gran parte de la discusión se limitará al caso de funciones de dos variables,  $f(x, y)$ .

### 5.1 Optimización libre. Criterio del Hessiano

A los valores máximos o mínimos de una función también se les denomina *valores extremos* de la función. Su definición en el caso de una función de varias variables es una generalización directa de las correspondientes a una función de una variable.

**Definición.** Sea  $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$  y sea  $\vec{x}_0 \in S$ . Se dice que

- a)  $f$  presenta un *máximo local o relativo* en  $\vec{x}_0$ , si existe una vecindad  $V$  de  $\vec{x}_0$  tal que  $f(\vec{x}_0) \geq f(\vec{x})$ , para todo  $\vec{x} \in V \cap S$ .
- b)  $f$  presenta un *máximo global o absoluto* en  $\vec{x}_0$ , si  $f(\vec{x}_0) \geq f(\vec{x})$  para todo  $\vec{x} \in S$ .

Cuando las desigualdades  $\geq$  se cambian por  $>$  se dice que el máximo (local o global) es *estricto o único*.

A  $\vec{x}_0$  se le llama el *punto máximo* de  $f$  y a  $f(\vec{x}_0)$  el *valor máximo* de  $f$ . Decimos que  $\vec{x}_0$  es un máximo local de  $f$  si no existen puntos cercanos en los que  $f$  alcance un valor mayor (puede ser igual); éste es estricto, si los puntos cercanos dan un valor menor que  $f(\vec{x}_0)$ . Asimismo,  $\vec{x}_0$  es un máximo global de  $f$  si en todo el dominio  $S$  no existe otro punto en el que  $f$  alcance un valor mayor; éste es estricto, si cualquier otro punto del dominio da un valor menor que  $f(\vec{x}_0)$ .

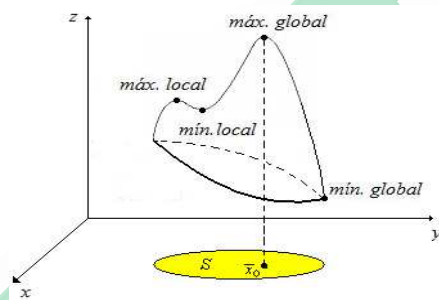
## 5.1 Optimización libre. Criterio del Hessiano

Invirtiendo las desigualdades anteriores se llega al concepto de punto mínimo, como se define a continuación.

**Definición.** Sea  $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$  y sea  $\vec{x}_0 \in S$ . Entonces,

- a)  $f$  presenta un *mínimo local o relativo* en  $\vec{x}_0$ , si existe una vecindad  $V$  de  $\vec{x}_0$  tal que  $f(\vec{x}_0) \leq f(\vec{x})$ , para todo  $\vec{x} \in V \cap S$ .
- b)  $f$  presenta un *mínimo global o absoluto* en  $\vec{x}_0$ , si  $f(\vec{x}_0) \leq f(\vec{x})$  para todo  $\vec{x} \in S$ .

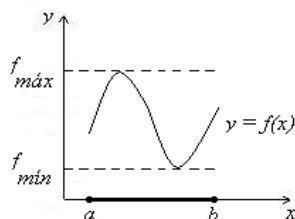
Cuando las desigualdades  $\leq$  se cambian por  $<$  se dice que el mínimo (local o global) es *estricto o único*.



Nota que todo extremo global es también un extremo local, pero no viceversa.

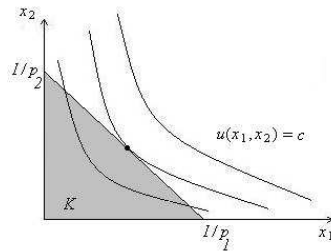
No toda función alcanza sus valores extremos globales (máximo o mínimo) en su dominio. Demostrar la existencia de esos extremos globales puede resultar bastante complejo en general, especialmente si la función no es continua, diferenciable, o su dominio no es convexo. El siguiente teorema establece una condición suficiente para garantizar su existencia.

**Teorema de Weierstrass.** Sea  $S \subset \mathbb{R}^n$  un conjunto compacto y sea  $f : S \rightarrow \mathbb{R}$ . Si  $f$  es continua en  $S$ , entonces  $f$  alcanza sus valores máximo global y mínimo global en  $S$ .



## Capítulo 5 Optimización

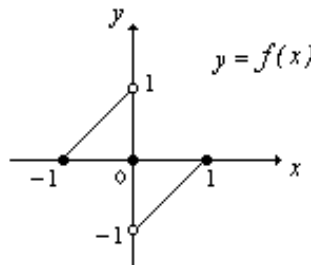
Por ejemplo, sea  $S = \{(x_1, x_2) \in \mathbb{R}_+^2 \mid p_1 x_1 + p_2 x_2 \leq I\}$  el conjunto de canastas compatibles con un ingreso menor o igual a  $I$ , para dos bienes con precios unitarios  $p_1, p_2$ . Este conjunto es cerrado y acotado, y por tanto, compacto. Si las preferencias del consumidor están dadas por una función continua  $u(x_1, x_2)$ , el teorema garantiza la existencia de una canasta que maximiza su utilidad.



Si el dominio de  $f$  no fuera compacto (es decir, si no fuera cerrado y acotado) o si  $f$  no fuera una función continua en su dominio, entonces no se garantiza que  $f$  alcance sus valores extremos globales en su dominio. Para ilustrar esta afirmación, considera la función discontinua  $f : [-1, 1] \rightarrow \mathbb{R}$ , dada por

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

En la figura se observa que esta función no alcanza sus valores máximo global ( $f_{\max} = 1$ ) y mínimo global ( $f_{\min} = -1$ ) en el intervalo  $-1 \leq x \leq 1$ .



Para encontrar los puntos extremos locales y globales de una función diferenciable es importante analizar sus propiedades de primer y segundo orden, dadas por su gradiente y su concavidad, como se discute en las siguientes subsecciones.

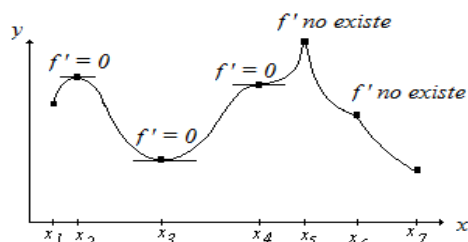
### 5.1.1 Condiciones necesarias de primer orden

Antes de iniciar la búsqueda de los valores extremos (locales y globales) de una función de varias variables, conviene recordar el procedimiento correspondiente a una función continua  $f : S \subset \mathbb{R} \rightarrow \mathbb{R}$  de una variable, con  $y = f(x)$ .

## 5.1 Optimización libre. Criterio del Hessiano

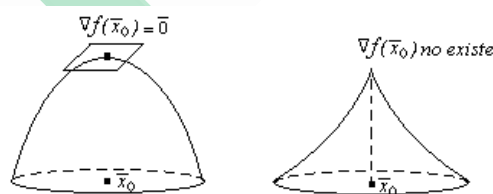
Como se ilustra en la siguiente figura, en este caso los candidatos a extremos son:

1. Los puntos interiores en donde  $f' = 0$  (puntos  $x_2, x_3$  y  $x_4$ ),
2. Los puntos interiores en donde  $f'$  no existe (puntos  $x_5$  y  $x_6$ ),
3. Los puntos frontera del dominio  $S$  (puntos  $x_1$  y  $x_7$ ).



Los puntos interiores del tipo 1 y 2 se conocen como los *puntos críticos* de  $f$ . En particular, cuando el dominio  $S$  es un conjunto abierto no existen puntos frontera, de modo que los únicos candidatos a óptimos son los puntos críticos. Adicionalmente, si  $f$  es diferenciable los únicos puntos críticos son los del tipo 1, que son aquellos en donde la derivada es cero, es decir, en donde la recta tangente a la curva  $y = f(x)$  es horizontal. Nota que los puntos críticos son sólo candidatos a óptimos, ya que no todos ellos dan origen a un extremo local, como es el caso de los puntos  $x_4$  y  $x_6$  en la figura. A continuación presentamos la generalización de estos resultados para funciones de varias variables.

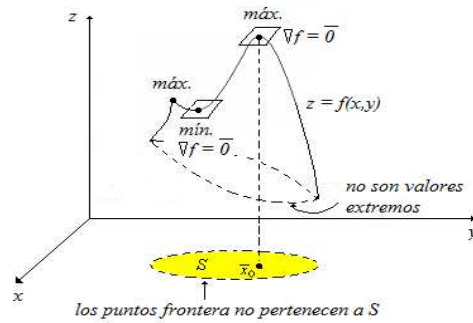
**Definición.** Sea  $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$  una función continua. Un punto interior  $\vec{x}_0 \in S$  se llama *punto crítico* de  $f$  si  $\nabla f(\vec{x}_0) = \vec{0}$  o  $\nabla f(\vec{x}_0)$  no existe.



En otras palabras, los puntos críticos de una función continua  $f$  son aquellos puntos interiores en donde el plano tangente a su gráfica es horizontal ( $\nabla f = \vec{0}$ ) o en donde ese plano tangente no existe (picos o cúspides, en donde  $\nabla f$  no está definido).

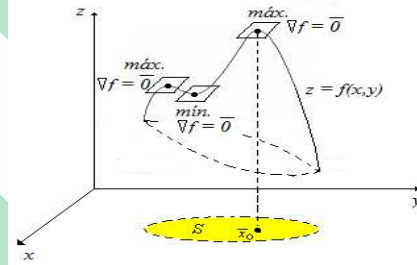
Cuando la función continua  $f$  está definida en un conjunto abierto  $S$ , la frontera de  $S$  no pertenece al dominio de la función, de modo que los únicos candidatos a máximos y mínimos de  $f$  son sus puntos críticos, como establece a continuación.

**Teorema (condiciones necesarias para un extremo local de funciones continuas).** Sea  $S \subset \mathbb{R}^n$  un conjunto abierto y sea  $f : S \rightarrow \mathbb{R}$  continua. Si  $f$  posee un extremo local en  $\vec{x}_0 \in S$ , entonces  $\vec{x}_0$  es un punto crítico de  $f$ .



En el caso particular de que  $f$  sea diferenciable en el abierto  $S$ , los candidatos a máximos y mínimos de  $f$  se reducen sólo a aquellos puntos en donde  $\nabla f = \vec{0}$ , como se establece a continuación.

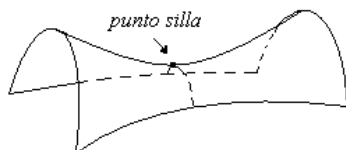
**Teorema (condiciones necesarias de primer orden para un extremo local de funciones diferenciables).** Si  $f$  es diferenciable en un punto interior  $\vec{x}_0$  de su dominio y  $f$  alcanza un extremo local en  $\vec{x}_0$ , entonces  $\nabla f(\vec{x}_0) = \vec{0}$ .



Es importante señalar que no todo punto crítico  $\vec{x}_0$  de  $f$  tal que  $\nabla f(\vec{x}_0) = \vec{0}$  es un extremo local, como es el caso de los puntos de inflexión para funciones de una variable. Estos últimos se denominan *puntos silla* en el caso multidimensional, como se define a continuación.

**Definición.** Una función  $f$  tiene un *punto silla* en un punto crítico  $\vec{x}_0$  de su dominio si para toda vecindad  $V$  de  $\vec{x}_0$  existen puntos  $\vec{x} \in V$  tales que  $f(\vec{x}) > f(\vec{x}_0)$  y puntos  $\vec{x}' \in V$  tales que  $f(\vec{x}') < f(\vec{x}_0)$ . En otras palabras, un punto silla es un punto crítico que no es un punto de extremo local.

## 5.1 Optimización libre. Criterio del Hessiano

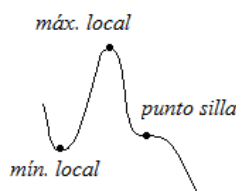


### 5.1.2 Condiciones suficientes de segundo orden

La condición  $\nabla f = \vec{0}$  para funciones diferenciables es una *condición necesaria* de primer orden para obtener los puntos críticos en la optimización libre de  $f$ . Para clasificar cada punto crítico como máximo, mínimo o punto silla, se debe analizar la concavidad o convexidad local de la función  $f$  en cada uno de ellos. En el caso particular de funciones doblemente diferenciables, el procedimiento correspondiente se simplifica considerablemente analizando el signo de la matriz hessiana  $H$ , definida en la sección 4.1. Este criterio constituye una *condición suficiente* de segundo orden para la optimización libre de una función, como se presenta a continuación.

#### Valores extremos locales

El criterio para clasificar los extremos locales de una función doblemente diferenciable es muy simple. La idea consiste en analizar la concavidad o convexidad local de la función  $f$  en cada punto crítico  $\vec{x}_0$ : si  $f$  es convexa en  $\vec{x}_0$ , se trata de un mínimo local; si  $f$  es cóncava en  $\vec{x}_0$ , se trata de un máximo local; si  $f$  no es cóncava ni convexa en  $\vec{x}_0$  se trata de un punto silla.



La concavidad de la función puede determinarse analizando el signo de la matriz hessiana  $H$  de  $f$ , cuyos elementos de matriz son las segundas derivadas parciales de la función  $f$  con respecto a sus variables independientes. En el caso simple de una función  $z = f(x, y)$  con dominio en  $\mathbb{R}^2$ , la matriz hessiana correspondiente está dada por

$$H(x, y) = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix},$$

para la que se establecen las siguientes condiciones suficientes de segundo orden.

## Capítulo 5 Optimización

**Teorema (condiciones suficientes de segundo orden).** Sea  $S \subset \mathbb{R}^2$  un conjunto abierto y convexo, y sea  $f : S \rightarrow \mathbb{R}$ , con  $f \in C^2(S)$ . Sea  $\vec{x}_0 \in S$  un punto crítico de  $f$ . Entonces

- a)  $f_{xx} > 0$  y  $f_{xx}f_{yy} - f_{xy}^2 > 0$  en  $\vec{x}_0$   
 $\Rightarrow f$  tiene un mínimo local estricto en  $\vec{x}_0$ ,
- b)  $f_{xx} < 0$  y  $f_{xx}f_{yy} - f_{xy}^2 > 0$  en  $\vec{x}_0$   
 $\Rightarrow f$  tiene un máximo local estricto en  $\vec{x}_0$ ,
- c)  $f_{xx}f_{yy} - f_{xy}^2 < 0$  en  $\vec{x}_0 \Rightarrow f$  tiene un punto silla en  $\vec{x}_0$ .

Nota que las condiciones a) y b) se refieren al caso de máximos o mínimos estrictos, es decir, en donde el valor extremo local es único. Más adelante discutiremos el caso de una matriz hessiana semidefinida, en donde los extremos locales no necesariamente son únicos.

Ejemplos:

1. Encuentra y clasifica los puntos críticos de  $f(x, y) = 6x^2 - 2x^3 + 3y^2 + 6xy$ .

Las condiciones de primer orden para  $f$  son

$$\begin{aligned}f_x(x, y) &= 12x - 6x^2 + 6y = 0, \\f_y(x, y) &= 6y + 6x = 0.\end{aligned}$$

Resolviendo este sistema de ecuaciones se obtiene que los puntos críticos son  $(0, 0)$  y  $(1, -1)$ . Para clasificarlos, establecemos las condiciones de segundo orden, a través del hessiano

$$H(x, y) = \begin{pmatrix} 12 - 12x & 6 \\ 6 & 6 \end{pmatrix}.$$

Para el punto  $(0, 0)$  se tiene

$$H(0, 0) = \begin{pmatrix} 12 & 6 \\ 6 & 6 \end{pmatrix}.$$

Como  $f_{xx} = 12 > 0$  y  $|H| = 36 > 0$ , por lo tanto  $f$  tiene un mínimo local estricto en  $(0, 0)$ . A su vez, para el punto  $(1, -1)$  se tiene

$$H(1, -1) = \begin{pmatrix} 0 & 6 \\ 6 & 6 \end{pmatrix}.$$

Como  $|H| = -36 < 0$ , por lo tanto  $f$  tiene un punto silla en  $(1, -1)$ .



## 5.1 Optimización libre. Criterio del Hessiano

2. Encuentra y clasifica los puntos críticos de  $f(x, y) = xy - x^2 - y^2 - 2x - 2y + 4$ .

Las condiciones de primer orden para  $f$  son

$$\begin{aligned}f_x(x, y) &= y - 2x - 2 = 0, \\f_y(x, y) &= x - 2y - 2 = 0.\end{aligned}$$

Resolviendo el sistema de ecuaciones se obtiene que el único punto crítico es  $(-2, -2)$ . El hessiano correspondiente es

$$H(x, y) = \begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix} = H(-2, -2).$$

Como  $f_{xx} = -2 < 0$  y  $|H| = 3 > 0$ , por lo tanto  $f$  tiene un máximo local estricto en  $(-2, -2)$ . Como veremos más adelante, este máximo local también es un máximo global de la función.

3. Encuentra y clasifica los puntos críticos de  $f(x, y) = x^5y + xy^5 + xy$ .

Las condiciones de primer orden para  $f$  son

$$\begin{aligned}f_x(x, y) &= 5x^4y + y^5 + y = y(5x^4 + y^4 + 1) = 0, \\f_y(x, y) &= x^5 + 5xy^4 + x = x(x^4 + 5y^4 + 1) = 0.\end{aligned}$$

Resolviendo el sistema de ecuaciones se obtiene que el único punto crítico es  $(0, 0)$ . El hessiano correspondiente es

$$H(x, y) = \begin{pmatrix} 20x^3y & 5x^4 + 5y^4 + 1 \\ 5x^4 + 5y^4 + 1 & 20xy^3 \end{pmatrix},$$

de modo que en el punto  $(0, 0)$  se tiene

$$H(0, 0) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Como  $|H| = -1 < 0$ , por lo tanto  $f$  tiene un punto silla en  $(0, 0)$ .

En el teorema anterior se ha excluido el caso en donde  $f_{xx} = 0$  con  $f_{xx}f_{yy} - f_{xy}^2 \geq 0$ , o bien  $f_{xx}f_{yy} - f_{xy}^2 = 0$ . En este último caso no es posible garantizar de que el extremo sea estricto, sino que puede (o no) corresponder a un problema de solución múltiple. Por esa razón, las anteriores condiciones suficientes de segundo orden deben reemplazarse por un conjunto de condiciones necesarias, como se enuncia en el siguiente teorema.

**Teorema (condiciones necesarias de segundo orden).** Sea  $S \subset \mathbb{R}^2$  un conjunto abierto y convexo, y sea  $f : S \rightarrow \mathbb{R}$ , con  $f \in C^2(S)$ . Sea  $\vec{x}_0 \in S$  un punto crítico de  $f$ . Entonces

- a)  $f$  tiene un mínimo local en  $\vec{x}_0$   
 $\Rightarrow f_{xx} \geq 0, f_{yy} \geq 0$  y  $f_{xx}f_{yy} - f_{xy}^2 \geq 0$  en  $\vec{x}_0$ ,
- b)  $f$  tiene un máximo local en  $\vec{x}_0$   
 $\Rightarrow f_{xx} \leq 0, f_{yy} \leq 0$  y  $f_{xx}f_{yy} - f_{xy}^2 \geq 0$  en  $\vec{x}_0$ .

Ejemplo:

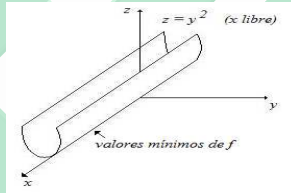
Encuentra y clasifica los puntos críticos de  $f(x, y) = y^2$ .

Para la función  $f(x, y) = y^2$  las condiciones de primer orden son

$$f_x(x, y) = 0$$

$$f_y(x, y) = 2y = 0,$$

de modo que los puntos críticos son todos los puntos del eje  $x$ , es decir,  $\{(x, y) \in \mathbb{R}^2 \mid y = 0\}$ . La siguiente figura muestra la gráfica de la superficie  $z = y^2$ , que corresponde a un cilindro parabólico que corre a lo largo del eje  $x$ . Ahí se observa que la función  $f(x, y) = y^2$  alcanza su valor mínimo (no único) a lo largo de todos los puntos del eje  $x$ .



Este resultado es consistente con las condiciones necesarias de segundo orden para la matriz hessiana correspondiente,

$$H = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix},$$

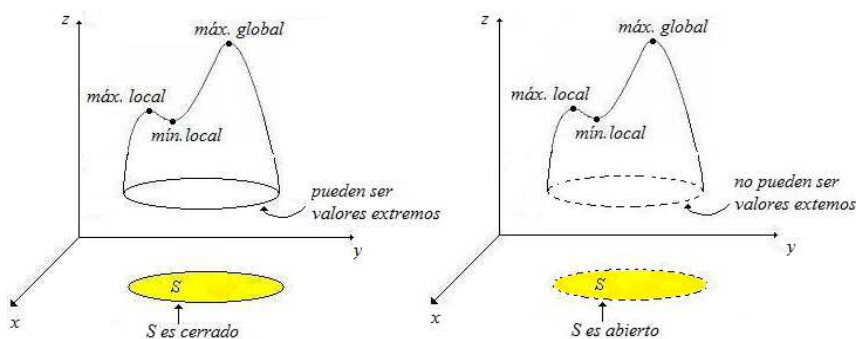
en donde  $f_{xx} = 0 \geq 0$ ,  $f_{yy} = 2 \geq 0$  y  $f_{xx}f_{yy} - f_{xy}^2 = 0 \geq 0$ .

### Valores extremos globales

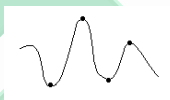
La búsqueda de los valores extremos globales de una función puede resultar bastante complicada en general, aun en el caso de funciones diferenciables. Una de esas dificultades puede surgir al tratar de determinar los extremos globales de una función cuyo dominio  $S$  sea un conjunto cerrado, ya que además de los puntos críticos interiores se tiene que considerar el valor de la función a lo largo de todos sus puntos frontera. Omitiremos aquí un ejemplo de este tipo, aunque en la sección

## 5.1 Optimización libre. Criterio del Hessiano

5.3 resolveremos problemas de optimización en regiones cerradas utilizando técnicas de optimización restringida (condiciones de Kuhn-Tucker). De este modo, en esta sección consideraremos sólo funciones definidas en conjuntos abiertos.



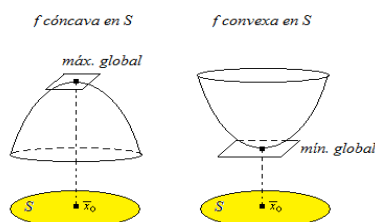
Otra dificultad surge cuando el número de extremos locales de una función es tan grande que puede resultar bastante engorroso, o incluso imposible, determinar cuáles de estos corresponden a sus valores extremos globales.



Para este caso no se cuenta con condiciones necesarias de segundo orden para encontrar los extremos globales. Sin embargo, si tú sabes de antemano que una función es cóncava (convexa) a lo largo de todo su dominio, y que ésta posee un punto crítico, este solo hecho es suficiente para garantizar que la función posee un máximo (mínimo) global.

**Teorema (condiciones suficientes).** Sea  $S \subset \mathbb{R}^n$  un conjunto abierto y convexo, y sea  $f : S \rightarrow \mathbb{R}$ . Sea  $\vec{x}_0 \in S$  un punto crítico de  $f$ . Entonces

- $f$  convexa en  $S \Rightarrow f$  tiene un mínimo global en  $\vec{x}_0$ ,
- $f$  estrictamente convexa en  $S \Rightarrow f$  tiene un mínimo global estricto en  $\vec{x}_0$ ,
- $f$  cóncava en  $S \Rightarrow f$  tiene un máximo global en  $\vec{x}_0$ ,
- $f$  estrictamente cóncava en  $S \Rightarrow f$  tiene un máximo global estricto en  $\vec{x}_0$ .



## Capítulo 5 Optimización

Este teorema es válido aun el caso general de funciones no diferenciables. Si la función es doblemente diferenciable, entonces su concavidad o convexidad puede determinarse directamente a partir de la matriz hessiana, como se muestra a continuación.

Ejemplos:

1. Demuestra que la función  $f(x, y) = -x^2 - xy - y^2 - 3y$  presenta un máximo global estricto en el punto  $(1, -2)$ .

Las condiciones de primer orden para  $f$  son

$$\begin{aligned}f_x(x, y) &= -2x - y = 0 \\f_y(x, y) &= -x - 2y - 3 = 0,\end{aligned}$$

que se satisfacen cuando  $x = 1$  y  $y = -2$ . Así, el único punto crítico de  $f$  es el  $(1, -2)$ . La matriz Hessiana de  $f$  es

$$H(x, y) = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix} = \begin{pmatrix} -2 & -1 \\ -1 & -2 \end{pmatrix}.$$

Como  $f_{xx} = -2 < 0$  y  $|H| = 3 > 0$ , para todo  $(x, y)$ , por lo tanto  $f$  es estrictamente cóncava en  $\mathbb{R}^2$ . Así,  $f$  tiene un máximo global único en  $(1, -2)$ .

2. Encuentra los extremos globales de la función  $f(x, y) = x^4 + y^4$ .

Las condiciones de primer orden para  $f$  son

$$\begin{aligned}f_x(x, y) &= 4x^3 = 0 \\f_y(x, y) &= 4y^3 = 0,\end{aligned}$$

que se satisfacen cuando  $x = y = 0$ . Así, el único punto crítico de  $f$  es el  $(0, 0)$ . La matriz Hessiana de  $f$  está dada por

$$H(x, y) = \begin{pmatrix} 12x^2 & 0 \\ 0 & 12y^2 \end{pmatrix}.$$

En el punto crítico, se tiene

$$H(0, 0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

que no presenta un signo definido, ni viola los criterios de concavidad/convexidad. Así, la prueba de los signos de  $H(0, 0)$  no es concluyente. En lugar de esto, para clasificar el punto crítico podemos utilizar el siguiente argumento alternativo, basado en la imagen de la función.

## 5.2 Optimización con restricciones de igualdad. Multiplicadores de Lagrange

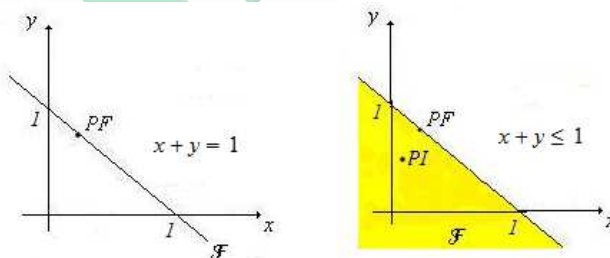
Sabemos que  $f(0,0) = 0$  y observamos que la imagen de  $f$  son los valores  $z \geq 0$ . De esta manera, concluimos que  $f$  tiene un mínimo global estricto en  $(0,0)$ .

Por último, en el apéndice B se presenta la generalización de los resultados de esta sección para funciones de varias variables.

### 5.2 Optimización con restricciones de igualdad. Multiplicadores de Lagrange

En muchas aplicaciones de interés la optimización de una función objetivo está sujeta a restricciones sobre las variables independientes o variables de decisión, como es el caso de restricciones presupuestales, laborales u operativas. Estas restricciones son un subconjunto del dominio de la función, conocido como la *región factible*  $\mathcal{F}$ .

En el caso de restricciones de igualdad, la región factible  $\mathcal{F}$  está formada solamente por puntos frontera, mientras que en el caso de restricciones de desigualdad,  $\mathcal{F}$  está constituida tanto por puntos frontera como por puntos interiores, como se ilustra en las siguientes figuras para restricciones en  $\mathbb{R}^2$ . La figura de la izquierda muestra la restricción de igualdad  $x + y = 1$ , en donde la recta  $\mathcal{F}$  está formada sólo por puntos frontera ( $PF$ ); la figura de la derecha muestra la restricción de desigualdad  $x + y \leq 1$ , en donde el semiplano  $\mathcal{F}$  está constituido por puntos frontera ( $PF$ ) y puntos interiores ( $PI$ ).



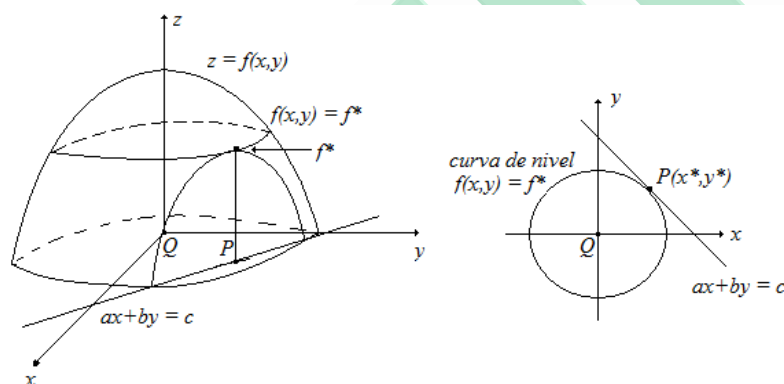
El método para encontrar la solución óptima es distinto en cada caso. En esta sección nos dedicaremos solamente al problema de optimización con restricciones de igualdad, conocido como el *método de multiplicadores de Lagrange*. El caso correspondiente a restricciones de desigualdad se tratará en la sección 5.3, en donde se incorporarán al método de Lagrange las llamadas *condiciones de Kuhn-Tucker*.

### 5.2.1 Condiciones necesarias de primer orden. Significado del multiplicador de Lagrange

Primero consideremos el caso simple de maximización de una función diferenciable de dos variables,  $f(x, y)$ , sujeto a una restricción de igualdad,  $g(x, y) = c$ , con  $g$  diferenciable y  $c$  constante, dado por

$$\begin{aligned} \text{máx. } & f(x, y) \\ \text{s.a. } & g(x, y) = c. \end{aligned}$$

Debido a la restricción, la solución óptima  $P(x^*, y^*)$  no necesariamente ocurre en los puntos en donde la superficie  $z = f(x, y)$  alcanza su altura máxima ( $\nabla f = \vec{0}$ ), sino en los puntos de la curva  $g(x, y) = c$  sobre los que  $f$  alcanza su máximo valor restringido. Para ilustrar esta idea, la siguiente figura muestra la maximización de una función cóncava  $z = f(x, y)$  en  $\mathbb{R}^3$  sujeta a una restricción lineal  $g(x, y) = ax + by = c$  en  $\mathbb{R}^2$ .



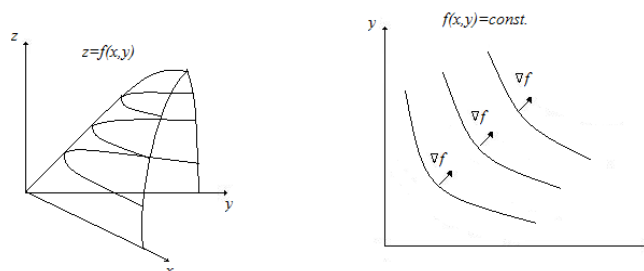
En la figura de la izquierda se observa que el máximo restringido  $f^*$  de la función  $f$  ocurre en el punto  $P(x^*, y^*)$  de la recta  $ax + by = c$ , y no en el punto  $Q$  en donde  $f$  se maximiza libremente. Nota que  $P$  pertenece a la curva de nivel  $f(x, y) = f^*$  correspondiente a  $z = f^*$ . En la figura de la derecha se muestra que el óptimo  $P$  es el punto de la restricción  $ax + by = c$  que está más cercano a  $Q$ . Esto ocurre en el punto de tangencia de la curva de restricción y la curva de nivel  $f(x, y) = f^*$ .

En general pueden existir varios candidatos a óptimos (locales o globales) para una función  $f$ , dados por los puntos donde la restricción  $g(x, y) = c$  es tangente a las curvas de nivel de  $f$ . Esta condición de tangencia puede expresarse formalmente en términos de los vectores gradiente de las funciones  $f$  y  $g$ , como se explica a continuación.

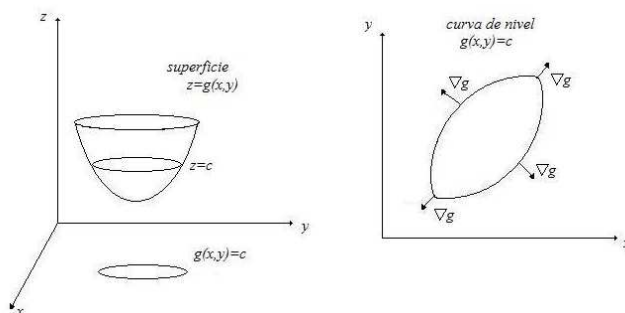
Para este fin, recordemos que el gradiente de una función diferenciable es un vector perpendicular a sus curvas de nivel y apunta en la dirección de su

## 5.2 Optimización con restricciones de igualdad. Multiplicadores de Lagrange

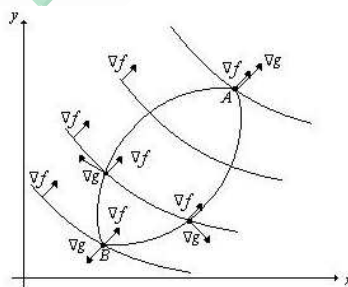
mayor crecimiento en cada punto. La siguiente figura muestra una posible función objetivo  $z = f(x, y)$  en  $\mathbb{R}^3$ . La figura de la derecha muestra algunas de sus curvas de nivel, en  $\mathbb{R}^2$ , y la dirección de los vectores gradiente  $\nabla f$ .



Por otra parte, la curva de restricción  $g(x, y) = c$  en  $\mathbb{R}^2$  puede considerarse como la curva de nivel  $z = c$  de una función  $z = g(x, y)$  en  $\mathbb{R}^3$ , como se ilustra en las siguientes figuras. En este ejemplo, el vector gradiente  $\nabla g$  apunta hacia afuera de la curva de nivel.



Así, el problema de maximización de  $f(x, y)$  sujeto a la restricción  $g(x, y) = c$  se representa gráficamente de la siguiente manera:



Aquí existen dos candidatos a óptimo, que son los puntos de tangencia denotados por  $A$  y  $B$ . La condición de tangencia implica que, en esos puntos, los vectores gradiente  $\nabla f$  y  $\nabla g$  son paralelos entre sí, es decir,

$$\nabla f \parallel \nabla g.$$

## Capítulo 5 Optimización

Por lo tanto, en los puntos en donde  $f$  alcanza sus valores extremos debe existir un número  $\lambda \in \mathbb{R}$  tal que

$$\nabla f = \lambda \nabla g.$$

El número  $\lambda$  se denomina el *multiplicador de Lagrange* asociado con la restricción  $g(x, y) = c$ . Aunque aquí  $\lambda$  juega el papel de una constante de proporcionalidad entre  $\nabla f$  y  $\nabla g$  en el óptimo, también presenta una interpretación muy interesante y útil, como discutiremos en breve.

Por lo general, en el óptimo restringido  $P$  de  $f$  se tiene  $\nabla g \neq \vec{0}$ , con  $\lambda \neq 0$ . Como  $\nabla f = \lambda \nabla g$  en  $P$ , en ese punto se tiene

$$\nabla f \neq \vec{0}.$$

Así, la condición  $\nabla f = \vec{0}$  para optimización libre, aquí deberá reemplazarse por las siguientes dos condiciones:

$$\begin{aligned}\nabla f &= \lambda \nabla g \\ g(x, y) &= c.\end{aligned}$$

Estas dos ecuaciones pueden conjuntarse dentro de un formalismo más elegante, de la siguiente manera. Para ello, reescribimos los gradientes de la primera ecuación en términos de sus componentes  $x$  y  $y$ , obteniendo

$$\begin{aligned}f_x(x, y) &= \lambda g_x(x, y) \\ f_y(x, y) &= \lambda g_y(x, y) \\ g(x, y) &= c.\end{aligned}$$

Éste es un sistema de 3 ecuaciones con 3 incógnitas, con solución  $x^*$ ,  $y^*$  y  $\lambda^*$ . Nota que esta solución no es el punto crítico de la función objetivo  $f(x, y)$ , ya que  $\nabla f = \lambda \nabla g \neq \vec{0}$ . Sin embargo,  $(x^*, y^*, \lambda^*)$  puede interpretarse como el punto crítico de una cierta función de las variables  $(x, y, \lambda)$ , a la que denominaremos la *función lagrangeana*,  $\mathcal{L}(x, y, \lambda)$ , definida como

$$\mathcal{L}(x, y, \lambda) = f(x, y) + \lambda(c - g(x, y)).$$

Nota que la función  $\mathcal{L}$  habita en un espacio de dimensión mayor que  $f$ , ya que no sólo tiene como variables independientes a  $x$  y  $y$ , sino también a  $\lambda$ . De esta manera, en lugar de considerar la optimización restringida de  $f$ , el método de Lagrange se basa en la optimización libre de la función lagrangeana, cuyas condiciones de primer orden son:

$$\begin{aligned}\mathcal{L}_x &= f_x(x, y) - \lambda g_x(x, y) = 0 \\ \mathcal{L}_y &= f_y(x, y) - \lambda g_y(x, y) = 0 \\ \mathcal{L}_\lambda &= c - g(x, y) = 0.\end{aligned}$$



## 5.2 Optimización con restricciones de igualdad. Multiplicadores de Lagrange

Estas tres ecuaciones representan la condición de tangencia,  $\nabla f = \lambda \nabla g$ , y el cumplimiento de la restricción,  $g(x, y) = c$ , antes discutidas.

**Teorema de Lagrange (condiciones necesarias de primer orden).** Sean  $f, g : S \rightarrow \mathbb{R}$  diferenciables en  $S \subset \mathbb{R}^2$  y sea  $(x^*, y^*) \in S$  una solución del problema

$$\begin{aligned} &\text{máx./mín. } f(x, y) \\ &\text{s.a. } g(x, y) = c. \end{aligned}$$

Si  $\nabla g(x^*, y^*) \neq \vec{0}$ , entonces existe  $\lambda^* \in \mathbb{R}$  tal que  $(x^*, y^*, \lambda^*)$  es un punto crítico de la función lagrangeana

$$\mathcal{L}(x, y, \lambda) = f(x, y) + \lambda(c - g(x, y)),$$

es decir, en ese punto  $\mathcal{L}_x = \mathcal{L}_y = \mathcal{L}_\lambda = 0$ .

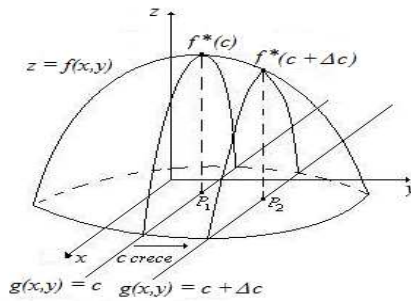
La condición  $\nabla g(x^*, y^*) \neq \vec{0}$  establece que  $(x^*, y^*)$  no debe ser un punto crítico de  $g$ , con el fin de que se cumpla la condición de tangencia  $\nabla f = \lambda \nabla g$  con  $\nabla f(x^*, y^*) \neq \vec{0}$ . Cuando  $(x^*, y^*)$  es un punto crítico de  $g$  el método de Lagrange puede fallar, como se discute en la sección 5.2.4.

### Significado del multiplicador $\lambda$

A cada valor del parámetro  $c$  le corresponde un punto óptimo,  $P(x^*(c), y^*(c))$ . En consecuencia, el valor óptimo  $f^*$  de la función  $f$ ,

$$f^*(c) = f(x^*(c), y^*(c)),$$

también depende de  $c$ , como se muestra en la siguiente figura.

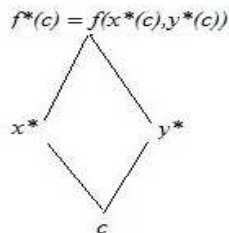


Tiene sentido, entonces, preguntarse cómo cambia el óptimo  $f^*$  cuando el parámetro  $c$  se incrementa en  $\Delta c$ . Si  $\Delta c$  es suficientemente pequeño, el cambio en  $f^*$  puede aproximarse por la derivada

$$\frac{df^*(c)}{dc} = \lim_{\Delta c \rightarrow 0} \frac{f^*(c + \Delta c) - f^*(c)}{\Delta c}.$$

## Capítulo 5 Optimización

Para encontrar esta derivada utilizamos la regla de la cadena, de acuerdo al siguiente diagrama:



Así, se tiene

$$\begin{aligned}
 \frac{df^*(c)}{dc} &= \frac{df(x^*(c), y^*(c))}{dc} \\
 &= \frac{\partial f}{\partial x^*} \frac{dx^*}{dc} + \frac{\partial f}{\partial y^*} \frac{dy^*}{dc} \dots\dots\dots(\text{regla de la cadena}) \\
 &= \left( \lambda^* \frac{\partial g}{\partial x^*} \right) \frac{dx^*}{dc} + \left( \lambda^* \frac{\partial g}{\partial y^*} \right) \frac{dy^*}{dc} \dots\dots(\text{en el óptimo, } f_x = \lambda g_x, f_y = \lambda g_y) \\
 &= \lambda^* \left( \frac{\partial g}{\partial x^*} \frac{dx^*}{dc} + \frac{\partial g}{\partial y^*} \frac{dy^*}{dc} \right) \\
 &= \lambda^* \left( \frac{dg(x^*(c), y^*(c))}{dc} \right) \dots\dots\dots(\text{regla de la cadena, al revés}) \\
 &= \lambda^* (1) \dots\dots\dots(\text{en el óptimo, } g(x, y) = c) \\
 &= \lambda^*.
 \end{aligned}$$

Concluimos entonces que

$$\lambda^* = \frac{df^*(c)}{dc}.$$

De acuerdo con este resultado, el multiplicador de Lagrange  $\lambda^*$  representa la razón de cambio instantánea del valor óptimo  $f^*$  (máximo o mínimo) de la función  $f$  al cambiar el parámetro  $c$ . Éste es un caso particular del llamado *Teorema de la Envoltente*, que estudiaremos en la sección 5.4.

Es importante señalar que, para el caso de optimización con restricciones de igualdad, el multiplicador  $\lambda^*$  puede tomar cualquier signo, independientemente de que se trate de un problema de maximización o minimización. Ante un pequeño incremento de  $c$ , si  $\lambda^* > 0$  se tiene que el valor óptimo de  $f$  (ya sea el máximo, o el mínimo) se incrementa, y si  $\lambda^* < 0$  éste decrece. Este resultado contrasta con el correspondiente al caso de optimización sujeta a restricciones de desigualdad, en donde  $\lambda^*$  no puede ser negativo, como veremos en la sección 5.3 correspondiente a restricciones de desigualdad.

## 5.2 Optimización con restricciones de igualdad. Multiplicadores de Lagrange

Ejemplos:

### 1. Resuelve el problema

$$\begin{aligned} \text{máx. } f(x, y) &= 9 - x^2 - y^2 \\ \text{s.a. } x + y &= 4. \end{aligned}$$

Luego estima el valor máximo de  $f$  si se utilizara  $x + y = 4.01$  como nueva restricción.

La función lagrangeana en este caso está dada por

$$\mathcal{L}(x, y, \lambda) = 9 - x^2 - y^2 + \lambda(4 - x - y),$$

y las condiciones de primer orden correspondientes son

$$\begin{aligned} \mathcal{L}_x &= -2x - \lambda = 0 \\ \mathcal{L}_y &= -2y - \lambda = 0 \\ \mathcal{L}_\lambda &= 4 - x - y = 0. \end{aligned}$$

Al resolver este sistema de ecuaciones se obtiene que  $f$  alcanza su valor máximo,  $f^* = 1$ , en el punto  $(x^*, y^*) = (2, 2)$ , con  $\lambda^* = -4$ .

Para estimar el nuevo valor máximo de  $f$  si la restricción se modifica a  $x + y = 4.01$ , utilizamos

$$\lambda^* = \frac{df^*(c)}{dc} \simeq \frac{\Delta f^*}{\Delta c},$$

de donde

$$\Delta f^* \simeq \lambda^* \Delta c.$$

Tomamos el multiplicador  $\lambda^* = -4$  evaluado en el óptimo inicial, y notamos que  $\Delta c = 4.01 - 4 = 0.01$ , obteniendo  $\Delta f^* \simeq (-4)(0.01) = -0.04$ .

Así, al incrementarse  $c$ , de 4 a 4.01, el máximo de  $f$  disminuye ( $\lambda^* < 0$ ) aproximadamente en 0.04. De esta manera, el nuevo máximo sería  $f^* \simeq 1 + \Delta f^* = 1 - 0.04 = 0.96$ , aproximadamente.

### 2. Resuelve el siguiente problema de maximización de la función de producción $P(L, K)$ sujeto a una restricción presupuestal,

$$\begin{aligned} \text{máx. } P(L, K) &= 50L^{2/3}K^{1/3} \\ \text{s.a. } 100L + 300K &= 45\,000, \end{aligned}$$

donde  $L$  denota el trabajo y  $K$  el capital. ¿Cómo afectaría a la producción máxima un ligero incremento presupuestal a partir de 45 000?

La función lagrangeana en este caso está dada por

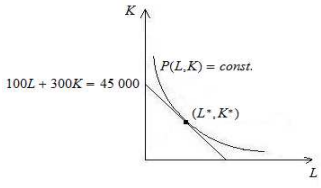
$$\mathcal{L}(L, K, \lambda) = 50L^{2/3}K^{1/3} + \lambda(45\,000 - 100L - 300K),$$

## Capítulo 5 Optimización

y las condiciones de primer orden correspondientes son

$$\begin{aligned}\mathcal{L}_L &= \frac{100}{3} \left( \frac{K}{L} \right)^{1/3} - 100\lambda = 0 \\ \mathcal{L}_K &= \frac{50}{3} \left( \frac{L}{K} \right)^{2/3} - 300\lambda = 0 \\ \mathcal{L}_\lambda &= 45\,000 - 100L - 300K = 0.\end{aligned}$$

Estas ecuaciones se conocen en economía como *condiciones de equimarginalidad*, que expresan que en el óptimo se da la tangencia de la ecuación de restricción presupuestal con alguna curva de nivel de la función de producción.



Al resolver este sistema de ecuaciones se obtiene que la producción máxima,  $P^* = 8254.8$ , se alcanza en el punto  $(L^*, K^*) = (300, 50)$ , con  $\lambda^* = 0.183\,44$ . Como  $\lambda^* > 0$ , un ligero incremento presupuestal a partir de 45 000 generaría un incremento en la producción máxima.

3. Resuelve el siguiente problema de minimización del costo  $C(L, K)$  sujeto a una restricción de producción,

$$\begin{aligned}\text{mín. } & C(L, K) = wL + rK \\ \text{s.a. } & L^{1/2}K^{1/2} = Q,\end{aligned}$$

donde  $L$  denota el trabajo,  $K$  el capital,  $w$  el salario,  $r$  la tasa de interés y  $Q$  el nivel de producción ( $w, r$  y  $Q$  constantes positivas). ¿Cómo afectaría al costo mínimo un ligero incremento en la producción a partir de  $Q$ ?

La función lagrangeana en este caso está dada por

$$\mathcal{L}(L, K, \lambda) = wL + rK + \lambda(Q - L^{1/2}K^{1/2}),$$

y las condiciones de primer orden correspondientes son

$$\begin{aligned}\mathcal{L}_L &= w - \frac{1}{2}\lambda \left( \frac{K}{L} \right)^{1/2} = 0 \\ \mathcal{L}_K &= r - \frac{1}{2}\lambda \left( \frac{L}{K} \right)^{1/2} = 0 \\ \mathcal{L}_\lambda &= Q - L^{1/2}K^{1/2} = 0.\end{aligned}$$

## 5.2 Optimización con restricciones de igualdad. Multiplicadores de Lagrange

Al resolver este sistema de ecuaciones se obtiene que el costo mínimo,  $C^* = 2Q\sqrt{wr}$ , se alcanza en el punto  $(L^*, K^*) = (Q\sqrt{\frac{r}{w}}, Q\sqrt{\frac{w}{r}})$ , con  $\lambda^* = 2\sqrt{wr}$ . Como  $\lambda^* > 0$ , un ligero incremento en la producción a partir de  $Q$  generaría un incremento en el costo mínimo.

### 5.2.2 Condiciones suficientes de segundo orden

Las condiciones  $\mathcal{L}_x = 0$ ,  $\mathcal{L}_y = 0$  y  $\mathcal{L}_\lambda = 0$ , son *condiciones necesarias* de primer orden para los niveles óptimos de una función sujeta a una restricción de igualdad. Existen ciertas condiciones bajo las cuales es posible asegurar que esos óptimos dan origen a un máximo o un mínimo de la función, conocidas como *condiciones suficientes* de segundo orden. Como se describe a continuación, los criterios correspondientes se basan en un análisis de la concavidad o convexidad de la función lagrangeana  $\mathcal{L}$  (¡no de  $f$ !), con respecto a las variables  $x$ ,  $y$  y  $\lambda$ .

Como se discutió en la sección 5.2.1, resolver el problema de optimización restringida de  $f$ ,

$$\begin{array}{ll} \text{máx.} & f(x, y) \\ \text{s.a.} & g(x, y) = c \end{array}$$

es equivalente a resolver el problema de optimización libre de la función lagrangeana asociada,

$$\mathcal{L}(x, y, \lambda) = f(x, y) + \lambda(c - g(x, y)).$$

En vista de ello, para utilizar los criterios de concavidad o convexidad inherentes a problemas de optimización libre, es claro que estos deben aplicarse a la función lagrangeana  $\mathcal{L}$  y no a la función objetivo  $f$ .

La búsqueda de extremos locales es un problema sencillo cuando  $f$  y  $g$  son doblemente diferenciables. En ese caso, es posible definir una matriz hessiana  $H_{\mathcal{L}}$  para la lagrangeana  $\mathcal{L}$  con respecto a  $\lambda$ ,  $x$  y  $y$ ,

$$H_{\mathcal{L}} = \begin{pmatrix} \mathcal{L}_{\lambda\lambda} & \mathcal{L}_{\lambda x} & \mathcal{L}_{\lambda y} \\ \mathcal{L}_{x\lambda} & \mathcal{L}_{xx} & \mathcal{L}_{xy} \\ \mathcal{L}_{y\lambda} & \mathcal{L}_{yx} & \mathcal{L}_{yy} \end{pmatrix} = \begin{pmatrix} 0 & -g_x & -g_y \\ -g_x & \mathcal{L}_{xx} & \mathcal{L}_{xy} \\ -g_y & \mathcal{L}_{yx} & \mathcal{L}_{yy} \end{pmatrix}.$$

Como se establece en el siguiente teorema, la clasificación de los puntos críticos de  $\mathcal{L}$  se basa en el signo del determinante  $|H_{\mathcal{L}}|$ ,

$$|H_{\mathcal{L}}| = \begin{vmatrix} 0 & -g_x & -g_y \\ -g_x & \mathcal{L}_{xx} & \mathcal{L}_{xy} \\ -g_y & \mathcal{L}_{yx} & \mathcal{L}_{yy} \end{vmatrix} = \begin{vmatrix} 0 & g_x & g_y \\ g_x & \mathcal{L}_{xx} & \mathcal{L}_{xy} \\ g_y & \mathcal{L}_{yx} & \mathcal{L}_{yy} \end{vmatrix},$$

evaluado en cada nivel óptimo  $(x^*, y^*, \lambda^*)$ .

## Capítulo 5 Optimización

**Teorema (condiciones suficientes para extremo local).** Considera el problema de optimización de  $f(x, y)$  sobre la restricción  $C_g = \{(x, y) \mid g(x, y) = c\}$ , con  $f$  y  $g$  funciones doblemente diferenciables en  $\mathbb{R}^2$ . Sea  $(x^*, y^*, \lambda^*)$  el punto crítico de la lagrangeana correspondiente,  $\mathcal{L}(x, y, \lambda) = f(x, y) + \lambda(c - g(x, y))$ , y sea

$$|H_{\mathcal{L}}(x^*, y^*, \lambda^*)| = \begin{vmatrix} 0 & g_x & g_y \\ g_x & \mathcal{L}_{xx} & \mathcal{L}_{xy} \\ g_y & \mathcal{L}_{yx} & \mathcal{L}_{yy} \end{vmatrix}$$

el determinante de la matriz hessiana de  $\mathcal{L}$  en  $(x^*, y^*, \lambda^*)$ . Entonces

- a)  $|H_{\mathcal{L}}(x^*, y^*, \lambda^*)| > 0 \Rightarrow (x^*, y^*)$  es un máximo local de  $f$  en  $C_g$ .
- b)  $|H_{\mathcal{L}}(x^*, y^*, \lambda^*)| < 0 \Rightarrow (x^*, y^*)$  es un mínimo local de  $f$  en  $C_g$ .

Ejemplo:

Clasifica los puntos críticos del problema de optimización

$$\begin{aligned} \text{optim. } & f(x, y) = x^2 + y^2 \\ \text{s.a. } & x^2 + xy + y^2 = 3. \end{aligned}$$

La función lagrangeana en este caso es

$$\mathcal{L}(x, y, \lambda) = x^2 + y^2 + \lambda(3 - x^2 - xy - y^2).$$

A partir de las condiciones de primer orden se obtienen 4 puntos críticos, a saber, los puntos  $(1, 1)$  y  $(-1, -1)$ , con  $\lambda^* = 2/3$ , y los puntos  $(\sqrt{3}, -\sqrt{3})$  y  $(-\sqrt{3}, \sqrt{3})$ , con  $\lambda^* = 2$ . Como

$$|H_{\mathcal{L}}(1, 1, 2/3)| = |H_{\mathcal{L}}(-1, -1, 2/3)| = -24,$$

$$|H_{\mathcal{L}}(\sqrt{3}, -\sqrt{3}, 2)| = |H_{\mathcal{L}}(-\sqrt{3}, \sqrt{3}, 2)| = 24,$$

concluimos que  $(1, 1)$  y  $(-1, -1)$  son mínimos locales, mientras que  $(\sqrt{3}, -\sqrt{3})$  y  $(-\sqrt{3}, \sqrt{3})$  son máximos locales.

Al igual que en el caso de optimización libre, la búsqueda de extremos globales puede resultar bastante más compleja en general. Sin embargo, si se sabe que la función lagrangeana es siempre cóncava o convexa en todo su dominio, se puede asegurar que el óptimo restringido es un máximo o un mínimo global, como establece el siguiente teorema.

## 5.2 Optimización con restricciones de igualdad. Multiplicadores de Lagrange

**Teorema (condiciones suficientes para extremo global).** Sea  $(x^*, y^*, \lambda^*)$  un punto crítico de la lagrangeana  $\mathcal{L}(x, y, \lambda)$ . Entonces

- a)  $\mathcal{L}$  es cóncava con respecto a  $(x, y) \Rightarrow f$  tiene un máximo global en  $(x^*, y^*)$ .
- b)  $\mathcal{L}$  es convexa con respecto a  $(x, y) \Rightarrow f$  tiene un mínimo global en  $(x^*, y^*)$ .

Ejemplos:

1. En el ejemplo de maximización de  $f(x, y) = 9 - x^2 - y^2$  sujeto a  $x + y = 4$ , la lagrangeana

$$\mathcal{L}(x, y, \lambda) = 9 - x^2 - y^2 + \lambda(4 - x - y)$$

es una función cóncava, ya que es la suma de la función cóncava  $9 - x^2 - y^2$  con la función lineal  $\lambda(4 - x - y)$ . Por lo tanto, en el punto óptimo  $f$  presenta un máximo global.

2. En el ejemplo de maximización de la producción  $P(L, K) = 50L^{2/3}K^{1/3}$  sujeto a  $100L + 300K = 45000$ , la lagrangeana

$$\mathcal{L}(L, K, \lambda) = 50L^{2/3}K^{1/3} + \lambda(45\,000 - 100L - 300K)$$

es una función cóncava, ya que es la suma de la función cóncava  $50L^{2/3}K^{1/3}$  con la función lineal  $\lambda(45\,000 - 100L - 300K)$ . Por lo tanto, en el punto óptimo la producción  $P$  presenta un máximo global.

3. En el ejemplo de minimización del costo  $C(L, K) = wL + rK$  sujeto a  $L^{1/2}K^{1/2} = Q$ , la lagrangeana

$$\mathcal{L}(L, K, \lambda) = wL + rK + \lambda(Q - L^{1/2}K^{1/2})$$

es una función convexa, ya que es la suma de la función lineal  $wL + rK$  con la función convexa  $\lambda(Q - L^{1/2}K^{1/2})$  (observa que esta última es convexa, ya que  $-L^{1/2}K^{1/2}$  es cóncava y  $\lambda > 0$ ). Por lo tanto, en el punto óptimo el costo  $C$  presenta un mínimo global.

### 5.2.3 El caso multidimensional

Es fácil generalizar los resultados anteriores al caso multidimensional, correspondiente a la optimización de una función de  $n$  variables sujeta a  $m < n$  restricciones de igualdad,

$$\begin{aligned} &\text{máx./mín.} && f(x_1, \dots, x_n) \\ &\text{s.a.} && g_1(x_1, \dots, x_n) = c_1 \\ &&& \vdots \\ &&& g_m(x_1, \dots, x_n) = c_m, \quad m < n. \end{aligned}$$

## Capítulo 5 Optimización

Es importante señalar que el número  $m$  de restricciones debe ser estrictamente menor al número  $n$  de variables. De otra manera, si  $m = n$  el sistema de ecuaciones podría tener una solución única, por lo que no habría grados de libertad para llevar a cabo la optimización, o bien, si  $m > n$  habrían más ecuaciones que incógnitas y el sistema podría ser inconsistente (no existiría solución posible).

La anterior condición de tangencia en el punto óptimo,  $\nabla f = \lambda \nabla g$ , se generaliza ahora requiriendo que, en ese punto, el gradiente  $\nabla f$  de la función  $f$  sea una combinación lineal del conjunto de gradientes  $\{\nabla g_1, \dots, \nabla g_m\}$  de todas las restricciones. En otras palabras, en el óptimo debe verificarse

$$\nabla f = \lambda_1 \nabla g_1 + \dots + \lambda_m \nabla g_m,$$

en donde  $\lambda_1, \dots, \lambda_m \in \mathbb{R}$  son los multiplicadores de Lagrange correspondientes a las restricciones  $g_1, \dots, g_m$ . La existencia de estos multiplicadores sólo está garantizada cuando el conjunto de gradientes  $\{\nabla g_j\}$  en el óptimo es linealmente independiente, lo que se conoce como la *cualificación de las restricciones*. Cuando esta condición no se cumple el método de Lagrange puede fallar, como se discute en la sección 5.2.4.

**Teorema de Lagrange (condiciones necesarias de primer orden).** Sean  $f : S \rightarrow \mathbb{R}$  y  $g_1, \dots, g_m : S \rightarrow \mathbb{R}$  funciones diferenciables en  $S \subset \mathbb{R}^n$ , con  $m < n$ . Sea  $\vec{x}^* \in S$  una solución del problema

$$\begin{aligned} &\text{máx./mín.} \quad f(\vec{x}) \\ &\text{s.a.} \quad g_j(\vec{x}) = c_j, \end{aligned}$$

con  $j = 1, \dots, m$ . Si el conjunto de gradientes  $\{\nabla g_j(\vec{x}^*)\}$  en el óptimo es linealmente independiente, entonces existen  $\lambda_1^*, \dots, \lambda_m^* \in \mathbb{R}$  tales que  $(\vec{x}^*, \vec{\lambda}^*)$  es un punto crítico de la función lagrangeana

$$\mathcal{L}(\vec{x}, \vec{\lambda}) = f(\vec{x}) + \sum_{j=1}^m \lambda_j (c_j - g_j(\vec{x})).$$

En este caso, las  $n + m$  condiciones de primer orden para la función lagrangeana son

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial x_1} = \frac{\partial f}{\partial x_1} - \lambda_1 \frac{\partial g_1}{\partial x_1} - \dots - \lambda_m \frac{\partial g_m}{\partial x_1} = 0 \\ \vdots \\ \frac{\partial \mathcal{L}}{\partial x_n} = \frac{\partial f}{\partial x_n} - \lambda_1 \frac{\partial g_1}{\partial x_n} - \dots - \lambda_m \frac{\partial g_m}{\partial x_n} = 0 \end{cases} \quad (n \text{ ecuaciones})$$

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial \lambda_1} = c_1 - g_1(x_1, \dots, x_n) = 0 \\ \vdots \\ \frac{\partial \mathcal{L}}{\partial \lambda_m} = c_m - g_m(x_1, \dots, x_n) = 0. \end{cases} \quad (m \text{ ecuaciones})$$



## 5.2 Optimización con restricciones de igualdad. Multiplicadores de Lagrange

Las primeras  $n$  ecuaciones equivalen a la condición  $\nabla f = \sum_{j=1}^m \lambda_j \nabla g_j$ , y las restantes son las  $m$  ecuaciones de restricción,  $g_j(\vec{x}) = c_j$ . Al resolver el sistema de  $n + m$  ecuaciones se obtienen las  $n$  coordenadas del punto óptimo,  $x_1^*, \dots, x_n^*$ , y los  $m$  multiplicadores de Lagrange,  $\lambda_1^*, \dots, \lambda_m^*$ .

El significado de los multiplicadores  $\lambda_1, \dots, \lambda_m$  es similar al del caso con una sola restricción. En el caso multidimensional, a cada valor del conjunto de parámetros  $\vec{c} = (c_1, \dots, c_m)$  le corresponde un punto óptimo,  $P(\vec{x}^*(\vec{c}), \vec{y}^*(\vec{c}))$ . En consecuencia, el valor óptimo  $f^*$  de la función  $f$ ,

$$f^*(\vec{c}) = f(\vec{x}^*(\vec{c}), \vec{y}^*(\vec{c})),$$

también depende de  $\vec{c}$ . Utilizando las  $n + m$  condiciones de primer orden anteriores, es posible demostrar que

$$\lambda_j^* = \frac{\partial f^*(\vec{c})}{\partial c_j},$$

para cada  $j = 1, \dots, m$ . Así,  $\lambda_j^*$  representa la razón de cambio instantánea del valor óptimo  $f^*$  de la función  $f$  al cambiar el parámetro  $c_j$ .

Por último, para clasificar los extremos locales y globales del problema puede utilizarse un criterio de signos para la matriz hessiana de  $\mathcal{L}$ , que es una matriz de  $(n + m) \times (n + m)$ . Debido al tamaño de esa matriz, este método de clasificación suele resultar bastante complejo. Sin embargo, en muchas de las aplicaciones de interés es fácil identificar un extremo global, simplemente argumentando sobre la concavidad o convexidad de  $\mathcal{L}$ , de acuerdo al siguiente teorema.

**Teorema (condiciones suficientes para extremo global).** Sea  $(\vec{x}^*, \vec{\lambda}^*)$  un punto crítico de la función lagrangeana  $\mathcal{L}(\vec{x}, \vec{\lambda})$ . Entonces

- a)  $\mathcal{L}$  es cóncava con respecto a  $\vec{x} \Rightarrow f$  tiene un máximo global en  $(\vec{x}^*)$ .
- b)  $\mathcal{L}$  es convexa con respecto a  $\vec{x} \Rightarrow f$  tiene un mínimo global en  $(\vec{x}^*)$ .

Ejemplo:

Resuelve el problema

$$\text{máx. } F(c_1, c_2, l_1, l_2) = \ln c_1 - l_1 + \beta (\ln c_2 - l_2)$$

$$\text{s.a. } 4l_1^{1/2} - c_1 = b_1$$

$$c_2 - 4l_2^{1/2} = b_1(1 + r),$$

donde  $\beta(1 + r) = 1$ , con  $\beta, r > 0$  parámetros del modelo.

## Capítulo 5 Optimización

Para simplificar nuestros desarrollos, multiplicamos la segunda restricción por  $\beta$  y utilizamos la condición  $\beta(1+r) = 1$ , obteniendo

$$\beta c_2 - 4\beta l_2^{1/2} = b_1.$$

La lagrangeana en este caso está dada por

$$\mathcal{L} = \ln c_1 - l_1 + \beta (\ln c_2 - l_2) + \lambda_1 (b_1 - 4l_1^{1/2} + c_1) + \lambda_2 (b_1 - \beta c_2 + 4\beta l_2^{1/2}),$$

que es función de las variables  $c_1, c_2, l_1, l_2, b_1, \lambda_1$  y  $\lambda_2$ . Las 7 condiciones de primer orden son

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial c_1} &= \frac{1}{c_1} + \lambda_1 = 0, & \frac{\partial \mathcal{L}}{\partial c_2} &= \frac{\beta}{c_2} - \beta \lambda_2 = 0, \\ \frac{\partial \mathcal{L}}{\partial l_1} &= -1 - 2\frac{\lambda_1}{l_1^{1/2}} = 0, & \frac{\partial \mathcal{L}}{\partial l_2} &= -\beta + 2\beta \frac{\lambda_2}{l_2^{1/2}} = 0, \\ \frac{\partial \mathcal{L}}{\partial b_1} &= \lambda_1 + \lambda_2 = 0, \\ \frac{\partial \mathcal{L}}{\partial \lambda_1} &= b_1 - 4l_1^{1/2} + c_1 = 0, & \frac{\partial \mathcal{L}}{\partial \lambda_2} &= b_1 - \beta c_2 + 4\beta l_2^{1/2} = 0.\end{aligned}$$

Resolviendo este sistema de ecuaciones se obtiene que el óptimo ocurre en  $l_1 = l_2 = \frac{1}{2}, c_1 = c_2 = 2^{3/2}, b_1 = 0, \lambda_1 = -2^{-3/2} = -\lambda_2$ . Por último, tomando en cuenta los signos de los multiplicadores  $\lambda_1$  y  $\lambda_2$  se tiene que  $\mathcal{L}$  es una función cóncava, de modo que se trata efectivamente de un máximo.

### 5.2.4 Cualificación de las restricciones: ¿cuándo falla el método de los multiplicadores de Lagrange?

El método de los multiplicadores de Lagrange para el problema de optimización restringida de  $f(x_1, \dots, x_n)$  sujeto a  $g_j(x_1, \dots, x_n) = c_j, j = 1, \dots, m$ , se basa en el cumplimiento de la llamada cualificación de las restricciones,

$$\nabla f^* = \lambda_1^* \nabla g_1^* + \dots + \lambda_m^* \nabla g_m^*.$$

Para que esta condición se cumpla es necesario que el conjunto de gradientes en el óptimo  $\{\nabla g_1^*, \dots, \nabla g_m^*\}$  sea linealmente independiente. Cuando esto no sucede es posible que el método de Lagrange no te permita obtener ninguno de los candidatos a óptimo, o bien, que no te dé todos los candidatos posibles. En ese caso, es necesario complementar el método de Lagrange con otro tipo de búsqueda de óptimos, como discutiremos a continuación.

En el caso particular de la optimización de una función de dos variables  $f(x, y)$  sujeto a una sola restricción  $g(x, y) = c$ , la cualificación de las restricciones se

## 5.2 Optimización con restricciones de igualdad. Multiplicadores de Lagrange

reduce a la condición de tangencia

$$\nabla f^* = \lambda^* \nabla g^*.$$

Esta condición no se cumple cuando el óptimo restringido  $(x^*, y^*)$  del problema coincide con un punto crítico de  $g$  ( $\nabla g = \vec{0}$ ), a menos que este óptimo también sea un punto crítico de  $f$  ( $\nabla f = \vec{0}$ ), como se ilustra en los siguientes ejemplos.

Ejemplos:

### 1. Encuentra la solución al problema

$$\begin{aligned} \text{máx. } & f(x, y) = -y \\ \text{s.a. } & y^3 - x^2 = 0. \end{aligned}$$

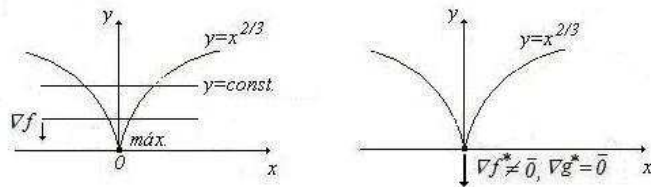
En este caso, la lagrangeana está dada por

$$\mathcal{L}(x, y, \lambda) = -y + \lambda (x^2 - y^3),$$

cuyas condiciones de primer orden son

$$\begin{aligned} \mathcal{L}_x &= 2\lambda x = 0 \\ \mathcal{L}_y &= -1 - 3\lambda y^2 = 0 \\ \mathcal{L}_\lambda &= x^2 - y^3 = 0. \end{aligned}$$

Es fácil verificar que no existe solución a este sistema de ecuaciones, por lo que erróneamente podríamos concluir que en este problema  $f$  no se maximiza. Sin embargo, si graficamos algunas curvas de nivel de  $f$ , dadas por  $-y = \text{const.}$ , y la curva de restricción  $y = x^{2/3}$ , observamos que la función  $f$  alcanza su máximo en el punto  $(0, 0)$ .



El método de Lagrange falla aquí, ya que no se verifica la condición  $\nabla f^* = \lambda^* \nabla g^*$ . En efecto, como  $\nabla f(x, y) = -\hat{j}$  y  $\nabla g(x, y) = -2x\hat{i} + 3y^2\hat{j}$ , por lo tanto en el óptimo se tiene

$$\nabla f^* = \nabla f(0, 0) = -\hat{j}, \quad \nabla g^* = \nabla g(0, 0) = \vec{0},$$

es decir, el óptimo  $(0, 0)$  es un punto crítico de  $g$ , pero no de  $f$ . Como en el óptimo  $\nabla f^* \neq \vec{0}$  y  $\nabla g^* = \vec{0}$ , por lo tanto no existe  $\lambda^*$  tal que  $\nabla f^* = \lambda^* \nabla g^*$ .

## Capítulo 5 Optimización

2. Encuentra la solución al problema

$$\text{máx. } f(x, y) = -x^2 - y^2$$

$$\text{s.a. } y^3 - x^2 = 0.$$

En este caso, la lagrangeana está dada por

$$\mathcal{L}(x, y, \lambda) = -x^2 - y^2 + \lambda(x^2 - y^3),$$

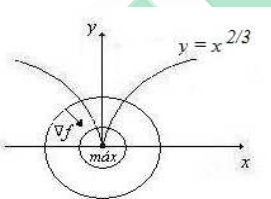
cuyas condiciones de primer orden son

$$\mathcal{L}_x = -2x + 2\lambda x = 2x(\lambda - 1) = 0,$$

$$\mathcal{L}_y = -2y - 3\lambda y^2 = y(-2 - 3\lambda y) = 0,$$

$$\mathcal{L}_\lambda = x^2 - y^3 = 0.$$

Es fácil verificar que este sistema de ecuaciones sí tiene solución, y ésta ocurre en el punto  $(0, 0)$ , que es la misma que se obtiene a partir de un análisis gráfico.



Aquí no falla el método, ya que sí se verifica la condición  $\nabla f^* = \lambda^* \nabla g^*$ . En efecto, como  $\nabla f(x, y) = -2x\hat{i} - 2y\hat{j}$  y  $\nabla g(x, y) = -2x\hat{i} + 3y^2\hat{j}$ , por lo tanto en el óptimo se tiene

$$\nabla f^* = \nabla f(0, 0) = \vec{0}, \quad \nabla g^* = \nabla g(0, 0) = \vec{0},$$

es decir, el óptimo  $(0, 0)$  es un punto crítico tanto de  $g$  como de  $f$ . Como en el óptimo  $\nabla f^* = \nabla g^* = \vec{0}$ , por lo tanto se cumple la condición  $\nabla f^* = \lambda^* \nabla g^*$  para todo valor de  $\lambda^*$ .

Este tipo de dificultades suele ocurrir cuando el óptimo restringido  $(x^*, y^*)$  de  $f$  coincide con algún punto cúspide de la curva de restricción  $g(x, y) = c$ , en donde no está definida la derivada  $dy/dx$ . La cúspide se origina en el hecho de la función  $z = g(x, y)$  tiene un punto crítico a lo largo de la curva de nivel  $g(x, y) = c$ . En efecto, a lo largo de la curva de nivel  $g(x, y) = c$  por el Teorema de la Función implícita se tiene

$$\frac{dy}{dx} = -\frac{g_x(x, y)}{g_y(x, y)}.$$

## 5.2 Optimización con restricciones de igualdad. Multiplicadores de Lagrange

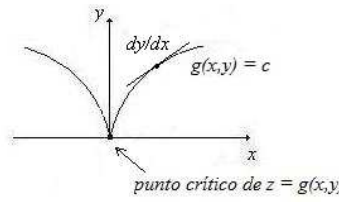
Cuando el óptimo  $(x^*, y^*)$  es un punto crítico de la función  $z = g(x, y)$ , entonces  $\nabla g(x^*, y^*) = \vec{0}$ . Por lo tanto,

$$g_x(x^*, y^*) = g_y(x^*, y^*) = 0.$$

Como  $(x^*, y^*)$  está en la curva  $g(x, y) = c$ , se tiene  $g(x^*, y^*) = c$ , de modo que

$$\frac{dy}{dx} = -\frac{g_x(x^*, y^*)}{g_y(x^*, y^*)} = -\frac{0}{0}.$$

Por lo tanto, la derivada  $dy/dx$  no está definida en  $(x^*, y^*)$  y la curva  $g(x, y) = c$  tiene una cúspide en  $(x^*, y^*)$ .



En resumen, los candidatos a óptimo para el problema de optimización restringida son los puntos críticos de la función lagrangeana  $\mathcal{L}(x, y, \lambda)$ , así como los puntos críticos de la función de restricción  $g(x, y)$ , en donde posiblemente se viole la cualificación de la restricción,  $\nabla f^* = \lambda^* \nabla g^*$ .

La extensión de los resultados anteriores para el caso multidimensional es más compleja. Solamente mencionaremos aquí que si se tiene  $m$  ecuaciones de restricción, de la forma  $g_j(\vec{x}) = c_j$ , la cualificación de las restricciones  $\nabla f^* = \sum_{j=1}^m \lambda_j^* \nabla g_j^*$  se viola cuando el rango de la matriz

$$Dg(\vec{x}^*) = \begin{pmatrix} \frac{\partial g_1^*}{\partial x_1} & \cdots & \frac{\partial g_1^*}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m^*}{\partial x_1} & \cdots & \frac{\partial g_m^*}{\partial x_n} \end{pmatrix}$$

es menor que  $m$  (ver, por ejemplo, el libro de Simon y Blume). En este caso, los candidatos a óptimo para el problema de optimización restringida son los puntos críticos de la función lagrangeana  $\mathcal{L}(\vec{x}, \vec{\lambda})$ , así como los puntos críticos de las funciones de restricción  $g_j(\vec{x})$ , en donde posiblemente se viole la cualificación de las restricciones,  $\nabla f^* = \sum_{j=1}^m \lambda_j^* \nabla g_j^*$ .

### 5.3 Optimización con restricciones de desigualdad. Condiciones de Kuhn-Tucker

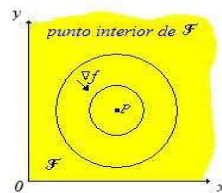
En esta sección estudiaremos cómo resolver un problema de optimización para una función sujeta a *restricciones de desigualdad*, en donde no todas las funciones son lineales, lo que se conoce como *Programación No Lineal*.

La solución a un problema de optimización depende de la geometría de la región factible  $\mathcal{F}$ . Diferentes tipos de restricciones dan origen a diferentes soluciones óptimas, como se muestra en los siguientes casos. En ellos se considera la maximización de una función cóncava  $f : \mathbb{R}_+^2 \rightarrow \mathbb{R}$ , representada por la superficie  $z = f(x, y)$  generada por los puntos  $(x, y) \in \mathbb{R}_+^2$ .

**Caso 1.** Primero consideremos el problema de maximización libre

$$\text{máx. } f(x, y).$$

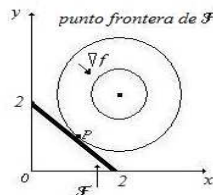
Aquí la región factible  $\mathcal{F}$  es simplemente  $\mathbb{R}_+^2$ . La siguiente figura muestra la región  $\mathcal{F}$ , algunas curvas de nivel de  $f$  y el vector gradiente  $\nabla f$ , que indica la dirección de crecimiento de  $f$ . En este caso, la solución óptima  $P$  corresponde al máximo no restringido de  $f$ . Como  $P$  está fuera de la frontera de  $\mathcal{F}$  (los ejes), se tiene que  $P$  es un *punto interior* de la región factible  $\mathcal{F}$ .



**Caso 2.** Ahora consideremos la maximización de  $f$  sujeta a una restricción de igualdad, tal como

$$\begin{aligned} \text{máx. } & f(x, y) \\ \text{s.a. } & x + y = 2. \end{aligned}$$

Aquí la región factible  $\mathcal{F}$  es el segmento de recta  $x + y = 2$  en  $\mathbb{R}_+^2$ . La solución óptima ocurre en el punto  $P$  de la curva  $\mathcal{F}$ . Como una curva en el plano está formada sólo de puntos frontera,  $P$  es un *punto frontera* de la región factible  $\mathcal{F}$ .

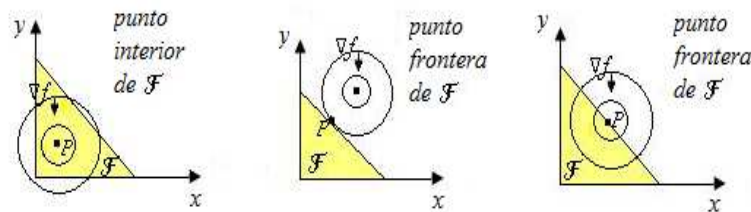


### 5.3 Optimización con restricciones de desigualdad. Condiciones de Kuhn-Tucker

**Caso 3.** Por último, consideremos la maximización de  $f$  sujeta a una restricción de desigualdad, tal como

$$\begin{aligned} \text{máx. } & f(x, y) \\ \text{s.a. } & x + y \leq 2. \end{aligned}$$

Aquí, la región factible  $\mathcal{F}$  es el triángulo definido por  $x + y \leq 2$ ,  $x \geq 0$  y  $y \geq 0$ . Dependiendo de la posición del máximo no restringido de  $f$  con respecto a la región  $\mathcal{F}$ , la solución óptima  $P$  puede ocurrir ya sea en un *punto interior* o en un *punto frontera* de  $\mathcal{F}$ .



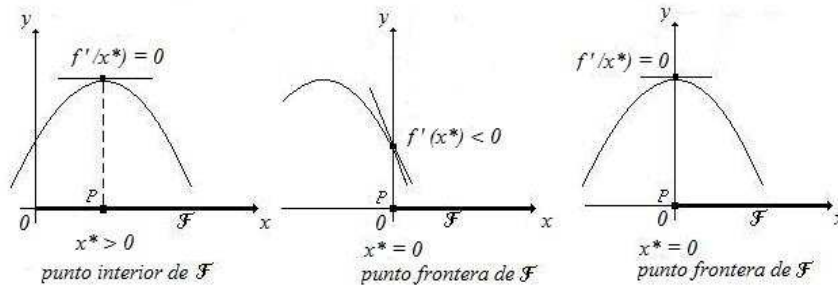
Esto es lo que se conoce como un *problema de Kuhn-Tucker*, y en las siguientes secciones deduciremos las condiciones de optimalidad correspondientes.

#### 5.3.1 Problemas de maximización

El siguiente ejemplo proporciona una idea intuitiva del problema de Kuhn-Tucker para una función de una variable. Sea  $f : \mathbb{R} \rightarrow \mathbb{R}$  una función cóncava, representada gráficamente por la curva  $y = f(x)$  en  $\mathbb{R}^2$ . Buscamos el máximo global del problema

$$\begin{aligned} \text{máx. } & f(x) \\ \text{s.a. } & x \geq 0. \end{aligned}$$

La siguiente figura muestra diferentes posibilidades para la función  $y = f(x)$  en  $\mathbb{R}^2$ . Dependiendo de la ubicación del máximo no restringido de  $f$  con respecto a la región factible  $\mathcal{F}$ , dada por  $x \geq 0$ , la solución óptima  $P$  puede ocurrir en un punto interior o un punto frontera de  $\mathcal{F}$ .



## Capítulo 5 Optimización

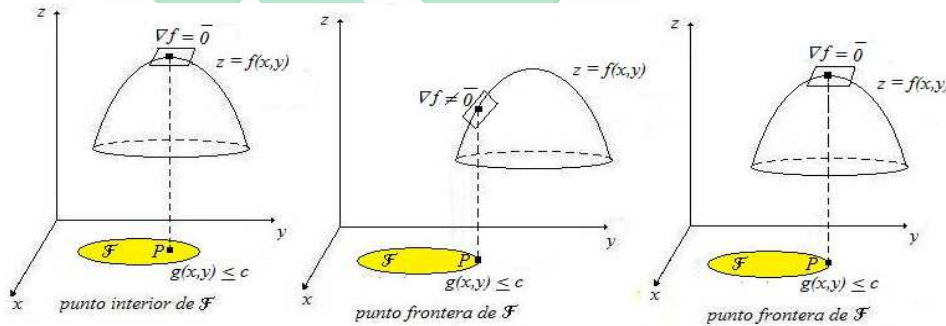
En la figura de la izquierda, la solución  $P$  coincide con el máximo no restringido de  $f$ , que es el punto  $x^* > 0$  en donde  $f'(x^*) = 0$ . En este caso,  $P$  es un *punto interior* de  $\mathcal{F}$ . En la figura central, el máximo no restringido de  $f$  cae fuera de la región factible, de modo que la mejor solución  $P$  ocurre en  $x^* = 0$ , con  $f'(x^*) < 0$ . Así,  $P$  es un *punto frontera* de  $\mathcal{F}$ . Por último, en la figura de la derecha el máximo no restringido de  $f$  ocurre exactamente en  $x^* = 0$ , con  $f'(x^*) = 0$ . Nuevamente,  $P$  es un *punto frontera* de  $\mathcal{F}$ . Del análisis gráfico podemos concluir que en el óptimo se verifican las siguientes tres condiciones

$$x^* \geq 0, \quad f'(x^*) \leq 0 \quad \text{y} \quad x^* f'(x^*) = 0.$$

Ahora consideremos el caso de una función de dos variables. Sea  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  una función cóncava, representada por la superficie  $z = f(x, y)$  en  $\mathbb{R}^3$ , que deberá maximizarse sujeto a una restricción de desigualdad,  $g(x, y) \leq c$ , con  $c$  una constante. Como se justificará más adelante, la restricción correspondiente a un problema de maximización deberá escribirse siempre en el formato dado ( $\leq$ ). De esta manera, considera el problema

$$\begin{aligned} \text{máx. } & f(x, y) \\ \text{s.a. } & g(x, y) \leq c. \end{aligned}$$

La siguiente figura muestra diferentes posibilidades para la función  $z = f(x, y)$  en  $\mathbb{R}^3$ . Dependiendo de la ubicación del máximo no restringido de  $f$  con respecto a la región factible  $\mathcal{F}$ , dada por  $g(x, y) \leq c$ , la solución óptima  $P$  puede ocurrir en un punto interior o un punto frontera de  $\mathcal{F}$ .

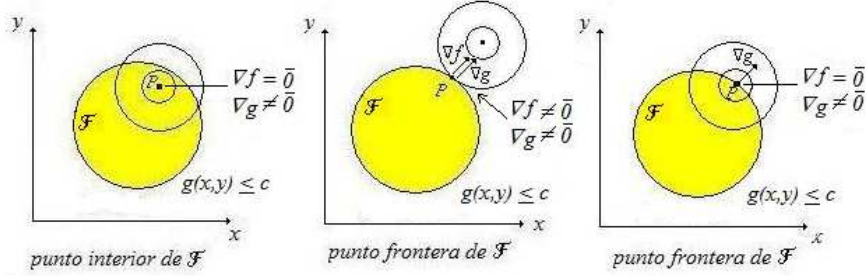


En la figura de la izquierda, el máximo no restringido de  $f$  ocurre en el interior de la región factible,  $g(x, y) < c$ ; la solución óptima  $P$  es un *punto interior* de  $\mathcal{F}$ . En la figura central, el máximo no restringido de  $f$  cae fuera de la región factible y la mejor opción es el punto  $P$  de la curva  $g(x, y) = c$  que se encuentra lo más cercano posible a ese máximo; la solución óptima  $P$  es un *punto frontera* de  $\mathcal{F}$ . En la figura de la derecha, el máximo no restringido de  $f$  ocurre exactamente sobre la curva  $g(x, y) = c$ , nuevamente, la solución óptima  $P$  es un *punto frontera* de  $\mathcal{F}$ . A



### 5.3 Optimización con restricciones de desigualdad. Condiciones de Kuhn-Tucker

continuación se muestran las proyecciones en el plano  $\mathbb{R}^2$  correspondientes a los tres escenarios anteriores. En ellas se puede observar la región factible  $\mathcal{F}$  y algunas curvas de nivel de cada una de las funciones  $f$ .



En el primer caso, la existencia de la restricción resulta irrelevante, y el problema puede resolverse a partir de la maximización libre de  $f$ . En el segundo caso, la restricción  $g(x, y) \leq c$  puede reemplazarse por la restricción de igualdad  $g(x, y) = c$ , y la solución del problema puede obtenerse con el método de los multiplicadores de Lagrange. El último caso es una combinación de los otros dos.

Con el fin de desarrollar un método general de solución que abarque los tres casos anteriores, partimos de una descripción basada en el método de los multiplicadores de Lagrange, pero que sea capaz de reducirse a un problema de optimización libre cuando ése sea el caso. Al tratarse de una descripción tipo Lagrange, pediremos que en el óptimo se verifique la condición de tangencia

$$\nabla f = \lambda \nabla g.$$

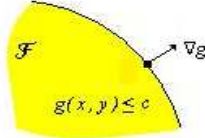
La diferencia con el método original de Lagrange consiste en que ahora el problema de optimización está sujeto a una restricción de desigualdad,  $g(x, y) \leq c$ . Por esa razón, debemos reemplazar la condición  $g(x, y) = c$  de solución frontera por algún otro criterio que permita la existencia de una solución interior, como se discutió en los ejemplos anteriores. Como se justifica a continuación, este nuevo criterio requiere el cumplimiento de las siguientes tres condiciones,

$$g(x, y) \leq c, \quad \lambda \geq 0 \quad \text{y} \quad \lambda(g(x, y) - c) = 0.$$

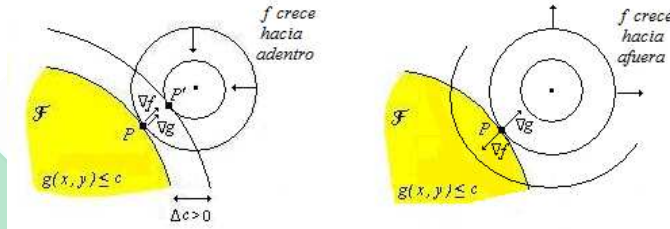
La condición  $g(x, y) \leq c$  establece que en el óptimo se debe satisfacer la restricción impuesta. Para entender la condición  $\lambda \geq 0$ , o condición de *no negatividad* del multiplicador  $\lambda$ , es necesario hacer la siguiente consideración. Primero observa que la restricción de desigualdad  $g(x, y) \leq c$  puede interpretarse como el contorno inferior  $CI_g(c)$  de la función  $z = g(x, y)$  correspondiente a  $z = c$ . Dado que el gradiente  $\nabla g$  apunta hacia el contorno superior  $CS_g(c)$  de

## Capítulo 5 Optimización

$g$ , por lo tanto  $\nabla g$  debe apuntar hacia afuera de la región factible  $\mathcal{F}$ , dada por  $g(x, y) \leq c$ , como se ilustra en la siguiente figura.



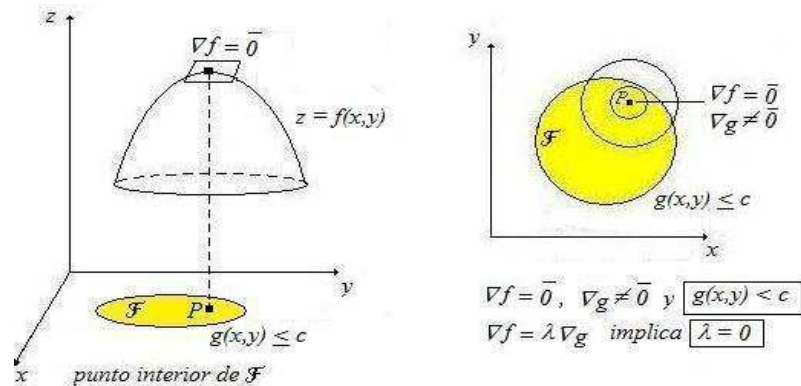
Imponer que  $\lambda \geq 0$  en la condición de tangencia  $\nabla f = \lambda \nabla g$  implica que, en el óptimo  $P$ , los gradientes  $\nabla f$  y  $\nabla g$  deben apuntar en el mismo sentido. Así, un pequeño incremento  $\Delta c$  en  $c$  hará que el nuevo óptimo  $P'$  esté en una curva de nivel con mayor valor de  $f$ , como se muestra en la figura de la izquierda. Si  $\lambda < 0$ , entonces  $\nabla f$  apuntaría en el sentido contrario a  $\nabla g$ , por lo que habría una curva de nivel con un mayor valor de  $f$  dentro de la región de restricción; en consecuencia,  $P$  no podría ser el punto que maximiza a  $f$ , ya que habría un mejor punto  $P'$  dentro de la región, como se muestra en la figura de la derecha.



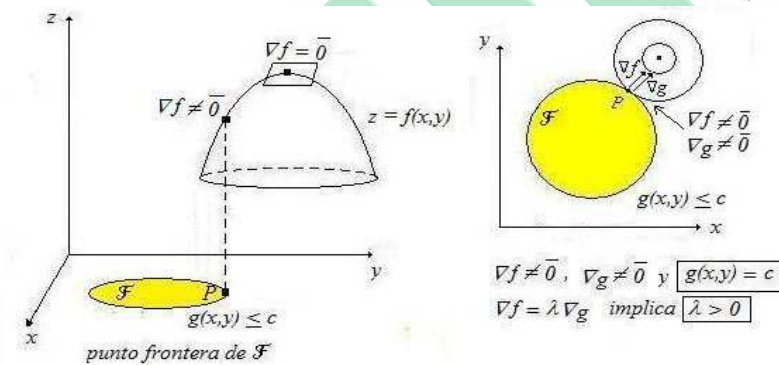
Nota que la no negatividad de  $\lambda$  en el problema de Kuhn-Tucker contrasta con el caso de una restricción de igualdad. En ese caso,  $\lambda$  podía tomar cualquier signo, ya que la optimización se lleva a cabo en la curva  $g(x, y) = c$ , sin posibilidad de desplazamiento hacia las regiones alrededor de la misma.

Para explicar la última condición,  $\lambda(g(x, y) - c) = 0$ , conviene analizar por separado los tres casos anteriores. Cuando el máximo no restringido de  $f$  ocurre en el interior de la región de restricción, el óptimo  $P$  satisface  $g(x, y) < c$ . Ahí  $\nabla f = \vec{0}$  y  $\nabla g \neq \vec{0}$ , por lo que la condición  $\nabla f = \lambda \nabla g$  implica que  $\lambda = 0$  en  $P$ .

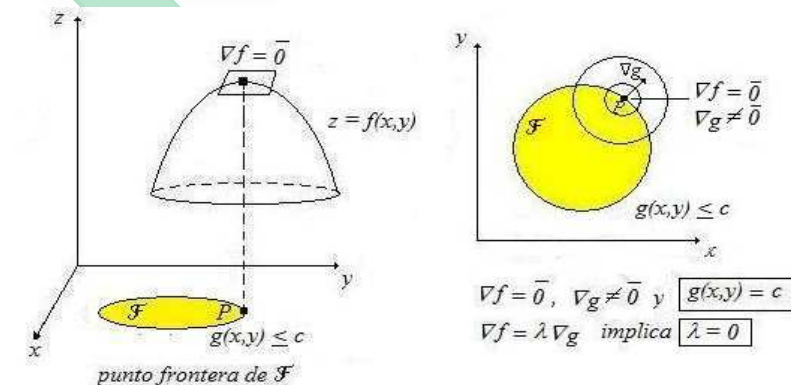
### 5.3 Optimización con restricciones de desigualdad. Condiciones de Kuhn-Tucker



Por otra parte, cuando el máximo no restringido de  $f$  cae fuera de la región de restricción, el óptimo  $P$  satisface  $g(x, y) = c$ . Ahí  $\nabla f \neq \vec{0}$  y  $\nabla g \neq \vec{0}$ , por lo que la condición  $\nabla f = \lambda \nabla g$  implica que  $\lambda > 0$  en  $P$ .



Por último, cuando el máximo no restringido de  $f$  ocurre exactamente en la frontera de la restricción, el óptimo  $P$  satisface  $g(x, y) = c$ . Como ahí  $\nabla f = \vec{0}$  y  $\nabla g \neq \vec{0}$ , la condición  $\nabla f = \lambda \nabla g$  implica que  $\lambda = 0$  en  $P$ .



## Capítulo 5 Optimización

Concluimos que en los tres casos se verifica la condición  $\lambda(g(x, y) - c) = 0$ , que establece que en el óptimo  $P$  debe cumplirse al menos una de las siguientes igualdades,

$$\lambda = 0 \quad \text{o} \quad g(x, y) = c.$$

### 5.3.1.1 Condiciones de Kuhn-Tucker para problemas de maximización

Los resultados de la discusión anterior para resolver el problema de Kuhn-Tucker

$$\text{máx. } f(x, y)$$

$$\text{s.a. } g(x, y) \leq c,$$

se pueden formalizar en términos del método de Lagrange, de la siguiente manera. Primeramente, se plantea la función lagrangeana

$$\mathcal{L}(x, y, \lambda) = f(x, y) + \lambda(c - g(x, y)).$$

En lugar de pedir  $\mathcal{L}_x = \mathcal{L}_y = \mathcal{L}_\lambda = 0$  como en el método de Lagrange, ahora las *condiciones necesarias de primer orden* son

$$\mathcal{L}_x = f_x(x, y) - \lambda g_x(x, y) = 0$$

$$\mathcal{L}_y = f_y(x, y) - \lambda g_y(x, y) = 0,$$

que representan la tangencia ( $\nabla f = \lambda \nabla g$ ) en el óptimo. La tercera igualdad,  $\mathcal{L}_\lambda = 0$ , equivalente a  $g(x, y) = c$ , debe sustituirse por las llamadas *condiciones de holgura complementaria*,

$$\mathcal{L}_\lambda \geq 0, \quad \lambda \geq 0 \quad \text{y} \quad \lambda \mathcal{L}_\lambda = 0,$$

o equivalentemente,

$$g(x, y) \leq c, \quad \lambda \geq 0 \quad \text{y} \quad \lambda(g(x, y) - c) = 0.$$

La primera condición impone el cumplimiento de la restricción en el óptimo. La condición  $\lambda \geq 0$  garantiza que al ampliar la región factible se obtendrá un valor mayor para el óptimo de  $f$ . La última condición establece que  $\lambda = 0$  (optimización libre) o  $g(x, y) = c$  (problema de Lagrange) en el óptimo.

Por último, como se justificará en la sección 5.3.1.5 sobre condiciones suficientes, la restricción en un problema de maximización siempre deberá estar expresada en el formato  $g(x, y) \leq c$ , con el fin de garantizar de que el óptimo obtenido se trate efectivamente de un máximo (para  $\lambda \geq 0$ ).

### 5.3 Optimización con restricciones de desigualdad. Condiciones de Kuhn-Tucker

**Condiciones de Kuhn-Tucker para el problema de maximización**  
máx.  $f(x, y)$ , s.a.  $g(x, y) \leq c$ .

1. Se construye la función lagrangeana correspondiente,

$$\mathcal{L}(x, y, \lambda) = f(x, y) + \lambda(c - g(x, y)).$$

2. Se establecen las condiciones de primer orden para  $\mathcal{L}$ , con respecto a las variables  $x$  y  $y$  solamente,

$$\mathcal{L}_x = f_x - \lambda g_x = 0$$

$$\mathcal{L}_y = f_y - \lambda g_y = 0.$$

3. Se establece las condiciones de holgura complementaria

$$\mathcal{L}_\lambda \geq 0, \quad \lambda \geq 0 \quad \text{y} \quad \lambda \mathcal{L}_\lambda = 0,$$

o equivalentemente,

$$g(x, y) \leq c, \quad \lambda \geq 0 \quad \text{y} \quad \lambda(g(x, y) - c) = 0.$$

4. Se resuelve, consistentemente, el sistema de ecuaciones y desigualdades.

Ejemplos:

1. Encuentra la solución al problema

$$\text{máx. } f(x, y) = x + y$$

$$\text{s.a. } x^2 + y^2 \leq 1.$$

En este caso, la lagrangeana correspondiente está dada por

$$\mathcal{L}(x, y, \lambda) = x + y + \lambda(1 - x^2 - y^2).$$

Las condiciones de primer orden en  $x$  y  $y$  son

$$\mathcal{L}_x = 1 - 2\lambda x = 0 \quad (1)$$

$$\mathcal{L}_y = 1 - 2\lambda y = 0, \quad (2)$$

que deberán resolverse junto con las condiciones de holgura complementaria

$$x^2 + y^2 \leq 1, \quad (3) \quad \lambda \geq 0, \quad (4) \quad \lambda(x^2 + y^2 - 1) = 0. \quad (5)$$

Para resolver el sistema (1)-(5) conviene comenzar por la igualdad (5), que establece que el óptimo debe cumplir al menos una de las condiciones  $\lambda = 0$  o  $x^2 + y^2 = 1$ . Analicemos cada caso por separado. i) Si  $\lambda = 0$ , las ecuaciones

## Capítulo 5 Optimización

(1)-(2) conducen a una inconsistencia ( $1 = 0$ ), por lo que esta opción se descarta. ii) Si  $x^2 + y^2 = 1$ , se obtiene el sistema de ecuaciones

$$1 - 2\lambda x = 0 \quad (1)$$

$$1 - 2\lambda y = 0 \quad (2)$$

$$x^2 + y^2 = 1, \quad (6)$$

correspondiente a un punto frontera ( $x^2 + y^2 = 1$ ) de la restricción. El sistema tiene dos soluciones posibles,

$$(x_1, y_1) = \left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) \quad \text{y} \quad (x_2, y_2) = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right).$$

Sustituyendo  $(x_1, y_1)$  en (1) se obtiene

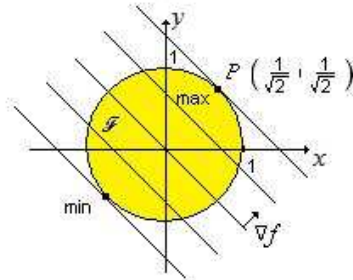
$$\lambda_1 = \frac{1}{2x_1} = -\frac{\sqrt{2}}{2} < 0.$$

Esto viola la condición (4), por lo que esta opción se descarta. Por otra parte, sustituyendo  $(x_2, y_2)$  en (1) se obtiene

$$\lambda_2 = \frac{1}{2x_2} = \frac{\sqrt{2}}{2} > 0,$$

que satisface la condición (4). Así, sólo los valores  $x_2, y_2, \lambda_2$ , satisfacen las condiciones (1)-(5) en su totalidad. Concluimos que el valor máximo de  $f$  ocurre en el punto frontera  $P\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$ , con  $\lambda_2 = \sqrt{2}/2$  y  $f^* = \frac{2}{\sqrt{2}}$ .

La siguiente figura muestra la región factible  $\mathcal{F}$ ,  $x^2 + y^2 \leq 1$ , y algunas curvas de nivel de  $f$ , dadas por  $x + y = \text{const}$ . Tomando en cuenta la dirección de crecimiento de  $f$ , dada por el vector gradiente  $\nabla f$ , es claro que la solución óptima ocurre en el punto frontera  $P\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$  de la restricción, en donde  $f$  toma su máximo valor posible.



### 5.3 Optimización con restricciones de desigualdad. Condiciones de Kuhn-Tucker

Nota que, al ser la solución un punto frontera de  $\mathcal{F}$ , este problema de Kuhn-Tucker es enteramente equivalente al problema de Lagrange

$$\begin{aligned} \text{máx. } & f(x, y) = x + y \\ \text{s.a. } & x^2 + y^2 = 1. \end{aligned}$$

2. Encuentra la solución al problema

$$\begin{aligned} \text{máx. } & f(x, y) = 9 - (x - 2)^2 - (y - 2)^2 \\ \text{s.a. } & 2x + 3y \leq 12. \end{aligned}$$

En este caso, la lagrangeana correspondiente es

$$\mathcal{L}(x, y, \lambda) = 9 - (x - 2)^2 - (y - 2)^2 + \lambda(12 - 2x - 3y)$$

Las condiciones de primer orden en  $x$  y  $y$  son

$$\mathcal{L}_x = -2(x - 2) - 2\lambda = 0 \quad (1)$$

$$\mathcal{L}_y = -2(y - 2) - 3\lambda = 0, \quad (2)$$

que deberán resolverse junto con las condiciones de holgura complementaria

$$2x + 3y \leq 12, \quad (3) \quad \lambda \geq 0, \quad (4) \quad \lambda(2x + 3y - 12) = 0. \quad (5)$$

De la igualdad (5) se sigue que  $\lambda = 0$  o  $2x + 3y = 12$ . i) Si  $\lambda = 0$ , las ecuaciones (1)-(2) implican

$$x = y = 2, \quad (6)$$

en donde todas las condiciones (1)-(5) se satisfacen. Nota que (6) constituye un punto interior de la restricción, ya que

$$2x + 3y = 2(2) + 3(2) = 10 < 12. \quad (7)$$

ii) Si  $2x + 3y = 12$ , se obtiene el sistema de ecuaciones

$$-2(x - 2) - 2\lambda = 0 \quad (1)$$

$$-2(y - 2) - 3\lambda = 0 \quad (2)$$

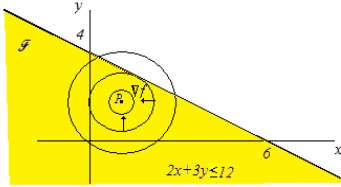
$$2x + 3y = 12, \quad (9)$$

que correspondería a un punto frontera. Al resolver el sistema obtenemos el punto  $(x, y) = \left(\frac{30}{13}, \frac{32}{13}\right)$ , con  $\lambda = -\frac{4}{13} < 0$ , que viola la condición (4). Así, esta opción se descarta. Concluimos que el valor máximo de  $f$  ocurre en el punto  $P(2, 2)$ , con  $\lambda = 0$  y  $f^* = 9$ .

La siguiente figura muestra la región factible  $\mathcal{F}$ ,  $2x + 3y \leq 12$ , y algunas curvas de  $f$ , dadas por  $9 - (x - 2)^2 - (y - 2)^2 = \text{const.}$  Tomando en cuenta la dirección del vector gradiente  $\nabla f$ , el máximo restringido de  $f$  coincide con su punto de

## Capítulo 5 Optimización

optimización libre,  $P(2, 2)$ . Como en ese punto  $2x + 3y < 12$ ,  $P$  es un punto interior de la restricción, y por tanto  $\lambda = 0$  ahí.



### 5.3.1.2 El caso multidimensional

Los resultados anteriores pueden generalizarse fácilmente para el caso de la optimización de una función  $f$  de varias variables, sujeta a una colección de restricciones de desigualdad, de la forma

$$\begin{aligned} &\text{máx. } f(x_1, \dots, x_n) \\ &\text{s.a. } g_j(x_1, \dots, x_n) \leq c_j, \quad j = 1, \dots, m. \end{aligned}$$

Aquí el número  $m$  de restricciones puede ser mayor, menor o igual al número  $n$  de variables, debido a que la región factible  $\mathcal{F}$  es la intersección de desigualdades (por ejemplo, un polígono en  $\mathbb{R}^2$ ) y no de igualdades, como en el caso de Lagrange (en donde  $m < n$ ). Para cada restricción  $g_j(\vec{x}) \leq c_j$  se introduce un multiplicador de Lagrange  $\lambda_j$ , y se establecen las condiciones que se enuncian a continuación.

**Condiciones de Kuhn-Tucker para el problema** máx.  $f(\vec{x})$  s.a.  $g_j(\vec{x}) \leq c_j$ ,  $j = 1, \dots, m$ ,  $\vec{x} \in \mathbb{R}^n$ .

1. Se construye la función lagrangeana correspondiente

$$\mathcal{L}(\vec{x}, \vec{\lambda}) = f(\vec{x}) + \sum_{j=1}^m \lambda_j (c_j - g_j(\vec{x})).$$

2. Se establecen  $n$  condiciones de primer orden para  $\mathcal{L}$ , en  $x_i$ ,  $i = 1, \dots, n$

$$\mathcal{L}_{x_i} = 0, \quad i = 1, \dots, n.$$

3. Se establecen  $m$  condiciones de holgura complementaria

$$g_j(\vec{x}) \leq c_j, \quad \lambda_j \geq 0 \quad \text{y} \quad \lambda_j (g_j(\vec{x}) - c_j) = 0, \quad j = 1, \dots, m,$$

o equivalentemente,

$$\mathcal{L}_{\lambda_j} \geq 0, \quad \lambda_j \geq 0 \quad \text{y} \quad \lambda_j \mathcal{L}_{\lambda_j} = 0.$$

4. Se resuelve consistentemente el sistema de ecuaciones y desigualdades.



### 5.3 Optimización con restricciones de desigualdad. Condiciones de Kuhn-Tucker

Ejemplos:

1. Encuentra la solución al problema

$$\text{máx. } f(x, y) = x^{1/2}y^{1/2}$$

$$\text{s.a. } x + y \leq 2 \quad (\text{I})$$

$$x + 2y \geq 2 \quad (\text{II})$$

$$x \geq 0. \quad (\text{III})$$

Primeramente, escribimos las restricciones en el formato  $g_j(\vec{x}) \leq c_j$ , es decir,

$$\text{máx. } f(x, y) = x^{1/2}y^{1/2}$$

$$\text{s.a. } x + y \leq 2$$

$$-x - 2y \leq -2$$

$$-x \leq 0$$

La lagrangeana correspondiente está dada por

$$\mathcal{L}(x, y, \lambda_1, \lambda_2, \lambda_3) = x^{1/2}y^{1/2} + \lambda_1(2 - x - y) + \lambda_2(-2 + x + 2y) + \lambda_3x,$$

que en el óptimo satisface las condiciones de igualdad

$$\mathcal{L}_x = \frac{y^{1/2}}{2x^{1/2}} - \lambda_1 + \lambda_2 + \lambda_3 = 0, \quad (1)$$

$$\mathcal{L}_y = \frac{x^{1/2}}{2y^{1/2}} - \lambda_1 + 2\lambda_2 = 0, \quad (2)$$

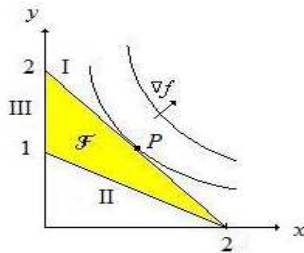
junto con las condiciones de holgura complementaria

$$x + y \leq 2, \quad \lambda_1 \geq 0, \quad \lambda_1(x + y - 2) = 0, \quad (3)$$

$$x + 2y \geq 2, \quad \lambda_2 \geq 0, \quad \lambda_2(x + 2y - 2) = 0, \quad (4)$$

$$x \geq 0, \quad \lambda_3 \geq 0, \quad \lambda_3x = 0. \quad (5)$$

De las igualdades en (3), (4) y (5) se siguen  $2^3 = 8$  casos, i)  $\lambda_1 = \lambda_2 = \lambda_3 = 0$ , ii)  $\lambda_1 = \lambda_2 = x = 0$ , etc, cuyo análisis puede resultar engorroso. En lugar de esto, conviene utilizar un análisis gráfico para reducir el número de casos. Para ello, dibujamos la región factible  $\mathcal{F}$  (intersección de las tres restricciones de desigualdad), algunas curvas de nivel de  $f$ , así como el gradiente  $\nabla f$ , para determinar la dirección de crecimiento de  $f$ .



## Capítulo 5 Optimización

En la gráfica se observa que el punto  $P$  que maximiza a  $f$  ocurre en donde la restricción I está activa,

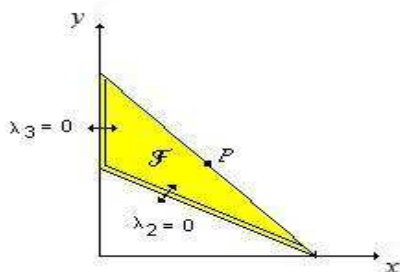
$$x + y = 2, \quad (6)$$

mientras que las restricciones II y III no lo están ( $x + 2y > 2$  y  $x > 0$ ). De acuerdo con las igualdades en (4) y (5), esta última afirmación implica que

$$\lambda_2 = 0 \text{ y } \lambda_3 = 0. \quad (7)$$

Por último, resolvemos el sistema de igualdades (1), (2), (6) y (7) para  $x, y, \lambda_1, \lambda_2$  y  $\lambda_3$ , y verificamos el cumplimiento de la condición  $\lambda_1 \geq 0$ . Con esto, se tiene que el valor máximo de  $f$  sucede en el punto  $P(1, 1)$ , con  $\lambda_1 = \frac{1}{2} > 0$ ,  $\lambda_2 = \lambda_3 = 0$  y  $f^* = 1$ .

Geométricamente,  $\lambda_2 = 0$  significa que si la restricción II se modificara ligeramente, digamos a  $x + 2y \geq 2.01$ , esto no afectaría la posición actual del óptimo,  $(x^*, y^*) = (1, 1)$ . Lo mismo sucede con la restricción III, ya que  $\lambda_3 = 0$ . En contraste,  $\lambda_1 \neq 0$  implica que un pequeño cambio en la restricción I sí produciría una nueva solución óptima. Por esta razón, a los multiplicadores de Lagrange  $\lambda_j$  se les denomina *variables de sensibilidad* ante cambios en los parámetros  $c_j$ .



Aquí la solución óptima es un punto frontera de la restricción I, y no es sensible a cambios infinitesimales en las restricciones II y III. En consecuencia, este problema de Kuhn-Tucker es enteramente equivalente al problema de Lagrange

$$\text{máx. } f(x, y) = x^{1/2}y^{1/2}$$

$$\text{s.a. } x + y = 2.$$

2. Resuelve el siguiente problema de maximización de la utilidad de un individuo,  $u(x, y)$ , en donde  $p, m > 0$ :

$$\text{máx. } u(x, y) = x + \ln(1 + y)$$

$$\text{s.a. } px + y \leq m \quad (\text{I})$$

$$x \geq 0 \quad (\text{II})$$

$$y \geq 0. \quad (\text{III})$$

### 5.3 Optimización con restricciones de desigualdad. Condiciones de Kuhn-Tucker

Primeramente, escribimos las restricciones en el formato  $g_j(\vec{x}) \leq c_j$ , es decir,

$$\begin{aligned} \text{máx. } & u(x, y) = x + \ln(1 + y) \\ \text{s.a. } & px + y \leq m \\ & -x \leq 0 \\ & -y \leq 0. \end{aligned}$$

La lagrangeana correspondiente está dada por

$$\mathcal{L}(x, y, \lambda_1, \lambda_2, \lambda_3) = x + \ln(1 + y) + \lambda_1(m - px - y) + \lambda_2x + \lambda_3y,$$

que en el óptimo satisface las condiciones de igualdad

$$\mathcal{L}_x = 1 - p\lambda_1 + \lambda_2 = 0, \quad (1)$$

$$\mathcal{L}_y = \frac{1}{1+y} - \lambda_1 + \lambda_3 = 0, \quad (2)$$

junto con las condiciones de holgura complementaria

$$px + y \leq m, \quad \lambda_1 \geq 0, \quad \lambda_1(px + y - m) = 0, \quad (3)$$

$$x \geq 0, \quad \lambda_2 \geq 0, \quad \lambda_2x = 0, \quad (4)$$

$$y \geq 0, \quad \lambda_3 \geq 0, \quad \lambda_3y = 0. \quad (5)$$

Es claro de (1) que  $\lambda_1 \neq 0$ , ya que en ese caso  $\lambda_{2,3} < 0$ . En consecuencia,  $\lambda_1 > 0$  y la condición (3) implica

$$px + y = m. \quad (6)$$

Nos quedan, entonces, sólo  $2^2 = 4$  casos:

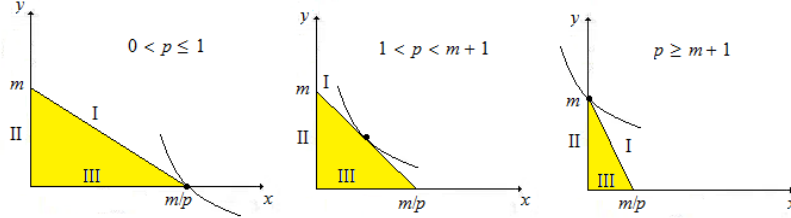
- i) Si  $x^* = 0$  y  $y^* = 0$ , se viola la condición (6), ya que  $m > 0$ .
- ii) Si  $x^* > 0$  y  $y^* = 0$ , de (6) se sigue que  $x^* = m/p$ . De (4) y (1) se obtiene  $\lambda_2 = 0$ ,  $\lambda_1 = 1/p$ . De (2) y (5), se tiene  $\lambda_3 = (1/p) - 1 \geq 0$ , de donde  $0 < p \leq 1$ .
- iii) Si  $x^* = 0$  y  $y^* > 0$ , de (6) se sigue que  $y^* = m$ . De (5) y (2) se obtiene  $\lambda_3 = 0$ ,  $\lambda_1 = 1/(1 + m) > 0$ . De (1) y (4) se tiene  $\lambda_2 = p/(1 + m) - 1 \geq 0$ , de donde  $p \geq m + 1$ .
- iv) Si  $x^* > 0$  y  $y^* > 0$ , de (4) y (5) se tiene  $\lambda_2 = \lambda_3 = 0$ . De (1) y (2) se obtiene  $\lambda_1 = 1$ ,  $y^* = p - 1$ . De (4), (5) y (6) se tiene  $x^* = (m + 1 - p)/p > 0$ ,  $y^* = p - 1 > 0$ , de donde  $1 < p < m + 1$ .

En resumen, la solución del problema es la siguiente:

- A. Si  $0 < p \leq 1$ , entonces  $(x^*, y^*) = (m/p, 0)$ , con  $\lambda_1 = 1/p$ ,  $\lambda_2 = 0$ ,  $\lambda_3 = (1 - p)/p$ .
- B. Si  $1 < p < m + 1$ , entonces  $(x^*, y^*) = ((m + 1 - p)/p, p - 1)$ , con  $\lambda_1 = 1$ ,  $\lambda_2 = \lambda_3 = 0$ .

## Capítulo 5 Optimización

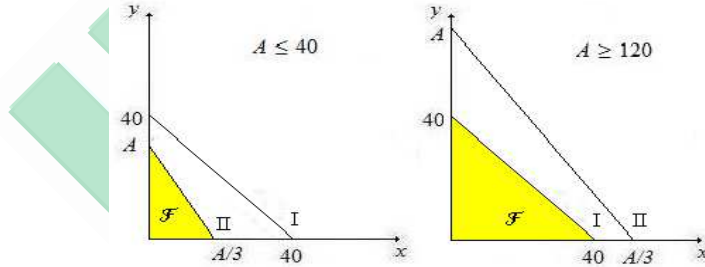
C. Si  $p \geq m + 1$ , entonces  $(x^*, y^*) = (0, m)$ , con  $\lambda_1 = 1/(1 + m)$ ,  $\lambda_2 = p/(1 + m) - 1$ ,  $\lambda_3 = 0$ .



3. Encuentra la solución al siguiente problema de maximización de la utilidad de un individuo,  $u(x, y)$ , en donde  $A \in (40, 120)$  es un parámetro:

$$\begin{aligned} \text{máx. } u(x, y) &= \frac{1}{3} \ln x + \frac{1}{3} \ln y \\ \text{s.a. } 3x + y &\leq A \quad (\text{I}) \\ x + y &\leq 40 \quad (\text{II}) \\ x &\geq 0 \quad (\text{III}) \\ y &\geq 0. \quad (\text{IV}) \end{aligned}$$

Para entender por qué se impone la condición  $A \in (40, 120)$ , considera las siguientes figuras. De acuerdo con la figura de la izquierda, si  $A \leq 40$ , entonces la restricción I sería irrelevante ( $\mathcal{F}$  es la intersección de I-IV). De acuerdo con la figura de la derecha, si  $A \geq 120$ , entonces sería irrelevante la restricción II.



Por otra parte, observa que en este ejemplo podemos ignorar las restricciones III y IV, ya que el dominio de la función objetivo está en  $\mathbb{R}^{++}$ , por lo que no puede existir solución a lo largo de los ejes coordenados. De esta manera, la función lagrangeana correspondiente está dada por

$$\mathcal{L} = \frac{1}{3} \ln x + \frac{1}{3} \ln y + \lambda_1(A - 3x - y) + \lambda_2(40 - x - y),$$

que en el óptimo debe satisfacer las condiciones de igualdad

$$\mathcal{L}_x = \frac{1}{3x} - 3\lambda_1 - \lambda_2 = 0 \quad (1)$$

### 5.3 Optimización con restricciones de desigualdad. Condiciones de Kuhn-Tucker

$$\mathcal{L}_y = \frac{1}{3y} - \lambda_1 - \lambda_2 = 0 \quad (2)$$

junto con las condiciones de holgura complementaria

$$3x + y \leq A, \quad \lambda_1 \geq 0, \quad \lambda_1(3x + y - A) = 0, \quad (3)$$

$$x + y \leq 40, \quad \lambda_2 \geq 0, \quad \lambda_2(x + y - 40) = 0. \quad (4)$$

De las igualdades en (3) y (4) se siguen  $2^2 = 4$  casos:

i) Si  $\lambda_1 = 0$  y  $\lambda_2 = 0$ , se violan las condiciones (1) y (2), ya que  $\nabla u \neq \vec{0}$ .

ii) Si en el óptimo está activa la restricción I ( $3x + y = A$ ) e inactiva la restricción II ( $x + y < 40$ ), entonces ahí se cumple

$$\begin{aligned} 3x + y &= A \\ \lambda_2 &= 0 \\ \frac{1}{3x} - 3\lambda_1 - \lambda_2 &= 0 \\ \frac{1}{3y} - \lambda_1 - \lambda_2 &= 0. \end{aligned}$$

La solución de este sistema es

$$x^* = \frac{A}{6}, \quad y^* = \frac{A}{2}, \quad \lambda_1 = \frac{2}{3A}, \quad \lambda_2 = 0.$$

Adicionalmente, como en este caso  $x^* + y^* < 40$ , por lo tanto

$$40 < A < 60.$$

iii) Si en el óptimo están activas la restricción I ( $3x + y = A$ ) y la restricción II ( $x + y = 40$ ), entonces ahí se cumple

$$\begin{aligned} 3x + y &= A \\ x + y &= 40 \\ \frac{1}{3x} - 3\lambda_1 - \lambda_2 &= 0 \\ \frac{1}{3y} - \lambda_1 - \lambda_2 &= 0. \end{aligned}$$

La solución de este sistema es

$$x^* = \frac{A - 40}{2}, \quad y^* = \frac{120 - A}{2}, \quad \lambda_1 = \frac{2(80 - A)}{3(A - 40)(120 - A)}, \quad \lambda_2 = \frac{4(A - 60)}{3(A - 40)(120 - A)}.$$

Adicionalmente, como  $\lambda_1 \geq 0$  y  $\lambda_2 \geq 0$ , por lo tanto

$$60 \leq A \leq 80.$$

## Capítulo 5 Optimización

iv) Por último, si en el óptimo está inactiva la restricción I ( $3x + y < A$ ) y activa la restricción II ( $x + y = 40$ ), entonces ahí se cumple

$$\begin{aligned}\lambda_1 &= 0 \\ x + y &= 40 \\ \frac{1}{3x} - 3\lambda_1 - \lambda_2 &= 0 \\ \frac{1}{3y} - \lambda_1 - \lambda_2 &= 0.\end{aligned}$$

La solución de este sistema es

$$x^* = 20, \quad y^* = 20, \quad \lambda_1 = 0, \quad \lambda_2 = \frac{1}{60}.$$

Adicionalmente, como en este caso  $3x^* + y^* < A$ , por lo tanto

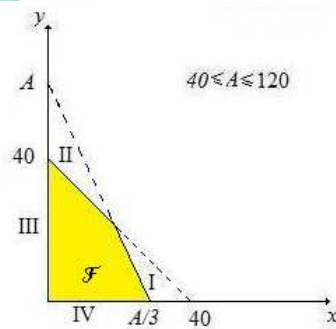
$$80 < A < 120.$$

En resumen, la solución del problema es la siguiente:

A. Si  $40 < A < 60$ , entonces  $(x^*, y^*) = \left(\frac{A}{6}, \frac{A}{2}\right)$ , con  $\lambda_1 = \frac{2}{3A}$ ,  $\lambda_2 = 0$ .

B. Si  $60 \leq A \leq 80$ , entonces  $(x^*, y^*) = \left(\frac{A-40}{2}, \frac{120-A}{2}\right)$ , con  $\lambda_1 = \frac{2(80-A)}{3(A-40)(120-A)}$ ,  $\lambda_2 = \frac{4(A-60)}{3(A-40)(120-A)}$ .

C. Si  $80 < A < 120$ , entonces  $(x^*, y^*) = (20, 20)$ , con  $\lambda_1 = 0$ ,  $\lambda_2 = \frac{1}{60}$ .



## 5.3 Optimización con restricciones de desigualdad. Condiciones de Kuhn-Tucker

### 5.3.1.3 No negatividad de las variables $x_i$

En economía, las variables independientes  $x_1, \dots, x_n$ , suelen representar cantidades, precios y otras variables que no pueden tomar valores negativos. En consecuencia, en los problemas de optimización aparecerá una o varias restricciones del tipo  $x_i \geq 0$ , además de las restricciones económicas (presupuestales, de producción, etc.).

Una manera de incorporar este tipo de restricción al método de Kuhn-Tucker consiste en introducir un multiplicador de Lagrange adicional, digamos  $\mu_i$ , por cada restricción  $x_i \geq 0$ , tal y como lo hicimos en el ejemplo 1, para la restricción  $x \geq 0$ . Así, por ejemplo, para un problema del tipo

$$\begin{aligned} \text{máx} \quad & f(x, y) \\ \text{s.a.} \quad & g(x, y) \leq c, \\ & x \geq 0, \quad (\text{es decir, } -x \leq 0) \\ & y \geq 0, \quad (\text{es decir, } -y \leq 0) \end{aligned}$$

se puede plantear una lagrangeana de la forma

$$\mathcal{L}(x, y, \lambda, \eta) = f(x, y) + \lambda(c - g(x, y)) + \mu_1 x + \mu_2 y,$$

que en el óptimo satisface las condiciones

$$\begin{aligned} \mathcal{L}_x &= f_x - \lambda g_x + \mu_1 = 0 \\ \mathcal{L}_y &= f_y - \lambda g_y + \mu_2 = 0 \\ g(x, y) &\leq c, \quad \lambda \geq 0, \quad \lambda(g(x, y) - c) = 0 \\ x &\geq 0, \quad \mu_1 \geq 0, \quad \mu_1 x = 0 \\ y &\geq 0, \quad \mu_2 \geq 0, \quad \mu_2 y = 0 \end{aligned}$$

Existe una manera alternativa de incorporar las condiciones  $x \geq 0, y \geq 0$ . Para ello, nota que la condición  $\mathcal{L}_x = 0$  implica  $\mu_1 = -(f_x - \lambda g_x)$ . Además, como  $\mu_1 \geq 0$  se tiene que  $-(f_x - \lambda g_x) \geq 0$ , es decir,

$$f_x - \lambda g_x \leq 0.$$

Esto equivale a reemplazar la condición  $\mathcal{L}_x = 0$  por las condiciones

$$\mathcal{L}_x \leq 0, \quad x \geq 0 \quad \text{y} \quad x\mathcal{L}_x = 0.$$

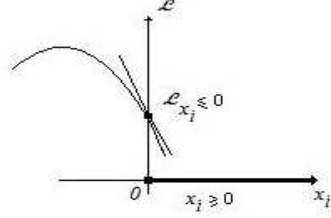
Un argumento similar se sigue para la condición de no negatividad en la variable  $y$ .

De hecho, si suponemos que la lagrangeana  $\mathcal{L}$  es una función cóncava en todas sus variables, nota que una condición de la forma

$$\mathcal{L}_{x_i} \leq 0, \quad x_i \geq 0 \quad \text{y} \quad x_i \mathcal{L}_{x_i} = 0, \quad i = 1, \dots, n$$

## Capítulo 5 Optimización

simplemente garantiza la existencia de un máximo global para  $\mathcal{L}$  en la región  $x_i \geq 0$ .



De esta manera, este segundo planteamiento consiste en proponer una lagrangeana de la siguiente forma (sin los multiplicadores  $\mu_1$  y  $\mu_2$ )

$$\mathcal{L}(x, y, \lambda) = f(x, y) + \lambda(c - g(x, y)),$$

que en el óptimo satisface las condiciones modificadas

$$\mathcal{L}_x = f_x - \lambda g_x \leq 0, \quad x \geq 0, \quad x\mathcal{L}_x = 0$$

$$\mathcal{L}_y = f_y - \lambda g_y \leq 0, \quad y \geq 0, \quad y\mathcal{L}_y = 0$$

$$g(x, y) \leq c, \quad \lambda \geq 0, \quad \lambda(g(x, y) - c) = 0,$$

como se resume a continuación.

**Condiciones de Kuhn-Tucker para el problema de variables no negativas:**  
 máx.  $f(\vec{x})$  s.a.  $g_j(\vec{x}) \leq c_j, x_i \geq 0, j = 1, \dots, m, \vec{x} \in \mathbb{R}^n$ .

1. Se construye la función lagrangeana correspondiente

$$\mathcal{L}(\vec{x}, \vec{\lambda}) = f(\vec{x}) + \sum_{j=1}^m \lambda_j(c_j - g_j(\vec{x})).$$

2. Se establecen  $n$  condiciones de holgura complementaria para las variables no negativas  $x_i$ :

$$\mathcal{L}_{x_i} \leq 0, \quad x_i \geq 0 \quad \text{y} \quad x_i \mathcal{L}_{x_i} = 0, \quad i = 1, \dots, n$$

3. Se establecen  $m$  condiciones de holgura complementaria para las restricciones  $g_j(\vec{x}) \leq c_j$ :

$$g_j(\vec{x}) \leq c_j, \quad \lambda_j \geq 0 \quad \text{y} \quad \lambda_j(g_j(\vec{x}) - c_j) = 0, \quad j = 1, \dots, m.$$

o, equivalentemente,

$$\mathcal{L}_{\lambda_j} \geq 0, \quad \lambda_j \geq 0 \quad \text{y} \quad \lambda_j \mathcal{L}_{\lambda_j} = 0, \quad j = 1, \dots, m.$$

4. Se resuelve consistentemente el sistema de ecuaciones y desigualdades.



### 5.3 Optimización con restricciones de desigualdad. Condiciones de Kuhn-Tucker

Ejemplo:

Encuentra la solución del problema

$$\text{máx. } f(x, y) = 9 - x^2 - (y - 2)^2$$

$$\text{s.a. } x + y \leq 1 \quad (\text{I})$$

$$x \geq 0 \quad (\text{II})$$

$$y \geq 0. \quad (\text{III})$$

La lagrangeana correspondiente es

$$\mathcal{L}(x, y, \lambda) = 9 - x^2 - (y - 2)^2 + \lambda(1 - x - y),$$

que en el óptimo satisface las condiciones de no negatividad

$$\mathcal{L}_x = -2x - \lambda \leq 0, \quad x \geq 0, \quad x(2x + \lambda) = 0, \quad (1)$$

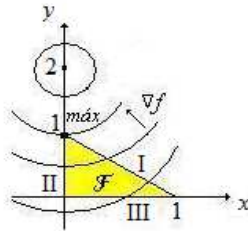
$$\mathcal{L}_y = -2(y - 2) - \lambda \leq 0, \quad y \geq 0, \quad y(2(y - 2) + \lambda) = 0, \quad (2)$$

$$x + y \leq 1, \quad \lambda \geq 0, \quad \lambda(x + y - 1) = 0. \quad (3)$$

De las tres igualdades en (1)-(3) se siguen  $2^3 = 8$  casos. Como se observa en la figura, de esos casos el único que nos interesa es aquél en donde la solución óptima ocurre simultáneamente en las fronteras de las restricciones I y II, es decir,

$$x + y = 1, \quad x = 0.$$

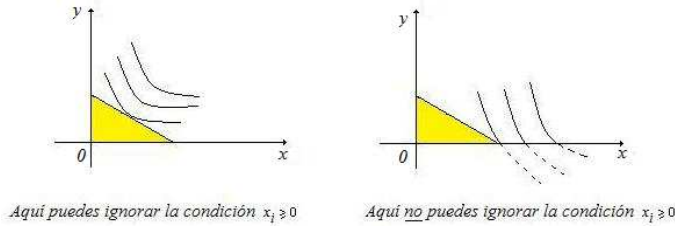
Sustituyendo esto en el sistema de desigualdades, se obtiene que la solución óptima es  $(x^*, y^*) = (0, 1)$ , con  $\lambda = 2$  y  $f^* = 8$ .



Por último, debes tener mucho cuidado en el manejo de la condición de no negatividad de las variables,  $x_i \geq 0$ , en problemas de economía. Ésta puede ser ignorada, sólo si sabes *a priori* que existe la restricción  $x_i > 0$  en el dominio de la función objetivo  $f$ , o bien, cuando la derivada parcial  $f_{x_i}$  diverge en  $x_i = 0$ . En cualquiera de estos casos, la solución óptima ocurrirá en  $x_i > 0$ . Si esto no sucede, al ignorar la restricción  $x_i \geq 0$  puedes obtener resultados erróneos cuando

## Capítulo 5 Optimización

el óptimo ocurre precisamente en  $x_i = 0$  (soluciones de esquina).



**5.3.1.4 Restricciones mixtas** En algunos problemas de optimización se tiene un conjunto de restricciones mixtas, en el sentido de que existen tanto restricciones de igualdad, del tipo  $g_j = c_j$ , como restricciones de desigualdad, del tipo  $g_l \leq c_l$ . Es posible demostrar (ver Simon & Blume) que en este caso el problema de optimización es una combinación de un problema de Lagrange y uno de Kuhn-Tucker, como se muestra en el siguiente ejemplo.

Ejemplo:

Plantea el siguiente problema de optimización:

$$\begin{aligned} \text{máx. } & f(x, y, z) \\ \text{s.a. } & g(x, y, z) = c \\ & h(x, y, z) \leq d. \end{aligned}$$

En este caso, podemos plantear una lagrangeana de la forma

$$\mathcal{L}(x, y, z, \lambda_1, \lambda_2) = f(x, y, z) + \lambda_1(c - g(x, y, z)) + \lambda_2(d - h(x, y, z)),$$

con condiciones de primer orden

$$\begin{aligned} \mathcal{L}_x &= 0 \\ \mathcal{L}_y &= 0 \\ \mathcal{L}_z &= 0 \\ g(x, y, z) &= c \\ h(x, y, z) &\leq d, \quad \lambda_2 \geq 0 \quad \text{y} \quad \lambda_2(h(x, y, z) - d) = 0. \end{aligned}$$

### 5.3.1.5 Condiciones suficientes para un máximo global

Una vez establecidas las condiciones necesarias de optimalidad en el problema de Kuhn-Tucker, buscamos condiciones que garanticen que el óptimo es un máximo global. Al igual que en un problema de Lagrange (sección 5.2), se puede demostrar que una condición suficiente para máximo es que sea cóncava la función lagrangeana  $\mathcal{L} = f(\vec{x}) + \sum_j \lambda_j(c_j - g_j(\vec{x}))$ . Dado que en este caso  $\lambda_j > 0$ , esto

### 5.3 Optimización con restricciones de desigualdad. Condiciones de Kuhn-Tucker

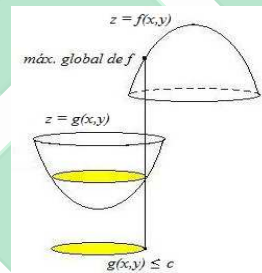
se satisface si la función objetivo  $f$  es cóncava y cada función de restricción  $g_j(\vec{x})$  es convexa, como se establece el siguiente teorema.

**Teorema de suficiencia para un máximo global.** Si el punto  $\vec{x}^*$  satisface las condiciones de Kuhn-Tucker para el problema

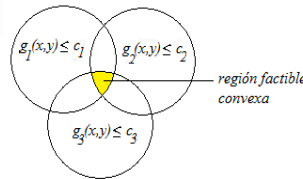
$$\begin{aligned} &\text{máx. } f(\vec{x}) \\ &\text{s.a. } g_j(\vec{x}) \leq c_j, \quad j = 1, \dots, m, \end{aligned}$$

donde  $f \in C^1$  es cóncava y cada función  $g_j \in C^1$  es convexa, entonces el máximo global de  $f$  se alcanza en  $\vec{x}^*$ .

El requisito de concavidad para la función objetivo  $f$  es claro en un problema de maximización. La convexidad de las funciones  $g_j$  se explica de la siguiente manera. Cuando hay una sola restricción de desigualdad,  $g(\vec{x}) \leq c$ , pedir que  $z = g(\vec{x})$  sea una función convexa te garantiza que su contorno inferior o región factible  $g(\vec{x}) \leq c$  sea una región convexa.



Cuando hay varias restricciones de desigualdad, la región factible es la intersección de todas las restricciones  $g_j \leq c_j$ . Pedir que cada función  $z = g_j(\vec{x})$  sea convexa garantiza que sus contornos inferiores  $g_j(\vec{x}) \leq c_j$  son conjuntos convexos, de modo que su intersección también será una región convexa.



Por último, sabemos que muchos problemas de optimización en economía se consideran funciones objetivo cuasicóncavas, en lugar de cóncavas. En vista de ello, a continuación presentamos un teorema de suficiencia alternativo, debido a K.J. Arrow y A.C. Enthoven, que constituye una versión menos restrictiva que la del teorema anterior.

**Teorema de Arrow-Enthoven para un máximo global.** Si el punto  $\vec{x}^*$  satisface las condiciones de Kuhn-Tucker para el problema

$$\begin{aligned} \text{máx. } & f(\vec{x}) \\ \text{s.a. } & g_j(\vec{x}) \leq c_j, \quad j = 1, \dots, m, \end{aligned}$$

donde  $f \in C^1$  es cuasicóncava y cada función  $g_j \in C^1$  es cuasiconvexa, entonces el máximo global de  $f$  se alcanza en  $\vec{x}^*$ .

Nota que la condición de cuasiconvexidad para las funciones de restricción  $g_j$  garantiza que sus contornos inferiores  $g_j(\vec{x}) \leq c_j$  sean regiones convexas, como ocurre con las funciones de restricción convexas.

### 5.3.2 Problemas de minimización

A diferencia de un problema de Lagrange, en un problema de Kuhn-Tucker el planteamiento para un caso de maximización difiere del de un caso de minimización. Por razones de claridad, en la discusión anterior nos enfocamos exclusivamente en problemas de maximización. A continuación presentamos cómo adaptar esos resultados para el caso de problemas de minimización.

Las condiciones de Kuhn-Tucker para el problema de minimización se basan en un razonamiento similar al del problema de maximización. En ambos casos, un aspecto importante se refiere al formato en el que debe expresarse la restricción de desigualdad. En el caso de maximización, el formato  $g(x, y) \leq c$  para la restricción, junto con la condición  $\lambda \geq 0$ , garantizan que al ampliarse la región factible se obtendrá el mejor máximo posible para el problema. Para que esto mismo ocurra en un problema de minimización, pero aún conservando la condición  $\lambda \geq 0$ , será necesario expresar la restricción en la forma  $g(x, y) \geq c$ , como se justificará más adelante, con las condiciones de segundo orden.

Nos interesa, entonces, resolver un problema de minimización, de la forma

$$\begin{aligned} \text{mín. } & f(x, y) \\ \text{s.a. } & g(x, y) \geq c. \end{aligned}$$

Para ello, se parte de la función lagrangeana

$$\mathcal{L}(x, y, \lambda) = f(x, y) + \lambda(c - g(x, y)),$$

que presenta la misma forma funcional que en el caso de maximización. Para esta función, se establecen las condiciones necesarias de primer orden,

$$\begin{aligned} \mathcal{L}_x &= f_x(x, y) - \lambda g_x(x, y) = 0 \\ \mathcal{L}_y &= f_y(x, y) - \lambda g_y(x, y) = 0, \end{aligned}$$

### 5.3 Optimización con restricciones de desigualdad. Condiciones de Kuhn-Tucker

que nuevamente representan la condición de tangencia ( $\nabla f = \lambda \nabla g$ ) en el óptimo. En relación con las condiciones de holgura complementaria, ahora deberá verificarse

$$\mathcal{L}_\lambda \leq 0, \quad \lambda \geq 0 \quad \text{y} \quad \lambda \mathcal{L}_\lambda = 0,$$

o equivalentemente,

$$g(x, y) \geq c, \quad \lambda \geq 0 \quad \text{y} \quad \lambda(g(x, y) - c) = 0.$$

La primera condición de holgura impone el cumplimiento de la restricción en el óptimo. La condición  $\lambda \geq 0$  garantiza que un pequeño incremento en  $c$  generará un menor valor para el óptimo de  $f$ . La última condición establece que  $\lambda = 0$  (optimización libre) o  $g(x, y) = c$  (problema de Lagrange) en el óptimo. Estas son las condiciones de Kuhn-Tucker para el problema de minimización.

#### Condiciones de Kuhn-Tucker para el problema de minimización

mín.  $f(x, y)$ , s.a.  $g(x, y) \geq c$ .

1. Se construye la función lagrangeana correspondiente

$$\mathcal{L}(x, y, \lambda) = f(x, y) + \lambda(c - g(x, y)).$$

2. Se establecen las condiciones de primer orden para  $\mathcal{L}$ , con respecto a las variables  $x$  y  $y$  solamente,

$$\mathcal{L}_x = f_x - \lambda g_x = 0,$$

$$\mathcal{L}_y = f_y - \lambda g_y = 0.$$

3. Se establece las condiciones de holgura complementaria

$$\mathcal{L}_\lambda \leq 0, \quad \lambda \geq 0 \quad \text{y} \quad \lambda(g(x, y) - c) = 0,$$

o equivalentemente,

$$g(x, y) \geq c, \quad \lambda \geq 0 \quad \text{y} \quad \lambda(g(x, y) - c) = 0.$$

4. Se resuelve, consistentemente, el sistema de ecuaciones y desigualdades.

Ejemplo:

Encuentra la solución al problema

$$\text{mín. } f(x, y) = -y$$

$$\text{s.a. } x^2 + y^2 \leq 1.$$

## Capítulo 5 Optimización

Primeramente escribimos el problema en un formato adecuado, es decir,

$$\text{mín. } f(x, y) = -y$$

$$\text{s.a. } -x^2 - y^2 \geq -1.$$

Así, la lagrangeana correspondiente está dada por

$$\mathcal{L}(x, y, \lambda) = -y + \lambda(-1 + x^2 + y^2).$$

Las condiciones de primer orden en  $x$  y  $y$  son

$$\mathcal{L}_x = 2\lambda x = 0 \quad (1)$$

$$\mathcal{L}_y = -1 + 2\lambda y = 0, \quad (2)$$

que deberán resolverse junto con las condiciones de holgura complementaria

$$x^2 + y^2 \leq 1, \quad (3) \quad \lambda \geq 0, \quad (4) \quad \lambda(x^2 + y^2 - 1) = 0. \quad (5)$$

Nota que la desigualdad en (3) está escrita en el formato original, sin que esto afecte el resultado; el formato  $\geq$  es importante sólo en las ecuaciones (1) y (2). De acuerdo con la igualdad (5), se tienen dos casos,  $\lambda = 0$  o  $x^2 + y^2 = 1$ . i) Si  $\lambda = 0$ , la ecuación (2) conduce a una inconsistencia ( $-1 = 0$ !), por lo que esta opción se descarta. ii) Si  $x^2 + y^2 = 1$ , se obtiene el sistema de ecuaciones

$$2\lambda x = 0 \quad (1)$$

$$-1 + 2\lambda y = 0 \quad (2)$$

$$x^2 + y^2 = 1, \quad (6)$$

correspondiente a un punto frontera ( $x^2 + y^2 = 1$ ) de la restricción. El sistema tiene dos soluciones posibles,

$$(x_1, y_1) = (0, -1) \quad \text{y} \quad (x_2, y_2) = (0, 1).$$

Sustituyendo  $(x_1, y_1)$  en (2) se obtiene

$$\lambda_1 = \frac{1}{2y_1} = -\frac{1}{2} < 0.$$

Esto viola la condición (4), por lo que esta opción se descarta. Por otra parte, sustituyendo  $(x_2, y_2)$  en (2) se obtiene

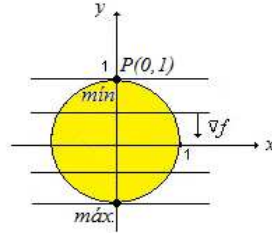
$$\lambda_2 = \frac{1}{2y_2} = \frac{1}{2} > 0,$$

que satisface la condición (4). Así, sólo los valores  $x_2, y_2, \lambda_2$ , satisfacen las condiciones (1)-(5) en su totalidad. Concluimos que el valor mínimo de  $f$  ocurre en el punto frontera  $P(0, 1)$ , con  $\lambda_2 = 1/2$  y  $f^* = -1$ .

La siguiente figura muestra la región factible  $\mathcal{F}$ ,  $x^2 + y^2 \leq 1$ , y algunas curvas de nivel de  $f$ , dadas por  $-y = \text{const}$ . Tomando en cuenta la dirección de

### 5.3 Optimización con restricciones de desigualdad. Condiciones de Kuhn-Tucker

crecimiento de  $f$ , dada por el vector gradiente  $\nabla f$ , es claro que la solución óptima ocurre en el punto frontera  $P(0, 1)$  de la restricción, en donde  $f$  toma su mínimo valor posible.



Para el caso con varias restricciones, el método de Kuhn-Tucker se generaliza de la siguiente manera.

**Condiciones de Kuhn-Tucker para el problema** mín.  $f(\vec{x})$  s.a.  $g_j(\vec{x}) \geq c_j$ ,  $j = 1, \dots, m$ ,  $\vec{x} \in \mathbb{R}^n$ .

1. Se construye la función lagrangeana correspondiente

$$\mathcal{L}(\vec{x}, \vec{\lambda}) = f(\vec{x}) + \sum_{j=1}^m \lambda_j (c_j - g_j(\vec{x})).$$

2. Se establecen  $n$  condiciones de primer orden para  $\mathcal{L}$ , en  $x_i$ ,  $i = 1, \dots, n$

$$\mathcal{L}_{x_i} = 0, \quad i = 1, \dots, n.$$

3. Se establecen  $m$  condiciones de holgura complementaria

$$g_j(\vec{x}) \geq c_j, \quad \lambda_j \geq 0 \quad \text{y} \quad \lambda_j (g_j(\vec{x}) - c_j) = 0, \quad j = 1, \dots, m,$$

o equivalentemente,

$$\mathcal{L}_{\lambda_j} \leq 0, \quad \lambda_j \geq 0 \quad \text{y} \quad \lambda_j \mathcal{L}_{\lambda_j} = 0.$$

4. Se resuelve consistentemente el sistema de ecuaciones y desigualdades.

Ejemplo:

Encuentra la solución al problema

$$\text{mín. } f(x, y) = x^2 + y^2$$

$$\text{s.a. } x + y \leq 2 \quad (\text{I})$$

$$x + 2y \geq 2 \quad (\text{II})$$

$$x \geq 0. \quad (\text{III})$$

## Capítulo 5 Optimización

Primeramente, escribimos las restricciones en el formato adecuado,

$$\begin{aligned} \text{mín. } & f(x, y) = x^2 + y^2 \\ \text{s.a. } & -x - y \geq -2 \\ & x + 2y \geq 2 \\ & x \geq 0. \end{aligned}$$

La lagrangeana correspondiente está dada por

$$\mathcal{L}(x, y, \lambda_1, \lambda_2, \lambda_3) = x^2 + y^2 + \lambda_1(-2 + x + y) + \lambda_2(2 - x - 2y) + \lambda_3(-x),$$

que satisface las condiciones de igualdad

$$\mathcal{L}_x = 2x + \lambda_1 - \lambda_2 - \lambda_3 = 0, \quad (1)$$

$$\mathcal{L}_y = 2y + \lambda_1 - 2\lambda_2 = 0, \quad (2)$$

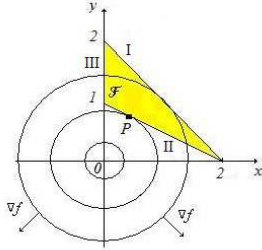
junto con las condiciones de holgura complementaria

$$x + y \leq 2, \quad \lambda_1 \geq 0, \quad \lambda_1(x + y - 2) = 0, \quad (3)$$

$$x + 2y \geq 2, \quad \lambda_2 \geq 0, \quad \lambda_2(x + 2y - 2) = 0, \quad (4)$$

$$x \geq 0, \quad \lambda_3 \geq 0, \quad \lambda_3 x = 0. \quad (5)$$

De las igualdades en (3), (4) y (5) se siguen  $2^3 = 8$  casos, algunos de los cuales pueden eliminarse mediante un análisis gráfico del problema. Para ello, dibujamos la región factible  $\mathcal{F}$ , algunas curvas de nivel de  $f$ , así como el gradiente  $\nabla f$ , para determinar la dirección de crecimiento de  $f$ .



En la gráfica se observa que el punto  $P$  que minimiza a  $f$  ocurre en donde la restricción II está activa,

$$x + 2y = 2, \quad (6)$$

mientras que las restricciones I y III no lo están ( $x + y < 2$  y  $x > 0$ ). De acuerdo con las igualdades en (3) y (5), esta última afirmación implica que

$$\lambda_1 = 0 \text{ y } \lambda_3 = 0. \quad (7)$$

Resolvemos el sistema de igualdades (1), (2), (6) y (7) para  $x, y, \lambda_1, \lambda_2$  y  $\lambda_3$ , y verificamos el cumplimiento de la condición  $\lambda_2 \geq 0$ . Con esto, se tiene que el



### 5.3 Optimización con restricciones de desigualdad. Condiciones de Kuhn-Tucker

valor mínimo de  $f$  sucede en el punto  $P\left(\frac{2}{5}, \frac{4}{5}\right)$ , con  $\lambda_2 = \frac{4}{5} > 0$ ,  $\lambda_1 = \lambda_3 = 0$  y  $f^* = \frac{20}{25}$ .

**Condiciones de Kuhn-Tucker para el problema de variables no negativas:**  
mín.  $f(\vec{x})$  s.a.  $g_j(\vec{x}) \geq c_j$ ,  $x_i \geq 0$ ,  $j = 1, \dots, m$ ,  $\vec{x} \in \mathbb{R}^n$ .

1. Se construye la función lagrangeana correspondiente

$$\mathcal{L}(\vec{x}, \vec{\lambda}) = f(\vec{x}) + \sum_{j=1}^m \lambda_j (c_j - g_j(\vec{x})).$$

2. Se establecen  $n$  condiciones de holgura complementaria para las variables no negativas  $x_i$ :

$$\mathcal{L}_{x_i} \geq 0, \quad x_i \geq 0 \quad \text{y} \quad x_i \mathcal{L}_{x_i} = 0, \quad i = 1, \dots, n$$

3. Se establecen  $m$  condiciones de holgura complementaria para las restricciones  $g_j \geq c_j$ :

$$\mathcal{L}_{\lambda_j} \leq 0, \quad \lambda_j \geq 0 \quad \text{y} \quad \lambda_j \mathcal{L}_{\lambda_j} = 0, \quad j = 1, \dots, m,$$

o equivalentemente,

$$g_j(\vec{x}) \geq c_j, \quad \lambda_j \geq 0 \quad \text{y} \quad \lambda_j (g_j(\vec{x}) - c_j) = 0, \quad j = 1, \dots, m.$$

4. Se resuelve consistentemente el sistema de igualdades y desigualdades.

Por último, las condiciones de suficiencia para un mínimo global son las siguientes.

**Teorema de suficiencia para un mínimo global.** Si el punto  $\vec{x}^*$  satisface las condiciones de Kuhn-Tucker para el problema

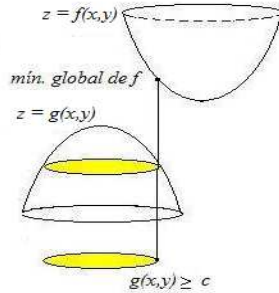
$$\begin{aligned} &\text{mín. } f(\vec{x}) \\ &\text{s.a. } g_j(\vec{x}) \geq c_j, \quad j = 1, \dots, m, \end{aligned}$$

donde  $f \in C^1$  es convexa y cada función  $g_j \in C^1$  es cóncava, entonces el mínimo global de  $f$  se alcanza en  $\vec{x}^*$ .

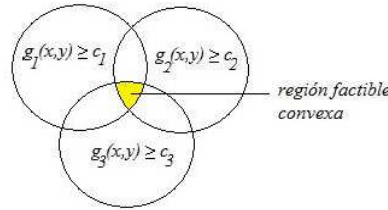
Es claro el requisito de convexidad para la función objetivo  $f$  en un problema de minimización. En el caso de una sola restricción de desigualdad,  $g(\vec{x}) \geq c$ , pedir

## Capítulo 5 Optimización

que  $z = g(\vec{x})$  sea una función cóncava te garantiza que su contorno superior o región factible  $g(\vec{x}) \geq c$  sea una región convexa.



Cuando hay varias restricciones de desigualdad, pedir que cada función  $z = g_j(\vec{x})$  sea cóncava garantiza que sus contornos superiores  $g_j(\vec{x}) \geq c_j$  son conjuntos convexos, de modo que su intersección también será una región convexa.



La generalización de este teorema para funciones cuasiconvexas es la siguiente.

**Teorema de Arrow-Enthoven para un mínimo global.** Si el punto  $\vec{x}^*$  satisface las condiciones de Kuhn-Tucker para el problema

$$\begin{aligned} \text{mín. } & f(\vec{x}) \\ \text{s.a. } & g_j(\vec{x}) \geq c_j, \quad j = 1, \dots, m, \end{aligned}$$

donde  $f \in C^1$  es cuasiconvexa y cada función  $g_j \in C^1$  es cuasicóncava, entonces el mínimo global de  $f$  se alcanza en  $\vec{x}^*$ .

### 5.3.3 Cualificación de las restricciones: ¿cuándo fallan las condiciones de Kuhn-Tucker?

Similarmente al caso de Lagrange, las condiciones de Kuhn-Tucker pueden fallar al ocurrir ciertas *irregularidades de frontera*, en las que se viola la cualificación de las restricciones  $\nabla f^* = \sum_{j=1}^m \lambda_j^* \nabla g_j^*$  en el óptimo, que en el caso particular de una única restricción se reduce a  $\nabla f^* = \lambda^* \nabla g^*$ . Esto se ilustra en los ejemplos que presentaremos a continuación.

### 5.3 Optimización con restricciones de desigualdad. Condiciones de Kuhn-Tucker

Ejemplos:

#### 1. Encuentra la solución al problema

$$\begin{aligned} \text{máx.} \quad & f(x, y) = -y \\ \text{s.a.} \quad & x^2 - y^3 \leq 0. \end{aligned}$$

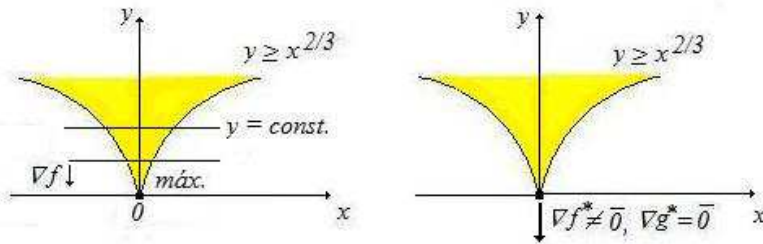
La lagrangeana correspondiente es la función

$$\mathcal{L}(x, y, \lambda) = -y + \lambda (y^3 - x^2),$$

cuyas condiciones de Kuhn-Tucker son

$$\begin{aligned} \mathcal{L}_x &= -2\lambda x = 0, \\ \mathcal{L}_y &= -1 + 3\lambda y^2 = 0, \\ x^2 - y^3 &\leq 0, \quad \lambda \geq 0, \quad \lambda (x^2 - y^3) = 0. \end{aligned}$$

Es fácil verificar que no existe solución a este sistema de ecuaciones y desigualdades. Sin embargo, el método gráfico sí nos permite obtener que la función  $f$  alcanza su máximo global,  $f^* = 0$ , en el punto  $(0, 0)$ .



Nuevamente, aquí la dificultad consiste en que en el óptimo  $\nabla f^* = -\hat{j} \neq \vec{0}$ , mientras que  $\nabla g^* = \vec{0}$ , de modo que no se verifica la condición  $\nabla f^* = \lambda^* \nabla g^*$ . Esto se debe a que la función  $z = g(x, y) = x^2 - y^3$  alcanza su óptimo en un punto frontera de su contorno inferior  $CI_g(0)$ ,  $x^2 - y^3 \leq 0$ , que precisamente coincide con el óptimo de  $f$  sujeto a esa restricción.

#### 2. Encuentra la solución al problema

$$\begin{aligned} \text{máx.} \quad & f(x, y) = x \\ \text{s.a.} \quad & y - (1 - x)^3 \leq 0, \\ & x, y \geq 0. \end{aligned}$$

La lagrangeana correspondiente es la función

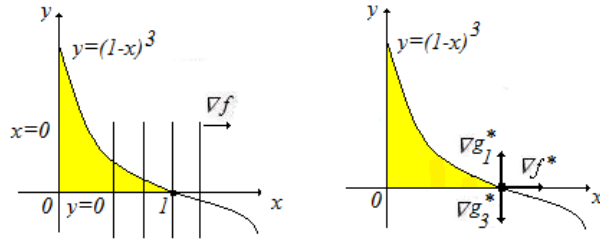
$$\mathcal{L}(x, y, \lambda_1, \lambda_2, \lambda_3) = x + \lambda_1 ((1 - x)^3 - y) + \lambda_2 x + \lambda_3 y,$$

## Capítulo 5 Optimización

cuyas condiciones de Kuhn-Tucker son

$$\begin{aligned}\mathcal{L}_x &= 1 - 3\lambda_1(1-x)^2 + \lambda_2 = 0 \\ \mathcal{L}_y &= \lambda_1 + \lambda_3 = 0 \\ y - (1-x)^3 &\leq 0, \quad \lambda_1 \geq 0, \quad \lambda_1(y - (1-x)^3) = 0 \\ x &\geq 0, \quad \lambda_2 \geq 0, \quad \lambda_2 x = 0, \\ y &\geq 0, \quad \lambda_3 \geq 0, \quad \lambda_3 y = 0.\end{aligned}$$

Con un poco de paciencia puedes verificar que no existe solución que satisfaga estas condiciones. Sin embargo, el método gráfico sí nos permite obtener que la función  $f$  alcanza su máximo global,  $f^* = 1$ , en el punto  $(1, 0)$ .



La dificultad se debe a que en el óptimo no se satisface la condición  $\nabla f^* = \lambda_1^* \nabla g_1^* + \lambda_2^* \nabla g_2^* + \lambda_3^* \nabla g_3^*$ , para las condiciones que están activas en ese punto, que son  $g_1$  y  $g_3$ . En efecto, es fácil comprobar que  $\nabla f^* = \hat{i}$ ,  $\nabla g_1^* = \hat{j}$  y  $\nabla g_3^* = -\hat{j}$ , de modo que no existen multiplicadores  $\lambda_1^*, \lambda_3^* \geq 0$  tales que

$$\begin{aligned}\hat{i} &= \lambda_1^* \hat{j} + \lambda_3^* (-\hat{j}) \\ &= (\lambda_1^* - \lambda_3^*) \hat{j}.\end{aligned}$$

Esto se debe a que al ser  $\nabla g_3^* = -\nabla g_1^*$  el conjunto de gradientes  $\{\nabla g_j^*\}$  en el óptimo no es linealmente independiente. Para una presentación más detallada de este tema te recomiendo consultar el libro de Simon & Blume.

### 5.4 Teorema de la envolvente

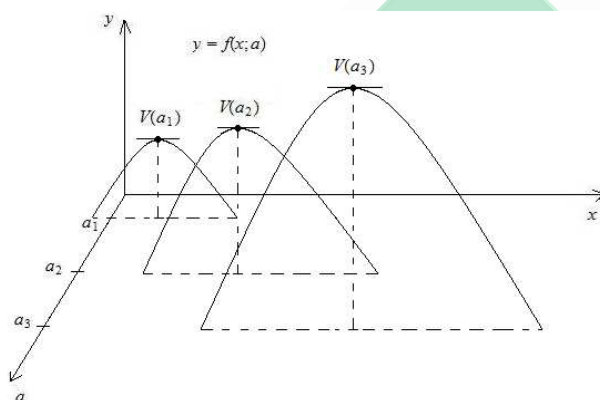
En muchas aplicaciones a economía, tanto la función objetivo  $f$  como el conjunto de restricciones  $\{g_j\}$  dependen no sólo de las variables independientes  $x_1, \dots, x_n$ , o *variables de decisión*, sino también de una colección de parámetros  $a_1, \dots, a_k$ , o *variables exógenas*. Una vez que se optimiza respecto a las primeras, el óptimo obtenido depende del valor de los parámetros. Tiene sentido entonces preguntarse qué efecto tiene sobre el valor óptimo de la función algún posible cambio en los

## 5.4 Teorema de la envolvente

parámetros. El *teorema de la envolvente* proporciona una respuesta a esta pregunta, y es válido en general para cualquier problema de optimización que involucre parámetros, ya sea en optimización libre, como en optimización restringida.

### 5.4.1 Optimización libre

Considera primero el caso más simple de maximización de una función diferenciable  $f(x; a)$  de la variable independiente,  $x$ , cuya forma funcional contiene un parámetro  $a$ . Aquí la notación para  $f$  indica que la optimización se lleva a cabo con respecto a la variable que está a la izquierda del punto y coma.



Se trata, entonces, de resolver el problema

$$\max_x f(x; a).$$

El punto  $x^*$  que maximiza  $f$  se obtiene de la condición de primer orden

$$\left. \frac{\partial f(x; a)}{\partial x} \right|_{x=x^*} = 0,$$

de modo que éste depende del parámetro  $a$ , es decir,

$$x^* = x^*(a).$$

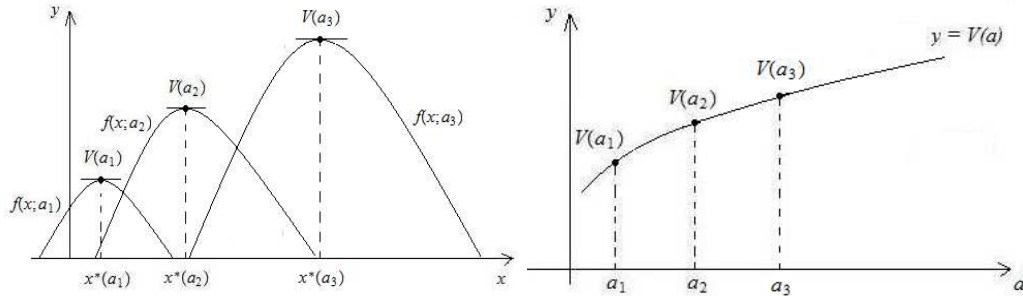
Al sustituir  $x^*$  en la función  $f(x; a)$  se obtiene su valor máximo  $V(a)$ , dado por

$$V(a) = f(x^*(a); a),$$

que es una función del parámetro  $a$ . La función  $V(a)$  se conoce como la *función valor*, y describe el comportamiento de los valores máximos de  $f$  al cambiar el parámetro  $a$ . De acuerdo con la figura anterior, en la figura de la izquierda se muestra las gráficas de  $y = f(x; a)$  como función de  $x$ , evaluada en dos valores arbitrarios  $a_2 > a_1$  del parámetro  $a$ . Podría suceder, por ejemplo, que el valor

## Capítulo 5 Optimización

máximo  $V(a)$  de  $f$  inicialmente se incrementa y luego disminuya, al incrementarse  $a$ , como se muestra en la figura de la derecha.



Suponiendo que la función valor  $V(a)$  es diferenciable, entonces el cambio en  $V$  al incrementarse el valor de  $a$  está dado por la derivada  $dV(a)/da$ . Como  $V(a) = f(x^*(a); a)$ , el cálculo de esta derivada involucra el efecto directo dado por la dependencia explícita  $f(\cdot; a)$  de  $f$  con  $a$ , así como el efecto indirecto dado por la dependencia implícita  $f(x^*(a); \cdot)$  a través de  $x^*(a)$ . De acuerdo con la regla de la cadena se tiene, entonces,

$$\begin{aligned} \frac{dV(a)}{da} &= \frac{df(x^*(a); a)}{da} \\ &= \frac{\partial f^*}{\partial x^*} \frac{dx^*}{da} + \frac{\partial f^*}{\partial a} \\ &= 0 + \frac{\partial f^*}{\partial a}, \end{aligned}$$

efecto indirecto      efecto directo

en donde se utilizó la condición de primer orden,  $\partial f^*/\partial x^* = 0$ . Nota que el término igual a 0 corresponde precisamente al efecto indirecto sobre  $V$  causado por el cambio en  $x^*$  al variar  $a$ , de modo que sólo sobrevive el efecto directo de  $a$  en  $f^*$ , dado por  $\partial f^*/\partial a$ , es decir,

$$\frac{dV(a)}{da} = \frac{\partial f(x^*; a)}{\partial a}.$$

Así, la condición de primer orden garantiza que en el óptimo se puede ignorar la dependencia implícita de  $f(x^*; a)$  con  $a$  a través de  $x^*(a)$ , considerando sólo la dependencia explícita de  $f(x^*; a)$  con  $a$ , como si  $x^*$  estuviera fijo. Esta igualdad constituye la versión más simple del teorema de la envolvente para optimización no restringida para el caso general de varios parámetros, que enunciaremos más adelante en esta sección. Antes de hacerlo, discutamos primero el significado geométrico del resultado obtenido para el caso de un solo parámetro  $a$ .

## 5.4 Teorema de la envolvente

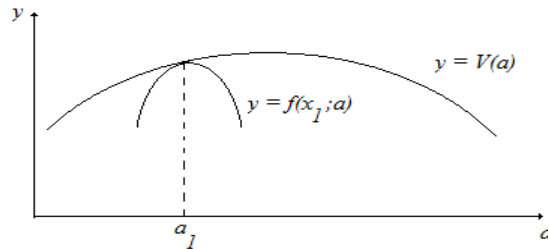
El teorema de la envolvente se refiere a dos funciones diferentes de  $a$ , dadas por  $V(a)$  y  $f(x^*; a)$ . La función valor  $V(a)$  resulta de la maximización con respecto a  $x$  de la función  $f(x; a)$ , para cada valor de  $a$ . Su forma funcional se obtiene al sustituir el nivel óptimo  $x = x^*(a)$  en  $f(x; a)$ , por lo que  $V$  depende de  $a$  solamente. En contraste, la función  $f(x^*; a)$  no proviene de una maximización. Su forma funcional se obtiene al evaluar  $x$  en un valor fijo  $x^*$ , por lo que  $f(x^*; a)$  queda expresada en términos de la constante  $x^*$  y el parámetro  $a$ . De acuerdo con el teorema de la envolvente, para cada valor del parámetro  $a$  se satisfacen las siguientes dos condiciones:

$$(1) \quad V(a) = f(x; a)|_{x=x^*}$$

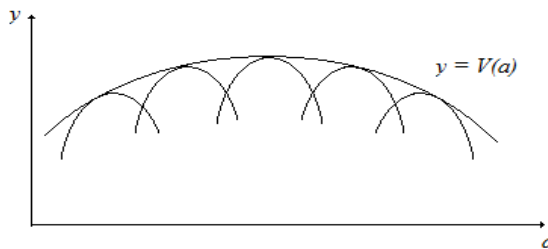
$$(2) \quad \frac{dV(a)}{da} = \left. \frac{\partial f(x; a)}{\partial a} \right|_{x=x^*},$$

que establecen la tangencia de estas dos funciones en ese punto.

La siguiente figura ilustra las condiciones de tangencia (1) y (2) para un problema de *maximización*. La curva superior representa la función  $V(a)$ , mientras que la curva inferior representa la función  $f(x_1; a)$  evaluada en un valor fijo  $x = x_1$ . Ambas son tangentes precisamente en aquel valor  $a = a_1$  que corresponde al óptimo  $x_1 = x^*(a_1)$ .



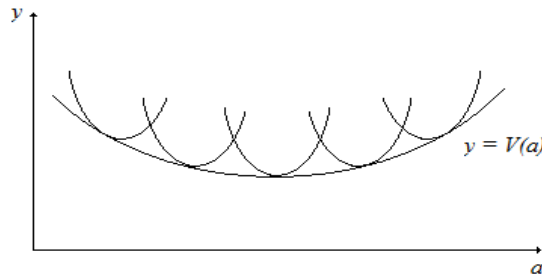
Para cada selección  $x = x_i$  se obtiene una curva distinta  $f(x_i; a)$ , como se ilustra en la siguiente figura. En ella se muestra una colección de curvas, envueltas superiormente por la función  $V(a)$  de los valores máximos de  $f$ , conocida también como la *curva envolvente*. El punto de tangencia de la envolvente con cada  $f(x_i; a)$  ocurre precisamente en aquel valor  $a_i$  que satisface la condición  $x_i = x^*(a_i)$ .



## Capítulo 5 Optimización

En problemas de maximización la envolvente  $V(a)$  nunca estará por debajo de las curvas  $y = f(x_i; a)$ , siendo además menos cóncava que éstas. Esto significa que  $V(a)$  representa el “mejor máximo posible” de la función  $f$  con respecto al parámetro  $a$ , y éste ocurre cuando se relaja la variable  $x$  de tal modo que se maximice  $f$ . En contraste, un máximo “menos bueno” se obtiene al evaluar la función  $f(x; a)$  en un valor fijo  $x = x_i$ , quedando ésta como función de  $a$ . Las curvas coinciden sólo cuando  $a$  satisface la condición de óptimo  $x_i = x^*(a_i)$ .

Para problemas de *minimización* se sigue un razonamiento similar, pero ahora las curvas son envueltas inferiormente por  $V(a)$ . En ese caso, la curva envolvente  $V(a)$  nunca estará por encima de las curvas  $f(x_i; a)$ , y además será menos convexa que éstas, como se muestra en la siguiente figura.



En este caso,  $V(a)$  representa el “mejor mínimo posible” de la función  $f$  con respecto al parámetro  $a$ , comparado con el obtenido al evaluar la función  $f(x; a)$  en un valor fijo  $x = x_i$ , con excepción de aquellos valores  $a_i$  correspondientes a la selección  $x_i = x^*(a_i)$ , en donde la  $i$ -ésima curva es tangente a  $V(a)$ .

Ejemplos:

1. Considera el problema de maximización

$$\max_x f(x; a) = 8x - \frac{x^2}{a^2},$$

con  $a > 0$ . En este caso, la condición de primer orden correspondiente es

$$\left. \frac{\partial f}{\partial x} \right|_{x=x^*} = 8 - \frac{2x}{a^2} \Big|_{x=x^*} = 0,$$

cuya solución está dada por

$$x^*(a) = 4a^2.$$

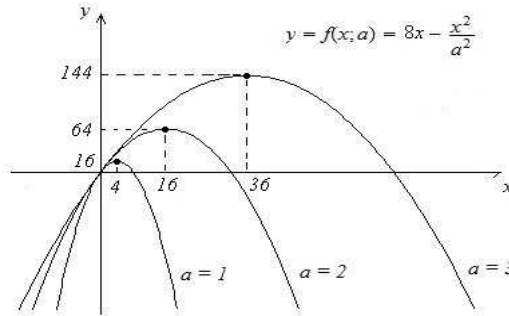
De este modo, el valor máximo de  $f$  es  $V(a) = 8(4a^2) - \frac{(4a^2)^2}{a^2}$ , es decir,

$$V(a) = 16a^2.$$



## 5.4 Teorema de la envolvente

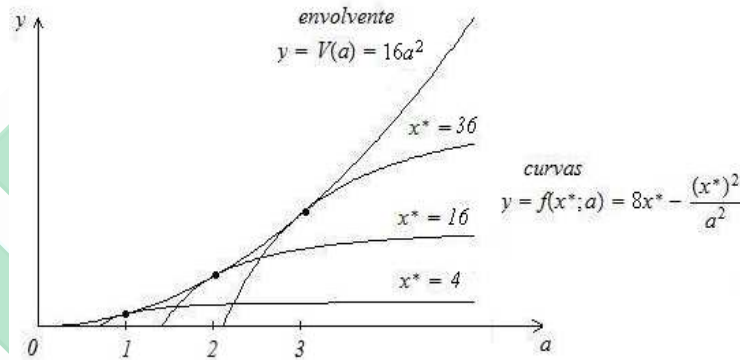
La siguiente figura muestra las gráficas de  $f(x; a) = 8x - x^2/a^2$  en función de  $x$ , para  $a = 1, 2, 3$ . En cada caso, el máximo ocurre en  $(x^*(a), V(a)) = (4a^2, 16a^2)$ .



Para ilustrar el teorema de la envolvente notamos que la función  $f(x^*; a)$  está dada simplemente por

$$f(x^*; a) = 8x^* - \frac{(x^*)^2}{a^2}.$$

La siguiente figura muestra las gráficas de la función  $V(a) = 16a^2$ , así como las curvas  $f(x^*; a) = 8x^* - (x^*)^2/a^2$  en función de  $a$ , para  $x^* = 4, 16, 36$ .



Aquí se observa que las curvas  $f(x^*; a)$  son tangentes a la envolvente  $V(a)$  en aquellos valores de  $a$  que satisfacen la condición de óptimo

$$a = \sqrt{\frac{x^*}{4}}.$$

En efecto, en esos puntos se cumple

$$(1) \quad f(x^*; a)|_{x^*=4a^2} = 8x^* - \frac{(x^*)^2}{a^2} \Big|_{x^*=4a^2} = 8(4a^2) - \frac{(4a^2)^2}{a^2} = 16a^2 = V(a)$$

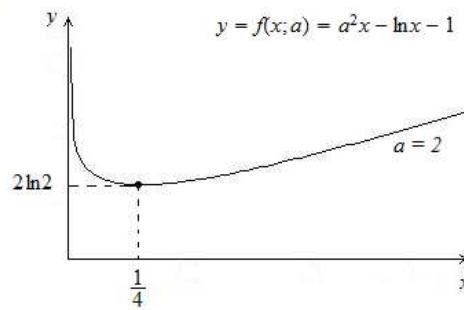
$$(2) \quad \frac{\partial f(x^*; a)}{\partial a} \Big|_{x^*=4a^2} = \frac{2(x^*)^2}{a^3} \Big|_{x^*=4a^2} = \frac{2(4a^2)^2}{a^3} = 32a = \frac{dV(a)}{da}.$$

## Capítulo 5 Optimización

2. Considera el problema de minimización

$$\min_x f(x; a) = a^2 x - \ln x - 1,$$

con  $a > 0$ . La siguiente figura muestra la gráfica de  $f(x; a) = a^2 x - \ln x - 1$  en función de  $x$ , para  $a = 2$ .



La condición de primer orden correspondiente está dada por

$$\left. \frac{\partial f}{\partial x} \right|_{x=x^*} = a^2 - \frac{1}{x} \Big|_{x=x^*} = 0,$$

cuya solución es

$$x^*(a) = \frac{1}{a^2}.$$

De este modo el valor mínimo de  $f$  es  $V(a) = a^2 \left( \frac{1}{a^2} \right) - \ln \left( \frac{1}{a^2} \right) - 1$ , es decir,

$$V(a) = 2 \ln a,$$

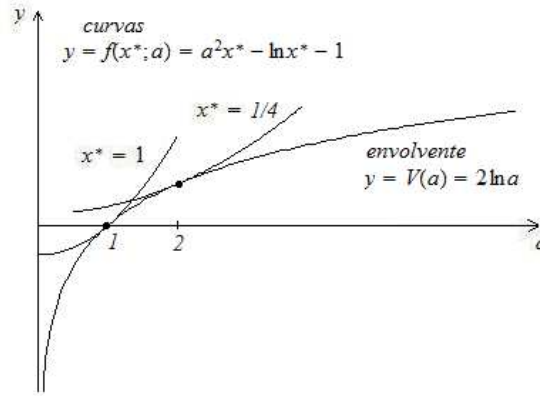
que representa la curva envolvente en este problema. Por otra parte, las curvitas  $f(x^*; a)$  se obtienen al evaluar  $f(x; a)$  en el punto óptimo  $x^*$ , es decir,

$$f(x^*; a) = a^2 x^* - \ln x^* - 1.$$

La siguiente figura muestra las gráficas de la función  $V(a) = 2 \ln a$ , así como

## 5.4 Teorema de la envolvente

las curvas  $f(x^*; a) = a^2 x^* - \ln x^* - 1$  en función de  $a$ , para  $x^* = 1, \frac{1}{4}$ .



Ahí se observa que las curvas  $f(x^*; a)$  son tangentes a la envolvente  $V(a)$  en aquellos valores de  $a$  que satisfacen la condición de óptimo

$$a = \frac{1}{\sqrt{x^*}}.$$

En efecto, en esos puntos se cumple

$$\begin{aligned} (1) \quad f(x^*; a)|_{x^*=1/a^2} &= a^2 x^* - \ln x^* - 1|_{x^*=1/a^2} = a^2 \left( \frac{1}{a^2} \right) - \ln \left( \frac{1}{a^2} \right) - 1 = 2 \ln a = V(a) \\ (2) \quad \frac{\partial f(x^*; a)}{\partial a} \Big|_{x^*=1/a^2} &= 2ax^*|_{x^*=1/a^2} = 2a \left( \frac{1}{a^2} \right) = \frac{2}{a} = \frac{dV(a)}{da}. \end{aligned}$$

Por lo general, puede resultar más complicado encontrar directamente la derivada  $dV(a)/da$  a partir de la función  $V(a)$ , que utilizando el teorema de la envolvente con  $f(x^*; a)$ , como se ilustra en el siguiente ejemplo.

Ejemplo:

Considera el problema de maximización

$$\max_x f(x; a) = [1 + (a^2 + 1)x] e^{-x},$$

con  $a \in \mathbb{R}$ . La condición de primer orden correspondiente es

$$\frac{\partial f}{\partial x} \Big|_{x=x^*} = \{(a^2 + 1) - [1 + (a^2 + 1)x]\} e^{-x} \Big|_{x=x^*} = 0,$$

## Capítulo 5 Optimización

cuya solución es

$$x^*(a) = \frac{a^2}{1+a^2}.$$

De este modo, el valor máximo de  $f$  es

$$V(a) = \left[ 1 + (a^2 + 1) \left( \frac{a^2}{1+a^2} \right) \right] e^{-\frac{a^2}{1+a^2}} = (1+a^2) e^{-\frac{a^2}{1+a^2}}.$$

Para encontrar la derivada  $dV(a)/da$  puedes proceder de dos maneras diferentes. La primera consiste en hacer caso omiso de toda la discusión anterior, y encontrar directamente la derivada de  $V(a)$ , es decir,

$$\begin{aligned} \frac{dV(a)}{da} &= \frac{d}{da} \left[ (1+a^2) e^{-\frac{a^2}{1+a^2}} \right] \\ &= 2ae^{-\frac{a^2}{1+a^2}} - (1+a^2) e^{-\frac{a^2}{1+a^2}} \frac{d}{da} \left[ \frac{a^2}{1+a^2} \right] \\ &= 2ae^{-\frac{a^2}{1+a^2}} - (1+a^2) e^{-\frac{a^2}{1+a^2}} \left[ \frac{2a}{(1+a^2)^2} \right] \\ &= 2a \left( 1 - \frac{1}{1+a^2} \right) e^{-\frac{a^2}{1+a^2}} \\ &= \left( \frac{2a^3}{1+a^2} \right) e^{-\frac{a^2}{1+a^2}}. \end{aligned}$$

La segunda consiste en utilizar el teorema de la envolvente, dado por la igualdad  $\frac{dV(a)}{da} = \frac{\partial f(x^*; a)}{\partial a}$ . Para ello, observa que  $f(x^*; a)$  es la función

$$f(x^*; a) = [1 + (a^2 + 1)x^*] e^{-x^*},$$

de modo que

$$\frac{\partial f(x^*; a)}{\partial a} = 2ax^* e^{-x^*}.$$

Sustituyendo la forma explícita de  $x^*$  en esta expresión obtenemos

$$\frac{\partial f(x^*; a)}{\partial a} = 2a \left( \frac{a^2}{1+a^2} \right) e^{-\frac{a^2}{1+a^2}}.$$

Nota entonces que, efectivamente,

$$\frac{dV(a)}{da} = \frac{\partial f(x^*; a)}{\partial a} = \left( \frac{2a^3}{1+a^2} \right) e^{-\frac{a^2}{1+a^2}},$$

pero la derivada  $dV/da$  se obtuvo de una manera más simple y directa con  $\partial f(x^*; a)/\partial a$ .

## 5.4 Teorema de la envolvente

Los resultados anteriores se pueden extender al caso de funciones diferenciables  $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$  con  $k$  parámetros  $\{a_1, \dots, a_k\}$ , de la siguiente manera. Considera el problema de maximización (o minimización)

$$\max_{x_1, \dots, x_n} f(\vec{x}; \vec{a}),$$

con  $\vec{x} = (x_1, \dots, x_n)$ . En este caso, el óptimo no restringido de esta función se obtiene de las condiciones de primer orden,

$$\left. \frac{\partial f}{\partial x_i} \right|_{\vec{x} = \vec{x}^*} = 0, \quad i = 1, \dots, n,$$

cuya solución es el vector

$$\vec{x}^* = \vec{x}^*(\vec{a}).$$

En consecuencia, el valor máximo  $f_{\max} \equiv V$  de la función  $f$  también dependerá de  $\vec{a}$ , a través de la relación

$$V(\vec{a}) \equiv f(\vec{x}^*(\vec{a}); \vec{a}).$$

Cuando alguno de los parámetros se modifica, digamos  $a_l$ ,  $l = 1, \dots, k$ , el óptimo  $V(\vec{a})$  de  $f$  cambia de acuerdo con

$$\begin{aligned} \frac{\partial V(\vec{a})}{\partial a_l} &= \frac{\partial f(x_1^*(\vec{a}), \dots, x_n^*(\vec{a}); \vec{a})}{\partial a_l} \\ &= \frac{\partial f^*}{\partial x_1^*} \frac{\partial x_1^*}{\partial a_l} + \dots + \frac{\partial f^*}{\partial x_n^*} \frac{\partial x_n^*}{\partial a_l} + \frac{\partial f^*}{\partial a_l} = \frac{\partial f^*}{\partial a_l}, \end{aligned}$$

en donde se han cancelado los primeros  $n$  términos por las  $n$  condiciones de primer orden,  $\partial f^* / \partial x_i^* = 0$ . Cada término cancelado,  $\partial f^* / \partial x_i^* \cdot \partial x_i^* / \partial a_l$ , corresponde al efecto indirecto sobre  $V$  causado por el cambio en  $x_i^*$  al cambiar  $a_l$ . El término sobreviviente,  $\partial f^* / \partial a_l$ , corresponde al efecto directo del cambio en  $a_l$  sobre el valor óptimo  $V$ . Se concluye entonces que

$$\frac{\partial V(\vec{a})}{\partial a_l} = \frac{\partial f(\vec{x}^*(\vec{a}); \vec{a})}{\partial a_l}, \quad l = 1, \dots, k.$$

**Teorema de la envolvente para optimización no restringida.** Sea  $f(\vec{x}; \vec{a})$  una función diferenciable de  $\vec{x} \in \mathbb{R}^n$  con parámetros  $\vec{a} \in \mathbb{R}^k$ . Para cada selección de  $\vec{a}$  considera el problema

$$\text{optimizar}_{x_1, \dots, x_n} f(\vec{x}; \vec{a}).$$

Sea  $\vec{x}^*(\vec{a})$  una solución del problema, con  $\vec{x}^*(\vec{a})$  una función diferenciable. Si  $V(\vec{a}) \equiv f(\vec{x}^*(\vec{a}); \vec{a})$  denota el valor óptimo de  $f$ , entonces

$$\frac{\partial V(\vec{a})}{\partial a_l} = \frac{\partial f(\vec{x}^*(\vec{a}); \vec{a})}{\partial a_l}, \quad l = 1, \dots, k.$$

Aquí la derivada parcial en el lado izquierdo se realiza sobre el valor óptimo  $V(a)$  de la función  $f$ , mientras que la derivada parcial en el lado derecho se realiza sobre la función  $f(\vec{x}^*(\vec{a}); \vec{a})$ , obtenida al evaluar  $f(\vec{x}; \vec{a})$  en  $\vec{x} = \vec{x}^*(\vec{a})$ , sin que  $f$  haya sido optimizada. En este último caso, sólo se considera la dependencia explícita de  $f(\vec{x}^*(\vec{a}); \vec{a})$  con  $a_l$ , ignorando la dependencia implícita de  $f$  con  $a_l$  a través de  $\vec{x}^*(\vec{a})$ , como si  $\vec{x}^*$  se mantuviera fijo, como se ilustra en los siguientes ejemplos.

Ejemplos:

1. Considera el problema

$$\max_x f(x; r_1, r_2) = x^{r_1} - r_2 x,$$

con  $0 < r_1 < 1$ . La condición de primer orden correspondiente está dada por

$$\left. \frac{\partial f}{\partial x} \right|_{x=x^*} = r_1 x^{r_1-1} - r_2 \Big|_{x=x^*} = 0,$$

cuya solución es

$$x^*(r_1, r_2) = \left( \frac{r_2}{r_1} \right)^{\frac{1}{r_1-1}}.$$

De esta manera, el valor máximo de  $f$  está dado por

$$V(r_1, r_2) = \left( \frac{r_2}{r_1} \right)^{\frac{r_1}{r_1-1}} - r_2 \left( \frac{r_2}{r_1} \right)^{\frac{1}{r_1-1}}.$$

El cálculo directo de las derivadas parciales de esta función resulta bastante complejo, particularmente  $\partial V / \partial r_1$ , que involucra derivadas del tipo  $\frac{dx^x}{dx}$ .

En contraste, el cálculo vía el teorema de la envolvente es directo. Para ello, primero evalúa  $f$  en  $x^*$ ,

$$f(x^*; r_1, r_2) = (x^*)^{r_1} - r_2 x^*,$$

cuyas derivadas parciales, bastante simples, quedan expresadas en términos de

$$x^*(r_1, r_2) = \left( \frac{r_2}{r_1} \right)^{\frac{1}{r_1-1}},$$

$$\begin{aligned} \frac{\partial V}{\partial r_1} &= \frac{\partial f(x^*; r_1, r_2)}{\partial r_1} = (x^*)^{r_1} \ln(x^*) \\ \frac{\partial V}{\partial r_2} &= \frac{\partial f(x^*; r_1, r_2)}{\partial r_2} = -x^*. \end{aligned}$$

## 5.4 Teorema de la envolvente

2. Considera el problema de minimizar el costo  $C(L, K) = wL + rK$  como función del trabajo  $L$  y el capital  $K$ , sujeto a una producción fija  $L^{1/2}K^{1/2} = Q$ . Los parámetros del sistema son el salario  $w$ , la tasa de interés  $r$  y el nivel de producción  $Q$ . Si despejamos la variable  $L$  de la ecuación de restricción,  $L = Q^2/K$ , y la sustituimos en la función de costos, el problema puede escribirse como el problema de optimización no restringida

$$\min_K C(K; r, w, Q) = \frac{wQ^2}{K} + rK.$$

La condición de primer orden correspondiente está dada por

$$\left. \frac{\partial C}{\partial K} \right|_{K=K^*} = -\frac{wQ^2}{K^2} + r \Big|_{K=K^*} = 0,$$

cuya solución es

$$K^*(w, r, Q) = Q\sqrt{\frac{w}{r}}.$$

El costo mínimo es, por tanto,  $C_{\min}(w, r, Q) = \frac{wQ^2}{(Q\sqrt{\frac{w}{r}})} + r\left(Q\sqrt{\frac{w}{r}}\right)$ , es decir,

$$C_{\min}(w, r, Q) = 2Q\sqrt{wr},$$

conocida como la *función de gasto*. Para encontrar sus derivadas parciales con respecto a los parámetros  $w, r$  y  $Q$ , podemos utilizar el teorema de la envolvente, de la siguiente manera. Primeramente evaluamos la función de costo en  $K^*$ ,

$$C(K^*; r, w, Q) = \frac{wQ^2}{K^*} + rK^*,$$

de modo que

$$\begin{aligned} \frac{\partial C_{\min}}{\partial w} &= \frac{\partial C(K^*; r, w, Q)}{\partial w} = \frac{Q^2}{K^*} = \frac{Q^2}{(Q\sqrt{\frac{w}{r}})} = Q\sqrt{\frac{r}{w}} \\ \frac{\partial C_{\min}}{\partial r} &= \frac{\partial C(K^*; r, w, Q)}{\partial r} = K^* = Q\sqrt{\frac{w}{r}} \\ \frac{\partial C_{\min}}{\partial Q} &= \frac{\partial C(K^*; r, w, Q)}{\partial Q} = \frac{2wQ}{K^*} = \frac{2wQ}{(Q\sqrt{\frac{w}{r}})} = 2\sqrt{wr}. \end{aligned}$$

La función  $C(K^*; r, w, Q)$  se conoce como la función de *costo de corto plazo SRC* (Short-Run Cost), puesto que en el corto plazo es de esperarse que el capital no cambie, manteniendo un valor fijo  $K^*$ . A su vez, la función de gasto  $C_{\min}(w, r, Q)$  se denomina la función de *costo de largo plazo LRC* (Long-Run Cost), ya que se obtiene permitiendo que el capital  $K$  varíe hasta que el costo

## Capítulo 5 Optimización

adquiera su valor mínimo. Esta última constituye, por tanto, el mejor mínimo posible, o curva envolvente en el problema de minimización de costos. De esta manera, para cada valor fijo  $\bar{K}_i$  del capital, se puede definir una familia de funciones de corto plazo, dadas por

$$SRC_i = \frac{wQ^2}{\bar{K}_i} + r\bar{K}_i,$$

que están envueltas inferiormente por la función de costo mínimo de largo plazo

$$LRC = 2Q\sqrt{wr}.$$

Equivalentemente, si en lugar del costo  $C$  consideramos en su lugar al costo promedio  $C/Q$ , o costo por unidad del bien, podemos definir las funciones de *costo promedio de corto plazo*  $SRAC$  (Short-Run Average Cost)

$$SRAC_i = \frac{SRC_i}{Q} = \frac{wQ}{\bar{K}_i} + \frac{r\bar{K}_i}{Q}$$

y de *costo promedio de largo plazo*  $LRAC$  (Long-Run Average Cost)

$$LRAC = \frac{LRC}{Q} = 2\sqrt{wr}.$$

Como funciones del parámetro  $Q$ , las curvas  $SRAC_i$  y  $LRAC$  son distintas entre sí. Sin embargo, nota que cuando  $SRAC_i$  es evaluada en aquel valor  $Q_i$  que satisface la condición del punto óptimo  $K^*(w, r, Q_i) = Q_i\sqrt{\frac{w}{r}} = \bar{K}_i$ , es decir

$$Q_i = \bar{K}_i\sqrt{\frac{r}{w}},$$

se obtiene

$$\begin{aligned} 1. \quad SRAC|_{Q_i} &= \frac{w(\bar{K}_i\sqrt{\frac{r}{w}})}{\bar{K}_i} + \frac{r\bar{K}_i}{(\bar{K}_i\sqrt{\frac{r}{w}})} = 2\sqrt{wr} = LRAC|_{Q_i} \\ 2. \quad \frac{d SRAC}{dQ} \Big|_{Q_i} &= \frac{w}{\bar{K}_i} - \frac{r\bar{K}_i}{Q^2} \Big|_{Q_i} = 0 = \frac{d LRAC}{dQ} \Big|_{Q_i}. \end{aligned}$$

Estas condiciones expresan la condición de tangencia entre la curva de costo medio de largo plazo  $LRAC$  y las curvas de corto plazo  $SRAC_i$ , en aquellos niveles de producción  $Q_i$  que corresponden a los correspondientes capitales fijos  $\bar{K}_i$ , de acuerdo con la condición de optimalidad

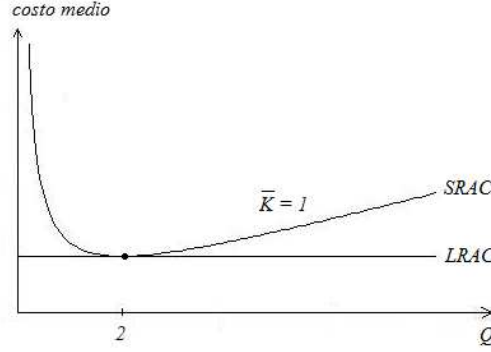
$$Q_i = \bar{K}_i\sqrt{\frac{r}{w}}.$$

Para ilustrar este concepto, la siguiente figura muestra las curvas  $SRAC_i$  y  $LRAC$  como funciones de  $Q$ , suponiendo que  $w = 1, r = 4$ . La función



## 5.4 Teorema de la envolvente

$SRAC_i$  fue evaluada en  $\bar{K} = 1$ , de modo que el punto de tangencia ocurre en  $Q = (1)\sqrt{\frac{4}{1}} = 2$ .



### 5.4.2 Optimización restringida

El teorema de la envolvente se aplica asimismo en problemas de optimización restringida, como se expone a continuación para el caso de optimización sujeta a restricciones de igualdad (método de los multiplicadores de Lagrange). Aunque no se demostrará aquí, los resultados que obtendremos son igualmente válidos para el caso correspondiente a restricciones de desigualdad (método de Kuhn-Tucker),

Considera el problema de maximización (o minimización) de una función diferenciable  $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$  con  $k$  parámetros  $\vec{a} = (a_1, \dots, a_k)$ , sujeto a un conjunto de  $m < n$  restricciones  $\{g_1, \dots, g_m\}$  de igualdad, es decir,

$$\max_{x_1, \dots, x_n} f(\vec{x}; \vec{a}),$$

$$\text{s.a. } g_j(\vec{x}; \vec{a}) = 0, \quad j = 1, \dots, m$$

con  $\vec{x} = (x_1, \dots, x_n)$ . Aquí la notación  $g_j = 0$  indica que todos los parámetros están contenidos en el lado izquierdo de la igualdad.

Para encontrar el óptimo restringido en este problema primero construimos la función lagrangeana correspondiente,

$$\mathcal{L}(\vec{x}, \vec{\lambda}; \vec{a}) = f(\vec{x}; \vec{a}) - \sum_{j=1}^m \lambda_j g_j(\vec{x}; \vec{a}),$$

## Capítulo 5 Optimización

con  $\vec{\lambda} = (\lambda_1, \dots, \lambda_m)$ . En este caso, el óptimo restringido de  $f$  se obtiene de las  $n + m$  condiciones de primer orden,

$$\begin{aligned} \left. \frac{\partial \mathcal{L}}{\partial x_i} \right|_{(\vec{x}, \vec{\lambda})=(\vec{x}^*, \vec{\lambda}^*)} &= 0, \quad i = 1, \dots, n \\ \left. \frac{\partial \mathcal{L}}{\partial \lambda_j} \right|_{(\vec{x}, \vec{\lambda})=(\vec{x}^*, \vec{\lambda}^*)} &= 0, \quad j = 1, \dots, m \end{aligned}$$

cuya solución está dada por

$$\begin{aligned} \vec{x}^* &= \vec{x}^*(\vec{a}) \\ \vec{\lambda}^* &= \vec{\lambda}^*(\vec{a}). \end{aligned}$$

En consecuencia, el valor máximo  $f_{\max} \equiv V$  de la función  $f$  también dependerá de  $\vec{a}$ , a través de la relación

$$V(\vec{a}) \equiv f(\vec{x}^*(\vec{a}); \vec{a}).$$

Cuando alguno de los parámetros se modifica, digamos  $a_l$ ,  $l = 1, \dots, k$ , el óptimo  $V(\vec{a})$  de  $f$  cambia de acuerdo con

$$\frac{\partial V(\vec{a})}{\partial a_l} = \frac{\partial f(\vec{x}^*(\vec{a}); \vec{a})}{\partial a_l}.$$

Como en el punto óptimo se satisfacen todas las restricciones  $g_j(\vec{x}; \vec{a}) = 0$ , ahí se satisface

$$f(\vec{x}^*(\vec{a}); \vec{a}) = \mathcal{L}(\vec{x}^*(\vec{a}), \vec{\lambda}^*(\vec{a}); \vec{a}),$$

es decir, la función objetivo toma el mismo valor que la lagrangeana óptima

$$\mathcal{L}^* \equiv \mathcal{L}(\vec{x}^*(\vec{a}), \vec{\lambda}^*(\vec{a}); \vec{a}) = f(\vec{x}^*; \vec{a}) - \sum_{j=1}^m \lambda_j^* g_j(\vec{x}^*; \vec{a}).$$

Así,

$$\begin{aligned} \frac{\partial V(\vec{a})}{\partial a_l} &= \frac{\partial \mathcal{L}^*}{\partial x_1^*} \frac{\partial x_1^*}{\partial a_l} + \dots + \frac{\partial \mathcal{L}^*}{\partial x_n^*} \frac{\partial x_n^*}{\partial a_l} \\ &\quad + \frac{\partial \mathcal{L}^*}{\partial \lambda_1^*} \frac{\partial \lambda_1^*}{\partial a_l} + \dots + \frac{\partial \mathcal{L}^*}{\partial \lambda_n^*} \frac{\partial \lambda_n^*}{\partial a_l} + \frac{\partial \mathcal{L}^*}{\partial a_l} = \frac{\partial \mathcal{L}^*}{\partial a_l}, \end{aligned}$$

en donde se han utilizado las condiciones de primer orden,  $\partial \mathcal{L}^* / \partial x_i^* = \partial \mathcal{L}^* / \partial \lambda_j^* = 0$ , en los primeros  $n + m$  términos. Los términos cancelados,  $\partial \mathcal{L}^* / \partial x_i^* \cdot \partial x_i^* / \partial a_l$  y  $\partial \mathcal{L}^* / \partial \lambda_j^* \cdot \partial \lambda_j^* / \partial a_l$ , corresponden al efecto indirecto sobre  $V$  causado por el cambio en cada  $x_i^*$  y cada  $\lambda_j^*$ , al cambiar  $a_l$ . El único término sobreviviente,  $\partial \mathcal{L}^* / \partial a_l$ ,

## 5.4 Teorema de la envolvente

corresponde al efecto directo del cambio en  $a_l$  sobre el valor óptimo  $V$ . De esta manera, se tiene

$$\frac{\partial V(\vec{a})}{\partial a_l} = \frac{\partial \mathcal{L}^*}{\partial a_l}, \quad l = 1, \dots, k.$$

**Teorema de la envolvente para restricciones de igualdad.** Sea  $f(\vec{x}; \vec{a})$  una función diferenciable de  $\vec{x} \in \mathbb{R}^n$ , con parámetros  $\vec{a} \in \mathbb{R}^k$ , y sea  $\vec{\lambda} \in \mathbb{R}^m$  un conjunto de multiplicadores de Lagrange correspondiente a  $m < n$  restricciones de igualdad. Considera el problema

$$\text{optimizar}_{x_1, \dots, x_n} f(\vec{x}; \vec{a})$$

$$\text{s.a. } g_j(\vec{x}; \vec{a}) = 0, \quad j = 1, \dots, m.$$

Sea  $\vec{x}^*(\vec{a})$  una solución del problema, con  $\vec{x}^*(\vec{a})$  una función diferenciable. Si  $\mathcal{L}(\vec{x}^*(\vec{a}), \vec{\lambda}^*(\vec{a}); \vec{a}) = V(\vec{a})$  denota el valor óptimo de la lagrangeana, entonces

$$\frac{\partial V(\vec{a})}{\partial a_l} = \frac{\partial \mathcal{L}(\vec{x}^*(\vec{a}), \vec{\lambda}^*(\vec{a}); \vec{a})}{\partial a_l}, \quad l = 1, \dots, k.$$

En este teorema, la derivada parcial en el lado izquierdo se realiza sobre el valor óptimo restringido  $V(a)$  de la función  $f$  ya optimizada. En contraste, la derivada parcial en el lado derecho se realiza sobre la función lagrangeana  $\mathcal{L}(\vec{x}^*(\vec{a}), \vec{\lambda}^*(\vec{a}); \vec{a})$ , obtenida al evaluar  $\mathcal{L}(\vec{x}, \vec{\lambda}; \vec{a})$  en el punto óptimo  $\vec{x} = \vec{x}^*(\vec{a})$ ,  $\vec{\lambda} = \vec{\lambda}^*(\vec{a})$ , sin que  $\mathcal{L}$  haya sido optimizada. Esta derivada parcial sólo toma en cuenta la dependencia explícita de  $\mathcal{L}(\vec{x}^*(\vec{a}), \vec{\lambda}^*(\vec{a}); \vec{a})$  con el parámetro  $a_l$ , manteniendo fijos  $\vec{x}^*$  y  $\vec{\lambda}^*$ , es decir, ignorando la dependencia implícita de  $f$  con  $a_l$  a través de  $\vec{x}^*(\vec{a})$  y  $\vec{\lambda}^*(\vec{a})$ , como se muestra en los siguientes ejemplos.

Ejemplos:

1. Considera el problema

$$\begin{aligned} \text{optim. } f(x, y) &= ye^x \\ \text{s.a. } a^2x + y &= 1, \end{aligned}$$

con  $a \neq 0$ . En este caso, la lagrangeana está dada por

$$\mathcal{L}(x, y, \lambda; a) = ye^x + \lambda(1 - a^2x - y).$$

## Capítulo 5 Optimización

Las condiciones de primer orden correspondientes son

$$\begin{aligned}\mathcal{L}_x &= ye^x - \lambda a^2 = 0 \\ \mathcal{L}_y &= e^x - \lambda = 0 \\ \mathcal{L}_\lambda &= 1 - a^2x - y = 0,\end{aligned}$$

de donde es fácil verificar que el punto óptimo es

$$\begin{aligned}x^*(a) &= a^{-2} - 1 \\ y^*(a) &= a^2 \\ \lambda^*(a) &= e^{a^{-2}-1}.\end{aligned}$$

De esta manera, el valor óptimo de  $f$  es

$$V(a) = a^2 e^{a^{-2}-1}.$$

En ese caso,

$$\begin{aligned}\frac{dV(a)}{da} &= a^2 e^{a^{-2}-1} (-2a^{-3}) + 2ae^{a^{-2}-1} \\ &= 2 \left( \frac{a^2 - 1}{a} \right) e^{a^{-2}-1}.\end{aligned}$$

Este mismo resultado puede obtenerse de una manera más simple con el teorema de la envolvente,  $dV/da = \partial \mathcal{L}^* / \partial a$ . Para ello, nota que en el óptimo

$$\mathcal{L}^* = y^* e^{x^*} + \lambda^* (1 - a^2 x^* - y^*),$$

de modo que

$$\frac{\partial \mathcal{L}^*}{\partial a} = -2a\lambda^* x^* = -2ae^{a^{-2}-1} \left( \frac{1 - a^2}{a^2} \right).$$

Así,

$$\frac{dV(a)}{da} = \frac{\partial \mathcal{L}^*}{\partial a} = 2 \left( \frac{a^2 - 1}{a} \right) e^{a^{-2}-1}.$$

2. Considera el problema de minimizar el costo sujeto a una producción dada,

$$\begin{aligned}\min_{L, K} C(L, K) &= wL + rK \\ \text{s.a. } P(L, K) &= Q,\end{aligned}$$

en donde el trabajo  $L$  y el capital  $K$  son las variables independientes, y el salario  $w$ , la tasa de interés  $r$  y el nivel de producción  $Q$  son los parámetros del sistema. La lagrangeana correspondiente es

$$\mathcal{L}(L, K, \lambda; w, r, Q) = wL + rK + \lambda (Q - P(L, K)).$$

## 5.4 Teorema de la envolvente

De las condiciones de primer orden  $\mathcal{L}_L = \mathcal{L}_K = \mathcal{L}_\lambda = 0$  se obtienen los valores óptimos

$$\begin{aligned} L^* &= L^*(w, r, Q) \\ K^* &= K^*(w, r, Q) \\ \lambda^* &= \lambda^*(w, r, Q). \end{aligned}$$

De esta manera, el costo mínimo está dado por

$$C_{\min}(w, r, Q) = wL^*(w, r, Q) + rK^*(w, r, Q).$$

Aparentemente, si no se tiene una forma explícita para la función de producción  $P(L, K)$  no se puede determinar la forma funcional de  $C_{\min}(w, r, Q)$  y, por lo tanto, tampoco se pueden encontrar sus derivadas  $\partial C_{\min}/\partial w$ ,  $\partial C_{\min}/\partial r$  y  $\partial C_{\min}/\partial Q$ . Sin embargo, el teorema de la envolvente permite obtener expresiones generales para estas derivadas. Para ello, se evalúa la lagrangeana en el óptimo

$$\mathcal{L}^* = wL^* + rK^* + \lambda^*(Q - P(L^*, K^*)),$$

de modo que

$$\begin{aligned} \frac{\partial C_{\min}}{\partial w} &= \frac{\partial \mathcal{L}^*}{\partial w} = L^* \\ \frac{\partial C_{\min}}{\partial r} &= \frac{\partial \mathcal{L}^*}{\partial r} = K^* \\ \frac{\partial C_{\min}}{\partial Q} &= \frac{\partial \mathcal{L}^*}{\partial Q} = \lambda^*. \end{aligned}$$

Así, en el óptimo, el costo marginal del salario es el trabajo  $L^*$ , el costo marginal de la tasa de interés es el capital  $K^*$  y el costo marginal de la producción es el multiplicador  $\lambda^*$ , en donde  $L^*$ ,  $K^*$  y  $\lambda^*$  son los valores de  $L$ ,  $K$  y  $\lambda$  que minimizan la función de costo bajo la restricción dada. Este resultado general se conoce como el *Lema de Shephard*.

Como caso particular de este lema, considera una función de producción Cobb-Douglas de la forma

$$P(L, K) = L^{1/2} K^{1/2}.$$

En ese caso, el óptimo ocurre en el punto

$$\begin{aligned} L^*(w, r, Q) &= Q\sqrt{\frac{r}{w}} \\ K^*(w, r, Q) &= Q\sqrt{\frac{w}{r}} \\ \lambda^*(w, r, Q) &= 2\sqrt{wr}, \end{aligned}$$

## Capítulo 5 Optimización

de modo que el costo mínimo  $C_{\min} = wL^* + rK^*$  está dado por

$$C_{\min}(w, r, Q) = 2Q\sqrt{wr}.$$

Se tiene, entonces,

$$\begin{aligned}\frac{\partial C_{\min}(w, r, Q)}{\partial w} &= Q\sqrt{\frac{r}{w}} = L^* \\ \frac{\partial C_{\min}(w, r, Q)}{\partial r} &= Q\sqrt{\frac{w}{r}} = K^* \\ \frac{\partial C_{\min}(w, r, Q)}{\partial Q} &= 2\sqrt{wr} = \lambda^*.\end{aligned}$$

## Capítulo 6

### Temas selectos de cálculo avanzado

#### 6.1 Funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$

Supongamos que una empresa produce  $m$  bienes utilizando  $n$  insumos,  $x_1, x_2, \dots, x_n$ , y hay una función de producción diferente  $f_j$  para la producción de cada una de las cantidades  $Q_j$ , con  $j = 1, \dots, m$ . En ese caso, se tiene

$$\begin{aligned} Q_1 &= f_1(x_1, x_2, \dots, x_n) \\ Q_2 &= f_2(x_1, x_2, \dots, x_n) \\ &\vdots \\ Q_m &= f_m(x_1, x_2, \dots, x_n). \end{aligned}$$

Desde un punto de vista formal, resulta más conveniente considerar este conjunto de  $m$  funciones con  $n$  variables como una sola función,  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , dada por

$$F(\vec{x}) = (f_1(x_1, x_2, \dots, x_n), f_2(x_1, x_2, \dots, x_n), \dots, f_m(x_1, x_2, \dots, x_n)).$$

Si denotamos por

$$\vec{Q} = (Q_1, Q_2, \dots, Q_m)$$

al vector de producción, decimos entonces que

$$\vec{Q} = F(\vec{x}).$$

**Definición.** Sea  $S \subset \mathbb{R}^n$ . Una *función*  $F : S \rightarrow \mathbb{R}^m$  es una regla de correspondencia que a cada elemento  $\vec{x} = (x_1, x_2, \dots, x_n)$  del dominio,  $S$ , le asigna un único elemento  $\vec{w} = (w_1, \dots, w_m)$  del contradominio,  $\mathbb{R}^m$ .

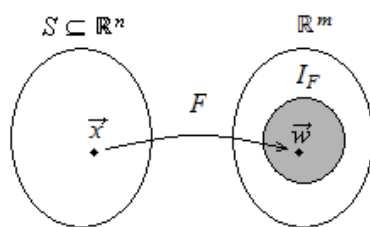
Aquí usaremos una letra mayúscula (tal como  $F$ ) para denotar una función con  $m > 1$  reglas de correspondencia, y conservaremos la notación usual de letra minúscula (tal como  $f$ ) para funciones con una sola regla de correspondencia. Así, por ejemplo, escribimos  $F = (f_1, \dots, f_m)$ .

Si denominamos por  $F(\vec{x})$  al elemento de  $\mathbb{R}^m$  que  $F$  le asigna al elemento  $\vec{x}$  de  $S$ , entonces decimos que

$$\vec{w} = F(\vec{x})$$

es la *imagen de  $\vec{x}$  bajo  $F$* . Asimismo, decimos que la *imagen de la función  $F$* , denotada por  $I_F$ , es el conjunto de elementos del contradominio obtenidos al aplicar la regla múltiple  $F$  a los elementos del dominio, es decir,

$$I_F = \{ \vec{w} \in \mathbb{R}^m \mid \vec{w} = F(x_1, x_2, \dots, x_n), \text{ para todo } \vec{x} = (x_1, x_2, \dots, x_n) \in S \}.$$



Una función  $F : \mathbb{R} \rightarrow \mathbb{R}^m$  (con  $n = 1$ )

$$\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{pmatrix} = \begin{pmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_m(x) \end{pmatrix}$$

se denomina *función vectorial de variable escalar*, ya que los elementos del dominio son escalares,  $x \in \mathbb{R}$ , y los elementos del contradominio son  $m$ -vectores,  $(w_1, w_2, \dots, w_m) \in \mathbb{R}^m$ . Geométricamente, las funciones de este tipo se representan mediante curvas paramétricas en  $\mathbb{R}^m$ , como las que estudiamos en la sección 1.2 (aquí  $x$  juega el papel del parámetro que antes llamamos  $t$ ). Por otra parte, una función  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  (con  $m = 1$ )

$$w = f(x_1, x_2, \dots, x_n)$$

es una *función escalar de variable vectorial*, ya que los elementos del dominio son  $n$ -vectores,  $(x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ , y los elementos del contradominio son escalares,  $w \in \mathbb{R}$ . Geométricamente, las funciones de este tipo se representan mediante hipersuperficies en  $\mathbb{R}^{n+1}$ , como las que estudiamos en los capítulos 2 al 5. Por



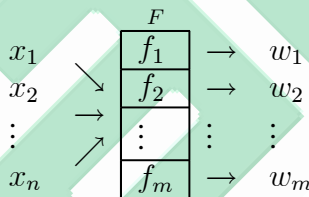
## 6.1 Funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$

último, una función  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  (con  $n, m > 1$ )

$$\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{pmatrix} = \begin{pmatrix} f_1(x_1, x_2, \dots, x_n) \\ f_2(x_1, x_2, \dots, x_n) \\ \vdots \\ f_m(x_1, x_2, \dots, x_n) \end{pmatrix}$$

es una *función vectorial de variable vectorial*, ya que los elementos del dominio son  $n$ -vectores,  $(x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ , y los elementos del contradominio son  $m$ -vectores,  $(w_1, w_2, \dots, w_m) \in \mathbb{R}^m$ . En general, no existe una representación geométrica simple para este tipo de funciones. Desde el punto de vista conceptual, sin embargo, es claro que una función de este tipo transforma vectores de  $\mathbb{R}^n$  en vectores de  $\mathbb{R}^m$ .

Para todos los casos anteriores, puede resultarnos útil visualizar la función  $F$  como una “caja negra” con  $n$  valores de entrada o *variables independientes*  $x_1, x_2, \dots, x_n$  y  $m$  valores de salida o *variables dependientes*,  $w_1, w_2, \dots, w_m$ , como se muestra en el siguiente diagrama.



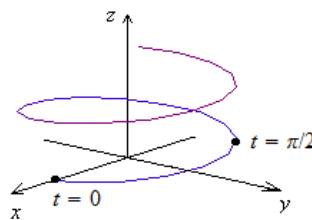
Ejemplos:

1. La función  $F = (f_1, f_2, f_3) : \mathbb{R} \rightarrow \mathbb{R}^3$  dada por

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} f_1(t) \\ f_2(t) \\ f_3(t) \end{pmatrix} = \begin{pmatrix} \cos t \\ \sin t \\ t \end{pmatrix}$$

asigna a cada valor de  $t \in \mathbb{R}$  un único punto  $(x, y, z)$  del espacio. El conjunto de puntos correspondientes en  $\mathbb{R}^3$  es la curva conocida como hélice.

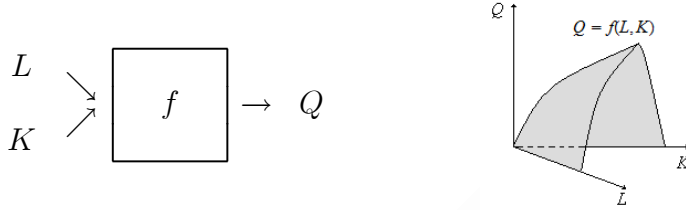
$$t \rightarrow \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} \rightarrow \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$



2. La función  $f : \mathbb{R}_+^2 \rightarrow \mathbb{R}$  dada por

$$Q = f(L, K)$$

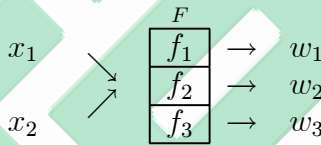
asigna a cada pareja de insumos  $(L, K) \in \mathbb{R}_+^2$  una única producción  $Q \in \mathbb{R}$ . El conjunto de puntos correspondientes es una superficie de producción.



3. La función  $F = (f_1, f_2, f_3) : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ , dada por

$$\begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \\ f_3(x_1, x_2) \end{pmatrix} = \begin{pmatrix} 5x_1 + 4x_2 \\ 2x_1 + x_2 \\ x_1 + 3x_2 \end{pmatrix},$$

transforma cada vector  $\vec{x} = (x_1, x_2) \in \mathbb{R}^2$  en otro vector  $\vec{w} = (w_1, w_2, w_3) \in \mathbb{R}^3$ .



Nota que, en este caso particular, se trata de un sistema lineal, de modo que puede escribirse como el producto de matrices

$$\begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} 5 & 4 \\ 2 & 1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

En otras palabras, se tiene

$$\vec{w} = A\vec{x},$$

en donde  $A$  es la matriz de  $3 \times 2$ , dada por

$$A = \begin{pmatrix} 5 & 4 \\ 2 & 1 \\ 1 & 3 \end{pmatrix}.$$

En este caso, identificamos la función vectorial lineal  $F$  como una transformación lineal, de la forma

$$F(\vec{x}) = A\vec{x}.$$

## 6.1 Funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$

4. En un problema de maximización de utilidad para dos bienes, la función  $F = (f_1, f_2) : \mathbb{R}_+^3 \rightarrow \mathbb{R}_+^2$  dada por

$$\begin{pmatrix} x^* \\ y^* \end{pmatrix} = \begin{pmatrix} f_1(I, p_1, p_2) \\ f_2(I, p_1, p_2) \end{pmatrix},$$

asigna a cada trío  $(I, p_1, p_2)$  de ingreso, precio del bien 1 y precio del bien 2, una única canasta óptima  $(x^*, y^*)$ .

$$\begin{array}{ccc} I & \searrow & \boxed{\begin{matrix} F \\ f_1 \\ f_2 \end{matrix}} \rightarrow x^* \\ p_1 & \rightarrow & \\ p_2 & \nearrow & \rightarrow y^* \end{array}$$

El análisis de la diferenciabilidad de una función  $F$  de  $\mathbb{R}^n$  en  $\mathbb{R}^m$  se facilita si observamos que cada componente de  $F = (f_1, \dots, f_m)$  es una función  $f_j : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$ , como las funciones que estudiamos en los capítulos anteriores. En consecuencia, podemos aplicar la teoría ya vista a cada componente  $f_j$  por separado, y luego escribir la información obtenida en un lenguaje matricial.

De acuerdo con la observación anterior, para estudiar la diferenciabilidad de la función

$$F = (f_1, \dots, f_m) : S \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$$

en un punto específico  $\vec{x}_0 \in S$ , podemos aplicar la aproximación por diferenciales de la sección 3.2 a cada componente  $f_j$ , obteniendo

$$\begin{aligned} f_1(\vec{x}_0 + \Delta \vec{x}) - f_1(\vec{x}_0) &\approx \frac{\partial f_1(\vec{x}_0)}{\partial x_1} \Delta x_1 + \dots + \frac{\partial f_1(\vec{x}_0)}{\partial x_n} \Delta x_n \\ f_2(\vec{x}_0 + \Delta \vec{x}) - f_2(\vec{x}_0) &\approx \frac{\partial f_2(\vec{x}_0)}{\partial x_1} \Delta x_1 + \dots + \frac{\partial f_2(\vec{x}_0)}{\partial x_n} \Delta x_n \\ &\vdots \\ f_m(\vec{x}_0 + \Delta \vec{x}) - f_m(\vec{x}_0) &\approx \frac{\partial f_m(\vec{x}_0)}{\partial x_1} \Delta x_1 + \dots + \frac{\partial f_m(\vec{x}_0)}{\partial x_n} \Delta x_n. \end{aligned}$$

Estos resultados pueden combinarse en una notación matricial, como

$$F(\vec{x}_0 + \Delta \vec{x}) - F(\vec{x}_0) \approx \begin{pmatrix} \frac{\partial f_1(\vec{x}_0)}{\partial x_1} & \frac{\partial f_1(\vec{x}_0)}{\partial x_2} & \dots & \frac{\partial f_1(\vec{x}_0)}{\partial x_n} \\ \frac{\partial f_2(\vec{x}_0)}{\partial x_1} & \frac{\partial f_2(\vec{x}_0)}{\partial x_2} & \dots & \frac{\partial f_2(\vec{x}_0)}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m(\vec{x}_0)}{\partial x_1} & \frac{\partial f_m(\vec{x}_0)}{\partial x_2} & \dots & \frac{\partial f_m(\vec{x}_0)}{\partial x_n} \end{pmatrix} \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \\ \vdots \\ \Delta x_n \end{pmatrix}.$$

Esta última expresión describe la *aproximación lineal de  $F$  en  $\vec{x}_0$* . En notación compacta, esta última expresión se escribe como

$$F(\vec{x}_0 + \Delta \vec{x}) - F(\vec{x}_0) \approx DF(\vec{x}_0) \Delta \vec{x},$$

en donde  $\Delta \vec{x}$  es el  $n$ -vector de incrementos, y en donde  $DF(\vec{x}_0)$  es la matriz de  $m \times n$  que está compuesta por las primeras derivadas parciales de primer orden de las funciones  $f_j$ 's con respecto a las variables  $x_i$ 's, como se define a continuación.

**Definición.** Sea  $F = (f_1, \dots, f_m) : S \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$  una función vectorial cuyas derivadas parciales  $\partial f_j / \partial x_i$  existan en una región  $R \subset S$ , para cada  $j = 1, \dots, m, i = 1, \dots, n$ . La *derivada* o *matriz jacobiana* de  $F$  con respecto al vector  $\vec{x} = (x_1, x_2, \dots, x_n)$  es la matriz  $DF(\vec{x})$  de  $m \times n$  dada por

$$DF(\vec{x}) = \begin{pmatrix} \frac{\partial f_1(\vec{x})}{\partial x_1} & \frac{\partial f_1(\vec{x})}{\partial x_2} & \dots & \frac{\partial f_1(\vec{x})}{\partial x_n} \\ \frac{\partial f_2(\vec{x})}{\partial x_1} & \frac{\partial f_2(\vec{x})}{\partial x_2} & \dots & \frac{\partial f_2(\vec{x})}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m(\vec{x})}{\partial x_1} & \frac{\partial f_m(\vec{x})}{\partial x_2} & \dots & \frac{\partial f_m(\vec{x})}{\partial x_n} \end{pmatrix}.$$

A la derivada también se le conoce como *gradiente generalizado* de  $f_1, \dots, f_m$  con respecto a  $x_1, x_2, \dots, x_n$ . Para denotarla se usan cualesquiera de los siguientes símbolos  $DF(\vec{x})$ ,  $\frac{dF}{d\vec{x}}$  o  $\nabla F$ .

Ejemplo:

Para las siguientes funciones de demanda para dos bienes identifica la función vectorial  $F$  y determina su derivada  $DF$ :

$$q_1 = 6p_1^{-2}p_2^{3/2}y, \quad q_2 = 4p_1p_2^{-1}y^2.$$

Se trata de la función  $F : \mathbb{R}_+^3 \rightarrow \mathbb{R}_+^2$ , dada por

$$F(p_1, p_2, y) = \begin{pmatrix} f_1(p_1, p_2, y) \\ f_2(p_1, p_2, y) \end{pmatrix} = \begin{pmatrix} 6p_1^{-2}p_2^{3/2}y \\ 4p_1p_2^{-1}y^2 \end{pmatrix}.$$

$$\begin{array}{ccc} p_1 & \searrow & \\ p_2 & \rightarrow & \boxed{\begin{array}{c} f_1 \\ f_2 \end{array}} \\ y & \nearrow & \end{array} \begin{array}{c} \xrightarrow{F} \\ \rightarrow \\ \rightarrow \end{array} \begin{array}{c} q_1 \\ q_2 \end{array}$$

De esta manera, su derivada  $DF$  es la siguiente matriz de  $2 \times 3$ :

$$DF(p_1, p_2, y) = \begin{pmatrix} \frac{\partial f_1}{\partial p_1} & \frac{\partial f_1}{\partial p_2} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial p_1} & \frac{\partial f_2}{\partial p_2} & \frac{\partial f_2}{\partial y} \end{pmatrix} = \begin{pmatrix} -12p_1^{-3}p_2^{3/2}y & 9p_1^{-2}p_2^{1/2}y & 6p_1^{-2}p_2^{3/2} \\ 4p_2^{-1}y^2 & -4p_1p_2^{-2}y^2 & 8p_1p_2^{-1}y \end{pmatrix}.$$

## 6.1 Funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$

En el caso particular de una función  $F = (f_1, f_2, \dots, f_m) : S \subset \mathbb{R} \rightarrow \mathbb{R}^m$ , con  $\vec{w} = F(t)$ , la derivada está dada por el  $m$ -vector columna

$$DF(t) = \begin{pmatrix} \frac{df_1(t)}{dt} \\ \frac{df_2(t)}{dt} \\ \vdots \\ \frac{df_m(t)}{dt} \end{pmatrix},$$

que representa el vector tangente a la curva  $\vec{w} = F(t)$  en cada  $t \in \mathbb{R}$  (ver sección 1.2). Asimismo, en el caso de una función  $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$ , con  $w = f(\vec{x})$ , la derivada es el  $n$ -vector renglón

$$Df(\vec{x}) = \nabla f(\vec{x}) = \left( \frac{\partial f(\vec{x})}{\partial x_1}, \frac{\partial f(\vec{x})}{\partial x_2}, \dots, \frac{\partial f(\vec{x})}{\partial x_n} \right),$$

que coincide con el vector gradiente de  $f$ . En consecuencia, la derivada representa un vector perpendicular al conjunto de nivel de la hipersuperficie  $w = f(\vec{x})$  que contiene al punto  $\vec{x} = (x_1, x_2, \dots, x_m) \in S$  (ver sección 3.5).

Por último, para el caso especial en que  $n = m$ , la matriz jacobiana es cuadrada y se puede calcular su determinante.

**Definición.** Sea  $F : S \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ , definida por  $F(\vec{x}) = (f_1(\vec{x}), f_2(\vec{x}), \dots, f_n(\vec{x}))$ , donde  $\vec{x} = (x_1, x_2, \dots, x_n)$ . Entonces, al determinante de su matriz jacobiana se le denomina *jacobiano* o *determinante jacobiano* de  $f_1, f_2, \dots, f_n$  con respecto a  $x_1, x_2, \dots, x_n$ , y se representa por  $J(f_1, f_2, \dots, f_n/x_1, x_2, \dots, x_n)$ , es decir,

$$J\left(\frac{f_1, f_2, \dots, f_n}{x_1, x_2, \dots, x_n}\right) = \begin{vmatrix} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \dots & \frac{\partial f_n}{\partial x_n} \end{vmatrix}.$$

Nota: En algunos textos el determinante jacobiano  $J(f_1, f_2, \dots, f_n/x_1, x_2, \dots, x_n)$  también se denota por

$$\frac{\partial(f_1, f_2, \dots, f_n)}{\partial(x_1, x_2, \dots, x_n)}.$$

El jacobiano tiene varias aplicaciones en cálculo, una de las cuales se discutirá en la sección 6.3.

## 6.2 Regla de la cadena en el caso general

La regla de la cadena tiene una extensión natural para funciones de  $\mathbb{R}^n$  en  $\mathbb{R}^m$ , como se establece en el siguiente teorema.

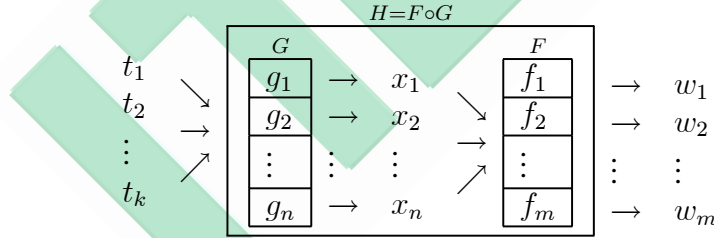
**Teorema.** Sean  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  y  $G : \mathbb{R}^k \rightarrow \mathbb{R}^n$  funciones diferenciables. Sean  $\vec{t} \in \mathbb{R}^k$  y  $\vec{x} \in \mathbb{R}^n$ . Considera la función compuesta

$$H = F \circ G : \mathbb{R}^k \rightarrow \mathbb{R}^m.$$

Sea  $DF(\vec{x})$  la matriz jacobiana de  $m \times n$  de las derivadas parciales de  $F$  en  $\vec{x}$  y sea  $DG(\vec{t})$  la matriz jacobiana de  $n \times k$  de las derivadas parciales de  $G$  en  $\vec{t}$ . Entonces, la matriz jacobiana  $DH(\vec{t})$  es la matriz de  $m \times k$  dada por el producto de las matrices jacobianas:

$$DH(\vec{t}) = DF(\vec{x}) \cdot DG(\vec{t}).$$

Más específicamente, sean  $\vec{t} = (t_1, \dots, t_k)$ ,  $\vec{x} = (x_1, \dots, x_n)$  y  $\vec{w} = (w_1, \dots, w_m)$ , tales que  $\vec{x} = G(\vec{t})$ ,  $\vec{w} = F(\vec{x})$ . Entonces, podemos representar la composición de funciones  $\vec{w} = F(G(\vec{t})) = H(\vec{t})$  mediante el siguiente diagrama:



De esta manera, la regla de la cadena  $DH(\vec{t}) = DF(\vec{x}) \cdot DG(\vec{t})$  establece que

$$\begin{pmatrix} \frac{\partial w_1}{\partial t_1} & \frac{\partial w_1}{\partial t_2} & \cdots & \frac{\partial w_1}{\partial t_k} \\ \frac{\partial w_2}{\partial t_1} & \frac{\partial w_2}{\partial t_2} & \cdots & \frac{\partial w_2}{\partial t_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial w_m}{\partial t_1} & \frac{\partial w_m}{\partial t_2} & \cdots & \frac{\partial w_m}{\partial t_k} \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix} \begin{pmatrix} \frac{\partial g_1}{\partial t_1} & \frac{\partial g_1}{\partial t_2} & \cdots & \frac{\partial g_1}{\partial t_k} \\ \frac{\partial g_2}{\partial t_1} & \frac{\partial g_2}{\partial t_2} & \cdots & \frac{\partial g_2}{\partial t_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_n}{\partial t_1} & \frac{\partial g_n}{\partial t_2} & \cdots & \frac{\partial g_n}{\partial t_k} \end{pmatrix}.$$

Nota que este producto matricial es consistente con los resultados de la sección 3.3.

## 6.3 Teorema general de la función implícita

### 6.3 Teorema general de la función implícita

En la sección 3.4 presentamos el teorema de la función implícita para una ecuación  $f(x_1, x_2, \dots, x_n, w) = 0$  que relaciona a varias variables de manera implícita. En esta sección extendemos el teorema de la función implícita al caso general,

$$\begin{aligned}f_1(x_1, \dots, x_n, w_1, \dots, w_m) &= 0 \\f_2(x_1, \dots, x_n, w_1, \dots, w_m) &= 0 \\&\vdots \\f_m(x_1, \dots, x_n, w_1, \dots, w_m) &= 0,\end{aligned}$$

en donde hay varias funciones relacionando implícitamente a varias variables.

Este tipo de sistemas aparece frecuentemente en economía. Por ejemplo, se sabe que en el problema de minimación del costo  $C = wL + rK$  sujeto a una producción fija  $P(L, K) = Q$ , las condiciones de primer orden están dadas por

$$\begin{aligned}w - \lambda P_L(L, K) &= 0 \\r - \lambda P_K(L, K) &= 0 \\P(L, K) - Q &= 0.\end{aligned}$$

Este es un sistema de 3 ecuaciones para 6 incógnitas: 3 variables endógenas, dadas por los niveles óptimos de trabajo  $L$ , capital  $K$  y el multiplicador  $\lambda$ , y 3 variables exógenas, dadas por el salario  $w$ , la tasa de interés  $r$  y la producción  $Q$ . Nos interesa determinar bajo qué condiciones este sistema define los niveles óptimos en función de las variables exógenas, y en ese caso, cómo se verían afectados esos niveles óptimos ante un pequeño cambio en las últimas. Sin embargo, por lo general no es posible determinar explícitamente esos niveles, ya que no se conoce la forma funcional de la función  $P$ , o bien,  $P$  puede ser una función compleja y esto no permite encontrar la solución deseada. El *teorema de la función implícita* establece bajo qué condiciones un sistema de ecuaciones de este tipo define las variables endógenas como funciones implícitas de las variables exógenas y, en ese caso, permite determinar cómo cambian las primeras ante un pequeño cambio en las últimas.

Por simplicidad, comenzaremos con un caso simple, dado por un sistema de 2 ecuaciones con 4 variables, de la forma

$$\begin{aligned}f(x, y, u, v) &= 0 \\g(x, y, u, v) &= 0.\end{aligned}$$

Como se trata de 2 ecuaciones, a lo más podemos tener 2 variables dependientes o *endógenas*, en términos de las 2 variables restantes, que son las variables

## Capítulo 6 Temas selectos de cálculo avanzado

independientes o *exógenas*. Por ejemplo, supongamos que el sistema define a las variables  $u$  y  $v$  como funciones implícitas de  $x$  y  $y$ , es decir,

$$\begin{aligned}u &= u(x, y) \\v &= v(x, y),\end{aligned}$$

y nos preguntamos cuánto valen las derivadas parciales  $u_x, u_y, v_x$  y  $v_y$ . Para ello, primero obtenemos las diferenciales totales de  $f$  y  $g$ , a saber,

$$\begin{aligned}df &= f_x dx + f_y dy + f_u du + f_v dv = 0 \\dg &= g_x dx + g_y dy + g_u du + g_v dv = 0.\end{aligned}$$

A su vez, como  $u = u(x, y)$  y  $v = v(x, y)$ , por lo tanto,

$$\begin{aligned}du &= u_x dx + u_y dy \\dv &= v_x dx + v_y dy.\end{aligned}$$

De esta manera,

$$\begin{aligned}df &= f_x dx + f_y dy + f_u (u_x dx + u_y dy) + f_v (v_x dx + v_y dy) = 0 \\dg &= g_x dx + g_y dy + g_u (u_x dx + u_y dy) + g_v (v_x dx + v_y dy) = 0.\end{aligned}$$

Ahora agrupamos términos con  $dx$  y términos con  $dy$ , es decir,

$$\begin{aligned}(f_x + f_u u_x + f_v v_x) dx + (f_y + f_u u_y + f_v v_y) dy &= 0 \\(g_x + g_u u_x + g_v v_x) dx + (g_y + g_u u_y + g_v v_y) dy &= 0.\end{aligned}$$

Como  $x$  y  $y$  son independientes, cada una de estas sumas es igual a cero sólo si sus sumandos se anulan por separado, es decir, sólo si

$$\begin{aligned}f_x + f_u u_x + f_v v_x &= 0 & f_y + f_u u_y + f_v v_y &= 0 \\g_x + g_u u_x + g_v v_x &= 0 & g_y + g_u u_y + g_v v_y &= 0.\end{aligned}$$

Para encontrar de aquí los valores de las derivadas parciales  $u_x, u_y, v_x$  y  $v_y$  es conveniente agrupar estas ecuaciones por pares. Un par está dado por

$$\begin{aligned}f_u u_x + f_v v_x &= -f_x \\g_u u_x + g_v v_x &= -g_x,\end{aligned}$$

que es un sistema de dos ecuaciones para las incógnitas,  $u_x$  y  $v_x$ . De acuerdo con



### 6.3 Teorema general de la función implícita

la regla de Cramer y las propiedades del determinante, su solución es

$$u_x = \frac{\partial u}{\partial x} = \frac{\begin{vmatrix} -f_x & f_v \\ -g_x & g_v \end{vmatrix}}{\begin{vmatrix} f_u & f_v \\ g_u & g_v \end{vmatrix}} = -\frac{\begin{vmatrix} f_x & f_v \\ g_x & g_v \end{vmatrix}}{\begin{vmatrix} f_u & f_v \\ g_u & g_v \end{vmatrix}}$$

$$v_x = \frac{\partial v}{\partial x} = \frac{\begin{vmatrix} f_u & -f_x \\ g_u & -g_x \end{vmatrix}}{\begin{vmatrix} f_u & f_v \\ g_u & g_v \end{vmatrix}} = -\frac{\begin{vmatrix} f_u & f_x \\ g_u & g_x \end{vmatrix}}{\begin{vmatrix} f_u & f_v \\ g_u & g_v \end{vmatrix}}.$$

El otro par está dado por

$$\begin{aligned} f_u u_y + f_v v_y &= -f_y \\ g_u u_y + g_v v_y &= -g_y, \end{aligned}$$

que es un sistema de dos ecuaciones para las incógnitas,  $u_y$  y  $v_y$ , cuya solución es

$$u_y = \frac{\partial u}{\partial y} = \frac{\begin{vmatrix} -f_y & f_v \\ -g_y & g_v \end{vmatrix}}{\begin{vmatrix} f_u & f_v \\ g_u & g_v \end{vmatrix}} = -\frac{\begin{vmatrix} f_y & f_v \\ g_y & g_v \end{vmatrix}}{\begin{vmatrix} f_u & f_v \\ g_u & g_v \end{vmatrix}}$$

$$v_y = \frac{\partial v}{\partial y} = \frac{\begin{vmatrix} f_u & -f_y \\ g_u & -g_y \end{vmatrix}}{\begin{vmatrix} f_u & f_v \\ g_u & g_v \end{vmatrix}} = -\frac{\begin{vmatrix} f_u & f_y \\ g_u & g_y \end{vmatrix}}{\begin{vmatrix} f_u & f_v \\ g_u & g_v \end{vmatrix}}.$$

Observa que las cuatro derivadas parciales son el cociente de dos determinantes. Todas ellas poseen el mismo denominador, dado por el jacobiano de la función vectorial  $F = (f, g) : \mathbb{R}^4 \rightarrow \mathbb{R}^2$ ,

$$J\left(\frac{f, g}{u, v}\right) = |DF(u, v)| = \begin{vmatrix} f_u & f_v \\ g_u & g_v \end{vmatrix},$$

con respecto a las variables dependientes  $u$  y  $v$ . Como  $J\left(\frac{f, g}{u, v}\right)$  está en el denominador, es claro que debe imponerse la condición

$$J\left(\frac{f, g}{u, v}\right) \neq 0.$$

Ahora observa que cada numerador está dado también por un determinante jacobiano, de la misma función  $F$ , pero con respecto a una de las variables

dependientes,  $u$  o  $v$ , mezclada con una de las independientes,  $x$  o  $y$ . Así, por ejemplo, se tiene

$$\frac{\partial u}{\partial x} = -\frac{J\left(\frac{f,g}{x,v}\right)}{J\left(\frac{f,g}{u,v}\right)},$$

en donde el determinante en el numerador intercambia la variable dependiente  $u$  del determinante en el denominador por la variable independiente  $x$ .

**Teorema.** Sean  $f(x, y, u, v)$  y  $g(x, y, u, v)$  funciones diferenciables. El sistema de ecuaciones

$$\begin{aligned} f(x, y, u, v) &= 0 \\ g(x, y, u, v) &= 0 \end{aligned}$$

define a las variables  $u$  y  $v$  como funciones implícitas, diferenciables de  $x$  y  $y$ , en todos los puntos en donde

$$J\left(\frac{f,g}{u,v}\right) \neq 0.$$

En ese caso,

$$\begin{aligned} \frac{\partial u}{\partial x} &= -\frac{J\left(\frac{f,g}{x,v}\right)}{J\left(\frac{f,g}{u,v}\right)} & \frac{\partial v}{\partial x} &= -\frac{J\left(\frac{f,g}{u,x}\right)}{J\left(\frac{f,g}{u,v}\right)} \\ \frac{\partial u}{\partial y} &= -\frac{J\left(\frac{f,g}{y,v}\right)}{J\left(\frac{f,g}{u,v}\right)} & \frac{\partial v}{\partial y} &= -\frac{J\left(\frac{f,g}{u,y}\right)}{J\left(\frac{f,g}{u,v}\right)}. \end{aligned}$$

Este teorema constituye una generalización de nuestro tiernísimo resultado  $\frac{dy}{dx} = -\frac{F_x}{F_y}$  de la sección 3.4, en donde  $F$  era una función de  $x$  y  $y$ .

Ejemplo:

Determina si el sistema de ecuaciones

$$\begin{aligned} u^2 - v - x^3 + 3y - 3 &= 0 \\ u + v - 2x - y^3 + 3 &= 0 \end{aligned}$$

define a las variables  $u$  y  $v$  como funciones implícitas, diferenciables de  $x$  y  $y$ , alrededor del punto  $P(x, y, u, v) = P(0, 2, 1, 4)$ . De ser así, calcula las derivadas parciales  $\frac{\partial u}{\partial x}\big|_P$ ,  $\frac{\partial u}{\partial y}\big|_P$ ,  $\frac{\partial v}{\partial x}\big|_P$  y  $\frac{\partial v}{\partial y}\big|_P$ .

### 6.3 Teorema general de la función implícita

Primero definimos las funciones  $f(x, y, u, v) = u^2 - v - x^3 + 3y - 3$  y  $g(x, y, u, v) = u + v - 2x - y^3 + 3$ , de modo que el determinante jacobiano  $J\left(\frac{f, g}{u, v}\right)$  en el punto  $P$  está dado por

$$J\left(\frac{f, g}{u, v}\right)\bigg|_P = \begin{vmatrix} f_u & f_v \\ g_u & g_v \end{vmatrix}\bigg|_P = \begin{vmatrix} 2u & -1 \\ 1 & 1 \end{vmatrix}\bigg|_P = 2u + 1|_P = 2(1) + 1 \neq 0.$$

Concluimos entonces que el sistema sí define a las variables  $u$  y  $v$  como funciones implícitas, diferenciables de  $x$  y  $y$ , alrededor del punto  $P$ . Determinemos ahora la derivada parcial  $\frac{\partial u}{\partial x}$ :

$$\begin{aligned} \frac{\partial u}{\partial x} &= -\frac{J\left(\frac{f, g}{x, v}\right)}{J\left(\frac{f, g}{u, v}\right)} = -\frac{\begin{vmatrix} f_x & f_v \\ g_x & g_v \end{vmatrix}}{\begin{vmatrix} f_u & f_v \\ g_u & g_v \end{vmatrix}} = -\frac{\begin{vmatrix} -3x^2 & -1 \\ -2 & 1 \end{vmatrix}}{\begin{vmatrix} 2u & -1 \\ 1 & 1 \end{vmatrix}} \\ &= -\frac{(-3x^2 - 2)}{(2u + 1)} = \frac{3x^2 + 2}{2u + 1}. \end{aligned}$$

De esta manera,

$$\frac{\partial u}{\partial x}\bigg|_P = \frac{3(0)^2 + 2}{2(1) + 1} = \frac{2}{3}.$$

El cálculo de las otras tres derivadas parciales queda como ejercicio para ti, entusiasta lector.

Concluimos esta sección enunciando el teorema general de la función implícita, correspondiente a  $m > 1$  ecuaciones con  $n + m$  variables.

#### Teorema general de la función implícita

Sean  $f_1(x_1, \dots, x_n, w_1, \dots, w_m), \dots, f_m(x_1, \dots, x_n, w_1, \dots, w_m)$  funciones diferenciables. El sistema de ecuaciones

$$\begin{aligned} f_1(x_1, \dots, x_n, w_1, \dots, w_m) &= 0 \\ f_2(x_1, \dots, x_n, w_1, \dots, w_m) &= 0 \\ &\vdots \\ f_m(x_1, \dots, x_n, w_1, \dots, w_m) &= 0. \end{aligned}$$

define a las variables  $w_1, \dots, w_m$  como funciones implícitas, diferenciables de las variables  $x_1, \dots, x_n$  en todos los puntos en donde

$$J\left(\frac{f_1, f_2, \dots, f_m}{w_1, w_2, \dots, w_m}\right) \neq 0.$$

En ese caso,

$$\frac{\partial w_j}{\partial x_i} = - \frac{J \left( \frac{f_1, f_2, \dots, f_j, \dots, f_m}{w_1, w_2, \dots, x_i, \dots, w_m} \right)}{J \left( \frac{f_1, f_2, \dots, f_j, \dots, f_m}{w_1, w_2, \dots, w_j, \dots, w_m} \right)},$$

para todos  $j = 1, \dots, m, i = 1, \dots, n$ .

## 6.4 Teorema del punto fijo

Este teorema se aplica para funciones de un conjunto compacto y convexo  $K \subset \mathbb{R}^n$  hacia el mismo compacto  $K \subset \mathbb{R}^n$ , y se utiliza en microeconomía, por ejemplo, para demostrar la existencia del equilibrio de Walras (equilibrios competitivos en una economía de intercambio).

**Teorema del punto fijo o teorema de Brouwer.** Sea  $K \subset \mathbb{R}^n$  un conjunto compacto y convexo y sea  $f : K \rightarrow K$  una función continua. Entonces  $f$  tiene un punto fijo, es decir, existe un punto  $\vec{x}^* \in K$  para el cual

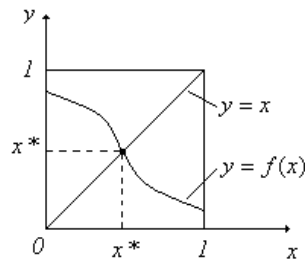
$$f(\vec{x}^*) = \vec{x}^*.$$

Ejemplo:

Sea  $f : [0, 1] \rightarrow [0, 1]$  una función continua. Demuestra que existe  $x^* \in [0, 1]$  tal que  $f(x^*) = x^*$ .

Como la imagen de la función está en el intervalo  $[0, 1]$ , por lo tanto, para todo  $x \in [0, 1]$  se tiene

$$0 \leq f(x) \leq 1.$$



Sea

$$h(x) = f(x) - x.$$

## 6.4 Teorema del punto fijo

Como  $f$  es continua, por lo tanto  $h$  es continua. Por otra parte, es claro que  $h(0) = f(0) - 0 \geq 0$  y  $h(1) = f(1) - 1 \leq 0$ . Así, por el teorema del valor intermedio sabemos que existe  $c \in [0, 1]$  tal que

$$h(c) = 0.$$

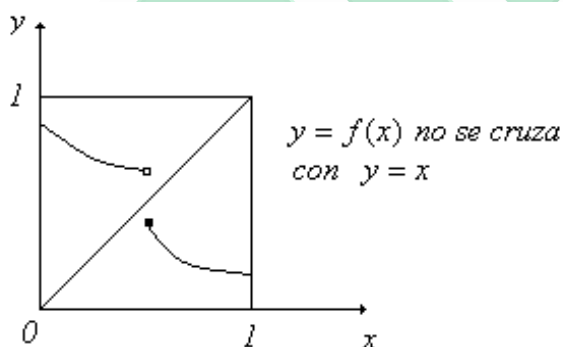
Por lo tanto,  $f(c) - c = 0$ , es decir,

$$f(c) = c.$$

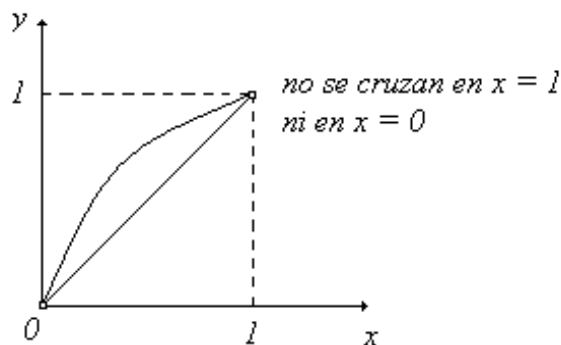
De modo que  $c$  es un punto fijo de  $f$ .

Si alguna de las condiciones del teorema no se satisficieran, entonces ya no necesariamente existiría un punto fijo, como se ilustra en los siguientes casos.

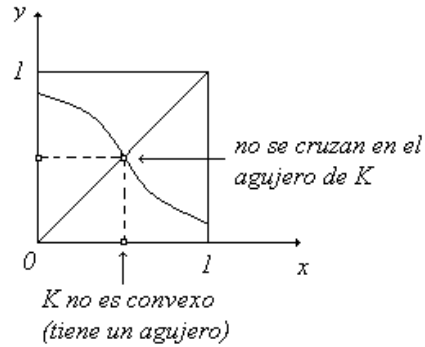
i) Si  $f$  no fuera continua:



ii) Si  $K$  fuera abierto y, por tanto, no fuera compacto:



iii) Si  $K$  no fuera convexo:

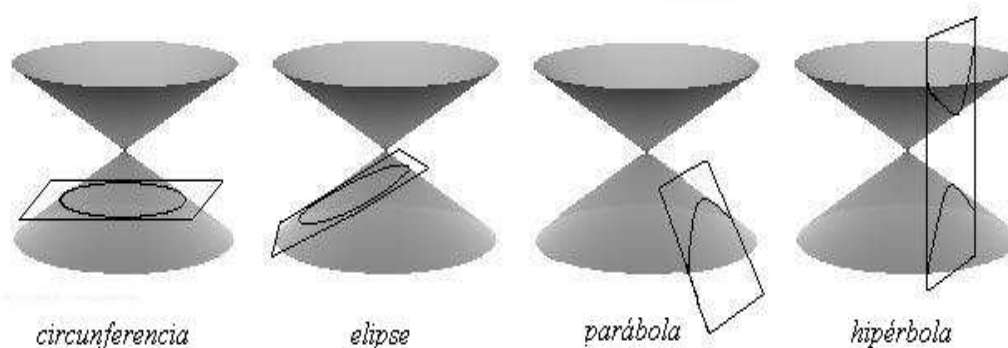


Por ejemplo, a partir de este teorema se puede demostrar la existencia de niveles de insumos  $(x_1, x_2)$  que maximizan una función de beneficio  $\Pi(x_1, x_2)$ , siempre y cuando el beneficio esté representado por una función continua y el conjunto de insumos factibles sea cerrado y acotado. El teorema no proporciona el valor de los insumos óptimos, pero sí garantiza su existencia.

# ApéndiceA

## Cónicas

Las *cónicas* son las curvas (no necesariamente funciones) que se obtienen al rebanar un cono doble con un plano.



La ecuación general de una cónica es una ecuación de segundo grado, de la forma

$$Ax^2 + Bxy + Cy^2 + Dx + Ey = F.$$

A partir de esta ecuación es posible decidir de qué curva se trata, dependiendo del signo del discriminante  $B^2 - 4AC$ :

- i) Cuando  $B^2 - 4AC < 0$  se trata de una elipse (o circunferencia).
- ii) Cuando  $B^2 - 4AC = 0$  se trata de una parábola.
- iii) Cuando  $B^2 - 4AC > 0$  se trata de una hipérbola.

En los casos que consideraremos a continuación supondremos que no aparece el término  $Bxy$ , lo cual implica que la cónica está alineada con los ejes de coordenadas, es decir, no está girada. Así, sólo consideraremos ecuaciones cuadráticas de la forma

$$Ax^2 + Cy^2 + Dx + Ey = F.$$

El objetivo es que puedas distinguir las cónicas más o menos “a ojo”, tomando en cuenta el tipo particular de ecuación.

### 1. Elipse:

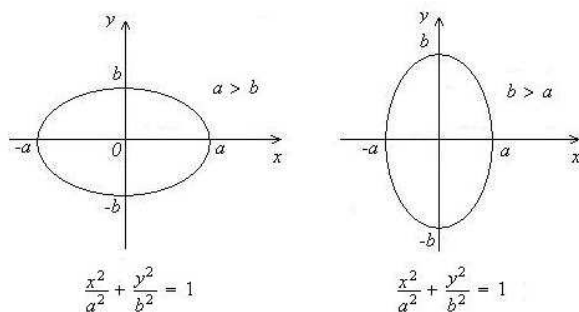
Está descrita por una ecuación cuadrática con  $A \neq 0$ ,  $C \neq 0$  y  $F \neq 0$ , en donde  $A$ ,  $C$  y  $F$  tienen todas el mismo signo. Por simplicidad, supondremos que

## ApéndiceA Cónicas

$D = E = 0$ , obteniendo.  $Ax^2 + Cy^2 = F$ . Al dividir por  $F$  ambos lados de la ecuación, se obtiene la *forma canónica* de la ecuación de la elipse,

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

A continuación se muestran las gráficas de la elipse, en los casos  $a > b$  y  $b > a$ .



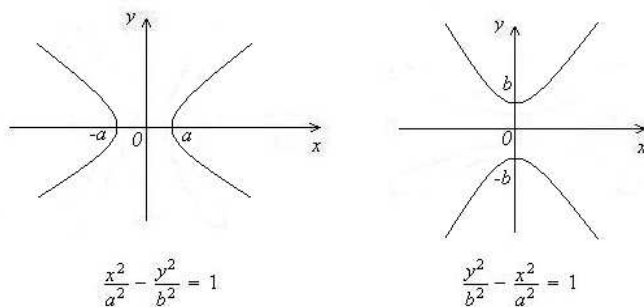
Nota que cuando  $a = b$  la elipse se convierte en la circunferencia  $x^2 + y^2 = a^2$ .

### 2. Hipérbola:

Está descrita por una ecuación cuadrática con  $A \neq 0$ ,  $C \neq 0$  y  $F \neq 0$ , en donde  $A$  y  $C$  tienen signos opuestos. Nuevamente tomamos  $D = E = 0$ . Expresadas en su *forma canónica*, las dos posibles ecuaciones de la hipérbola están dadas por

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 \quad \text{y} \quad \frac{y^2}{b^2} - \frac{x^2}{a^2} = 1.$$

A continuación se muestran las gráficas de las hipérbolas correspondientes a cada una de estas ecuaciones.



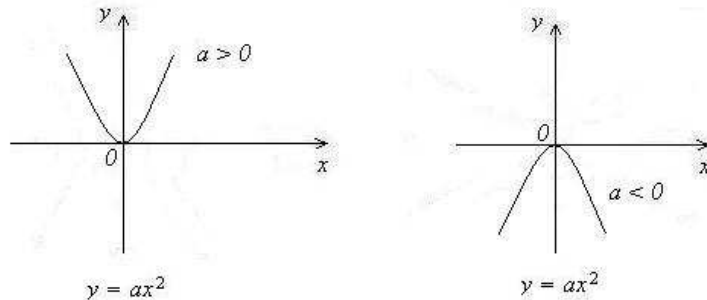
### 3. Parábola:

Está descrita por una ecuación cuadrática tal que  $A = 0$  o  $C = 0$ , pero no ambos cero. En su forma más simple, las *ecuaciones canónicas* de la parábola son

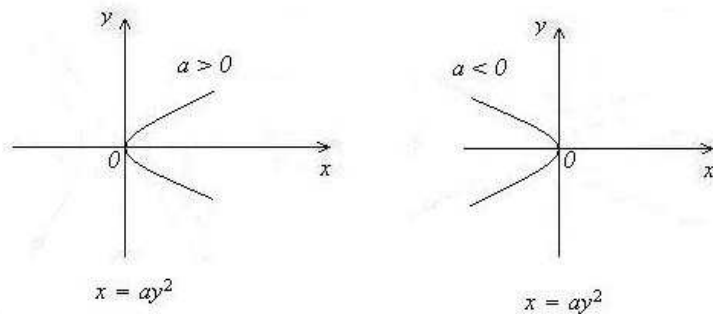
$$y = ax^2 \quad \text{y} \quad x = ay^2.$$



A continuación se muestran las gráficas de las parábola correspondiente a la primer ecuación, para los casos  $a > 0$  y  $a < 0$ .

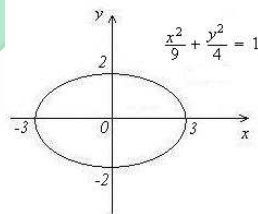


A continuación se muestran las gráficas de las parábola correspondiente a la segunda ecuación, para los casos  $a > 0$  y  $a < 0$ .

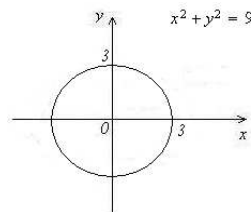


Ejemplos:

1. La ecuación  $4x^2 + 9y^2 = 36$  describe a la elipse  $\frac{x^2}{9} + \frac{y^2}{4} = 1$ .

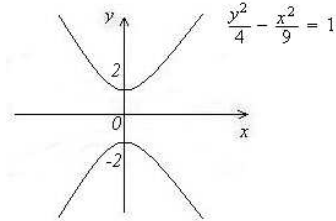


2. La ecuación  $4x^2 + 4y^2 = 36$  describe a la circunferencia  $x^2 + y^2 = 9$ .

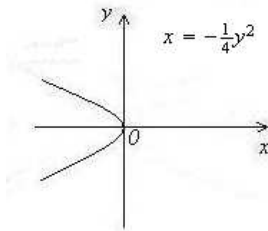


## ApéndiceA Cónicas

3. La ecuación  $4x^2 - 9y^2 = -36$  describe a la hipérbola  $\frac{y^2}{4} - \frac{x^2}{9} = 1$ .



4. La ecuación  $4x + y^2 = 0$  describe a la parábola  $x = -\frac{1}{4}y^2$ .



Por último, es importante señalar que existen algunos casos degenerados de cónicas, como se muestra a continuación.

Ejemplos:

1. La ecuación  $4x^2 + 9y^2 = 0$  no define una elipse, sino más bien un punto en el plano  $\mathbb{R}^2$ , a saber, el origen.
2. La ecuación  $4x^2 - 9y^2 = 0$  no define una hipérbola, sino más bien dos rectas en el plano, a saber, las rectas  $y = \pm \frac{2x}{3}$ .

## ApéndiceB

### Teoremas de concavidad para funciones en $\mathbb{R}^n$

Aquí se presenta la generalización de algunos de los teoremas de optimización para el caso de funciones diferenciables en  $\mathbb{R}^n$ . Para ello, es necesario introducir antes algunas definiciones importantes.

**Definición.** Dada una matriz  $A$  de  $n \times n$ , una *submatriz principal dominante de orden  $k$* , con  $k = 1, \dots, n$ , es la matriz  $A_k$  de  $k \times k$  que se obtiene al eliminar en  $A$  los últimos  $n - k$  renglones y las últimas  $n - k$  columnas. El determinante  $|A_k|$  de la submatriz principal dominante  $A_k$  de orden  $k$  se conoce como el *menor principal dominante de orden  $k$* .

Ejemplos:

1. De la matriz  $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$  de  $2 \times 2$  se puede construir una submatriz principal dominante de orden  $k = 1$ :

$$A_1 = (a),$$

obtenida al eliminar en  $A$  el renglón 2 y la columna 2, y una submatriz principal dominante de orden  $k = 2$ :

$$A_2 = A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

en donde no se ha eliminado renglón ni columna alguna. Los menores principales dominantes correspondientes son

$$|A_1| = a \quad \text{y} \quad |A_2| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - cb.$$

2. De la matriz  $A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$  de  $3 \times 3$  se puede construir una submatriz principal dominante de orden  $k = 1$ :

$$A_1 = (a),$$

obtenida al eliminar en  $A$  los renglones 2 y 3, y sus correspondientes columnas, una submatriz principal dominante de orden  $k = 2$ :

$$A_2 = \begin{pmatrix} a & b \\ d & e \end{pmatrix},$$

obtenida al eliminar el renglón 3 y la columna 3, y una submatriz principal dominante de orden  $k = 3$ :

$$A_3 = A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix},$$

en donde no se ha eliminado renglón ni columna alguna. Los menores principales dominantes correspondientes son los determinantes

$$|A_1| = a, \quad |A_2| = \begin{vmatrix} a & b \\ d & e \end{vmatrix} \quad \text{y} \quad |A_3| = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix}.$$

**Definición.** Dada una matriz  $A$  de  $n \times n$ , las *submatrices principales de orden  $k$* , con  $k = 1, \dots, n$ , son todas las matrices de  $k \times k$  que se obtienen al eliminar en  $A$  cualesquiera  $n - k$  renglones y sus correspondientes  $n - k$  columnas. El determinante de cada submatriz principal de orden  $k$  se conoce como *menor principal de orden  $k$* .

Ejemplos:

1. De la matriz  $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$  de  $2 \times 2$  se puede construir dos submatrices principales de orden  $k = 1$ :

$$(a) \quad \text{y} \quad (d),$$

obtenidas al eliminar en  $A$  el renglón 2 y la columna 2, o bien, el renglón 1 y la columna 1. Hay una sola submatriz principal de orden  $k = 2$ :

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

en donde no se ha eliminado renglón ni columna alguna.

2. De la matriz  $A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$  de  $3 \times 3$  se puede construir tres submatrices principales de orden  $k = 1$ :

$$(a), \quad (e) \quad \text{y} \quad (i),$$

obtenidas al eliminar en  $A$  los renglones 2 y 3 y sus correspondientes columnas, los renglones 1 y 3 y sus correspondientes columnas, o bien, los renglones 1 y 2 y sus correspondientes columnas. Hay tres submatrices principales de orden  $k = 2$ :

$$\begin{pmatrix} a & b \\ d & e \end{pmatrix}, \quad \begin{pmatrix} a & c \\ g & i \end{pmatrix} \quad \text{y} \quad \begin{pmatrix} e & f \\ h & i \end{pmatrix},$$

obtenidas al eliminar el renglón 3 y la columna 3, el renglón 2 y la columna 2, o bien, el renglón 1 y la columna 1. Hay una sola submatriz principal de orden  $k = 3$ :

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix},$$

en donde no se ha eliminado renglón ni columna alguna.

**Definición.** Sea  $A$  una matriz simétrica de  $n \times n$ . Se dice que:

- a)  $A$  es *definida positiva*  $\Leftrightarrow$  todos los menores principales dominantes de  $A$  son estrictamente positivos ( $> 0$ ).
- b)  $A$  es *semidefinida positiva*  $\Leftrightarrow$  todos los menores principales de  $A$  son no negativos ( $\geq 0$ ).
- c)  $A$  es *definida negativa*  $\Leftrightarrow$  todos los menores principales dominantes de  $A$  de orden impar son negativos ( $< 0$ ) y todos los de orden par son estrictamente positivos ( $> 0$ ).
- d)  $A$  es *semidefinida negativa*  $\Leftrightarrow$  todos los menores principales de  $A$  de orden impar son no positivos ( $\leq 0$ ) y todos los de orden par son no negativos ( $\geq 0$ ).
- e)  $A$  es *indefinida*, si no se satisfacen los patrones de signo anteriores.

Ejemplo:

La matriz

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

es indefinida, ya que sus menores principales violan los patrones de signo a)-d). En efecto, los menores principales de orden 2 (par) de  $A$  son

$$\begin{vmatrix} 1 & 0 \\ 0 & 0 \end{vmatrix} = 0, \quad \begin{vmatrix} 1 & 1 \\ 1 & 0 \end{vmatrix} = -1 \quad \text{y} \quad \begin{vmatrix} 0 & 0 \\ 0 & 0 \end{vmatrix} = 0,$$

uno de los cuales es negativo.

## ApéndiceB Teoremas de concavidad para funciones en $\mathbb{R}^n$

Con esta última definición podemos extender los conceptos de concavidad y convexidad para una función general  $f(x_1, x_2, \dots, x_n)$  doblemente diferenciable. En ese caso, la matriz hessiana  $H$  de  $f$  es la matriz simétrica

$$H = \begin{pmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ f_{n1} & f_{n2} & \cdots & f_{nn} \end{pmatrix},$$

en donde  $f_{ij} \equiv \frac{\partial^2 f}{\partial x_j \partial x_i}$ .

Comenzamos con el caso correspondiente a concavidad o convexidad estricta, en donde todos los menores principales dominantes son distintos de cero, es decir,  $|H_k| \neq 0$ .

**Teorema.** Sea  $S \subset \mathbb{R}^n$  un conjunto abierto y convexo, y sea  $f : S \rightarrow \mathbb{R}$ , con  $f \in C^2(S)$ . Sea  $H$  la matriz hessiana de  $f$ . Entonces

- a)  $H$  es definida positiva  $\Rightarrow f$  es estrictamente convexa en  $S$ ,
- b)  $H$  es definida negativa  $\Rightarrow f$  es estrictamente cóncava en  $S$ ,
- c)  $H$  es indefinida  $\Rightarrow f$  no es ni cóncava ni convexa en  $S$ .

Los resultados de este teorema se resumen en la siguiente tabla:

|          | $f$ estric. convexa | $f$ estric. cóncava                    |
|----------|---------------------|--|
| $ H_1 $  | +                   | −                                      |
| $ H_2 $  | +                   | +                                      |
| $ H_3 $  | +                   | −                                      |
| $ H_4 $  | +                   | +                                      |
| $\vdots$ | $\vdots$            | $\vdots$                               |
| $ H_n $  | +                   | +, si $n$ es par<br>−, si $n$ es impar |

Ejemplos:

1. La función  $f(x, y, z) = x^2 + y^2 + z^2$  es estrictamente convexa, ya que la matriz hessiana de  $f$ ,

$$H = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix},$$

es definida positiva. En efecto, sus menores principales dominantes son

$$|H_1| = 2, \quad |H_2| = \begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix} = 4 \quad \text{y} \quad |H_3| = \begin{vmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{vmatrix} = 8,$$

que satisfacen el patrón de signos  $|H_1| > 0$ ,  $|H_2| > 0$  y  $|H_3| > 0$ .

2. La función  $f(x, y, z) = -x^2 - y^2 - z^2$  es estrictamente cóncava, ya que la matriz hessiana de  $f$ ,

$$H = \begin{pmatrix} -2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{pmatrix},$$

es definida negativa. En efecto, sus menores principales dominantes son

$$|H_1| = -2, \quad |H_2| = \begin{vmatrix} -2 & 0 \\ 0 & -2 \end{vmatrix} = 4 \quad \text{y} \quad |H_3| = \begin{vmatrix} -2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{vmatrix} = -8,$$

que satisfacen el patrón de signos  $|H_1| < 0$ ,  $|H_2| > 0$  y  $|H_3| < 0$ .

### Teorema (condiciones suficientes de segundo orden)

Sea  $S \subset \mathbb{R}^n$  un conjunto abierto y convexo, y sea  $f : S \rightarrow \mathbb{R}$ , con  $f \in C^2(S)$ . Sea  $\vec{x}_0 \in S$  un punto crítico de  $f$ . Sea  $H(\vec{x}_0)$  la matriz hessiana de  $f$  evaluada en  $\vec{x}_0$ , con menores principales dominantes  $|H_1|, |H_2|, \dots, |H_n|$ . Entonces

- a)  $|H_k| > 0$ , para toda  $k = 1, \dots, n$   
 $\Rightarrow f$  tiene un mínimo local estricto en  $\vec{x}_0$ ,
- b)  $(-1)^k |H_k| > 0$ , para toda  $k = 1, \dots, n$   
 $\Rightarrow f$  tiene un máximo local estricto en  $\vec{x}_0$ ,
- c) Si  $\{|H_1|, |H_2|, \dots, |H_n|\}$  viola la secuencia anterior de signos  
 $\Rightarrow f$  tiene un punto silla en  $\vec{x}_0$ .

El valor mínimo o máximo local es un extremo global de  $f$ , cuando los patrones de signo a) y b) se satisfacen en todo el dominio de  $f$ .

Ejemplos:

1. La función  $f(x, y, z) = x^2 + y^2 + z^2$  tiene un único punto crítico en  $(x, y, z) = (0, 0, 0)$ . Como se demostró anteriormente, los menores principales dominantes de  $f$  satisfacen  $|H_1| > 0$ ,  $|H_2| > 0$  y  $|H_3| > 0$ , de modo que  $f$  es estrictamente convexa en  $\mathbb{R}^3$ . Concluimos que  $f$  tiene un mínimo global estricto en  $(0, 0, 0)$ .

## ApéndiceB Teoremas de concavidad para funciones en $\mathbb{R}^n$

2. La función  $f(x, y, z) = -x^2 - y^2 - z^2$  tiene un único punto crítico en  $(x, y, z) = (0, 0, 0)$ . Como se demostró anteriormente, los menores principales dominantes de  $f$  satisfacen  $|H_1| < 0$ ,  $|H_2| > 0$  y  $|H_3| < 0$ , de modo que  $f$  es estrictamente cóncava en  $\mathbb{R}^3$ . Concluimos que  $f$  tiene un máximo global estricto en  $(0, 0, 0)$ .

A continuación se enuncian los teoremas correspondientes al caso en donde alguno(s) de los menores principales dominantes  $|H_k|$  de la función  $f$  es igual a cero.

**Teorema.** Sea  $S \subset \mathbb{R}^n$  un conjunto abierto y convexo, y sea  $f : S \rightarrow \mathbb{R}$ , con  $f \in C^2(S)$ . Sea  $H$  la matriz hessiana de  $f$ . Entonces

- a)  $H$  es semidefinida positiva  $\Leftrightarrow f$  es convexa en  $S$ ,
- b)  $H$  es semidefinida negativa  $\Leftrightarrow f$  es cóncava en  $S$ ,

Nota las implicaciones del tipo  $\Leftrightarrow$  en este último teorema, que contrastan con las del tipo  $\Rightarrow$  para funciones estrictamente convexas (cóncavas).

Ejemplo:

La función  $f(x, y, z) = x^2 + y + z^2$  es convexa no estricta, ya que la matriz hessiana de  $f$ ,

$$H = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix},$$

es semidefinida positiva. En efecto, sus menores principales de orden 1 son

$$2, \quad 0 \quad \text{y} \quad 2,$$

sus menores principales de orden 2 son

$$\begin{vmatrix} 2 & 0 \\ 0 & 0 \end{vmatrix} = 0, \quad \begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix} = 4 \quad \text{y} \quad \begin{vmatrix} 0 & 0 \\ 0 & 2 \end{vmatrix} = 0,$$

y su menor principal de orden 3 es

$$\begin{vmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{vmatrix} = 0.$$



### Teorema (condiciones necesarias de segundo orden)

Sea  $S \subset \mathbb{R}^n$  un conjunto abierto y convexo, y sea  $f : S \rightarrow \mathbb{R}$ , con  $f \in C^2(S)$ . Sea  $\vec{x}_0 \in S$  un punto crítico de  $f$ . Sea  $H(\vec{x}_0)$  la matriz hessiana de  $f$  evaluada en  $\vec{x}_0$ . Entonces

- a)  $f$  tiene un mínimo local en  $\vec{x}_0$   
 $\Rightarrow$  todos los menores principales de  $H$  son no negativos ( $\geq 0$ ) en  $\vec{x}_0$ ,
- b)  $f$  tiene un máximo local en  $\vec{x}_0$   
 $\Rightarrow$  todos los menores principales de  $H$  de orden impar son no positivos ( $\leq 0$ ) en  $\vec{x}_0$  y todos los de orden par son no negativos ( $\geq 0$ ) en  $\vec{x}_0$ .

El valor mínimo o máximo local es un extremo global de  $f$ , cuando los patrones de signo a) y b) se satisfacen en todo el dominio de  $f$ .

Nota que aquí la concavidad o convexidad de  $f$  es una condición necesaria, mas no suficiente, para un extremo local. En otras palabras, no basta con demostrar la concavidad o convexidad de la función para garantizar la existencia de un máximo o un mínimo.

## Bibliografía

1. B.R. Binger, E. Hoffman, *Microeconomics with Calculus*, 2nd. edition, Addison Wesley, 1997.
2. A.C. Chiang, *Métodos Fundamentales de Economía Matemática*, 3a. edición, McGraw-Hill Interamericana de México, 1987.
3. O. Estrada, P. García y Colomé, G. Monsivais, *Cálculo Vectorial y Aplicaciones*, Grupo Editorial Iberoamérica, 2003.
4. D.S. Kaplan, *A Practical Guide to Lagrangeans*, Centro de Investigación Económica, ITAM, 2005.
5. H. Lomelí, B. Rumbos, *Métodos Dinámicos en Economía: Otra Búsqueda del Tiempo Perdido*, 2a. edición, Jit Press, 2010.
6. J. E. Marsden, A.J. Tromba, *Cálculo Vectorial*, 5a. edición, Pearson, 2004.
7. M.J. Osborne, *Mathematical Methods for Economic Theory: A Tutorial*, <http://www.economics.utoronto.ca/osborne/MathTutorial>, 2007.
8. G. Pastor, *Matemáticas IV*, ITAM, 1993.
9. E. Silberberg, W. Suen, *The Estructure of Economics: A Mathematical Analysis*, 3rd. edition, McGraw-Hill, 2001.
10. C.P. Simon, L. Blume, *Mathematics for Economists*, Norton, 1994.
11. K. Sydsaeter, P.J. Hammond, A. Carvajal, *Matemáticas para el Análisis Económico*, Pearson, 2a. edición, 2012.
12. K. Sydsaeter, P.J. Hammond, *Essential Mathematics for Economic Analysis*, 2nd. edition, Prentice Hall, 2006.
13. K. Sydsaeter, P.J. Hammond, A. Seierstad, A. Strom, *Further Mathematics for Economic Analysis*, 2nd. edition, Prentice Hall, 2008.
14. G.B. Thomas, R.L. Finney, *Cálculo*, Vols. I y II, 12a. edición, Adisson Wesley, 2004.