

# CÁLCULO DE FUNCIONES DE VARIAS VARIABLES

## 1. INTRODUCCIÓN

Consideramos en este curso funciones definidas sobre el espacio  $\mathbb{R}^N$ , el conjunto de las  $N$ -tuplas ordenadas de números reales,

$$x = (x_1, \dots, x_N), \quad x_i \in \mathbb{R} \quad \forall i = 1, \dots, N.$$

Equivalentemente, puede caracterizarse  $\mathbb{R}^N$  como el conjunto de las funciones

$$x : \{1, \dots, N\} \rightarrow \mathbb{R}, \quad i \mapsto x_i.$$

Dotado de las operaciones básicas de suma y producto por escalar,

$$\alpha(x_1, \dots, x_N) + \beta(y_1, \dots, y_N) := (\alpha x_1 + \beta y_1, \dots, \alpha x_N + \beta y_N),$$

$\mathbb{R}^N$  es un *espacio vectorial*. El curso de Álgebra lineal fue destinado fundamentalmente al estudio de funciones *lineales*  $f : \mathbb{R}^N \rightarrow \mathbb{R}^m$ . Este curso continúa este estudio, ahora para funciones no-necesariamente lineales, en torno a los conceptos de continuidad y derivabilidad que serán apropiadamente definidas. Como en el cálculo de funciones de una variable, las funciones *diferenciables*  $\mathbb{R}^N \rightarrow \mathbb{R}^m$  serán aquellas que pueden *aproximarse bien* por funciones lineales-afines, localmente en torno a cada punto del dominio.

El concepto de límite se generalizará a  $\mathbb{R}^N$ , lo que conllevará la extensión de las nociones topológicas básicas ya conocidas en la recta real, al espacio  $\mathbb{R}^N$ , especialmente aquella de *continuidad* de funciones  $f : \mathbb{R}^N \rightarrow \mathbb{R}^m$ . Con la ayuda del álgebra lineal, la noción de diferenciabilidad aparecerá en modo natural.

El Cálculo integral también se extenderá a funciones de  $\mathbb{R}^N$  a  $\mathbb{R}^m$ . En interpretación geométrica, cuando  $N = 2$ ,  $m = 1$ , se trata de obtener una noción apropiada de *volumen bajo el gráfico* de una función de dos variables a valores reales.

La extensión de las nociones del cálculo diferencial e integral a funciones de más de una variable, será a veces directa, a veces merecedora de un análisis más profundo que aquél llevado a cabo en una variable. La buena comprensión de los conceptos del cálculo en una variable es condición necesaria para entender aquellos en este curso, pero al mismo tiempo, la mayor generalidad permitirá una comprensión más profunda de los conceptos básicos.

El Cálculo en varias variables es fundamental en el desarrollo de la física y en la aplicación de la matemática al modelamiento de una amplia diversidad de fenómenos en ingeniería, química, biología, economía y otras ciencias. No es el rol de este curso el análisis de modelos en los cuales el cálculo se aplica, sino la profundización en los conceptos matemáticos inherentes al Cálculo, que son por si mismos delicados y profundos, en parte por ello sus posibilidades virtualmente ilimitadas.

## 2. CONCEPTOS PRELIMINARES

**2.1.  $\mathbb{R}^N$  como espacio vectorial normado.** El propósito de esta sección es la introducción de algunos elementos topológicos básicos asociados a la estructura de *espacio normado* con la que naturalmente cuenta  $\mathbb{R}^N$ . El modo más natural de medir la distancia desde el origen a un punto  $x = (x_1, \dots, x_N) \in \mathbb{R}^N$  es mediante la *norma Euclídeana* de  $x$ , definida como

$$\|x\| =: \left( \sum_{i=1}^N |x_i|^2 \right)^{\frac{1}{2}} . \quad (2.1)$$

La norma Euclídeana está vinculada al *producto interno canónico* de vectores  $x, y \in \mathbb{R}^N$ , dado por

$$x \cdot y = \sum_{i=1}^N x_i y_i .$$

En efecto, vemos que

$$\|x\| = \sqrt{x \cdot x} .$$

Resumimos las propiedades principales de la norma Euclídeana en el siguiente resultado.

**Proposición 2.1.** *Para todo  $x, y \in \mathbb{R}^N$  y todo  $\alpha \in \mathbb{R}$  se tiene la validez de las siguientes propiedades.*

(1)

$$\|\alpha x\| = |\alpha| \|x\| .$$

(2)

$$\|x\| \geq 0 \quad y \quad \|x\| = 0 \iff x = 0 .$$

(3) *Desigualdad triangular:*

$$\|x + y\| \leq \|x\| + \|y\| .$$

(4) *Desigualdad de Cauchy-Schwartz:*

$$|x \cdot y| \leq \|x\| \|y\| .$$

**Demostración.** Las propiedades (1) y (2) son evidentes. Para verificar (4), suponemos que  $x \neq 0$ ,  $y \neq 0$ , pues de lo contrario no hay nada que probar. Observemos que dado cualquier  $\lambda > 0$ ,

$$\|\lambda^{\frac{1}{2}}x - \lambda^{-\frac{1}{2}}y\|^2 = \sum_{i=1}^N (\lambda x_i - \lambda^{-1}y_i)^2 = \lambda \sum_{i=1}^N x_i^2 + 2 \sum_{i=1}^N x_i y_i + \lambda^{-1} \sum_{i=1}^N y_i^2,$$

de modo que

$$0 \leq \lambda \|x\|^2 + \lambda^{-1} \|y\|^2 + 2 \sum_{i=1}^N x_i y_i.$$

Escojamos  $\lambda = \|y\| \|x\|^{-1} > 0$ . Evaluando en la desigualdad anterior obtenemos

$$0 \leq 2\|x\| \|y\| + 2 \sum_{i=1}^N x_i y_i.$$

Reemplazando  $x$  por  $-x$  obtenemos también que

$$0 \leq 2\|x\| \|y\| - 2 \sum_{i=1}^N x_i y_i.$$

Concluimos que

$$\pm \sum_{i=1}^N x_i y_i \leq \|x\| \|y\|$$

y por ende la validez de (4). Verifiquemos ahora la desigualdad triangular (3). Tenemos, gracias a la definición de la norma y la desigualdad (4), que

$$\begin{aligned} \|x + y\|^2 &= \|x\|^2 + 2x \cdot y + \|y\|^2 \leq \\ \|x\|^2 + 2\|x\| \|y\| + \|y\|^2 &= (\|x\| + \|y\|)^2, \end{aligned}$$

y por lo tanto

$$\|x + y\| \leq \|x\| + \|y\|,$$

y la propiedad (3) está entonces demostrada.  $\square$

Naturalmente asociada a la norma Euclídeana, la distancia entre dos puntos  $x, y \in \mathbb{R}^N$  se define entonces como

$$d(x, y) =: \|x - y\| = \left( \sum_{i=1}^k |x_i - y_i|^2 \right)^{\frac{1}{2}}. \quad (2.2)$$

**2.2. Sucesiones en  $\mathbb{R}^N$ , convergencia.** Asociada a la noción de distancia (2.2) está la de límite. Introducimos, para comenzar, el concepto de límite de una sucesión en  $\mathbb{R}^N$ .

Una sucesión en  $\mathbb{R}^N$  es una función  $x : \mathbb{N} \rightarrow \mathbb{R}^N$ ,  $n \mapsto x_n$ . Anotamos usualmente

$$x = (x_n)_{n \in \mathbb{N}} \quad \text{o simplemente} \quad x_n, \quad n \in \mathbb{N}.$$

Por ejemplo,

$$x_n = \left(\frac{1}{n}, n^2, e^{-n}\right), \quad n \in \mathbb{N}, \quad y_n = (\sin n, \cos n), \quad n \in \mathbb{N},$$

representan respectivamente sucesiones en  $\mathbb{R}^3$  y  $\mathbb{R}^2$ .

Sean  $x_n$ ,  $n \in \mathbb{N}$  una sucesión en  $\mathbb{R}^N$  y  $x$  un punto en  $\mathbb{R}^N$ . Decimos que  $x_n$  *converge a  $x$*  si

$$\lim_{n \rightarrow \infty} \|x_n - x\| = 0,$$

esto es, si la sucesión de numeros reales dada por la distancia de  $x_n$  a  $x$ , tiende a 0. Escribimos en tal caso,

$$\lim_{n \rightarrow \infty} x_n = x \quad \text{o también} \quad x_n \rightarrow x \quad \text{cuando} \quad n \rightarrow \infty.$$

Observemos que, escribiendo la definición del límite de la sucesión real  $\|x_n - x\|$  a cero, obtenemos que  $x_n \rightarrow x$  si y solo si

$$(\forall \varepsilon > 0) (\exists n_0 \in \mathbb{N}) (\forall n \geq n_0) : \|x_n - x\| < \varepsilon, \quad (2.3)$$

que corresponde a la definición de convergencia en  $\mathbb{R}$  con la norma Euclidea reemplazando al valor absoluto.

Una criterio práctico para la convergencia de sucesiones es la siguiente afirmación:

**Proposición 2.2.** Sean  $x_n$ ,  $n \in \mathbb{N}$  una sucesión en  $\mathbb{R}^N$  y  $x$  un punto en  $\mathbb{R}^N$ . Escribamos

$$x_n = (x_{n1}, \dots, x_{nN}), \quad x = (x_1, \dots, x_N).$$

Entonces, las siguientes afirmaciones son equivalentes:

- (a)  $x_n \rightarrow x$  cuando  $n \rightarrow \infty$ .
- (b) Para todo  $i = 1, \dots, N$  se tiene que  $x_{ni} \rightarrow x_i$  cuando  $n \rightarrow \infty$ .

En otras palabras, la convergencia de una sucesión en  $\mathbb{R}^N$  es equivalente a la convergencia de todas y cada una de sus coordenadas.

**Demostración.** Demostremos que  $(a) \implies (b)$ . Supongamos que  $x_n \rightarrow x$  cuando  $n \rightarrow \infty$ . Consideremos  $i \in \{1, \dots, N\}$  y observemos que

$$0 \leq |x_{ni} - x_i| \leq \sqrt{\sum_{k=1}^N |x_{nk} - x|^2} \rightarrow 0.$$

Por el teorema del sandwich, se sigue entonces que  $|x_{ni} - x_i| \rightarrow 0$  cuando  $n \rightarrow +\infty$ , lo cual equivale a

$$\lim_{n \rightarrow \infty} x_{ni} = x_i. \quad (2.4)$$

Probemos ahora que  $(b) \implies (a)$ . Supongamos ahora la validez de (b), esto es que (2.4) se cumple para todo  $i$ . Entonces  $|x_{ni} - x_i| \rightarrow 0$  cuando  $n \rightarrow +\infty$ , y por ende  $|x_{ni} - x_i|^2 \rightarrow 0$ . Se sigue que  $\sum_{i=1}^N |x_{ni} - x_i|^2 \rightarrow 0$  y por lo tanto

$$\|x_n - x\| = \sqrt{\sum_{i=1}^N |x_{ni} - x_i|^2} \rightarrow 0,$$

esto es,  $x_n \rightarrow x$  y (a) por ende se cumple. La demostración ha sido concluida.  $\square$

A partir de la caracterización anterior puede fácilmente calcularse el límite de sucesiones concretas en  $\mathbb{R}^N$

**Ejemplo 2.1.** Consideremos la sucesión en  $\mathbb{R}^3$

$$x_n = (x_{n1}, x_{n2}, x_{n3}) = \left( \frac{1}{n}, 2e^{\frac{1}{n^2}}, \cos e^{-n} \right).$$

Sabemos, a partir de lo conocido para sucesiones reales que:

$$\lim_{n \rightarrow \infty} \frac{1}{n} = 0, \quad \lim_{n \rightarrow \infty} \frac{1}{n^2} = 0, \quad \lim_{n \rightarrow \infty} \frac{1}{e^n} = 0.$$

Como también sabemos, las funciones  $t \mapsto 2e^t$  y  $t \mapsto \cos t$  son continuas en  $t = 0$ . Por lo tanto, tenemos que

$$\lim_{n \rightarrow \infty} 2e^{\frac{1}{n^2}} = 2e^0 = 2, \quad \lim_{n \rightarrow \infty} \cos e^{-n} = \cos 0 = 1.$$

Así,

$$\lim_{n \rightarrow \infty} x_{1n} = 0, \quad \lim_{n \rightarrow \infty} x_{2n} = 2, \quad \lim_{n \rightarrow \infty} x_{3n} = 1,$$

y por lo tanto,

$$\lim_{n \rightarrow \infty} x_n = (0, 2, 1).$$

A partir de la caracterización de la convergencia en la proposición 2.2, varias propiedades de la convergencia de sucesiones en  $\mathbb{R}^N$  se deducen en modo relativamente simple a partir de las correspondientes propiedades de sucesiones en  $\mathbb{R}$ . Por ejemplo, tenemos la validez del hecho siguiente.

**Proposición 2.3.** *Sean  $x_n, y_n$  sucesiones en  $\mathbb{R}^N$ . Supongamos que se tiene  $x_n \rightarrow x, y_n \rightarrow y$ . Entonces, si  $\alpha, \beta \in \mathbb{R}$ , la sucesión  $\alpha x_n + \beta y_n$  es convergente y su límite es igual a  $\alpha x + \beta y$ .*

**Demostración.** Supongamos que

$$x_n = (x_{n1}, \dots, x_{nN}), \quad x = (x_1, \dots, x_N),$$

$$y_n = (y_{n1}, \dots, y_{nN}), \quad y = (y_1, \dots, y_N).$$

Sea  $z_n = \alpha x_n + \beta y_n$ . Por definición de las operaciones de suma y ponderación por escalar, se tiene que

$$z_n = (\alpha x_{n1} + \beta y_{n1}, \dots, \alpha x_{nN} + \beta y_{nN}).$$

Por el lema anterior, se tiene que

$$x_{ni} \rightarrow x_i, \quad y_{ni} \rightarrow y_i \quad \text{para todo } i = 1, \dots, N.$$

Por la propiedad conocida para convergencia de sucesiones reales se tiene que

$$\alpha x_{ni} + \beta y_{ni} \rightarrow \alpha x_i + \beta y_i,$$

Esto es, cada coordenada de la sucesión  $z_n$  converge a la correspondiente coordenada del punto  $\alpha x + \beta y$ . Nuevamente en virtud de la proposición 2.2, se sigue que  $z_n \rightarrow \alpha x + \beta y$ , y hemos completado la demostración.  $\square$

**2.3. Interior, adherencia, conjuntos abiertos y cerrados.** Introduciremos a continuación ciertas nociones básicas asociadas a la estructura de *espacio vectorial normado* de  $\mathbb{R}^N$ .

Consideremos un número  $R > 0$  y un punto  $x_0 \in \mathbb{R}^N$ . La *bola abierta* de centro  $x_0$  y radio  $R > 0$  es el conjunto

$$B(x_0, R) = \{x \in \mathbb{R}^N / \|x - x_0\| < R\}.$$

Por ejemplo en  $\mathbb{R}^3$  se tiene que

$$B((0, 1, -1), 2) = \{(x, y, z) / x^2 + (y - 1)^2 + (z + 1)^2 < 4\}$$

que representa una esfera sólida centrada en el punto  $(0, 1, -1)$  de radio 2, que no incluye su periferia.

Sea  $A \subset \mathbb{R}^N$ . Definimos el *interior de A* como el conjunto  $\text{Int}(A)$  dado por

$$\text{Int}(A) = \{x \in \mathbb{R}^N / \exists \delta > 0 : B(x, \delta) \subset A\}.$$

$x \in \text{Int}(A)$  se denomina *punto interior de A*. Definimos paralelamente la noción de *adherencia de A*,  $\text{Adh}(A)$  como

$$\text{Adh}(A) = \{x \in \mathbb{R}^N / \exists x_n \rightarrow x : x_n \in A \forall n \in \mathbb{N}\}.$$

Similarmente,  $x \in \text{Adh}(A)$  se denomina *punto de adherencia de A*. Así, en palabras,  $\text{Int}(A)$  es el conjunto de aquellos puntos para los cuales existe alguna bola centrada en el punto completamente contenida en el conjunto  $A$ .

Por otra parte  $\text{Adh}(A)$  es el conjunto *de todos aquellos puntos del espacio que pueden aproximarse por una sucesión de puntos en A*. Observemos que siempre se tiene la cadena de inclusiones

$$\text{Int}(A) \subset A \subset \text{Adh}(A).$$

Para la primera inclusión, observemos que si  $x \in \text{Int}(A)$ , entonces para algún  $\delta > 0$ ,  $x \in B(x, \delta) \subset A$ , por ende  $x \in A$ .

Intuitivamente,  $\text{Int}(A)$  es  $A$  sin el borde de  $A$  mientras que  $\text{Adh}(A)$  está constituido por  $A$  unido con estos puntos de borde.

Para la segunda inclusión, basta observar que si  $x \in A$  entonces la sucesión constante  $x_n = x$ ,  $n \in \mathbb{N}$ , está contenida en  $A$  y converge a  $x$ . Por lo tanto,  $x \in \text{Adh}(A)$ .

Los *operadores* interior y adherencia de subconjuntos de  $\mathbb{R}^N$  están relacionados del modo siguiente:

$$\text{Adh}(A) = \mathbb{R}^N \setminus \text{Int}(\mathbb{R}^N \setminus A). \quad (2.5)$$

En efecto,  $x \in \mathbb{R}^N \setminus \text{Int}(\mathbb{R}^N \setminus A)$  si y solo si  $x \notin \text{Int}(\mathbb{R}^N \setminus A)$ , esto es,

$$\forall \delta > 0 : B(x, \delta) \not\subset \mathbb{R}^N \setminus A,$$

o sea

$$\forall \delta > 0 : B(x, \delta) \cap A \neq \emptyset. \quad (2.6)$$

Afirmamos que la relación (2.6) es equivalente a  $x \in \text{Adh}(A)$ . En efecto, si  $x$  satisface la relación (2.6), se sigue que para todo  $n \in \mathbb{N}$  la bola  $B(x, \frac{1}{n})$  contiene algún punto que  $x_n \in A$  con  $\|x_n - x\| < \frac{1}{n}$ . Por lo tanto existe una sucesión de puntos de  $A$  que converge a  $x$ . Recíprocamente, si  $x \in \text{Adh}(A)$  entonces existe  $x_n \in A$  con  $\|x_n - x\| \rightarrow 0$ . Por definición de límite, dado cualquier  $\delta > 0$ , se tiene que existe  $n_0$  tal que para todo  $n \geq n_0$ ,  $\|x_n - x\| < \delta$ . Por lo tanto para todo  $n$  suficientemente grande,  $x_n \in B(x, \delta)$  y entonces  $B(x, \delta) \cap A \neq \emptyset$ . Por lo tanto la relación (2.6) se satisface. Esto demuestra la afirmación (2.5).

Notemos también que, cambiando  $A$  por  $\mathbb{R}^N \setminus A$  en (2.5), obtenemos la relación dual,

$$\text{Int}(A) = \mathbb{R}^N \setminus \text{Adh}(\mathbb{R}^N \setminus A). \quad (2.7)$$

Un conjunto  $A$  en  $\mathbb{R}^N$  se dice *abierto* si  $\text{Int}(A) = A$ .  $A$  se denomina *cerrado* si  $\text{Adh}(A) = A$ . Las relaciones (2.5) y (2.7) implican entonces que  $A$  es abierto si y solo si su complemento  $\mathbb{R}^N \setminus A$  es cerrado. Se define también la *frontera* de  $A$ ,  $\text{Fr}(A)$ , como el conjunto

$$\text{Fr}(A) = \text{Adh}(A) \setminus \text{Int}(A).$$

Así un conjunto es cerrado si y solo si contiene a su frontera, y es abierto si y solo si no intersecta su frontera.

Por otra parte, notemos que, de la definición de adherencia se sigue que un conjunto  $A$  es cerrado si y solo si contiene a los límites de sucesiones convergentes de elementos de  $A$ .

A manera de ejemplo, consideremos el conjunto

$$A = \{(x, y) \in \mathbb{R}^2 / x, y > 0, \quad y \leq e^x\}.$$

Se tiene que:

$$\begin{aligned} \text{Int}(A) &= \{(x, y) \in \mathbb{R}^2 / x, y > 0, \quad y < e^x\}, \\ \text{Adh}(A) &= \{(x, y) \in \mathbb{R}^2 / x, y \geq 0, \quad y \leq e^x\}. \end{aligned}$$

**Ejemplo 2.2.** Se tiene que

$$\text{Adh}(B(x_0, R)) = \{x \in \mathbb{R}^N / \|x - x_0\| \leq R\} := \bar{B}(x_0, R).$$

En efecto, si  $x \in \text{Adh}(B(x_0, R))$ , existe una sucesión  $x_n$  con  $\|x_n - x_0\| < R$  y  $\|x_n - x\| \rightarrow 0$ . Entonces

$$\sum_{i=1}^k |x_i - x_{0i}|^2 < R^2$$

y  $x_{ni} \rightarrow x_i$  para todo  $i$ , de donde  $|x_{ni} - x_{0i}| \rightarrow |x_i - x_{0i}|$ , y por lo tanto

$$\|x_n - x\|^2 = \lim_{n \rightarrow \infty} \sum_{i=1}^N |x_{ni} - x_{0i}|^2 \leq R^2.$$

Recíprocamente, si  $\|x - x_0\| \leq R$  consideremos la sucesión

$$x_n = x - \frac{1}{n}(x - x_0).$$

Entonces,

$$\|x_n - x\| = \frac{1}{n}\|x - x_0\| \rightarrow 0$$



y además

$$\|x_n - x_0\| = (1 - \frac{1}{n})\|x - x_0\| < R$$

para todo  $n \in \mathbb{N}$ . Por lo tanto,  $x_n \in B(x_0, R)$  y  $x_n \rightarrow x_0$ , esto es,  $x \in \text{Adh}(B(x_0, R))$ . Esto concluye la demostración.  $\square$

Se propone como ejercicio demostrar que, por otra parte, la bola abierta  $B(x_0, R)$  es, en efecto, un conjunto abierto.

Un subconjunto de la adherencia de un conjunto  $A$  importante para nuestros propósitos posteriores es aquél de sus *puntos de acumulación*, que es el de los puntos  $x$  de  $\text{Adh}(A)$  que no están aislados del conjunto  $A \setminus \{x\}$ .

Así, decimos que  $x$  es un punto de acumulación de  $A$  si existe una sucesión  $x_n \rightarrow x$  con  $x_n \in A$  y  $x_n \neq x$  para todo  $n \in \mathbb{N}$ . El conjunto de los puntos de acumulación de  $A$  se denota comúnmente  $\text{Der}(A)$ .

Demos un ejemplo en  $\mathbb{R}$ . Si

$$A = [0, 1[ \cup 2 \cup [3, 1[$$

entonces  $x = 2$  no es punto de acumulación de  $A$  pero sí de adherencia. Tenemos

$$\text{Der}(A) = [0, 1] \cup [3, 1], \quad \text{Adh}(A) = [0, 1] \cup \{2\} \cup [3, 1].$$

#### 2.4. Subsucesiones, Teorema de Bolzano-Weierstrass.

Una propiedad muy importante de la convergencia en  $\mathbb{R}^N$  es el teorema de Bolzano-Weierstrass, que dice que *toda sucesión acotada posee una subsucesión convergente*. Demostraremos este hecho, nuevamente haciendo uso del lema anterior y suponiendo ya conocido este hecho para sucesiones reales.

Recordemos que una subsucesión de una sucesión  $x_n$  es una sucesión de la forma  $x_{k(n)}$  donde  $k : \mathbb{N} \rightarrow \mathbb{N}$  es una función estrictamente creciente. Por ejemplo: Si  $x_n = (\sin \frac{1}{n}, e^n)$  entonces son subsucesiones de  $x_n$  las siguientes:

$$y_n = (\sin 2^{-n}, e^{2^n}), \quad z_n = (\sin n^{-2}, e^{n^2}).$$

En efecto,  $y_n = x_{2^n}$ ,  $z_n = x_{n^2}$ , y las funciones  $k(n) = 2^n$ ,  $k(n) = n^2$  son estrictamente crecientes.

Una propiedad inmediata es la siguiente: Si  $x_n \rightarrow x$ , entonces para toda subsucesión  $x_{k(n)}$  de  $x_n$  se tiene que

$$x_{k(n)} \rightarrow x \quad \text{cuando } n \rightarrow \infty.$$

Una sucesión  $x_n$  en  $\mathbb{R}^N$  se dice *acotada* si existe una constante  $M > 0$  tal que

$$\|x_n\| \leq M .$$

Esto quiere decir que todos los elementos de la sucesión están contenidos en una bola de radio suficientemente grande. En efecto,  $x_n \in \bar{B}(0, M)$  para todo  $n \in \mathbb{N}$ . Tenemos la validez del siguiente importante resultado.

**Teorema 2.1.** (*Bolzano-Weierstrass*). *Sea  $x_n$  una sucesión acotada en  $\mathbb{R}^N$ . Entonces existe una subsucesión  $x_{k(n)}$ , y un punto  $x \in \mathbb{R}^N$  tales que  $x_{k(n)} \rightarrow x$  cuando  $n \rightarrow \infty$ .*

**Demostración.** Supongamos que la sucesión

$$x_n = (x_{n1}, \dots, x_{nN})$$

es acotada. Entonces para cierto número  $M > 0$  se tiene que, para cada  $i = 1, \dots, N$ ,

$$|x_{ni}| \leq \sqrt{\sum_{k=1}^N |x_{nk}|^2} = \|x_n\| \leq M . \quad (2.8)$$

Así, la sucesión de números reales  $x_{n1}$  es acotada. Se sigue, por el Teorema de Bolzano-Weierstrass para sucesiones reales, que esta sucesión posee una subsucesión convergente, digamos  $x_{k_1(n)1} \rightarrow x_1 \in \mathbb{R}$ . En modo similar se tiene, a partir de (2.8), que la sucesión  $x_{k_1(n)2}$  posee una subsucesión convergente, digamos  $x_{k_1(k_2(n))2} \rightarrow x_2$ . Notemos que la sucesión  $x_{k_1(k_2(n))1}$  es una subsucesión de  $x_{k_1(n)1} \rightarrow x_1$  y por lo tanto  $x_{k_1(k_2(n))1} \rightarrow x_1$ . Del mismo modo, (si  $N \geq 3$ ),  $x_{k_1(k_2(n))3}$  es una sucesión real acotada, gracias a (2.8), y se sigue que posee una subsucesión convergente, digamos,  $x_{k_1(k_2(k_3(n)))3} \rightarrow x_3$ , y se tiene también que  $x_{k_1(k_2(k_3(n)))l} \rightarrow x_l$  para  $l = 1, 2$ . Iterando, este procedimiento  $N$  veces, construimos una subsucesión de  $x_n$  de la forma  $x_{k_1(k_2(\dots(k_N(n))\dots))}$  tal que para ciertos números reales  $x_1, \dots, x_N$  se tiene que

$$x_{k_1(k_2(\dots(k_N(n))\dots))l} \rightarrow x_l, \text{ cuando } n \rightarrow \infty, \quad \forall l = 1, \dots, N .$$

Gracias al lema anterior se sigue que

$$x_{k(n)} \rightarrow x = (x_1, \dots, x_N), \text{ cuando } n \rightarrow \infty ,$$

donde  $k(n) = k_1 \circ k_2 \circ \dots \circ k_N(n)$ , es una composición sucesiva de funciones estrictamente crecientes  $\mathbb{N} \mapsto \mathbb{N}$ , siendo por tanto  $k(n)$  también estrictamente creciente, constituyendo entonces  $x_{k(n)}$  una subsucesión de la sucesión original  $x_n$ . Esto concluye la demostración.  $\square$

**2.5. Sucesiones de Cauchy, completitud de  $\mathbb{R}^N$ .** Una sucesión  $(x_n)_{n \in \mathbb{N}}$  en  $\mathbb{R}^N$  se dice *de Cauchy* si la siguiente propiedad se cumple:

$$(\forall \varepsilon > 0) (\exists n_0 \in \mathbb{N}) (\forall n, m \geq n_0) : \|x_n - x_m\| < \varepsilon ,$$

Intuitivamente, una sucesión de Cauchy es una sucesión cuyos elementos tienden a acumularse en una región cuyo tamaño puede tomarse tan pequeño como se quiera. Como este es el caso de una sucesión convergente, no es sorprendente la validez de la siguiente propiedad:

Si  $x_n$  es una sucesión convergente, entonces es de Cauchy. Probemos esto. Supongamos que  $x_n \rightarrow x$  cuando  $n \rightarrow \infty$ . Sabemos entonces que dado  $\varepsilon > 0$  puede encontrarse un índice  $n_0(\varepsilon)$  tal que si  $n, m \geq n_0$  entonces

$$\|x_n - x\| < \frac{\varepsilon}{2} , \quad \|x_m - x\| < \frac{\varepsilon}{2} .$$

Por otra parte, por la desigualdad triangular para la norma, tenemos también que

$$\|x_n - x_m\| \leq \|x_n - x\| + \|x_m - x\| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2}$$

Por lo tanto, se tiene que dado cualquier  $\varepsilon > 0$ , existe un índice  $n_0(\varepsilon)$  tal que para todo  $n, m \geq n_0(\varepsilon)$ ,  $\|x_n - x_m\| < \varepsilon$ , esto es,  $x_n$  es de Cauchy.

Un hecho mucho menos obvio, de hecho una importante y profunda propiedad de  $\mathbb{R}^N$ , es el hecho que, recíprocamente, las sucesiones de Cauchy son convergentes. Esta propiedad se denomina *completitud*. Que toda sucesión “que se acumula” tiene en realidad un límite quiere decir que  $\mathbb{R}^N$  no tiene “hoyos”: no hay nada “pegado a  $\mathbb{R}^N$ ” que no esté en realidad en  $\mathbb{R}^N$ . Suponiendo la validez de esta propiedad en  $\mathbb{R}$ , discutida en el Cálculo de una variable, probaremos entonces el siguiente resultado.

**Teorema 2.2.** (*Completitud de  $\mathbb{R}^N$* ). Si  $x_n$ ,  $n \in \mathbb{N}$ , es una sucesión de Cauchy, entonces es convergente, esto es existe  $x \in \mathbb{R}^N$  con  $x_n \rightarrow x$ .

**Demostración.** Supongamos  $x_n = (x_{n1}, \dots, x_{nN})$ . Sabemos que dado  $\varepsilon > 0$ , existe  $n_0$  tal que si  $n, m \geq n_0$  entonces  $\|x_n - x_m\| < \varepsilon$ . Por otra parte, para tales índices  $n, m$  y cada componente  $l = 1, \dots, N$  se tiene que

$$|x_{nl} - x_{ml}| \leq \left( \sum_{i=1}^N |x_{ni} - x_{mi}|^2 \right)^{\frac{1}{2}} = \|x_n - x_m\| < \varepsilon .$$

Por lo tanto, para todo  $l = 1, \dots, N$ , la sucesión en  $\mathbb{R}$   $x_{nl}$ ,  $n \in \mathbb{N}$  es de Cauchy. En consecuencia  $x_{nl}$  es convergente, esto es, existe un número

$x_l \in \mathbb{R}$  tal que  $x_{nl} \rightarrow x_l$ . En virtud de la proposición 2.2, se sigue entonces que la sucesión  $x_n$  es convergente en  $\mathbb{R}^N$  y

$$\lim_{n \rightarrow \infty} x_n = x := (x_1, x_2, \dots, x_N) .$$

Esto concluye la demostración.  $\square$

### 3. LÍMITES Y CONTINUIDAD

Sea  $\Omega$  un subconjunto de  $\mathbb{R}^N$ . Queremos definir la noción de continuidad de una función

$$f : \Omega \subset \mathbb{R}^N \rightarrow \mathbb{R}^m$$

en un punto  $x_0 \in \Omega$ . Para funciones de una variable definidas en un intervalo  $I = [a, b]$ , esta definición se escribía para un punto interior de  $I$ ,  $x_0 \in ]a, b[$ , diciendo que para todo  $x$  cercano a  $x_0$ , el valor de  $f(x)$  está cercano a  $f(x_0)$ . Similarmente, se definieron continuidad por la izquierda en  $b$  y continuidad por la derecha en  $a$ , como  $f(x)$  está cercano a  $f(b)$  si  $x$  lo está de  $b$ , con  $x < b$ , similarmente en para  $a$ . Los términos izquierda y derecha en el espacio  $\mathbb{R}^N$  carecen en principio de un sentido claro, pero puede verse que una definición de continuidad en general de  $f$  en  $x_0 \in [a, b]$  *relativa a*  $[a, b]$  podría haberse enunciado como “ $f(x)$  está cercano a  $f(x_0)$  si  $x$  está cercano a  $x_0$  con  $x \in [a, b]$ ”. Esta es la noción general que utilizaremos,  $f : \Omega \subset \mathbb{R}^N \rightarrow \mathbb{R}^m$  como *continuidad relativa a*  $\Omega$ .

**Definición.** Consideremos un conjunto  $\Omega \subset \mathbb{R}^N$ , una función  $f : \Omega \rightarrow \mathbb{R}^m$  y un punto  $x_0 \in \Omega$ . Decimos que  $f$  es *continua en*  $x_0$ , *relativamente a*  $\Omega$  si para *toda* sucesión  $x_n \rightarrow x_0$  con  $x_n \in \Omega \forall n \in \mathbb{N}$  se tiene que

$$\lim_{n \rightarrow \infty} f(x_n) = f(x_0) .$$

Cuando  $f$  esté definida a partir de una fórmula cuyo dominio maximal de definición es una región  $\Omega$ , diremos simplemente que  $f$  es continua en  $x_0$ , omitiendo la partícula *relativamente a*  $\Omega$ .

Si  $f$  es continua en todo  $x_0 \in \Omega$  diremos simplemente que  $f$  es *continua en*  $\Omega$ .

Esta caracterización ayuda a verificar en modo simple la continuidad de una gran cantidad de funciones.

**Ejemplo 3.1.** Consideremos  $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  dada por

$$f(x, y) = (x^2 + 2e^{xy}, 5 \cos(xy^2)).$$

Afirmamos que esta función es continua en  $(1, 0)$ . En efecto, consideremos una sucesión cualquiera  $(x_n, y_n) \rightarrow (1, 0)$ . Debemos probar que  $f(x_n, y_n) \rightarrow f(1, 0) = (3, 5)$ . Como  $x_n \rightarrow 1$ ,  $y_n \rightarrow 0$ , se sigue, por las propiedades de sucesiones en  $\mathbb{R}$ , que

$$x_n^2 \rightarrow 1, \quad x_n y_n \rightarrow 0, \quad x_n y_n^2 \rightarrow 0.$$

Como las funciones  $t \mapsto e^t$  y  $t \mapsto \cos t$  son continuas en  $\mathbb{R}$  se sigue que

$$e^{x_n y_n} \rightarrow e^0 = 1, \quad \cos(x_n y_n^2) \rightarrow \cos 0 = 1.$$

Nuevamente por el álgebra de límites de sucesiones se sigue que

$$x_n^2 + 2e^{x_n y_n} \rightarrow 1 + 2 \cdot 1 = 3, \quad 5 \cos(x_n y_n^2) \rightarrow 5 \cos 0 = 5,$$

y entonces de acuerdo a la proposición 2.2,

$$f(x_n, y_n) = (x_n^2 + 2e^{x_n y_n}, 5 \cos(x_n y_n^2)) \rightarrow (3, 5) = f(1, 0),$$

y como la sucesión  $(x_n, y_n) \rightarrow (1, 0)$  es arbitraria, la continuidad de  $f$  en  $(1, 0)$  ha sido demostrada.

**Ejemplo 3.2.** Consideremos la función  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  definida por

$$f(x, y) = \frac{xy}{\sqrt{x^2 + y^2}}, \text{ si } (x, y) \neq (0, 0), \quad f(0, 0) = 0.$$

Afirmamos que  $f$  es continua en  $(0, 0)$ . En efecto, consideremos una sucesión  $(x_n, y_n) \rightarrow (0, 0)$  cualquiera. Tenemos que,

$$|f(x_n, y_n)| = \frac{|x_n y_n|}{\sqrt{x_n^2 + y_n^2}}, \text{ si } (x_n, y_n) \neq (0, 0).$$

Por otra parte, tenemos la validez de la desigualdad

$$2|a||b| \leq (|a| + |b|)^2,$$

para cualquier par de números  $a, b$ . Entonces,

$$|x_n y_n| \leq \frac{1}{2}(|x_n| + |y_n|)^2,$$

por ende,

$$|f(x_n, y_n)| \leq \frac{1}{2} \sqrt{x_n^2 + y_n^2},$$

desigualdad, que es obviamente también válida si  $(x_n, y_n) = (0, 0)$ . Deducimos entonces que

$$f(x_n, y_n) \rightarrow 0 = f(0, 0).$$

Como la sucesión  $(x_n, y_n) \rightarrow (0, 0)$  es arbitraria, concluimos que  $f$  es continua en  $(0, 0)$ .  $\square$

**Ejemplo 3.3.** Consideremos la función  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  definida por

$$f(x, y) = \frac{xy}{x^2 + y^2}, \text{ si } (x, y) \neq (0, 0), \quad f(0, 0) = 0.$$

Afirmamos que  $f$  no es continua en  $(0, 0)$ . En efecto, consideremos la sucesión  $(x_n, y_n) = (\frac{1}{n}, \frac{1}{n})$ . Entonces,  $\forall n \in \mathbb{N}$ ,

$$f(x_n, y_n) = \frac{1}{2},$$

Por ende  $f(x_n, y_n) \not\rightarrow f(0, 0) = 0$ , y ya existiendo una sola sucesión con esta propiedad, se tiene que  $f$  no es continua en  $(0, 0)$ .

Observemos que para cualquier valor  $L$  que demos a  $f$  en  $(0, 0)$ , la función resultante,

$$f(x, y) = \frac{xy}{x^2 + y^2}, \text{ si } (x, y) \neq (0, 0), \quad f(0, 0) = L,$$

resulta ser discontinua en  $(0, 0)$ . En efecto, por ejemplo para la sucesión  $(x_n, y_n) = (\frac{1}{n}, 0)$  resulta que  $\forall n \in \mathbb{N}$ ,

$$f(x_n, y_n) = 0,$$

por ende para dos sucesiones tendiendo a  $(0, 0)$ ,  $(\frac{1}{n}, \frac{1}{n})$  y  $(\frac{1}{n}, 0)$ , tenemos que  $f$  a lo largo de éstas aproxima a dos límites distintos. Uno de los límites por cierto será distinto de  $L$ , y por ende la función no es continua en  $(0, 0)$ .  $\square$

Ejemplos como los dos anteriores motivan a definir *límite* de una función. Sea  $\Omega \subset \mathbb{R}^N$ ,  $f : \Omega \rightarrow \mathbb{R}^m$ ,  $x_0 \in \mathbb{R}^N$  un *punto de acumulación* de  $\Omega$ . Decimos que  $L \in \mathbb{R}^m$  es el límite de  $f(x)$  cuando  $x$  tiende a  $x_0$  relativamente a  $\Omega$  si la siguiente propiedad se cumple: para toda sucesión  $x_n \rightarrow x_0$  con  $x_n \in \Omega$  y  $x_n \rightarrow x_0$  se tiene que

$$\lim_{n \rightarrow \infty} f(x_n) = L.$$

Escribimos en tal caso

$$\lim_{x \rightarrow x_0, x \in \Omega} f(x) = L,$$

o, típicamente, si el dominio de  $f(x)$  está subentendido,

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

De este modo, para la función  $f(x, y)$  del ejemplo 3.2, se tiene que

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

mientras que para la función  $f(x, y)$  del ejemplo 3.3 el límite

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

*no existe.*

**Proposición 3.1.** *Sea  $\Omega \subset \mathbb{R}^N$ ,  $f : \Omega \rightarrow \mathbb{R}^m$ ,  $x_0 \in \Omega$  un punto de acumulación de  $\Omega$ . Las siguientes afirmaciones son equivalentes:*

(a)  *$f$  es continua en  $x_0$  relativamente a  $\Omega$ .*

(b)

$$\lim_{x \rightarrow x_0} f(x) = f(x_0).$$

La demostración de este resultado la proponemos como un ejercicio. También dejamos al lector la demostración del siguiente resultado.

**Proposición 3.2.** (*Álgebra de límites*). *Sean  $\Omega \subset \mathbb{R}^N$ ,  $f, g : \Omega \rightarrow \mathbb{R}^m$ ,  $x_0 \in \mathbb{R}^N$  un punto de acumulación de  $\Omega$ ,  $\alpha, \beta \in \mathbb{R}$ . Supongamos que  $f$  y  $g$  son tales que los límites*

$$\lim_{x \rightarrow x_0} f(x), \quad \lim_{x \rightarrow x_0} g(x)$$

*existen. Entonces, los siguientes límites existen y pueden calcularse como se expresa:*

(a)

$$\lim_{x \rightarrow x_0} (\alpha f + \beta g)(x) = \alpha \lim_{x \rightarrow x_0} f(x) + \beta \lim_{x \rightarrow x_0} g(x).$$

(b) *Si  $m = 1$ ,*

$$\lim_{x \rightarrow x_0} f(x)g(x) = \left( \lim_{x \rightarrow x_0} f(x) \right) \left( \lim_{x \rightarrow x_0} g(x) \right).$$

Una propiedad útil para el análisis de continuidad de una función a valores en  $\mathbb{R}^m$ , es que su continuidad equivale a aquella de sus  $m$  funciones coordenadas.

**Proposición 3.3.** *Sean  $\Omega \subset \mathbb{R}^N$ ,  $f : \Omega \rightarrow \mathbb{R}^m$ ,  $x_0 \in \Omega$ ,*

$$f(x) = (f_1(x), f_2(x), \dots, f_m(x)).$$

*Entonces las siguientes afirmaciones son equivalentes.*

(a)  *$f$  es continua en  $x_0$  relativamente a  $\Omega$ .*

(b) *Para todo  $i = 1, \dots, m$ , las funciones  $f_i$  son continuas en  $x_0$  relativamente a  $\Omega$ .*

**Demostración.** (a)  $\implies$  (b). Sea  $x_n \rightarrow x_0$  con  $x_n \in \Omega$ . Entonces  $f(x_n) \rightarrow f(x_0)$ . Gracias a la proposición 2.2 se sigue que  $f_i(x_n) \rightarrow f_i(x_0)$  para todo  $i = 1, \dots, N$ . Como la sucesión  $x_n$  es arbitraria, se sigue que  $f_i$  es continua en  $x_0$  relativamente a  $\Omega$ , esto es, (b) se cumple.

(b)  $\implies$  (a). Sea  $x_n \rightarrow x_0$  con  $x_n \in \Omega$ . Entonces  $f_i(x_n) \rightarrow f_i(x_0)$  para todo  $i = 1, \dots, N$ . Nuevamente, por la proposición 2.2 tenemos que  $f(x_n) \rightarrow f(x_0)$ . Como la sucesión  $x_n$  es arbitraria, se sigue que  $f$  es continua en  $x_0$  relativamente a  $\Omega$ , y la demostración queda concluida.  $\square$

Las propiedades habituales del álgebra de funciones continuas se cumplen, en modo similar a funciones de una variable. Resumimos éstas en el siguiente resultado.

**Proposición 3.4.** Sean  $\Omega \subset \mathbb{R}^N$ ,  $f, g : \Omega \rightarrow \mathbb{R}^m$ ,  $x_0 \in \mathbb{R}^N$ ,  $\alpha, \beta \in \mathbb{R}$ . Supongamos que  $f$  y  $g$  son continuas en  $x_0$  relativamente a  $\Omega$ . Entonces

(a) La función  $\alpha f + \beta g : \Omega \rightarrow \mathbb{R}^m$  es continua en  $x_0$  relativamente a  $\Omega$ .

(b) Si  $m = 1$ , la función  $f \cdot g : \Omega \rightarrow \mathbb{R}$  es continua en  $x_0$  relativamente a  $\Omega$ .

**Demostración.** Si  $x_n \rightarrow x_0$  en  $\mathbb{R}^N$ , con  $x_n \in \Omega$  para todo  $n \in \mathbb{N}$ , entonces, por hipótesis, se tiene que  $f(x_n) \rightarrow f(x_0)$  y  $g(x_n) \rightarrow g(x_0)$ . Así, en virtud de la proposición 2.3 se sigue que

$$\alpha f(x_n) + \beta g(x_n) \rightarrow \alpha f(x_0) + \beta g(x_0) .$$

Como la sucesión  $x_n \rightarrow x_0$  es arbitraria, esto nos dice que  $\alpha f + \beta g$  es continua en  $x_0$  relativamente a  $\Omega$ . La demostración de (b) la proponemos como ejercicio.  $\square$

**Ejemplo 3.4.** Las funciones

$$\pi_i : \mathbb{R}^N \rightarrow \mathbb{R}, \quad \pi_i(x_1, \dots, x_N) = x_i$$

son continuas sobre todo  $\mathbb{R}^N$ , pues si  $x_n \rightarrow x_0$  en  $\mathbb{R}^N$ , se sigue de la proposición 2.2 que  $\pi_i(x_n) \rightarrow \pi_i(x_0)$ . Un polinomio en  $\mathbb{R}^N$  es una función que puede expresarse en la forma

$$P(x) = \sum_{i_1=1}^N \sum_{i_2=1}^N \cdots \sum_{i_s=1}^N a_{i_1 i_2, \dots, i_s} x_{i_1}^{m_{i_1}} x_{i_2}^{m_{i_2}} \cdots x_{i_s}^{m_{i_s}} .$$

Deducimos entonces de la proposición anterior que los polinomios son funciones continuas en  $\mathbb{R}^N$  pues pueden ser escritos como productos



sucesivos y combinaciones lineales de las funciones continuas  $\pi_i$ ,  $i = 1, \dots, N$ .  $\square$

Con frecuencia, la definición de continuidad se realiza en el lenguaje  $\varepsilon$ - $\delta$ . Ésta es en realidad equivalente a aquella con sucesiones que hemos utilizado.

**Proposición 3.5.** (*Caracterización  $\varepsilon$ - $\delta$  de la continuidad*). Sean  $\Omega \subset \mathbb{R}^N$ ,  $f : \Omega \rightarrow \mathbb{R}^m$ ,  $x_0 \in \mathbb{R}^N$ . Entonces  $f$  es continua en  $x_0$  relativamente a  $\Omega$  si y solo si la siguiente propiedad se cumple:

$$(\forall \varepsilon > 0) (\exists \delta > 0) (\forall x \in \Omega) : \|x - x_0\| < \delta \implies \|f(x) - f(x_0)\| < \varepsilon . \quad (3.9)$$

**Demostración.** Supongamos que  $f$  es continua en  $x_0$  relativamente a  $\Omega$ . Supongamos por contradicción que (3.9) no se cumple. Entonces,

$$(\exists \varepsilon_0 > 0) (\forall \delta > 0) (\exists x_\delta \in \Omega) : \|x_\delta - x_0\| < \delta \text{ y } \|f(x_\delta) - f(x_0)\| \geq \varepsilon_0 . \quad (3.10)$$

Escojamos en (3.10)  $\delta = \frac{1}{n}$ . Existe entonces  $\tilde{x}_n := x_{\frac{1}{n}} \in \Omega$  tal que

$$\|\tilde{x}_n - x_0\| < \frac{1}{n} \text{ y } \|f(\tilde{x}_n) - f(x_0)\| \geq \varepsilon_0 .$$

Así, la sucesión  $\tilde{x}_n$  satisface que  $\tilde{x}_n \rightarrow x_0$  pero  $\|f(\tilde{x}_n) - f(x_0)\| \geq \varepsilon_0$ , de modo que  $f(\tilde{x}_n) \not\rightarrow f(x_0)$ . Por lo tanto  $f$  no es continua en  $x_0$  relativamente a  $\Omega$ , una contradicción que prueba entonces que la condición (3.9) se cumple.

Supongamos ahora que (3.9) se cumple. Queremos probar que  $f$  es continua en  $x_0$  relativamente a  $\Omega$ . Consideremos entonces una sucesión  $x_n \in \Omega$  con  $x_n \rightarrow x_0$ . Debemos probar que  $f(x_n) \rightarrow f(x_0)$ . Sea  $\varepsilon > 0$  y escojamos  $\delta = \delta(\varepsilon)$  de modo que

$$x \in \Omega \text{ y } \|x - x_0\| < \delta(\varepsilon) \implies \|f(x) - f(x_0)\| < \varepsilon . \quad (3.11)$$

Como  $x_n \rightarrow x_0$ , se tiene que, gracias a la caracterización de la convergencia (2.3), existe  $n_0$  tal que  $\forall n \geq n_0$  se tiene que  $\|x_n - x_0\| < \delta(\varepsilon)$ . Así, en virtud de (3.11), se concluye que  $\|f(x_n) - f(x_0)\| < \varepsilon$ . Hemos probado que

$$(\forall \varepsilon > 0) (\exists n_0 \in \mathbb{N}) (\forall n \geq n_0) : \|f(x_n) - f(x_0)\| < \varepsilon ,$$

lo que significa precisamente que  $f(x_n) \rightarrow f(x_0)$ , gracias a (2.3). Como la sucesión  $x_n$  escogida es arbitraria, tenemos entonces que  $f$  es continua en  $x_0$  relativamente a  $\Omega$ . Esto concluye la demostración.  $\square$

Tal como en funciones de una variable, se tiene que la composición de funciones continuas es continua, como enuncia el siguiente resultado.

**Proposición 3.6.** (*Regla de la composición*). Sean  $\Omega \subset \mathbb{R}^N$ ,  $\Lambda \subset \mathbb{R}^m$ ,  $f : \Omega \rightarrow \mathbb{R}^m$ ,  $g : \Lambda \rightarrow \mathbb{R}^k$ . Supongamos que  $f$  es continua en  $x_0$  relativamente a  $\Omega$ , que  $f(x) \in \Lambda \ \forall x \in \Omega$ , y que  $g$  es continua en  $f(x_0)$  relativamente a  $\Lambda$ . Entonces la composición

$$g \circ f : \Omega \subset \mathbb{R}^N \rightarrow \mathbb{R}^k$$

es continua en  $x_0$  relativamente a  $\Omega$ .

**Demostración.** Supongamos que  $x_n \rightarrow x_0$ , con  $x_n \in \Omega \ \forall n \in \mathbb{N}$ . La continuidad de  $f$  en  $x_0$  implica entonces que

$$y_n := f(x_n) \rightarrow y_0 = f(x_0).$$

Pero  $y_n \in \Lambda \ \forall n \in \mathbb{N}$ , por ende la continuidad de  $g$  en  $y_0$  relativa a  $\Lambda$  implica que  $g(y_n) \rightarrow g(y_0)$ . En otras palabras, hemos demostrado que para una sucesión  $x_n$  arbitraria en  $\Omega$  con  $x_n \rightarrow x_0$ , se tiene que

$$(g \circ f)(x_n) = g(f(x_n)) \rightarrow g(f(x_0)) = (g \circ f)(x_0),$$

esto es, que  $g \circ f$  es continua en  $x_0$  relativamente a  $\Omega$ .  $\square$

**Ejemplo 3.5.** Las reglas operacionales de la continuidad ya establecidas, nos permiten verificar que las funciones construidas a través de fórmulas algebraicas basadas en las funciones habituales del Cálculo: polinomios, trigonométricas, exponencial, etc., son típicamente continuas relativamente a sus dominios de definición. Consideremos por ejemplo la función  $f : \mathbb{R}^2 \rightarrow \mathbb{R}^3$  dada por

$$f(x, y) = (x^2 \cos y, 2ye^{xy} + xy^2e^{xy}, 2xy) .$$

Afirmamos que esta función es continua en todo punto  $(x, y) \in \mathbb{R}^2$ . En efecto, sabemos que la función  $(x, y) \mapsto y$  es continua, y que  $t \mapsto \cos t$  también lo es, por ende la composición  $(x, y) \mapsto \cos y$  también lo es. Como también  $(x, y) \mapsto x^2y$  lo es, al ser un polinomio, deducimos que la función  $(x, y) \mapsto x^2y \cos y$  es continua, por ser producto de funciones continuas. Las otras dos funciones coordenadas de  $f$  son también continuas, por argumentos similares, y concluimos entonces, de la Proposición 3.3 que  $f$  es continua en  $\mathbb{R}^2$ .  $\square$

**3.1. Máximo y mínimo de una función continua.** Como se discutió ampliamente en el curso de Cálculo de funciones de una variable, una característica hasta cierto punto sorprendente de la continuidad, es que siendo una propiedad *local*, esto es, referida a una pequeña región en torno al punto de interés, resulta que su validez *en todo punto* conlleva la validez de propiedades *globales* de la función, algunas de estas

de extraordinaria importancia, como lo es el teorema a continuación en cuanto a existencia de máximos y mínimos de una función continua en todo punto de una región cerrada y acotada.

**Teorema 3.1.** *Sea  $\mathcal{K} \subset \mathbb{R}^N$  un conjunto cerrado y acotado y  $f : \mathcal{K} \rightarrow \mathbb{R}$  una función continua relativamente a  $\mathcal{K}$ , esto es, continua en todo punto  $x_0 \in \mathcal{K}$ , relativamente a  $\mathcal{K}$ . Existen entonces puntos  $x_*$ ,  $x^* \in \mathcal{K}$  con la siguiente propiedad:*

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

*En otras palabras,  $f$  alcanza sus valores máximo y mínimo en  $\mathcal{K}$ :*

$$f(x_*) = \min_{x \in \mathcal{K}} f(x) \quad f(x^*) = \max_{x \in \mathcal{K}} f(x).$$

**Demostración.** Demostraremos la existencia de un punto  $x^*$  donde  $f$  alcanza su máximo en  $\mathcal{K}$ . Recordemos que dado cualquier conjunto  $A \subset \mathbb{R}$ , no-vacío y acotado superiormente, existe el supremo de  $A$ ,  $\sup A \in \mathbb{R}$ , la menor de sus cotas superiores, y además puede encontrarse una sucesión  $a_n$ , con

$$a_n \in A \quad \forall n \in \mathbb{N} \quad \text{y} \quad a_n \rightarrow \sup A.$$

Si  $A$  no es acotado superiormente, existe una sucesión

$$a_n \in A \quad \forall n \in \mathbb{N} \quad \text{y} \quad a_n \rightarrow +\infty.$$

En este último caso, escribimos, de todos modos,  $\sup A = +\infty$ .

Apliquemos esto al conjunto de números reales

$$A = \{ f(x) / x \in \mathcal{K} \}.$$

De acuerdo a lo anterior, existe una sucesión

$$a_n := f(x_n) \in A \quad \text{con} \quad f(x_n) \rightarrow \sup A.$$

Así,  $x_n \in \mathcal{K} \quad \forall n \in \mathbb{N}$ . De acuerdo al Teorema de Bolzano-Weierstrass,  $x_n$  posee una subsucesión convergente, con límite en  $\mathcal{K}$ , digamos

$$x_{k(n)} \rightarrow x^* \in \mathcal{K} \quad \text{cuando} \quad n \rightarrow \infty.$$

Como  $f$  es continua en  $x^*$  relativamente a  $\mathcal{K}$ , se sigue entonces que  $f(x_n) \rightarrow f(x^*)$ . Pero se tiene también que  $f(x_n) \rightarrow \sup A$ . De este modo, necesariamente  $\sup A < +\infty$  y  $f(x^*) = \sup A$ . Como  $\sup A$  es cota superior de  $A$ , se tiene entonces que

$$f(x) \leq f(x^*) \quad \forall x \in \mathcal{K},$$

y por ende  $f$  se maximiza en  $x^*$  sobre  $\mathcal{K}$ . La existencia de un punto de mínimo  $x_*$  en  $\mathcal{K}$  se prueba en modo similar, y proponemos la demostración como un ejercicio.  $\square$

El teorema anterior puede aplicarse para probar la existencia de máximos o mínimos de funciones continuas sobre conjuntos que no son necesariamente cerrados y acotados. Para una función  $f : \mathbb{R}^N \rightarrow \mathbb{R}$  decimos que

$$\lim_{\|x\| \rightarrow \infty} f(x) = +\infty$$

si para cualquier sucesión  $x_n \in \mathbb{R}^N$  con  $\|x_n\| \rightarrow +\infty$  se tiene que  $f(x_n) \rightarrow +\infty$ .

El siguiente resultado es una aplicación clásica del Teorema 3.1.

**Teorema 3.2.** *Sea  $f : \mathbb{R}^N \rightarrow \mathbb{R}$  una función continua en  $\mathbb{R}^N$  con la propiedad que*

$$\lim_{\|x\| \rightarrow \infty} f(x) = +\infty.$$

*Entonces existe un punto  $x_* \in \mathbb{R}^N$  tal que*

$$f(x_*) \leq f(x) \quad \forall x \in \mathbb{R}^N,$$

*esto es,  $f$  alcanza su mínimo en  $\mathbb{R}^N$ .*

**Demostración.** Consideremos el conjunto

$$\mathcal{K} = \{x \in \mathbb{R}^N / f(x) \leq f(0)\}.$$

Afirmamos que  $\mathcal{K}$  es cerrado y acotado en  $\mathbb{R}^N$ .

Probemos primero que  $\mathcal{K}$  es **acotado**. Por contradicción, si  $\mathcal{K}$  no fuese acotado, existiría una sucesión  $x_n \in \mathcal{K}$  tal que  $\|x_n\| \rightarrow +\infty$ . Debe tenerse entonces, por hipótesis, que  $f(x_n) \rightarrow +\infty$ . Pero esto es imposible, pues  $f(x_n) \leq f(0)$  para todo  $n \in \mathbb{N}$ . Por lo tanto,  $\mathcal{K}$  es acotado.

Para ver que  $\mathcal{K}$  es **cerrado**, consideremos un punto cualquiera  $x_0 \in \text{Adh}(\mathcal{K})$ . Por definición de adherencia, existe una sucesión  $x_n \in \mathcal{K}$  con  $x_n \rightarrow x_0$ . Como  $f$  es continua en  $x_0$ , tenemos entonces que  $f(x_n) \rightarrow f(x_0)$ . Pero como  $x_n \in \mathcal{K}$  se tiene que  $f(x_n) \leq f(0) \forall n \in \mathbb{N}$ , y por lo tanto también  $f(x_0) \leq f(0)$ . Esto significa precisamente que  $x_0 \in \mathcal{K}$ . Hemos probado que todo punto de  $\text{Adh}(\mathcal{K})$  está en realidad en  $\mathcal{K}$ , lo que quiere decir  $\text{Adh}(\mathcal{K}) \subset \mathcal{K}$ . Como siempre se tiene la inclusión opuesta, concluimos que  $\text{Adh}(\mathcal{K}) = \mathcal{K}$ , esto es que  $\mathcal{K}$  es cerrado.

Así,  $\mathcal{K}$  es cerrado y acotado, y por lo tanto, en virtud del teorema 3.1, existe  $x_* \in \mathcal{K}$  con

$$f(x_*) \leq f(x) \quad \forall x \in \mathcal{K}.$$

En particular,  $f(x_*) \leq f(0)$ . Como, por definición de  $\mathcal{K}$ , tenemos  $f(0) < f(x) \forall x \in \mathbb{R}^N \setminus \mathcal{K}$ , se sigue que, también,

$$f(x_*) \leq f(x) \quad \forall x \in \mathbb{R}^N \setminus \mathcal{K},$$

y por ende  $f(x_*) \leq f(x) \forall x \in \mathbb{R}^N$ . Esto concluye la demostración.  $\square$

**Ejemplo 3.6.** Consideremos la función  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$

$$f(x, y) = x^2 + y^2 - \log(1 + x^2 + y^2) + \frac{x^2}{2} \cos(xy).$$

Afirmamos que  $f$  alcanza su mínimo en  $\mathbb{R}^2$ . En efecto, observemos que

$$\frac{f(x, y)}{x^2 + y^2} = 1 - \frac{\log(1 + x^2 + y^2)}{x^2 + y^2} + \frac{x^2}{x^2 + y^2} \cos(xy).$$

Por la regla de l'hôpital, tenemos que

$$\lim_{s \rightarrow +\infty} \frac{\log s}{s} = 0,$$

de modo que, en particular, existe  $R_0 > 0$  tal que

$$\frac{\log(1 + x^2 + y^2)}{x^2 + y^2} < \frac{1}{4} \quad \text{si } \|(x, y)\|^2 = x^2 + y^2 > R_0.$$

Por otra parte,

$$\frac{1}{2} \frac{x^2}{x^2 + y^2} \cos(xy) \geq -\frac{x^2}{x^2 + y^2} \geq -\frac{1}{2}$$

y entonces, si  $\|(x, y)\|^2 = x^2 + y^2 > R_0$ , tenemos que

$$\frac{f(x, y)}{x^2 + y^2} \geq 1 - \frac{1}{4} - \frac{1}{2} = \frac{1}{4},$$

de modo que, para todo punto  $(x, y)$  con  $\|(x, y)\| > \sqrt{R_0}$ , tenemos

$$f(x, y) \geq \frac{1}{4} \|(x, y)\|^2,$$

de lo que se deduce, en particular,

$$\lim_{\|(x, y)\| \rightarrow +\infty} f(x, y) = +\infty.$$

Así, de acuerdo al teorema 3.2, el resultado deseado se concluye.  $\square$

### 3.2. Funciones Lipschitz, Teorema del punto fijo de Banach.

Un tipo particular de funciones continuas son las llamadas *funciones Lipschitz*. Decimos que una función  $f : \Omega \subset \mathbb{R}^N \rightarrow \mathbb{R}^m$  es *Lipschitz en*  $\Omega$  de constante  $K > 0$ , si

$$(\forall x_1, x_2 \in \Omega) : \|f(x_1) - f(x_2)\| \leq K \|x_1 - x_2\|. \quad (3.12)$$

Tal función es automáticamente continua relativamente a  $\Omega$ . En efecto, Si  $x_n \rightarrow x_0 \in \Omega$  con  $x_n \in \Omega \forall n$ , entonces se tiene que  $\|x_n - x_0\| \rightarrow 0$ . Se sigue que

$$0 \leq \|f(x_n) - f(x_0)\| \leq K \|x_n - x_0\| \rightarrow 0,$$

de modo que  $\|f(x_n) - f(x_0)\| \rightarrow 0$ , lo que significa  $f(x_n) \rightarrow f(x_0)$ . Como la sucesión  $x_n \rightarrow x_0$  escogida es arbitraria, concluimos que  $f$  es continua en  $x_0$ , esto en todo  $x_0 \in \Omega$ .

**Ejemplo 3.7.** Consideremos la función  $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  dada por

$$f(x, y) = \left(1 + \frac{x}{2} + \frac{1}{2} \sin y, 3 + \frac{1}{2} e^{-x^2}\right).$$

Afirmamos que  $f$  es Lipschitz en  $\mathbb{R}^2$  para cierto  $K > 0$ . En efecto,

$$\|f(x_1, y_1) - f(x_2, y_2)\|^2 = \frac{1}{4} |(x_1 - x_2) + (\sin y_1 - \sin y_2)|^2 + \frac{1}{4} |e^{-x_1^2} - e^{-x_2^2}|^2,$$

Tenemos que  $|a + b|^2 \leq 2(a^2 + b^2)$ , por lo tanto

$$|(x_1 - x_2) + (\sin y_1 - \sin y_2)|^2 \leq 2|x_1 - x_2|^2 + 2|\sin y_1 - \sin y_2|^2,$$

de modo que

$$\|f(x_1, y_1) - f(x_2, y_2)\|^2 \leq \frac{1}{2} |x_1 - x_2|^2 + \frac{1}{2} |\sin y_1 - \sin y_2|^2 + \frac{1}{4} |e^{-x_1^2} - e^{-x_2^2}|^2. \quad (3.13)$$

Ahora, por el teorema del valor medio, tenemos que existe  $\xi$  entre  $y_1$  e  $y_2$  con

$$(\sin y_1 - \sin y_2) = (\cos \xi) (y_1 - y_2).$$

Entonces, como  $|\cos \xi| \leq 1$ , se sigue que

$$|\sin y_1 - \sin y_2| \leq |y_1 - y_2|. \quad (3.14)$$

Por otra parte, para cierto  $\xi$  entre  $x_1$  y  $x_2$  tenemos

$$(e^{-x_1^2} - e^{-x_2^2}) = (-2\xi e^{-\xi^2}) (x_1 - x_2),$$

y por lo tanto

$$|e^{-x_1^2} - e^{-x_2^2}| \leq 2|\xi| e^{-|\xi|^2} |x_1 - x_2|.$$

La función  $t \mapsto 2te^{-t^2}$  se maximiza en  $[0, \infty)$  en  $t = \frac{1}{\sqrt{2}}$ , como se verifica fácilmente. Así,

$$\max_{t \geq 0} 2te^{-t^2} = \sqrt{2}e^{-\frac{1}{2}},$$

de lo cual obtenemos

$$|e^{-x_1^2} - e^{-x_2^2}| \leq \sqrt{2}e^{-\frac{1}{2}} |x_1 - x_2|. \quad (3.15)$$

Sustituyendo las desigualdades (3.14), (3.15) en (3.13), obtenemos que

$$\|f(x_1, y_1) - f(x_2, y_2)\|^2 \leq \frac{1}{2}|x_1 - x_2|^2 + \frac{1}{2}|y_1 - y_2|^2 + \frac{1}{2}e^{-1}|x_1 - x_2|^2,$$

de donde

$$\|f(x_1, y_1) - f(x_2, y_2)\|^2 \leq \frac{1}{2}(1 + e^{-1}) (|x_1 - x_2|^2 + |y_1 - y_2|^2),$$

y por lo tanto,

$$\|f(x_1, y_1) - f(x_2, y_2)\| \leq K \|(x_1, y_1) - (x_2, y_2)\|, \quad K = \sqrt{\frac{1}{2}(1 + e^{-1})}.$$

Un Teorema fundamental para determinar existencia de soluciones de ecuaciones no-lineales en  $\mathbb{R}^N$  es el llamado *Teorema del punto fijo de Banach*. Consideremos una función  $f : \Omega \subset \mathbb{R}^N \rightarrow \mathbb{R}^N$ . Escribamos

$$f(x) = (f_1(x), \dots, f_N(x)), \quad x = (x_1, \dots, x_N).$$

El teorema del punto fijo de Banach trata de la resolución del sistema de  $N$  ecuaciones y  $N$  incógnitas,

$$x_j - f_j(x_1, \dots, x_N) = 0, \quad j = 1, \dots, N,$$

esto es, la ecuación en  $\mathbb{R}^N$   $x = f(x)$ . Si  $x$  satisface esta igualdad, decimos que  $x$  es un *punto fijo* de  $f$ .

Decimos, por otra parte que  $f$  es *contractante en  $\Omega$*  si es Lipschitz de constante  $K < 1$ , esto es, si existe  $0 < K < 1$  tal que la relación (3.12) se cumple.

**Teorema 3.3.** (*Teorema del punto fijo de Banach*) Sea  $f : \Omega \subset \mathbb{R}^N \rightarrow \mathbb{R}^N$  una función contractante. Supongamos además que  $\Omega$  es cerrado, y que  $f(x) \in \Omega$  para todo  $x \in \Omega$ . Entonces existe un único  $\bar{x} \in \Omega$  tal que  $\bar{x} = f(\bar{x})$ , en otras palabras,  $f$  posee un único punto fijo en  $\Omega$ .

**Demostración.** Para probar este resultado consideramos la sucesión de puntos de  $\Omega$ , definida recursivamente como sigue: Dado  $x_0 \in \Omega$  cualquiera,

$$x_{n+1} = f(x_n) \quad n = 0, 1, 2, \dots$$

Para probar la existencia de un punto fijo, demostraremos que esta sucesión es convergente, y que su límite es el punto fijo buscado. Observemos que, como  $f$  es Lipschitz de constante  $K$  en  $\Omega$ , entonces para todo  $j \geq 1$ ,

$$\|x_{j+1} - x_j\| = \|f(x_j) - f(x_{j-1})\| \leq K\|x_j - x_{j-1}\|.$$

Así, iterando esta relación obtenemos

$$\|x_{j+1} - x_j\| \leq K\|x_j - x_{j-1}\| \leq K^2\|x_{j-1} - x_{j-2}\| \leq \cdots \leq K^j\|x_1 - x_0\|. \quad (3.16)$$

Demostraremos que  $(x_n)_{n \in \mathbb{N}}$  es una sucesión de Cauchy. Supongamos que  $1 \geq n < m$ . Por la propiedad telescópica de la suma, tenemos

$$x_m - x_n = \sum_{j=n}^{m-1} (x_{j+1} - x_j).$$

Por la desigualdad triangular para la norma, tenemos entonces que

$$\|x_m - x_n\| = \left\| \sum_{j=n}^{m-1} (x_{j+1} - x_j) \right\| \leq \sum_{j=n}^{m-1} \|x_{j+1} - x_j\|.$$

Así, de la relación (3.16), y del hecho que  $K < 1$ , obtenemos

$$\begin{aligned} \|x_m - x_n\| &\leq \sum_{j=n}^{m-1} K^j \|x_1 - x_0\| \leq \\ \sum_{j=n}^{\infty} K^j \|x_1 - x_0\| &= \sum_{j=0}^{\infty} K^{j+n} \|x_1 - x_0\| = \frac{K^n}{1-K} \|x_1 - x_0\|. \end{aligned}$$

Ahora, como  $K < 1$  tenemos que  $\frac{K^n}{1-K} \|x_1 - x_0\| \rightarrow 0$ . Así, dado  $\varepsilon > 0$ , existe  $n_0 \geq 1$  tal que

$$\frac{K^n}{1-K} \|x_1 - x_0\| < \varepsilon.$$

De este modo, tenemos que para  $n, m \geq n_0$ ,  $m > n$ ,

$$\|x_m - x_n\| < \varepsilon.$$

Hemos probado que la sucesión  $x_n$  es de Cauchy. Por el teorema 2.2, deducimos que  $x_n$  es convergente en  $\mathbb{R}^N$ , digamos

$$\lim_{n \rightarrow \infty} x_n = \bar{x}.$$



Por otra parte, como  $x_n \in \Omega$  para todo  $n$ , se tiene que  $\bar{x} \in \text{Adh}(\Omega)$ . Pero  $\Omega$  es cerrado, por lo tanto  $\bar{x} \in \Omega$ . Además, como  $f$  es Lipschitz, es continua relativamente a  $\Omega$ , y se sigue que, también,

$$\lim_{n \rightarrow \infty} f(x_n) = f(\bar{x}) .$$

Pero  $x_{n+1}$  es una subsucesión de  $x_n$ , posee por lo tanto el mismo límite  $\bar{x}$ . Concluimos que

$$f(\bar{x}) = \lim_{n \rightarrow \infty} f(x_n) = \lim_{n \rightarrow \infty} x_{n+1} = \bar{x} ,$$

y entonces  $\bar{x}$  es un punto fijo de  $f$  en  $\Omega$ . Hemos demostrado la existencia del punto fijo. Para probar unicidad, supongamos que existen  $\bar{x}_1, \bar{x}_2 \in \Omega$  con  $\bar{x}_1 = f(\bar{x}_1)$ ,  $\bar{x}_2 = f(\bar{x}_2)$ . Entonces

$$\|\bar{x}_1 - \bar{x}_2\| = \|f(\bar{x}_1) - f(\bar{x}_2)\| \leq K\|\bar{x}_1 - \bar{x}_2\| ,$$

y se sigue que  $(1 - K)\|\bar{x}_1 - \bar{x}_2\| \leq 0$ . Como  $K < 1$ , deducimos que  $\|\bar{x}_1 - \bar{x}_2\| = 0$ , es decir  $\bar{x}_1 = \bar{x}_2$ . Hemos probado que solo un punto fijo de  $f$  existe.  $\square$

**Ejemplo 3.8.** Consideremos la función  $f$  del ejemplo 3.7. Según vimos,  $f$  es Lipschitz en  $\mathbb{R}^2$  con constante

$$K = \sqrt{\frac{1}{2}(1 + e^{-1})} < 1.$$

De acuerdo al teorema anterior,  $f$  posee un único punto fijo en  $\mathbb{R}^2$  ( $\mathbb{R}^2$  es obviamente un conjunto cerrado!). Esto significa que el sistema no-lineal de ecuaciones

$$2 - x + \sin y = 0 ,$$

$$6 + e^{-x^2} - 2y = 0 .$$

posee una única solución  $(x, y) \in \mathbb{R}^2$ .

#### 4. DIFERENCIABILIDAD DE FUNCIONES DE VARIAS VARIABLES

En las secciones anteriores hemos analizado las nociones de límite y continuidad de funciones. En modo análogo a la secuencia lógica utilizada en el Cálculo de una variable, introducimos a continuación la noción de diferenciabilidad. El álgebra lineal nos será de gran utilidad en este análisis. En efecto, tal como en el caso de una variable, diferenciabilidad en un punto corresponderá al hecho que la función podrá *aproximarse bien* por una función lineal cerca de este punto.

Recordemos que toda función lineal  $L : \mathbb{R}^N \rightarrow \mathbb{R}^m$  puede representarse en modo matricial como

$$L(x) = Ax$$

donde  $A = [a_{ij}]$  es una matriz  $m \times N$ , y  $x = (x_1, \dots, x_N)$  se escribe como un vector columna. Así,

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & \cdot & \cdots & a_{2N} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{m1} & a_{m2} & \cdots & a_{mN} \end{bmatrix} \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_N \end{bmatrix} \quad Ax = \begin{bmatrix} \sum_{j=1}^N a_{1j}x_j \\ \sum_{j=1}^N a_{2j}x_j \\ \cdot \\ \cdot \\ \cdot \\ \sum_{j=1}^N a_{mj}x_j \end{bmatrix}.$$

Una función *lineal afín* es una  $f : \mathbb{R}^N \rightarrow \mathbb{R}^m$  de la forma

$$f(x) = Ax + b$$

donde  $b \in \mathbb{R}^m$ . Notemos que si  $x_0 \in \mathbb{R}^N$ , entonces tenemos que  $b = f(x_0) - Ax_0$ , y por lo tanto vale la igualdad

$$f(x) = f(x_0) + A(x - x_0). \quad (4.1)$$

Para una función  $f$  cualquiera, diremos que  $f$  es diferenciable en el punto  $x_0$  si existe una matriz  $A$  tal que para todo  $x$  cercano a  $x_0$  vale que  $f(x) \sim f(x_0) + A(x - x_0)$ , esto es, si  $f$  es *aproximadamente* una función afín cerca de  $x_0$ . Precisaremos este concepto a continuación, primero recordando que para  $f : \mathbb{R} \rightarrow \mathbb{R}$  decimos que  $f$  es diferenciable en  $x_0$  si el límite

$$a = \lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

existe. En tal caso, por cierto, llamamos  $a = f'(x_0)$ . Esta relación puede reescribirse en el modo siguiente:

$$\lim_{x \rightarrow x_0} \frac{|f(x_0 + h) - f(x_0) - ah|}{|h|} = 0.$$

Así,  $f$  es diferenciable en  $x_0$  si y solo si existe  $a \in \mathbb{R}$  tal que

$$f(x_0 + h) = f(x_0) + ah + \theta(h)$$

donde la función  $\theta(h)$  satisface

$$\lim_{h \rightarrow 0} \frac{|\theta(h)|}{|h|} = 0.$$

Por cierto,  $\theta(h) = f(x_0 + h) - f(x_0) - ah$ . Expresada en esta forma que la noción de diferenciabilidad puede ser naturalmente extendida a funciones de varias variables. Para evitar complicaciones que surgen

en intentar extender nociones análogas a derivadas laterales, como fue hecho en el caso de una variable, es conveniente suponer que  $f$  está definida en un conjunto *abierto* que contiene al punto  $x_0$  de nuestro interés.

**4.1. Definición de diferenciabilidad y derivada.** Consideremos entonces una función  $f : \Omega \subset \mathbb{R}^N \rightarrow \mathbb{R}^m$ , donde  $\Omega$  es un abierto en  $\mathbb{R}^N$ , y  $x_0 \in \Omega$ . Decimos que  $f$  es *diferenciable en  $x_0$*  si existe una matriz  $A$ ,  $m \times N$ , tal que

$$f(x_0 + h) = f(x_0) + Ah + \theta(h) \quad (4.2)$$

donde la función  $\theta(h)$  satisface

$$\lim_{h \rightarrow 0} \frac{\|\theta(h)\|}{\|h\|} = 0. \quad (4.3)$$

Esta última condición expresa que  $f(x_0 + h)$  difiere de una función afín en un término que va a cero más rápido que el orden lineal cuando  $h$  va a 0.

Afirmamos que si existe una matriz  $A$  tal que las relaciones (4.2), (4.3) se satisfacen, entonces ésta es única. En efecto, supongamos que  $A_1$  y  $A_2$  son dos matrices tales que

$$\lim_{h \rightarrow 0} \frac{\|\theta_i(h)\|}{\|h\|} = 0, \quad i = 1, 2,$$

donde

$$\theta_i(h) = f(x_0 + h) - f(x_0) - A_i h.$$

Notemos que

$$0 \leq \frac{\|\theta_1(h) - \theta_2(h)\|}{\|h\|} \leq \frac{\|\theta_1(h)\|}{\|h\|} + \frac{\|\theta_2(h)\|}{\|h\|} \rightarrow 0$$

cuando  $h \rightarrow 0$ , y entonces

$$0 = \lim_{h \rightarrow 0} \frac{\|\theta_1(h) - \theta_2(h)\|}{\|h\|} = \lim_{h \rightarrow 0} \frac{\|(A_1 - A_2)h\|}{\|h\|}.$$

Fijemos cualquier vector  $g \in \mathbb{R}^N$  con  $\|g\| = 1$  y consideremos la sucesión  $h_n = \frac{1}{n}g$ . Entonces, por definición de límite  $h \rightarrow 0$  tenemos en particular que

$$0 = \lim_{n \rightarrow \infty} \frac{\|(A_1 - A_2)h_n\|}{\|h_n\|} = \|(A_1 - A_2)g\|,$$

esto es,  $(A_1 - A_2)g = 0$ . Esto de inmediato implica que  $A_1 = A_2$ , pues escogiendo  $g = e_i$ , el  $i$ -ésimo elemento de la base canónica, obtenemos que la  $i$ -ésima columna de  $A$  es igual a 0, esto para todo  $i = 1, \dots, N$ .

En caso de existir esta matriz  $A$  le llamamos, sin ambigüedad en virtud de su unicidad, *la derivada de  $A$  en  $x_0$* , y denotamos

$$A = f'(x_0).$$

$A$  se denomina también a veces *la matriz Jacobiana de  $f$  en  $x_0$* .

La función

$$T(x) = f(x_0) + f'(x_0)(x - x_0)$$

se denomina, naturalmente, *aproximación afín de  $f$  en torno a  $x_0$* .

Como en el caso de funciones de una variable, la diferenciabilidad de  $f$  en  $x_0$  implica su continuidad. En efecto, tenemos que

$$\lim_{h \rightarrow 0} \frac{\|f(x_0 + h) - f(x_0) - f'(x_0)h\|}{\|h\|} = 0,$$

lo que implica que existe  $\delta > 0$  tal que si  $\|h\| < \delta$  entonces

$$\frac{\|f(x_0 + h) - f(x_0) - f'(x_0)h\|}{\|h\|} \leq 1. \quad (4.4)$$

Anora, por desigualdad triangular

$$\|f(x_0 + h) - f(x_0)\| \leq \|f(x_0 + h) - f(x_0) - f'(x_0)h\| + \|f'(x_0)h\|. \quad (4.5)$$

Escribamos  $f'(x_0) = [a_{ij}]$ . Entonces

$$\begin{aligned} \|f'(x_0)h\|^2 &= \sum_{i=1}^m \left( \sum_{j=1}^N a_{ij}h_j \right)^2 \leq \\ &\sum_{i=1}^m \left( \sum_{j=1}^N |a_{ij}| |h_j| \right)^2. \end{aligned}$$

Como para todo  $j$  tenemos  $|h_j| \leq \|h\|$ , entonces

$$\|f'(x_0)h\| \leq \left( \sum_{i=1}^m \left( \sum_{j=1}^N |a_{ij}| |h_j| \right)^2 \right)^{\frac{1}{2}} \|h\|. \quad (4.6)$$

Usando las desigualdades (4.4) y (4.6) para estimar el lado derecho en (4.5), obtenemos que

$$(\forall h \in B(0, \delta)) : \quad \|f(x_0 + h) - f(x_0)\| \leq C\|h\|, \quad (4.7)$$

donde

$$C = 1 + \left( \sum_{i=1}^m \left( \sum_{j=1}^N |a_{ij}|^2 \right) \right)^{\frac{1}{2}}.$$

De (4.7), concluimos en particular

$$\lim_{h \rightarrow 0} f(x_0 + h) = f(x_0),$$

o equivalentemente

$$\lim_{x \rightarrow x_0} f(x) = f(x_0),$$

lo que significa que  $f$  es continua en  $x_0$ .

**Ejemplo 4.1.** Consideremos la función  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  dada por

$$f(x, y) = 2x^2 + 3xy + 5y^2 + 3x + 2y + 10.$$

Probaremos que  $f$  es diferenciable en este punto y encontraremos su derivada. Tenemos

$$f(x_0 + h, y_0 + k) = 2(x_0 + h)^2 + 3(x_0 + h)(y_0 + k) + 5(y_0 + k)^2 + 3(x_0 + h) + 2(y_0 + k) + 10.$$

Expandiendo los cuadrados y reagrupando términos, vemos que

$$\begin{aligned} f(x_0 + h, y_0 + k) &= (2x_0^2 + 5y_0^2 + 3x_0y_0 + 3x_0 + 2y_0 + 10) + \\ &\quad (4x_0 + 3y_0 + 3)h + (3x_0 + 10y_0 + 2)k + (2h^2 + 3hk + 5k^2). \end{aligned}$$

Así, vemos que

$$f(x_0 + h, y_0 + k) = f(x_0, y_0) + \ell(h, k) + \theta(h, k)$$

donde  $\ell(h, k)$  es lineal en  $(h, k)$  y  $\theta(h, k)$  lleva solo términos cuadráticos. Tenemos

$$\ell(h, k) = [4x_0 + 3y_0 + 3, 3x_0 + 10y_0 + 2] \begin{bmatrix} h \\ k \end{bmatrix}.$$

Ahora,

$$|\theta(h, k)| = |2h^2 + 3hk + 5k^2| \leq 2h^2 + 5k^2 + 3|h||k| \leq 7(h^2 + k^2)$$

y entonces

$$0 \leq \lim_{(h,k) \rightarrow (0,0)} \frac{|\theta(h, k)|}{\|(h, k)\|} \leq 7 \lim_{(h,k) \rightarrow (0,0)} \|(h, k)\| = 0.$$

Concluimos entonces que  $f$  es diferenciable en  $(x_0, y_0)$  y que su derivada está dada por la matriz fila  $1 \times 2$ ,

$$f'(x_0, y_0) = [4x_0 + 3y_0 + 3, 3x_0 + 10y_0 + 2].$$

**Ejemplo 4.2.** Sea  $f : \mathbb{R}^N \rightarrow \mathbb{R}$  dada por la forma cuadrática

$$f(x) = x^T A x$$

donde  $A$  es una matriz  $N \times N$ . Dado  $x_0 \in \mathbb{R}^N$ , observemos que

$$f(x_0 + h) = (x_0 + h)^T A (x_0 + h) = x_0^T A x_0 + x_0^T A h + h^T A x_0 + h^T A h.$$

Como en el ejemplo anterior, reconocemos en esta expansión inmediatamente términos lineales y cuadráticos en  $h$ . Notemos que

$$h^T A x_0 = (h^T A x_0)^T = x_0^T A^T h$$

y entonces, para el término lineal tenemos

$$\ell(h) := x_0^T A h + h^T A x_0 = [x_0^T (A + A^T)] h.$$

Por otra parte, tenemos que

$$\theta(h) := h^T A h = \sum_{i=1}^N \sum_{j=1}^N a_{ij} h_i h_j.$$

Notemos que, para todo  $i, j$ ,

$$|h_i| |h_j| \leq |h_i|^2 + |h_j|^2 \leq 2 \sum_{l=1}^N |h_l|^2 = 2 \|h\|^2$$

Por lo tanto,

$$|\theta(h)| \leq 2 \|h\|^2 \sum_{i=1}^N \sum_{j=1}^N |a_{ij}|,$$

de modo que

$$0 \leq \lim_{h \rightarrow 0} \frac{|\theta(h)|}{\|h\|} \leq 2 \lim_{h \rightarrow 0} \|h\| \sum_{i,j} |a_{ij}| = 0.$$

Concluimos entonces que  $f$  es diferenciable en  $x_0$  y que su derivada está dada por la matriz (fila)  $1 \times N$ ,

$$f'(x_0) = x_0^T (A + A^T).$$

**4.2. Algunas propiedades operacionales de la diferenciabilidad.** La sola definición, como la utilizamos en los ejemplos anteriores, no es por si sola una herramienta suficientemente poderosa en la identificación de la derivada en caso que una función sea diferenciable. Desarrollaremos a continuación una serie de reglas operacionales que nos permitan probar diferenciabilidad de una función dada, y encontrar la matriz derivada si es que ésta existe.

Para comenzar, una propiedad básica de la diferenciabilidad es su respeto al álgebra de  $\mathbb{R}^N$ , tal como en una variable. Tenemos el siguiente resultado.

**Proposición 4.1.** (*Álgebra de la diferenciabilidad*). Sean  $\Omega \subset \mathbb{R}^N$  abierto,  $f, g : \Omega \rightarrow \mathbb{R}^m$ ,  $x_0 \in \mathbb{R}^N$ ,  $\alpha, \beta \in \mathbb{R}$ . Supongamos que  $f$  y  $g$  son diferenciables en  $x_0$ . Entonces

(a) La función  $\alpha f + \beta g : \Omega \rightarrow \mathbb{R}^m$  es diferenciable en  $x_0$  y

$$(\alpha f + \beta g)'(x_0) = \alpha f'(x_0) + \beta g'(x_0).$$

(b) Si  $m = 1$ , la función  $f \cdot g : \Omega \rightarrow \mathbb{R}$  es diferenciable en  $x_0$  y se tiene la validez de la regla del producto,

$$(f \cdot g)'(x_0) = g(x_0)f'(x_0) + f(x_0)g'(x_0)$$

**Demostración.** Realizaremos solo la demostración de la propiedad del producto (b), dejando la parte (a) como un ejercicio al lector. Denotemos

$$\theta_1(h) = f(x_0+h) - f(x_0) - f'(x_0)h, \quad \theta_2(h) = g(x_0+h) - g(x_0) - g'(x_0)h.$$

Observemos que

$$\begin{aligned} \theta(h) &:= (fg)(x_0+h) - (fg)(x_0) - [g(x_0)f'(x_0) + f(x_0)g'(x_0)]h = \\ &= [g(x_0) + g'(x_0)h]\theta_1(h) + f(x_0+h)\theta_2(h) + f'(x_0)h g'(x_0)h. \end{aligned}$$

Tenemos que

$$\lim_{h \rightarrow 0} [g(x_0) + g'(x_0)h] \frac{\theta_1(h)}{\|h\|} = g(x_0) \cdot 0 = 0.$$

Por otra parte,

$$\lim_{h \rightarrow 0} f(x_0+h) \frac{\theta_2(h)}{\|h\|} = f(x_0) \cdot 0 = 0.$$

Finalmente, gracias a la desigualdad de Cauchy-Schwartz,

$$|f'(x_0)h| |g'(x_0)h| \leq \|f'(x_0)\| \|g'(x_0)\| \|h\|^2,$$

y se sigue que

$$\lim_{h \rightarrow 0} \frac{f'(x_0)h g'(x_0)h}{\|h\|} = 0.$$

La conclusión es entonces que

$$\lim_{h \rightarrow 0} \frac{\theta(h)}{\|h\|} = 0,$$

y la afirmación (b) entonces se cumple.  $\square$

Si bien la introducción de varias variables es una variante altamente no-trivial de la diferenciabilidad de funciones de una variable real, no es en realidad el caso en lo que concierne a varias funciones coordenadas: la diferenciabilidad de una función  $f : \mathbb{R}^N \rightarrow \mathbb{R}^m$  se reduce a la diferenciabilidad de cada una de sus  $m$  funciones coordenadas, como

enuncia el siguiente resultado, análogo a la proposición 3.3 en cuanto a continuidad.

**Proposición 4.2.** Sean  $\Omega \subset \mathbb{R}^N$  un abierto,  $f : \Omega \rightarrow \mathbb{R}^m$ ,  $x_0 \in \Omega$ ,

$$f(x) = (f_1(x), f_2(x), \dots, f_m(x)).$$

Entonces las siguientes afirmaciones son equivalentes.

(a)  $f$  es diferenciable en  $x_0$ .

(b) Para todo  $i = 1, \dots, m$ , las funciones  $f_i : \Omega \rightarrow \mathbb{R}$  son diferenciables en  $x_0$  relativamente a  $\Omega$ .

En tal caso se tiene la igualdad

$$f'(x_0) = \begin{bmatrix} f'_1(x_0) \\ f'_2(x_0) \\ \vdots \\ f'_m(x_0) \end{bmatrix}.$$

**Demostración.** Notemos que para una matriz  $A$ ,  $m \times N$ , que expresamos

$$A = \begin{bmatrix} A_1. \\ A_2. \\ \vdots \\ A_m. \end{bmatrix}.$$

donde los  $A_i$  son vectores fila  $N \times 1$ , se tiene la igualdad

$$\begin{aligned} \theta(h) &:= f(x_0 + h) - f(x_0) - Ah = \\ &= \begin{bmatrix} f_1(x_0 + h) - f_1(x_0) - A_1.h \\ f_2(x_0 + h) - f_2(x_0) - A_2.h \\ \vdots \\ f_m(x_0 + h) - f_m(x_0) - A_m.h \end{bmatrix} =: \begin{bmatrix} \theta_1(h) \\ \theta_2(h) \\ \vdots \\ \theta_m(h) \end{bmatrix}. \end{aligned}$$

De este modo,

$$\lim_{h \rightarrow 0} \frac{\theta(h)}{\|h\|} = 0 \quad \Longleftrightarrow \quad \lim_{h \rightarrow 0} \frac{\theta_i(h)}{\|h\|} = 0 \quad \forall i = 1, \dots, m.$$

Así,  $A = f'(x_0)$  si y solo si  $A_i = f'_i(x_0)$  para todo  $i$ . Esto concluye la demostración.  $\square$



A manera de ejemplo, consideremos una función  $\phi : ]a, b[ \rightarrow \mathbb{R}^m$ . Así,  $\phi$  es diferenciable en  $t$  si y solo si sus coordenadas lo son y se tiene la relación natural

$$\phi'(t) = \begin{bmatrix} \phi'_1(t) \\ \phi'_2(t) \\ \vdots \\ \phi'_m(t) \end{bmatrix}.$$

Observemos en particular que la derivada puede calcularse a partir de la fórmula habitual,

$$\phi'(t) = \lim_{h \rightarrow 0} \frac{1}{h} (\phi(t+h) - \phi(t)).$$

**4.3. Derivadas direccionales, parciales y diferenciabilidad.** Una propiedad sencilla, al mismo tiempo poderosa para el cálculo de la derivada de una función de varias variables, es su vínculo con derivadas de funciones de una variable, en el modo que enuncia el siguiente resultado.

**Proposición 4.3.** Sean  $\Omega \subset \mathbb{R}^N$  un abierto,  $f : \Omega \rightarrow \mathbb{R}^m$ ,  $x_0 \in \Omega$ . Supongamos que  $f$  es diferenciable en  $x_0$ . Entonces para todo  $e \in \mathbb{R}^N \setminus \{0\}$  la función  $t \mapsto f(x_0 + te)$  es diferenciable en  $t = 0$ , y se cumple que

$$\frac{d}{dt} f(x_0 + te) \Big|_{t=0} = f'(x_0)h.$$

**Demostración.** A lo largo de cualquier sucesión  $h_n \rightarrow 0$  tenemos que

$$\lim_{n \rightarrow \infty} \frac{\|f(x_0 + h_n) - f(x_0) - f'(x_0)h_n\|}{\|h_n\|} = 0.$$

Escojamos  $h_n = t_n e$ , con  $t_n \rightarrow 0$ . Entonces

$$\lim_{n \rightarrow \infty} \frac{\|f(x_0 + t_n e) - f(x_0) - t_n f'(x_0)e\|}{|t_n| \|e\|} = 0,$$

lo que implica

$$\lim_{n \rightarrow \infty} \left\| \frac{1}{t_n} (f(x_0 + t_n e) - f(x_0)) - f'(x_0)e \right\| = 0,$$

es decir

$$\lim_{n \rightarrow \infty} \frac{1}{t_n} (f(x_0 + t_n e) - f(x_0)) = f'(x_0)e.$$

Como la sucesión  $t_n \rightarrow 0$  es arbitraria, se sigue que

$$\lim_{t \rightarrow 0} \frac{1}{t} (f(x_0 + te) - f(x_0)) = f'(x_0)e,$$

y la demostración ha sido concluida.  $\square$

El resultado anterior nos permite entregar una interpretación geométrica de la derivada de una función de varias variables en el caso  $m = 1$ . Supongamos que  $\|e\| = 1$ . Entonces  $t \mapsto x_0 + te$  define la recta que pasa por  $x_0$  y tiene a  $e$  como vector director. Así, la función  $t \mapsto f(x_0 + te)$  corresponde a la restricción de la función  $f$  a esta recta, y su derivada en  $t = 0$ , el número  $f'(x_0)e$ , corresponde entonces a la *pendiente* del gráfico de  $f$  en el punto  $x_0$ , medida en la dirección del vector unitario  $e$  o a la *tasa de crecimiento* de la función  $f$  en este punto y en esta dirección.

Lo anterior motiva la siguiente definición general: Sean  $\Omega \subset \mathbb{R}^N$  un abierto,  $f : \Omega \rightarrow \mathbb{R}^m$ ,  $x_0 \in \Omega$ ,  $e \in \mathbb{R}^N$  con  $\|e\| = 1$ . En caso de existir, el límite

$$f'(x_0; e) := \frac{d}{dt} f(x_0 + te) \Big|_{t=0} = \lim_{t \rightarrow 0} \frac{1}{t} (f(x_0 + te) - f(x_0))$$

se denomina *derivada direccional* en  $x_0$ , en la dirección  $e$ .

En virtud de la proposición anterior, se tiene entonces que la diferenciabilidad de  $f$  en  $x_0$  implica la existencia de derivadas direccionales en  $x_0$  en toda dirección  $e$ , y además en tal caso  $f'(x_0; e) = f'(x_0)e$ .

De especial relevancia son las derivadas direccionales de  $f$  en la dirección de los elementos de la base canónica,

$$e_j = (0, 0, \dots, 0, 1, 0 \dots 0),$$

en el cual todas las componentes excepto la  $j$ -ésima son iguales a cero. Las derivadas direccionales en  $x_0$  en las direcciones  $e_j$  se denominan *derivadas parciales* de  $f$  en  $x_0$ . Así, la  $j$ -ésima derivada parcial de  $f$  se define, en caso de existir, como

$$\frac{\partial f}{\partial x_j}(x_0) := \lim_{t \rightarrow 0} \frac{1}{t} (f(x_0 + te_j) - f(x_0)).$$

Es frecuente también denotar

$$\frac{\partial f}{\partial x_j}(x_0) =: f_{x_j}(x_0).$$

Analicemos más precisamente esta cantidad. Tenemos que

$$\frac{\partial f}{\partial x_j}(x_0) = \frac{d}{dt} f(x_{01}, \dots, x_{0j} + t, \dots, x_{0N}) \Big|_{t=0} =$$

$$\frac{d}{dx_j} f(x_{01}, \dots, x_j, \dots, x_{0N}) \Big|_{x_j=x_{0j}}.$$

Así, derivar parcialmente, corresponde a derivar en la  $j$ -ésima variable, considerando a las restantes variables como constantes.

**Ejemplo 4.3.** Sea  $f(x, y) = e^{xy} \cos(x^2 + y^2)$ . Las derivadas parciales respecto a las variables  $x$  e  $y$  calculadas en un punto arbitrario  $(x, y)$  están dadas por

$$\frac{\partial f}{\partial x}(x, y) = ye^{xy} \cos(x^2 + y^2) - 2xe^{xy} \sin(x^2 + y^2),$$

$$\frac{\partial f}{\partial y}(x, y) = xe^{xy} \cos(x^2 + y^2) - 3y^2 e^{xy} \sin(x^2 + y^2).$$

Con este ejemplo vemos que en el caso de funciones dadas por fórmulas explícitas, el cálculo de derivadas parciales se realiza simplemente de acuerdo a las reglas de la derivación de una variable. En caso de una función diferenciable, estas cantidades de hecho *determinan* a la matriz derivada. En efecto, si  $f$  es diferenciable en  $x_0$  tenemos que

$$\frac{\partial f}{\partial x_j}(x_0) = f'(x_0)e_j$$

que corresponde exactamente a la  $j$ -ésima columna de la matriz  $f'(x_0)$ . Por otra parte, sabemos también, en virtud de la proposición 4.2 que la  $i$ -ésima fila de la matriz  $f'(x_0)$  corresponde precisamente a la derivada de la función coordenada  $f_i$ ,  $f'_i(x_0)$ . La fórmula anterior nos dice que

$$\frac{\partial f_i}{\partial x_j}(x_0) = f'_i(x_0)e_j.$$

Así, este último número es exactamente la  $j$ -ésima coordenada de la fila  $i$  de la matriz  $f'(x_0)$ , esto es, la entrada  $ij$  de esta matriz. Tenemos entonces la validez del siguiente importante resultado para el cálculo de la matriz derivada.

**Proposición 4.4.** Sean  $\Omega \subset \mathbb{R}^N$  un abierto,  $f : \Omega \rightarrow \mathbb{R}^m$ ,  $x_0 \in \Omega$ . Supongamos que  $f = (f_1, \dots, f_m)$  es diferenciable en  $x_0$ . Tenemos entonces que la matriz  $f'(x_0)$  puede calcularse como

$$[f'(x_0)]_{ij} = \frac{\partial f_i}{\partial x_j}(x_0), \quad i = 1, \dots, m, \quad j = 1, \dots, N,$$

o

$$f'(x_0) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x_0) & \frac{\partial f_1}{\partial x_2}(x_0) & \cdots & \frac{\partial f_1}{\partial x_N}(x_0) \\ \frac{\partial f_2}{\partial x_1}(x_0) & \frac{\partial f_2}{\partial x_2}(x_0) & \cdots & \frac{\partial f_2}{\partial x_N}(x_0) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x_0) & \frac{\partial f_m}{\partial x_2}(x_0) & \cdots & \frac{\partial f_m}{\partial x_N}(x_0) \end{bmatrix}.$$

Además

$$\frac{\partial f}{\partial x_j}(x_0) = \begin{bmatrix} \frac{\partial f_1}{\partial x_j}(x_0) \\ \frac{\partial f_2}{\partial x_j}(x_0) \\ \vdots \\ \frac{\partial f_m}{\partial x_j}(x_0) \end{bmatrix}.$$

Es importante destacar que la sola existencia de las derivadas parciales, o incluso la de todas las derivadas direccionales, no implica por si sola la diferenciabilidad de  $f$ . Veamos dos ejemplos de este hecho.

**Ejemplo 4.4.** Consideremos la función

$$f(x, y) = \begin{cases} \frac{|x|y|}{\sqrt{x^2+y^2}} & \text{si } (x, y) \neq (0, 0), \\ 0 & \text{si } (x, y) = (0, 0). \end{cases}$$

Afirmamos que esta función es continua en  $(0, 0)$ , que todas sus derivadas direccionales existen en este punto, pero que  $f$  no es diferenciable.

Para la continuidad, observemos que si  $(x_n, y_n) \rightarrow (0, 0)$  con  $(x_n, y_n) \neq (0, 0)$ , se tiene que

$$|f(x_n, y_n) - f(0, 0)| = \frac{|x_n||y_n|}{\sqrt{x_n^2 + y_n^2}} \leq \frac{1}{2} \sqrt{x_n^2 + y_n^2} \rightarrow 0.$$

Así, tenemos que

$$\lim_{n \rightarrow \infty} f(x_n, y_n) = f(0, 0),$$

y como la sucesión escogida es arbitraria,

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

por lo tanto  $f$  es continua en  $(0, 0)$ .

Sea ahora  $e = (e_1, e_1)$  con  $\|e\| = 1$ . Tenemos entonces que para  $t \neq 0$ ,

$$\frac{1}{t} [f((0, 0) + t(e_1, e_2)) - f(0, 0)] = \frac{e_1|e_2|}{\sqrt{e_1^2 + e_2^2}},$$

y por lo tanto

$$f'((0,0); e) = \frac{e_1|e_2|}{\sqrt{e_1^2 + e_2^2}}.$$

En particular, notemos que evaluando en  $e = (1,0)$  y en  $e = (0,1)$  obtenemos

$$\frac{\partial f}{\partial x}(0,0) = 0 = \frac{\partial f}{\partial y}(0,0).$$

Si  $f$  fuese diferenciable en  $(0,0)$ , debiésemos entonces tener  $f'(0,0) = [0 \ 0]$ , y por lo tanto para cualquier  $e$ ,

$$f'((0,0); e) = f'(0,0)e = 0.$$

Pero la fórmula obtenida nos dice por ejemplo que para  $e = 2^{-\frac{1}{2}}(1,1)$ ,  $f'((0,0); e) = 1$ , una contradicción que nos muestra que  $f$  no es diferenciable en  $(0,0)$ .

**Ejemplo 4.5.** Consideremos ahora la función

$$f(x,y) = \begin{cases} 1 & \text{si } 0 < y < x^2, \\ 0 & \text{si no.} \end{cases}$$

Esta función no es continua en  $(0,0)$ , pues, escogiendo la sucesión

$$(x_n, y_n) = \left( \frac{1}{n}, \frac{1}{2n^2} \right) \rightarrow (0,0)$$

obtenemos que

$$f(x_n, y_n) \equiv 1 \not\rightarrow f(0,0) = 0.$$

Por otra parte, fijemos cualquier vector  $e = (e_1, e_2)$  con  $e \neq 0$ . Si  $e_1 = 0$  o  $e_2 = 0$ , obviamente  $f(te) = 0$  para todo  $t$ . Si  $e_1, e_2 \neq 0$ , se tiene que la desigualdad  $0 < te_2 < t^2 e_1^2$  solo puede tenerse si  $|t| > e_1^{-2}|e_2|$ . Así, independientemente de  $e$ , tenemos que  $f(te) = 0$  para todo  $t$  suficientemente pequeño. Entonces

$$f'((0,0); e) = \lim_{t \rightarrow 0} \frac{f(te) - 0}{t} = 0,$$

por lo tanto todas las derivadas direccionales en  $(0,0)$  existen y son iguales a 0. Por cierto, siendo  $f$  discontinua en  $(0,0)$ , no puede ser diferenciable.

#### 4.4. Continuidad de derivadas parciales y diferenciabilidad.

Los ejemplos anteriores nos dicen que la existencia de derivadas parciales no implica diferenciabilidad, ni siquiera continuidad de la función en cuestión. Como veremos a continuación, por fortuna sí es cierto que la existencia de derivadas parciales en un entorno del punto, más su continuidad como función de su argumento garantizan diferenciabilidad. Esta condición suficiente es la principal herramienta para decidir cuándo una función concreta es diferenciable.

**Teorema 4.1.** *Sea  $\Omega \subset \mathbb{R}^N$  un abierto,  $f : \Omega \rightarrow \mathbb{R}^m$ ,  $x_0 \in \Omega$ . Supongamos que la derivadas parciales*

$$\frac{\partial f}{\partial x_j}(x) \quad \text{existen} \quad \forall x \in \Omega, \quad \forall j = 1, \dots, N,$$

*y que además las funciones*

$$\frac{\partial f}{\partial x_j} : \Omega \rightarrow \mathbb{R}^m, \quad x \mapsto \frac{\partial f}{\partial x_j}(x)$$

*son continuas en  $x_0$ . Entonces  $f$  es diferenciable en  $x_0$ .*

**Demostración.** Basta probar el teorema para cada una de las funciones coordenadas de  $f$  en virtud de la proposición 4.2. Suponemos entonces que  $m = 1$ . Llamemos  $A$  la matriz  $1 \times N$  que tiene a las derivadas parciales de  $f$  en  $x_0$  como sus entradas, esto es

$$A = \left[ \frac{\partial f}{\partial x_1}(x_0) \quad \cdot \quad \cdot \quad \cdot \quad \frac{\partial f}{\partial x_N}(x_0) \right].$$

Sea

$$\theta(h) = f(x_0 + h) - f(x_0) - Ah.$$

Debemos probar que

$$\lim_{h \rightarrow 0} \frac{\theta(h)}{\|h\|} = 0.$$

Para ello, escribamos

$$a_j(h) = f(x_{01}, \dots, x_{0j-1}, x_{0j} + h_j, \dots, x_{0N} + h_N), \quad j = 1, \dots, N,$$

de modo que, en particular  $a_1(h) = f(x_0 + h)$ . Definimos también,  $a_{N+1}(h) = f(x_0)$ . Notemos que, entonces

$$f(x_0 + h) - f(x_0) = \sum_{i=1}^N [a_i(h) - a_{i+1}(h)].$$

Por otra parte,

$$\begin{aligned} a_j(h) - a_{j+1}(h) &= f(x_{01}, \dots, x_{0j-1}, x_{0j} + h_j, x_{0j+1} + h_{j+1}, \dots, x_{0N} + h_N) - \\ &\quad f(x_{01}, \dots, x_{0j-1}, x_{0j}, x_{0j+1} + h_{j+1}, \dots, x_{0N} + h_N) \end{aligned}$$

que es un incremento de  $f$  en la  $i$ -ésima variable. Aplicando entonces el teorema del valor medio para esta función de una variable (y a valores reales!), encontramos que existe  $\xi_j = \xi_j(h)$  entre 0 y  $h_j$  tal que

$$a_j(h) - a_{j+1}(h) = \frac{\partial f}{\partial x_j}(x_{01}, \dots, x_{0j-1}, x_{0j} + \xi_j, x_{0j+1} + h_{j+1}, \dots, x_{0N} + h_N) h_j .$$

Escribamos por conveniencia

$$\mathbf{h}^j = (0, \dots, 0, \xi_j, h_{j+1}, \dots, h_N) .$$

Se tiene ciertamente que  $\mathbf{h}^j \rightarrow 0$  si  $h \rightarrow 0$ . Notemos también que

$$Ah = \sum_{j=1}^N \frac{\partial f}{\partial x_j}(x_0) h_j ,$$

y por lo tanto

$$\theta(h) = \sum_{i=1}^N [a_i(h) - a_{i+1}(h) - \frac{\partial f}{\partial x_i}(x_0) h_i] ,$$

esto es,

$$\theta(h) = \sum_{i=1}^N \left[ \frac{\partial f}{\partial x_i}(x_0 + \mathbf{h}^i) - \frac{\partial f}{\partial x_i}(x_0) \right] h_i ,$$

de modo que, por desigualdad triangular,

$$|\theta(h)| \leq \sum_{i=1}^N \left| \frac{\partial f}{\partial x_i}(x_0 + \mathbf{h}^i) - \frac{\partial f}{\partial x_i}(x_0) \right| |h_i| ,$$

y por Cauchy-Swchartz,

$$|\theta(h)| \leq \left( \sum_{i=1}^N \left| \frac{\partial f}{\partial x_i}(x_0 + \mathbf{h}^i) - \frac{\partial f}{\partial x_i}(x_0) \right|^2 \right)^{\frac{1}{2}} \left( \sum_{i=1}^N h_i^2 \right)^{\frac{1}{2}} ..$$

Así,

$$0 \leq \lim_{h \rightarrow 0} \frac{|\theta(h)|}{\|h\|} \leq \lim_{h \rightarrow 0} \left( \sum_{i=1}^N \left| \frac{\partial f}{\partial x_i}(x_0 + \mathbf{h}^i) - \frac{\partial f}{\partial x_i}(x_0) \right|^2 \right)^{\frac{1}{2}} = 0 ,$$

este último límite pues para todo  $j$ ,

$$\lim_{h \rightarrow 0} \frac{\partial f}{\partial x_j}(x_0 + \mathbf{h}^j) = \frac{\partial f}{\partial x_j}(x_0)$$

por la continuidad de la función  $\frac{\partial f}{\partial x_j}(x)$  en  $x = x_0$ . Esto concluye la demostración.  $\square$

Decimos que una función  $f$  es *continuamente diferenciable en*  $\Omega$  si todas sus derivadas parciales existen en todo  $\Omega$  y definen funciones

continuas en  $\Omega$ . Se dice a veces en tal caso que  $f$  es de clase  $C^1$  en  $\Omega$ . Se denota por  $C^1(\Omega)$  el espacio vectorial de las funciones de de clase  $C^1$  en  $\Omega$ .

**Ejemplo 4.6.** Las funciones construidas a través de fórmulas algebraicas basadas en las funciones habituales del Cálculo: polinomios, trigonométricas, exponencial, etc., son típicamente diferenciables dentro de sus dominios de definición. Consideremos por ejemplo  $f : \mathbb{R}^2 \rightarrow \mathbb{R}^3$  dada por

$$f(x, y) = (x^2 \sin y, y^2 e^{xy}, x^2 y)$$

Notemos que

$$\frac{\partial f}{\partial x}(x, y) = (2x \sin y, y^3 e^{xy}, 2xy),$$

$$\frac{\partial f}{\partial y}(x, y) = (x^2 \cos y, 2ye^{xy} + xy^2 e^{xy}, x^2).$$

Estas funciones están definidas en todo  $\mathbb{R}^2$ , y son evidentemente continuas en todo punto  $(x, y)$ , al ser constituidas por sumas, producto y composición de funciones continuas. Por ejemplo,  $\frac{\partial f}{\partial y}(x, y)$  es precisamente la función del ejemplo 3.5. Concluimos entonces, del teorema 4.1, que  $f$  es diferenciable en todo punto  $(x, y)$ . De acuerdo a la proposición 4.4, tenemos además que

$$f'(x, y) = \begin{bmatrix} 2x \sin y & x^2 \cos y \\ y^3 e^{xy} & 2ye^{xy} + xy^2 e^{xy} \\ 2xy & x^2 \end{bmatrix}.$$

Encontremos la aproximación afín de  $f(x, y)$  cerca del punto  $(0, \pi)$ . Esta está dada por

$$T(x, y) = f(0, 0) + f'(0, 0) \begin{bmatrix} x \\ y - \pi \end{bmatrix},$$

esto es,

$$T(x, y) = - \begin{bmatrix} 0 \\ \pi^2 \\ 0 \end{bmatrix} + x \begin{bmatrix} 0 \\ \pi^3 \\ 0 \end{bmatrix} + y \begin{bmatrix} 0 \\ 2\pi \\ 0 \end{bmatrix}.$$



**4.5. Definición de gradiente.** Consideremos el caso de  $f$  a valores reales:  $\Omega \subset \mathbb{R}^N$  un abierto,  $f : \Omega \rightarrow \mathbb{R}$ ,  $x_0 \in \Omega$ ,  $f$  diferenciable en  $x_0$ . En este caso, como sabemos, la derivada de  $f$  en  $x_0$   $f'(x_0)$  es una matriz fila,  $1 \times N$ , y por ello puede identificarse en si misma con un vector de  $\mathbb{R}^N$ . Llamamos a este vector el *gradiente de  $f$  en  $x_0$*  y le denotamos  $\nabla f(x_0)$ . Identificando, como hemos hecho a lo largo de este capítulo, los vectores de  $\mathbb{R}^N$  con las matrices columna, tenemos entonces la relación

$$\nabla f(x_0) = f'(x_0)^T,$$

aunque en realidad no haremos diferencia entre uno y otro si no conlleva ambigüedad. Las coordenadas de este vector son precisamente las derivadas parciales de  $f$  en  $x_0$ . Así,

$$\nabla f(x_0) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x_0) \\ \frac{\partial f}{\partial x_2}(x_0) \\ \vdots \\ \frac{\partial f}{\partial x_N}(x_0) \end{bmatrix}. \quad (4.8)$$

El vector gradiente tiene a su vez una interesante interpretación geométrica. Si  $e$  es tal que  $\|e\| = 1$  entonces

$$f'(x_0; e) = f'(x_0) e = \nabla f(x_0)^T e = \nabla f(x_0) \cdot e,$$

donde  $\cdot$  denota el producto interno canónico. Supongamos que  $\nabla f(x_0) \neq 0$ . Por la desigualdad de Cauchy-Schwartz tenemos entonces que

$$\nabla f(x_0) \cdot e \leq |\nabla f(x_0) \cdot e| \leq \|\nabla f(x_0)\| \|e\| = \|\nabla f(x_0)\|.$$

Escojamos  $e_* = \frac{\nabla f(x_0)}{\|\nabla f(x_0)\|}$ . Entonces

$$\nabla f(x_0) \cdot e_* = \frac{\|\nabla f(x_0)\|^2}{\|\nabla f(x_0)\|} = \|\nabla f(x_0)\|.$$

La conclusión es entonces que

$$f'(x_0; e) \leq f'(x_0; e_*)$$

para toda dirección  $e$ , esto es,  $e_*$ , la dirección del gradiente  $\nabla f(x_0)$ , es aquella de *máximo crecimiento de  $f$* .

**4.6. Plano tangente.** Para una función diferenciable de dos variables, consideremos su gráfico definido del modo siguiente:

$$f : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}$$

Consideremos su grafo, el subconjunto de  $\mathbb{R}^3$  dado por

$$\{(x, y, z) / z = f(x, y), (x, y) \in \Omega\} .$$

Como tenemos que

$$f(x, y) \sim f(x_0, y_0) + \nabla f(x_0, y_0) \cdot (x - x_0, y - y_0),$$

cerca de  $(x_0, y_0)$ , es entonces también el caso que su grafo se asemeja a aquél de la función afín al lado derecho de la expresión anterior. Este último grafo es el conjunto

$$\{(x, y, z) / z = f(x_0, y_0) + \nabla f(x_0, y_0) \cdot (x - x_0, y - y_0)\} ,$$

o

$$\{(x, y, z) / 0 = (\nabla f(x_0, y_0), -1) \cdot (x - x_0, y - y_0, z - f(x_0, y_0))\} ,$$

o

$$f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) + (-1)(z - f(x_0, y_0)) = 0.$$

Este conjunto es un *plano* en  $\mathbb{R}^3$ , que pasa por el punto  $(x_0, y_0, f(x_0, y_0))$  que está en la *superficie* definida por el grafo. Este plano aproximante de la superficie se denomina *plano tangente* al grafo en el punto  $(x_0, y_0, f(x_0, y_0))$ . Notemos que su vector normal está dado por

$$\mathbf{n} = (\nabla f(x_0, y_0), -1) = (f_x(x_0, y_0), f_y(x_0, y_0), -1),$$

lo cual nos entrega otra interpretación geométrica del gradiente: determina el *vector normal* a la superficie dada por el grafo. En modo similar, para una función de  $N$  variables  $f : \Omega \subset \mathbb{R}^N \rightarrow \mathbb{R}$ , su grafo se define como el conjunto de los puntos

$$(x, x_{N+1}) \in \Omega \times \mathbb{R} \subset \mathbb{R}^{N+1}, \quad \text{tales que } x_{N+1} = f(x).$$

El *hiperplano tangente* al grafo en el punto  $(x_0, f(x_0))$  está dado entonces como el conjunto de puntos  $(x, x_{N+1})$  que satisfacen

$$x_{N+1} = f(x_0) + \nabla f(x_0) \cdot (x - x_0),$$

o

$$(f_{x_1}(x_0), \dots, f_{x_N}(x_0), -1) \cdot (x - x_0, x_N - f(x_0)) = 0$$

**Ejemplo 4.7.** Consideremos la esfera en  $\mathbb{R}^3$  dada por el conjunto de puntos  $(x, y, z)$  que satisfacen

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

esto es, la esfera con centro en el origen y radio  $\sqrt{2}$ . Encontremos la ecuación del plano tangente a esta esfera, respectivamente en los puntos

$(0, 1, -1)$  y  $(1, 1, 0)$ . Notemos que los puntos  $(x, y, z)$  de la esfera con  $z \leq 0$  satisfacen

$$z = -\sqrt{2 - x^2 - y^2} =: f(x, y)$$

Así, la esfera corresponde, cerca del punto  $(0, 1, -1)$ , precisamente al grafo de la función  $f(x, y)$ . Calculamos

$$f_x(x, y) = \frac{x}{\sqrt{2 - x^2 - y^2}}, \quad f_y(x, y) = \frac{y}{\sqrt{2 - x^2 - y^2}},$$

de modo que el plano tangente en este punto está dado por la ecuación

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

esto es, el plano en  $\mathbb{R}^3$ ,  $y - z = 2$ .

Si bien el punto  $(1, 1, 0)$  está también en el grafo de la función  $f$  anterior, sus derivadas se hacen infinitas. Podemos sin embargo visualizar la esfera en torno a este punto también como un grafo, pues ésta puede expresarse por la ecuación

$$y = \sqrt{2 - x^2 - z^2} =: g(x, z) .$$

En este caso tenemos

$$f_x(x, z) = -\frac{x}{\sqrt{2 - x^2 - z^2}}, \quad f_z(x, z) = -\frac{z}{\sqrt{2 - x^2 - z^2}},$$

y el plano tangente está entonces dado por la ecuación

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

esto es, el plano  $x + y = 2$ .

**4.7. Teorema del valor medio.** el teorema del valor medio para funciones de una variable admite una generalización a funciones de variables en el modo siguiente.

**Teorema 4.2.** (*Teorema del valor medio*) Sea  $\Omega \subset \mathbb{R}^N$  abierto,  $f : \Omega \rightarrow \mathbb{R}$  diferenciable sobre todo  $\Omega$ . Sean  $x, y \in \Omega$  puntos tales que el segmento entre  $x$  e  $y$  está contenido en  $\Omega$ , esto es

$$[x, y] := \{x + t(y - x) \mid t \in [0, 1]\} \subset \Omega .$$

Existe entonces  $\xi \in ]0, 1[$  tal que

$$f(y) - f(x) = \nabla f(x + \xi(y - x)) \cdot (y - x) .$$

**Demostración.** Consideremos la función  $\varphi : [0, 1] \rightarrow \mathbb{R}$  definida por  $\varphi(t) = f(x + t(y - x))$ . Por el teorema del valor mmedio en  $\mathbb{R}$ , sabemos que

$$\varphi(1) - \varphi(0) = \varphi'(\xi)(1 - 0). \quad (4.9)$$

para cierto  $\xi \in ]0, 1[$ . Por otra parte,

$$\varphi'(\xi) = \frac{d}{dt}\varphi(\xi + t)\big|_{t=0},$$

de modo que

$$\varphi'(\xi) = \frac{d}{dt}f(x + \xi(y - x) + t(y - x))\big|_{t=0}.$$

De acuerdo a la proposición 4.3, tenemos entonces que

$$\varphi'(\xi) = f'(x + \xi(y - x))(y - x) = \nabla f(x + \xi(y - x)) \cdot (y - x).$$

Usando esto y el hecho que  $\varphi(1) = f(y)$ ,  $\varphi(0) = f(x)$ , obtenemos de la igualdad (4.9) la validez del resultado deseado.  $\square$

**4.8. Regla de la cadena.** Como en funciones de una variable, tenemos la validez del hecho que la composición de funciones diferenciables es diferenciable, y de una fórmula para el cálculo de la derivada de esta composición.

**Teorema 4.3.** (*Regla de la cadena*). Sean  $\Omega \subset \mathbb{R}^N$ ,  $\Lambda \subset \mathbb{R}^m$ , abiertos,  $f : \Omega \rightarrow \mathbb{R}^m$ ,  $g : \Lambda \rightarrow \mathbb{R}^k$ . Supongamos que  $f$  es diferenciable en  $x_0$ , que  $f(x) \in \Lambda \ \forall x \in \Omega$ , y que  $g$  es diferenciable en  $f(x_0)$ . Entonces la composición

$$g \circ f : \Omega \subset \mathbb{R}^N \rightarrow \mathbb{R}^k$$

es diferenciable en  $x_0$  y además

$$(g \circ f)'(x_0) = g'(f(x_0)) f'(x_0).$$

**Demostración.** Consideremos

$$\theta(h) = (g \circ f)(x_0 + h) - (g \circ f)(x_0) - g'(f(x_0)) f'(x_0) h.$$

Denotemos

$$q(h) = [f(x_0 + h) - f(x_0)], \quad y_0 = f(x_0).$$

Ciertamente tenemos  $q(h) \rightarrow 0$  cuando  $h \rightarrow 0$ . En realidad, recordemos que, de la relación (4.7) tenemos que existen  $\delta > 0$  y  $C > 0$  tales que para todo  $h$  con  $\|h\| < \delta$ ,

$$\|q(h)\| \leq C\|h\|, \quad (4.10)$$

Por otra parte, la función

$$\beta(k) = \begin{cases} \frac{g(y_0+k)-g(y_0)-g'(y_0)k}{\|k\|} & \text{si } k \neq 0, \\ 0 & \text{si } k = 0, \end{cases}$$

es continua en  $k = 0$  gracias a la diferenciabilidad de  $g$  en  $y_0$ . Podemos escribir entonces

$$\frac{\theta(h)}{\|h\|} = g'(y_0) \frac{f(x_0+h) - f(x_0) - f'(x_0)h}{\|h\|} + \frac{\|q(h)\|}{\|h\|} \beta(g(h)) .$$

El primer termino del lado derecho de la expresión anterior va a cero si  $h \rightarrow 0$  por la diferenciabilidad de  $f$  en  $x_0$ . Como  $g(h) \rightarrow 0$  cuando  $h \rightarrow 0$  y  $\beta$  es continua, tenemos que  $\beta(g(h)) \rightarrow 0$ . Por otra parte, de (4.10), sabemos que para todo  $h$  es suficientemente pequeño  $\frac{\|q(h)\|}{\|h\|} \leq C$ . Deducimos que

$$\frac{\|q(h)\|}{\|h\|} \beta(g(h)) \rightarrow 0 \quad \text{cuando } h \rightarrow 0,$$

y por lo tanto

$$\frac{\|\theta(h)\|}{\|h\|} \rightarrow 0 \quad \text{cuando } h \rightarrow 0,$$

y la demostración está concluida.  $\square$

Seamos más específicos en lo que concierne al cálculo de derivadas parciales.

**Corolario 4.1.** Sean  $f$  y  $g$  como en el teorema 4.3. Escribamos  $f(x) = (f_1(x), \dots, f_m(x))$ . Entonces si

$$h(x) = g(f_1(x), \dots, f_m(x))$$

se tiene que

$$\frac{\partial h}{\partial x_j}(x_0) = \sum_{i=1}^m \frac{\partial g}{\partial y_i}(f(x_0)) \frac{\partial f_i}{\partial x_j}(x_0) . \quad (4.11)$$

**Demostración.** Sabemos que

$$\frac{\partial h}{\partial x_j}(x_0) = h'(x_0) e_j.$$

De acuerdo al teorema 4.3, escribiendo  $y_0 = f(x_0)$ , tenemos entonces que

$$\frac{\partial h}{\partial x_j}(x_0) = g'(y_0) f'(x_0) e_j = g'(y_0) \frac{\partial f}{\partial x_j}(x_0) =$$

$$\begin{aligned}
&= \begin{bmatrix} \frac{\partial g_1}{\partial y_1}(y_0) & \frac{\partial g_1}{\partial y_2}(y_0) & \cdots & \frac{\partial g_1}{\partial y_m}(y_0) \\ \frac{\partial g_2}{\partial y_1}(y_0) & \frac{\partial g_2}{\partial y_2}(y_0) & \cdots & \frac{\partial g_2}{\partial y_m}(y_0) \\ \vdots & \vdots & & \vdots \\ \frac{\partial g_k}{\partial y_1}(y_0) & \frac{\partial g_k}{\partial y_2}(y_0) & \cdots & \frac{\partial g_k}{\partial y_m}(y_0) \end{bmatrix} \begin{bmatrix} \frac{\partial f_1}{\partial x_j}(x_0) \\ \frac{\partial f_2}{\partial x_j}(x_0) \\ \vdots \\ \frac{\partial f_m}{\partial x_j}(x_0) \end{bmatrix} = \\
&\sum_{i=1}^m \frac{\partial f_i}{\partial x_j}(x_0) \begin{bmatrix} \frac{\partial g_1}{\partial y_i}(y_0) \\ \frac{\partial g_2}{\partial y_i}(y_0) \\ \vdots \\ \frac{\partial g_k}{\partial y_i}(y_0) \end{bmatrix} = \sum_{i=1}^m \frac{\partial f_i}{\partial x_j}(x_0) \frac{\partial g}{\partial y_i}(y_0) .
\end{aligned}$$

Esto concluye la demostración.  $\square$

**Ejemplo 4.8.** En aplicaciones de la regla de la cadena, la situación más frecuente que se encuentra es que una de las funciones es explícita y la otra no.

Consideremos por ejemplo una función  $T : \mathbb{R}^2 \rightarrow \mathbb{R}$ ,  $(x, y) \mapsto T(x, y)$ .  $T$  puede representar una cantidad física medida en el plano, digamos temperatura de cada punto, que no conocemos en principio en modo explícito, aunque eventualmente satisface una ecuación que relaciona sus derivadas parciales, esto es una *ecuación diferencial en derivadas parciales*. Por alguna razón asociada al problema específico que se trate, puede ser más conveniente expresar la cantidad representada por  $T$  en términos de un sistema de coordenadas que no sea el Cartesiano  $(x, y)$ . Por ejemplo, un sistema alternativo está constituido por las coordenadas polares  $(r, \theta)$  de modo que  $x = r \cos \theta$ ,  $y = r \sin \theta$ . La cantidad representada por  $T$  expresada en coordenadas  $(r, \theta)$  deviene entonces la función

$$h(r, \theta) = T(r \cos \theta, r \sin \theta),$$

y nos interesa conocer las derivadas parciales de  $h$  en términos de aquellas de  $T$ .  $T$  tiene la forma

$$T(r, \theta) = T(f_1(r, \theta), f_2(r, \theta)).$$

De este modo, aplicamos la fórmula (4.11) y encontramos, bajo las hipótesis de diferenciabilidad requeridas,

$$\frac{\partial T}{\partial r}(r, \theta) = \frac{\partial T}{\partial x}(f_1(r, \theta), f_2(r, \theta)) \frac{\partial f_1}{\partial r}(r, \theta) + \frac{\partial T}{\partial y}(f_1(r, \theta), f_2(r, \theta)) \frac{\partial f_2}{\partial r}(r, \theta),$$

de modo que, haciendo uso de

$$\frac{\partial f_1}{\partial r} = \frac{\partial}{\partial r}(r \cos \theta) = \cos \theta, \quad \frac{\partial f_2}{\partial r} = \frac{\partial}{\partial r}(r \sin \theta) = \sin \theta ,$$

encontramos

$$\frac{\partial h}{\partial r}(r, \theta) = \frac{\partial T}{\partial x}(r \cos \theta, r \sin \theta) \cos \theta + \frac{\partial T}{\partial y}(r \cos \theta, r \sin \theta) \sin \theta .$$

Con un argumento similar, obtenemos que

$$\frac{\partial h}{\partial \theta}(r, \theta) = -\frac{\partial T}{\partial x}(r \cos \theta, r \sin \theta) r \sin \theta + \frac{\partial T}{\partial y}(r \cos \theta, r \sin \theta) r \cos \theta .$$

Despejando, esto conduce también a las fórmulas

$$\frac{\partial T}{\partial x}(r \cos \theta, r \sin \theta) = \frac{\partial h}{\partial r} \cos \theta - \frac{\partial h}{\partial \theta} \frac{\sin \theta}{r} \quad (4.12)$$

$$\frac{\partial T}{\partial y}(r \cos \theta, r \sin \theta) = \frac{\partial h}{\partial r} \sin \theta - \frac{\partial h}{\partial \theta} \frac{\cos \theta}{r} \quad (4.13)$$

**Ejemplo 4.9.** Consideremos la para una función  $T(x, y)$  la ecuación diferencial

$$\left[ \left( \frac{\partial T}{\partial x} \right)^2 + \left( \frac{\partial T}{\partial y} \right)^2 \right](x, y) = \frac{1}{(1 + x^2 + y^2)^2} \quad \forall (x, y) \in \mathbb{R}^2 . \quad (4.14)$$

Se pide encontrar una solución  $T(x, y)$  tal que

$$\lim_{\|(x, y)\| \rightarrow +\infty} T(x, y) = 0. \quad (4.15)$$

En lugar de resolver esta ecuación directamente para  $T$ , consideramos el cambio de variables  $h(r, \theta) = T(r \cos \theta, r \sin \theta)$ . Notemos que  $T$  satisface la ecuación (4.14) si y solo si,

$$\left[ \left( \frac{\partial T}{\partial x} \right)^2 + \left( \frac{\partial T}{\partial y} \right)^2 \right](r \cos \theta, r \sin \theta) = \frac{1}{(1 + r^2)^2} \quad (4.16)$$

para todo  $(r, \theta) \in [0, \infty) \times [0, 2\pi)$ . Así, sustituyendo las expresiones (4.12), (4.13) en (4.16), obtenemos, luego de algunas operaciones,

$$\left( \frac{\partial h}{\partial r} \right)^2(r, \theta) + \frac{1}{r^2} \left( \frac{\partial h}{\partial \theta} \right)^2(r, \theta) = \frac{1}{(1 + r^2)^2}. \quad (4.17)$$

Esta expresión sugiere que el modo más simple de encontrar una solución, es buscar  $h$  independiende de  $\theta$ ,  $h(r, \theta) = g(r)$ . Sustituyendo en (4.17) obtenemos la ecuación para  $g$ ,

$$g'(r)^2 = \frac{1}{(1 + r^2)^2} \quad \forall r \in [0, \infty) . \quad (4.18)$$

Así, encontramos una solución si resolvemos la ecuación

$$g'(r) = \frac{1}{1+r^2} \quad \forall r \in [0, \infty) . \quad (4.19)$$

las soluciones de esta última ecuación diferencial son las primitivas de  $\frac{1}{1+r^2}$ , de este modo, obtenemos las soluciones

$$g(r) = \arctan(r) + C.$$

En terminos de la función original, tenemos entonces

$$T(r \cos \theta, r \sin \theta) = \arctan(r) + C.$$

de modo que,

$$T(x, y) = \arctan(\sqrt{x^2 + y^2}) + C.$$

La condición de anulamiento en  $\infty$  (4.15) nos fuerza a escoger la constante  $C = -\frac{\pi}{2}$  y una solución de (4.14) como se requiere es

$$T(x, y) = \arctan(\sqrt{x^2 + y^2}) - \frac{\pi}{2} .$$

**4.9. Derivadas parciales de orden superior.** Ciertamente, si las derivadas parciales de  $f$  definen en si mismas funciones en  $\Omega$  está la posibilidad de que ellas mismas puedan derivarse. Supongamos que  $\Omega \subset \mathbb{R}^N$  es un abierto,  $f : \Omega \rightarrow \mathbb{R}^m$  y que la derivada parcial

$$\frac{\partial f}{\partial x_j}(x) \quad \text{existe} \quad \forall x \in \Omega.$$

Consideremos la función

$$\frac{\partial f}{\partial x_j} : \Omega \rightarrow \mathbb{R}^m, \quad x \mapsto \frac{\partial f}{\partial x_j}(x) .$$

En caso de existir, la derivada parcial respecto a  $x_i$  en un punto  $x_0$  de esta función se denota del modo siguiente:

$$\frac{\partial}{\partial x_i} \left( \frac{\partial f}{\partial x_j} \right) (x_0) =: \frac{\partial^2 f}{\partial x_i \partial x_j} (x_0).$$

Si  $i = j$ , denotamos también

$$\frac{\partial^2 f}{\partial x_i \partial x_i} (x_0) =: \frac{\partial^2 f}{\partial x_i^2} (x_0).$$

Esta definición se extiende inductivamente a derivadas parciales de cualquier orden  $k$ . Así, si  $(i_1, i_2, \dots, i_s)$  es una  $s$ -tupla de índices  $i_l \in \{1, \dots, N\}$  con  $\sum_{l=1}^s i_l = k$ , denotamos, en caso de existir,

$$\frac{\partial}{\partial x_{i_1}} \left( \frac{\partial}{\partial x_{i_2}} \left( \dots \frac{\partial}{\partial x_{i_{k-1}}} \left( \frac{\partial f}{\partial x_{i_k}} \right) \dots \right) \right) (x_0) =: \frac{\partial^k f}{\partial x_{i_1} \partial x_{i_2} \dots \partial x_{i_s}} (x_0) .$$



Esta es la expresión en notación de Leibnitz. Es también común utilizar la notación

$$\frac{\partial^k f}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_s}}(x_0) =: f_{x_{i_s} x_{i_{s-1}} \cdots x_{i_1}}(x_0).$$

**Ejemplo 4.10.** Consideremos la función  $f(x, y) = e^{xy^2} \sin y$ . Entonces

$$\frac{\partial f}{\partial x} = y^2 e^{xy^2} \sin y, \quad \frac{\partial f}{\partial y} = 2xy e^{xy^2} \sin y + e^{xy^2} \cos y,$$

y tenemos, para las segundas derivadas parciales,

$$\begin{aligned} \frac{\partial^2 f}{\partial x^2} &= y^4 e^{xy^2} \sin y, \\ \frac{\partial^2 f}{\partial y^2} &= e^{xy^2} [(4x^2 y^2 + 2x - 1) \sin y + 4xy \cos y], \\ \frac{\partial^2 f}{\partial y \partial x} &= e^{xy^2} [(2y + 2xy^3) \sin y + y^2 \cos y], \\ \frac{\partial^2 f}{\partial x \partial y} &= e^{xy^2} [(2y + 2xy^3) \sin y + y^2 \cos y]. \end{aligned}$$

Observemos que

$$\frac{\partial^2 f}{\partial y \partial x}(x, y) = \frac{\partial^2 f}{\partial x \partial y}(x, y)$$

para todo  $(x, y)$ . Esto parece una coincidencia ya que los valores de las derivadas sucesivas fueron obtenidos en modos bastante distintos, aun cuando la siempre “provocativa” notación de Leibnitz sugiere que los “diferenciales”  $\partial y$  y  $\partial x$  podrían ser intercambiados. Esto es en realidad cierto en gran generalidad, como enuncia el siguiente resultado.

**Teorema 4.4.** (*Teorema de Schwartz*). Supongamos que  $\Omega \subset \mathbb{R}^N$  es un abierto,  $f : \Omega \rightarrow \mathbb{R}^m$  y que las segundas derivadas parciales

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x), \quad \frac{\partial^2 f}{\partial x_j \partial x_i}(x), \quad \text{existen} \quad \forall x \in \Omega$$

y definen funciones continuas en  $x_0 \in \Omega$ . Entonces,

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x_0).$$

**Demostración.** Supongamos primero que  $N = 2$ ,  $m = 1$ . Consideremos la siguiente cantidad:

$$\theta(h_1, h_2) = f(x_{01} + h_1, x_{02} + h_2) - f(x_{01} + h_1, x_{02}) - f(x_{01}, x_{02} + h_2) + f(x_{01}, x_{02}).$$

Podemos entonces escribir Sea

$$\phi(t) = [f(x_{01} + t, x_{02} + h_2) - f(x_{01}, x_{02} + h_2)] - [f(x_{01} + t, x_{02}) - f(x_{01}, x_{02})].$$

Por el teorema del valor medio de funciones de una variable, tenemos que existe un  $t_h$  entre 0 y  $h_1$  tal que

$$\phi(h_1) - \phi(0) = \phi'(t_h)h_1,$$

lo que quiere decir exactamente

$$\theta(h_1, h_2) = h_1 \left[ \frac{\partial f}{\partial x_1}(x_{01} + t_h, x_{02} + h_2) - \frac{\partial f}{\partial x_1}(x_{01} + t_h, x_{02}) \right]$$

Nuevamente por el teorema del valor medio, existe un valor  $s_h$  entre 0 y  $h_2$  tal que

$$\frac{\partial f}{\partial x_1}(x_{01} + t_h, x_{02} + h_2) - \frac{\partial f}{\partial x_1}(x_{01} + t_h, x_{02}) = \frac{\partial^2 f}{\partial x_2 \partial x_1}(x_{01} + t_h, x_{02} + s_h)h_2,$$

y por lo tanto, como  $(t_h, s_h) \rightarrow (0, 0)$  si  $(h_1, h_2) \rightarrow (0, 0)$ , obtenemos,

$$\lim_{(h_1, h_2) \rightarrow (0, 0)} \frac{\theta(h_1, h_2)}{h_1 h_2} = \lim_{(h_1, h_2) \rightarrow (0, 0)} \frac{\partial^2 f}{\partial x_2 \partial x_1}(x_{01} + t_h, x_{02} + s_h) = \frac{\partial^2 f}{\partial x_2 \partial x_1}(x_{01}, x_{02}),$$

gracias a la continuidad de  $\frac{\partial^2 f}{\partial x_2 \partial x_1}$  en  $(x_{01}, x_{02})$ . Por otra parte, observemos que  $\theta(h_1, h_2) = \theta(h_2, h_1)$ , de modo que el intercambio de nombre de las variables nos permite concluir que también

$$\lim_{(h_1, h_2) \rightarrow (0, 0)} \frac{\theta(h_1, h_2)}{h_1 h_2} = \frac{\partial^2 f}{\partial x_1 \partial x_2}(x_{01}, x_{02}),$$

y el resultado deseado se cumple. Si  $N > 2$ ,  $m = 1$ , el resultado se sigue de aquél para dos variables, pues en la derivación parcial solo las variables  $x_i$  y  $x_j$  están en juego, permaneciendo las otras constantes. Si  $m > 1$ , basta aplicar el resultado a cada una de las funciones coordenadas. Esto concluye la demostración.  $\square$

El resultado anterior se generaliza por inducción a derivadas de mayor orden. Por ejemplo, si las terceras derivadas parciales

$$\frac{\partial^3 f}{\partial x_1 \partial x_2 \partial x_3}(x), \quad \frac{\partial^3 f}{\partial x_2 \partial x_1 \partial x_3}(x), \quad \frac{\partial^3 f}{\partial x_3 \partial x_1 \partial x_2}(x), \quad \frac{\partial^3 f}{\partial x_3 \partial x_2 \partial x_1}(x),$$

existen en todo  $x \in \Omega$  y definen funciones continuas en  $\Omega$  entonces todas coinciden. En efecto, por ejemplo

$$\begin{aligned} \frac{\partial^3 f}{\partial x_2 \partial x_1 \partial x_3}(x) &= \frac{\partial}{\partial x_2} \left( \frac{\partial^2 f}{\partial x_1 \partial x_3} \right)(x) = \\ \frac{\partial}{\partial x_2} \left( \frac{\partial^2 f}{\partial x_3 \partial x_1} \right)(x) &= \frac{\partial^2}{\partial x_2 \partial x_3} \left( \frac{\partial f}{\partial x_1} \right)(x) = \end{aligned}$$

$$\frac{\partial^2}{\partial x_3 \partial x_2} \left( \frac{\partial f}{\partial x_1} \right) (x) = \frac{\partial^3 f}{\partial x_3 \partial x_2 \partial x_1} (x) .$$

En general, si todas las derivadas parciales de orden  $k$  de  $f$  existen en  $\Omega$  y definen funciones continuas, entonces pueden expresarse todas en la forma

$$\frac{\partial^k f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \dots \partial x_N^{\alpha_N}} , \quad \alpha_i \geq 0, \quad \sum_{i=1}^N \alpha_i = k,$$

con la convención que  $\alpha_i = 0$  significa que no hay derivación en la variable  $x_i$ .

Si todas las derivadas parciales de orden  $k$  de  $f$  existen en  $\Omega$  y definen funciones continuas, decimos que la función  $f$  es  $k$  veces continuamente diferenciable en  $\Omega$  o que  $f$  es de clase  $C^k$  en  $\Omega$ . El espacio vectorial de estas funciones se denota  $C^k(\Omega)$ . Si  $f \in C^k(\Omega)$  para todo  $k$  decimos que la función es de clase  $C^\infty(\Omega)$ .

**4.10. Matriz Hessiana.** En el importante caso de funciones a valores reales:  $\Omega \subset \mathbb{R}^N$  abierto,  $f : \Omega \rightarrow \mathbb{R}$  tal que todas sus segundas derivadas parciales existen y son continuas en  $\Omega$  (esto es  $f$  es de clase  $C^2(\Omega)$ ), la noción de *segunda derivada* se extiende del modo siguiente. La función  $\nabla f : \Omega \rightarrow \mathbb{R}^N$ ,  $x \mapsto \nabla f(x)$  es de clase  $C^1$ , y llamamos a su derivada en  $x_0$ , *segunda derivada de  $f$  en  $x_0$* . Así, denotamos naturalmente

$$f''(x_0) = (\nabla f)'(x_0) .$$

$f''(x_0)$  es una matriz cuadrada  $N \times N$ , a la que también se le llama comúnmente *matriz Hessiana de  $f$  en  $x_0$* . Describámosla en modo más explícito. Gracias a la fórmula (4.8), tenemos que

$$\nabla f(x) = f'(x)^T = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \frac{\partial f}{\partial x_2}(x) \\ \vdots \\ \frac{\partial f}{\partial x_N}(x) \end{bmatrix} .$$

y que, por la proposición 4.4

$$f''(x_0) = (\nabla f)'(x_0) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x_0) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(x_0) & \cdots & \frac{\partial^2 f}{\partial x_N \partial x_1}(x_0) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(x_0) & \frac{\partial^2 f}{\partial x_2^2}(x_0) & \cdots & \frac{\partial^2 f}{\partial x_N \partial x_2}(x_0) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_N}(x_0) & \frac{\partial^2 f}{\partial x_2 \partial x_N}(x_0) & \cdots & \frac{\partial^2 f}{\partial x_N^2}(x_0) \end{bmatrix}. \quad (4.20)$$

Entonces

$$f''(x_0) = \left[ \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) \right].$$

Observemos que gracias al Teorema de Schwartz, esta matriz es *simétrica*. A manera de ejemplo, consideremos la función  $f(x, y) = e^{xy^2} \sin y$  del ejemplo 4.10. Entonces,

$$f''(x, y) = e^{xy^2} \begin{bmatrix} y^4 \sin y & (2y + 2xy^3) \sin y + y^2 \cos y \\ (2y + 2xy^3) \sin y + y^2 \cos y & (4x^2 y^2 + 2x - 1) \sin y + 4xy \cos y \end{bmatrix}.$$

**4.11. El Teorema de Taylor.** El Teorema de Taylor para funciones de una variable admite una extensión al contexto presente. Recordemos que si  $\phi : ]a, b[ \rightarrow \mathbb{R}$  es derivable  $m$  veces en  $]a, b[$  y  $x_0, x_0 + h \in ]a, b[$ , entonces vale la siguiente expresión, extensión del teorema del valor medio.

$$f(x_0 + h) = \sum_{k=0}^{m-1} f^{(k)}(x_0) \frac{h^k}{k!} + f^{(m)}(x_0 + \xi h) \frac{h^m}{m!}$$

donde  $\xi$  depende de  $h$  y  $\xi \in ]0, 1[$ . Sean

$$T_k(h) = f^{(k)}(x_0) \frac{h^k}{k!}, \quad R_m(h) = f^{(m)}(x_0 + \xi h) \frac{h^m}{m!}.$$

Usamos aquí la convención  $T_0(h) = f(x_0)$ . Así,

$$f(x_0 + h) = \sum_{k=0}^{m-1} T_k(h) + R_m(h). \quad (4.21)$$

El *polinomio de Taylor* en  $h$  de grado  $m - 1$ ,

$$P_{m-1}(h) = \sum_{k=0}^{m-1} T_k(h)$$

es una aproximación de  $f(x_0 + h)$  que para  $h$  pequeño deja un resto  $R_m(h)$  de tamaño comparable a  $|h|^m$ . Extenderemos la fórmula (4.21) al caso de varias variables, donde  $T_k(h)$  es un polinomio en  $h = (h_1, \dots, h_N)$  de grado  $k$ , que además es *homogéneo*, lo que quiere decir que  $T_k(th) =$

$t^k T_k(h)$  para todo  $t$ . En otras palabras, todos los términos de este polinomio tienen grado exactamente  $k$ . Sus coeficientes están determinados por las derivadas parciales de orden  $k$  de  $f$ .

**Teorema 4.5.** (*Teorema de Taylor*) Sea  $f : \Omega \subset \mathbb{R}^N \rightarrow \mathbb{R}$ ,  $\Omega$  abierto. Supongamos que  $f$  es de clase  $C^m$  en  $\Omega$ ,  $m \geq 1$ . Sean  $x_0 \in \Omega$ ,  $h$  tal que  $x_0 + th \in \Omega$  para todo  $t \in [0, 1]$ .

Vale entonces la siguiente expansión.

$$f(x_0 + h) = \sum_{k=0}^{m-1} T_k(h) + R_m(h) , \quad (4.22)$$

donde  $T_0(h) = f(x_0)$ , y para  $k \geq 1$ ,

$$T_k(h) = \sum_{i_1=1}^N \sum_{i_2=1}^N \cdots \sum_{i_k=1}^N \frac{\partial^k f}{\partial x_{i_1} \cdots \partial x_{i_k}}(x_0) h_{i_1} \cdots h_{i_k} , \quad (4.23)$$

$$R_m(h) = \frac{1}{m!} \sum_{i_1=1}^N \sum_{i_2=1}^N \cdots \sum_{i_m=1}^N \frac{\partial^m f}{\partial x_{i_1} \cdots \partial x_{i_m}}(x_0 + \xi h) h_{i_1} \cdots h_{i_m} ,$$

con  $\xi = \xi_h \in ]0, 1[$ .

**Demostración.** Consideremos la función de una variable  $\phi(t) = f(x_0 + th)$ . El teorema de Taylor en una variable nos dice que

$$\phi(1) = \phi(0) + \sum_{k=1}^{m-1} \frac{1}{k!} \phi^{(k)}(0) + \frac{1}{m!} \phi^{(m)}(\xi) \quad (4.24)$$

con  $\xi \in ]0, 1[$ . Tenemos que  $\phi(1) = f(x_0 + h)$ ,  $\phi(0) = f(x_0)$ . Investiguemos las derivadas de  $\phi$ . Tenemos que

$$\phi(t) = f(x_{01} + th_1, \dots, x_{0N} + th_N).$$

Así, por la regla de la cadena tenemos que

$$\phi'(t) = \sum_{i_1=1}^N f_{x_{i_1}}(x_{01} + th_1, \dots, x_{0N} + th_N) h_{i_1}.$$

Derivando esta expresión una vez más encontramos

$$\begin{aligned} \phi''(t) &= \sum_{i_1=1}^N \frac{d}{dt} \frac{\partial f}{\partial x_{i_1}}(x_{01} + th_1, \dots, x_{0N} + th_N) h_{i_1} = \\ &= \sum_{i_1=1}^N \sum_{i_2=1}^N \frac{\partial^2 f}{\partial x_{i_2} \partial x_{i_1}}(x_{01} + th_1, \dots, x_{0N} + th_N) h_{i_1} h_{i_2}. \end{aligned}$$

Iteramos este procedimiento y encontramos

$$\phi'''(t) = \sum_{i_1=1}^N \sum_{i_2=1}^N \sum_{i_3=1}^N \frac{\partial^3 f}{\partial x_{i_3} \partial x_{i_2} \partial x_{i_1}}(x_{01} + th_1, \dots, x_{0N} + th_n) h_{i_1} h_{i_2} h_{i_3} .$$

Continuando, encontramos en general

$$\phi^{(k)}(t) = \sum_{i_1=1}^N \sum_{i_2=1}^N \cdots \sum_{i_k=1}^N \frac{\partial^k f}{\partial x_{i_1} \cdots \partial x_{i_k}}(x_0 + th) h_{i_1} \cdots h_{i_k} .$$

El resultado deseado se obtiene entonces de inmediato a partir de la expansión (4.24).  $\square$

Analicemos explícitamente el caso especial del segundo orden  $m = 2$ . En este caso la fórmula (4.22)-(4.23) se reduce simplemente a

$$f(x_0 + h) = f(x_0) + \sum_{i=1}^N \frac{\partial f}{\partial x_i}(x_0) h_i + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0 + \xi h) h_i h_j .$$

Recordando las fórmulas que definen el vector gradiente  $\nabla f(x_0)$  en (4.8) y la matriz Hessiana  $f''(x_0)$  en (4.20), obtenemos el teorema de Taylor para una función  $f : \Omega \rightarrow \mathbb{R}$  de clase  $C^2(\Omega)$ ,

$$f(x_0 + h) = f(x_0) + \nabla f(x_0) \cdot h + \frac{1}{2} h^T f''(x_0 + \xi h) h \quad (4.25)$$

Para el caso de una función de dos variables, es posible expresar la fórmula de Taylor (4.22)-(4.23) en un modo más eficiente. Observemos que el  $k$ -ésimo termino en la expansión queda en tal caso dado por

$$T_k(h) = \frac{1}{k!} \sum_{i_1=1}^2 \sum_{i_2=1}^2 \cdots \sum_{i_k=1}^2 \frac{\partial^k f}{\partial x_{i_1} \cdots \partial x_{i_k}}(x_0) h_{i_1} \cdots h_{i_k} .$$

Como la derivación parcial en distintas variables conmuta, muchos términos en la expresión anterior coinciden. Es conveniente reescribir esta expresión en el modo siguiente:

$$T_k(h) = \frac{1}{k!} \left[ \sum_{i_1=1}^2 \sum_{i_2=1}^2 \cdots \sum_{i_k=1}^2 \left( h_{i_1} \frac{\partial}{\partial x_{i_1}} \right) \left( h_{i_2} \frac{\partial}{\partial x_{i_2}} \right) \cdots \left( h_{i_k} \frac{\partial}{\partial x_{i_k}} \right) \right] f(x_0) .$$

Los operadores diferenciales en esta expresión conmutan, por lo que pueden reagruparse en su aplicación tal como lo haríamos con el producto de números. Recordemos que (para números) tenemos la forma

del binomio

$$(a_1 + a_2)^m = \sum_{i_1=1}^2 \sum_{i_2=1}^2 \cdots \sum_{i_k=1}^2 a_{i_1} \cdots a_{i_k} = \sum_{j=0}^k \binom{k}{j} a_1^j a_2^{k-j},$$

$$\binom{k}{j} = \frac{k!}{(k-j)!j!},$$

y obtenemos entonces en modo similar

$$T_k(h) = \frac{1}{k!} \left( h_1 \frac{\partial}{\partial x_1} + h_2 \frac{\partial}{\partial x_2} \right)^k f(x_0),$$

lo que quiere decir exactamente

$$T_k(h) = \frac{1}{k!} \sum_{j=0}^k \binom{k}{j} \frac{\partial^k f}{\partial x_1^j \partial x_2^{k-j}}(x_0) h_1^j h_2^{k-j},$$

Así, para una función de dos variables, la fórmula de Taylor (4.22) puede escribirse compactamente como

$$f(x_0 + h) = \sum_{k=0}^{m-1} \sum_{j=0}^k \frac{h_1^j h_2^{k-j}}{(k-j)!j!} \frac{\partial^k f}{\partial x_1^j \partial x_2^{k-j}}(x_0) + R_m(h), \quad (4.26)$$

$$R_m(h) = \sum_{j=0}^m \frac{h_1^j h_2^{m-j}}{(m-j)!j!} \frac{\partial^m f}{\partial x_1^j \partial x_2^{m-j}}(x_0).$$

Para el caso de una función de  $N$  variables,  $N > 2$ , es posible también obtener expresiones más “económicas” para el Teorema de Taylor. Puede decirse en general, que el término  $T_k(h)$  puede ser expresado operacionalmente como

$$T_k(h) = \frac{1}{k!} \left( h_1 \frac{\partial}{\partial x_1} + h_2 \frac{\partial}{\partial x_2} + \cdots + h_N \frac{\partial}{\partial x_N} \right)^k f(x_0),$$

y puede recurrirse a la formula del multinomio para expresar en modo más explícito esta cantidad.

**Ejemplo 4.11.** Consideremos la función

$$f(x, y) = x^2 y + x \cos y$$

y calculemos su expansión de Taylor de orden 3 en torno al punto  $x_0 = (1, 0)$ . Tenemos:

$$f_x = 2xy + \cos y, \quad f_y = x^2 - x \sin y,$$

$$f_{xy} = f_{yx} = 2x - \sin y, \quad f_{xx} = 2y, \quad f_{yy} = -x \cos y,$$

$$f_{xxy} = 0, \quad f_{xyy} = -\cos y, \quad f_{xxx} = 0, \quad f_{yyy} = x \sin y$$

Tenemos entonces, en virtud de la fórmula (4.25),

$$f(1+h, 0+k) = 1 + (1)h + (1)k + \frac{1}{2}((0)h^2 + 2(2)hk + (-1)k^2) + \frac{1}{6}((0)h^3 + 3(0)h^2k + 3(-\cos(\xi k)hk^2 + (1+\xi h)\sin(\xi k)k^3), \quad \xi \in ]0, 1[.$$

En conclusión, tenemos la validez de la fórmula

$$f(1+h, k) = 1 + h + k + 2hk - \frac{1}{2}k^2 - \frac{1}{2}\cos(\xi k)hk^2 + \frac{1}{6}(1+\xi h)\sin(\xi k)k^3,$$

para cierto  $\xi \in ]0, 1[$  dependiente de  $h$  y  $k$ .

## 5. PUNTOS CRÍTICOS DE FUNCIONES DIFERENCIABLES

En este capítulo consideramos una función  $f : \Omega \subset \mathbb{R}^N \rightarrow \mathbb{R}$  con  $\Omega$  un conjunto abierto.

La definición básica de interés en lo que sigue es la de punto crítico de  $f$ . Decimos que  $x_0 \in \Omega$  es un *punto crítico de  $f$*  si  $f$  es diferenciable en  $x_0$  y

$$\nabla f(x_0) = 0.$$

Para una función de dos variables  $f(x, y)$ , este hecho se interpreta geométricamente como el hecho que el plano tangente al grafo de  $f$ ,  $z = f(x, y)$  en el punto  $(x_0, y_0, f(x_0, y_0))$  es *horizontal*. En efecto, el vector normal a este plano es  $(\nabla f(x_0, y_0), -1)$ , esto es  $(0, 0, -1)$ , que tiene la dirección del eje  $z$ .

De gran importancia son los casos de un mínimo y un máximo locales de  $f$ . Decimos que  $x_0 \in \Omega$  es un *mínimo local*, si existe  $\delta > 0$  tal que la bola  $B(x_0, \delta) \subset \Omega$  y

$$f(x_0) \leq f(x) \quad \forall x \in B(x_0, \delta), \quad \text{es decir } f(x_0) = \min_{x \in B(x_0, \delta)} f(x).$$

Similarmente, decimos que  $x_0 \in \Omega$  es un *máximo local de  $f$* , si existe  $\delta > 0$  tal que la bola  $B(x_0, \delta) \subset \Omega$  y

$$f(x_0) \geq f(x) \quad \forall x \in B(x_0, \delta), \quad \text{es decir } f(x_0) = \max_{x \in B(x_0, \delta)} f(x).$$

Mínimo local y máximo local se dicen *estrictos* si estas desigualdades son estrictas para  $x \neq x_0$ .

Como en el caso de funciones de una variable, máximos y mínimos locales son puntos críticos de  $f$ .

**Proposición 5.1.** *Sea  $f : \Omega \subset \mathbb{R}^N \rightarrow \mathbb{R}$ ,  $\Omega$  un abierto. Supongamos que  $x_0$  es un mínimo local de  $f$  y que  $f$  es diferenciable en  $x_0$ . Entonces  $\nabla f(x_0) = 0$ . Lo mismo se tiene si  $x_0$  es un máximo local.*



**Demostración.** Supongamos que  $f$  tiene un mínimo local en  $x_0$ , digamos para cierto  $\delta > 0$ ,

$$f(x_0) \leq f(x) \quad \forall x \in B(x_0, \delta).$$

Para  $j \in \{1, \dots, N\}$ , consideremos la función

$$t \in ]-\delta, \delta[ \mapsto \phi(t) := f(x_0 + te_j) .$$

Entonces  $\phi(t) \geq \phi(0)$  para todo  $t \in ]-\delta, \delta[$ . Por lo tanto

$$\frac{\phi(t) - \phi(0)}{t} \geq 0 \quad \forall t \in ]0, \delta[ ,$$

y entonces

$$\phi'(0) = \lim_{t \rightarrow 0^+} \frac{\phi(t) - \phi(0)}{t} \geq 0.$$

Del mismo modo,

$$\frac{\phi(t) - \phi(0)}{t} \leq 0 \quad \forall t \in ]-\delta, 0[ ,$$

y entonces  $\phi'(0) \leq 0$ , esto es  $\phi'(0) = 0$ . Pero, por definición,

$$0 = \phi'(0) = \frac{\partial f}{\partial x_i}(x_0) .$$

Como esto se tiene para todo  $i = 1, \dots, N$ , concluimos que  $\nabla f(x_0) = 0$ .

Para el caso en que  $x_0$  sea un máximo local, nos basta observar que la función  $-f$  tiene un mínimo local en  $x_0$ . Por lo tanto  $\nabla(-f)(x_0) = -\nabla f(x_0) = 0$ .  $\square$

Es importante discriminar si un eventual punto crítico de  $f$  se trata de un máximo local, un mínimo local, o ninguno de estos tipos. Como en el caso una variable, tenemos condiciones sobre la segunda derivada. Recordemos por ejemplo que si  $x_0$  es un punto crítico de una función de una variable dos veces derivable, que es a su vez un mínimo local, entonces  $f''(x_0) \geq 0$ . Recíprocamente, si  $f''(x_0) > 0$ , el punto crítico es un mínimo local. En el contexto de varias variables,  $f''(x_0)$  es una matriz  $N \times N$ , cabe entonces preguntarse si alguna noción análoga a positividad de esta matriz juega un rol similar. De hecho, este es el caso.

Sea  $A$  una matriz  $N \times N$ . Decimos que  $A$  es *semidefinida positiva* si

$$\forall h \in \mathbb{R}^N \quad h^T A h \geq 0 .$$

Por otra parte, decimos que  $A$  es *definida positiva* si

$$\forall h \in \mathbb{R}^N \setminus \{0\} \quad h^T A h > 0 .$$

Decimos que  $A$  es *semidefinida negativa*, respectivamente *definida negativa*, si la matriz  $-A$  es semidefinida positiva, respectivamente definida positiva.

Si  $A$  es simétrica, la positividad y semipositividad está vinculada a sus valores propios. Recordemos que si  $A = A^T$ , entonces existe una base ortonormal de  $\mathbb{R}^N$  constituida por vectores propios, digamos  $\{v_1, \dots, v_N\}$  con

$$Av_i = \lambda_i v_i.$$

$\lambda_1, \dots, \lambda_N$  son los valores propios de  $A$  (posiblemente repetidos). Como  $A$  es simétrica, éstos son todos reales. Notemos que

$$\lambda_i = \lambda_i \|v_i\|^2 = v_i^T A v_i, \quad \forall i = 1, \dots, N. \quad (5.27)$$

Todo  $h \in \mathbb{R}^N \setminus \{0\}$  puede escribirse en la forma

$$h = \sum_{i=1}^N \alpha_i v_i,$$

Como los  $v_i$  constituyen una base ortonormal se tiene también que

$$\|h\|^2 = \sum_{i=1}^N \alpha_i^2. \quad (5.28)$$

Observemos que

$$h^T A h = \left[ A \left( \sum_{i=1}^N \alpha_i v_i \right) \right] \cdot \left( \sum_{j=1}^N \alpha_j v_j \right) = \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j (A v_i) \cdot v_j.$$

Ahora,

$$(A v_i) \cdot v_j = \lambda_i v_i \cdot v_j = \lambda_i \delta_{ij}$$

donde

$$\delta_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

y por lo tanto

$$h^T A h = \sum_{i=1}^N \lambda_i \alpha_i^2. \quad (5.29)$$

De las relaciones (5.29) y (5.28) se sigue que

$$h^T A h \geq \beta \|h\|^2, \quad \beta = \min_{i=1, \dots, N} \lambda_i. \quad (5.30)$$

Deducimos entonces que si  $\lambda_i > 0$  para todo  $i$ , entonces  $h^T A h > 0$ . Como  $h$  es arbitrario, entonces  $A$  es definida positiva.

Recíprocamente, si  $A$  es definida positiva, (5.27) implica que  $\lambda_i > 0$  para todo  $i$ . En modo similar se tiene que  $A$  es semidefinida positiva si y solo si sus valores propios son todos  $\geq 0$ .

Lo anterior se aplica en particular a la matriz Hessiana  $f''(x_0)$  en caso que las segundas derivadas parciales de  $f$  sean continuas en  $x_0$ .

**Teorema 5.1.** (*Optimalidad y segundo orden*). Sea  $f : \Omega \subset \mathbb{R}^N \rightarrow \mathbb{R}$  una función de clase  $C^2(\Omega)$ ,  $\Omega$  un abierto, y  $x_0 \in \Omega$  un punto crítico de  $f$ . Se tienen entonces la validez de las siguientes afirmaciones:

(a) Si  $x_0$  es un mínimo local de  $f$  entonces la matriz simétrica  $f''(x_0)$  es semidefinida positiva. Si  $x_0$  es un máximo local, entonces  $f''(x_0)$  es semidefinida negativa.

(b) Si  $f''(x_0)$  es definida positiva, entonces  $x_0$  es un mínimo local estricto de  $f$ . Del mismo modo,  $x_0$  es un máximo local estricto si  $f''(x_0)$  es definida negativa.

**Demostración.** Sea  $\phi(t) = f(x_0 + th)$ . Si  $x_0$  es un mínimo local, entonces  $\phi'(0) = \nabla f(x_0) \cdot h = 0$  y entonces para todo  $t$  pequeño.

$$0 \leq \phi(t) - \phi(0) - \phi'(0)t.$$

Entonces, gracias a la regla de l'hôpital tenemos que

$$0 \leq \lim_{t \rightarrow 0} \frac{\phi(t) - \phi(0) - \phi'(0)t}{t^2} = \lim_{t \rightarrow 0} \frac{\phi'(t) - \phi'(0)}{2t} = \frac{1}{2}\phi''(0).$$

y entonces  $\phi''(0) \geq 0$ . Como hemos calculado en la deducción de la fórmula (4.25), se tiene que

$$\phi''(0) = h^T f''(x_0)h$$

y como  $h$  es arbitrario, se sigue que  $f''(x_0)$  es semidefinida positiva. Hemos probado (a) para un mínimo local. La aseveración correspondiente a máximo local se sigue aplicando este resultado a la función  $-f$ .

Probemos (b). Como  $\nabla f(x_0) = 0$  tenemos de la fórmula de Taylor de segundo orden (4.25) que

$$f(x_0 + h) - f(x_0) = \frac{1}{2}h^T f''(x_0 + \xi_h h)h \quad (5.31)$$

con  $\xi_h \in ]0, 1[$ . Supongamos que  $f''(x_0)$  es definida positiva. Se sigue entonces de la relación (5.30) que existe  $\beta > 0$  tal que para todo  $h \in \mathbb{R}^N$ ,

$$h^T f''(x_0)h \geq \beta \|h\|^2. \quad (5.32)$$

Ahora, para una matriz cualquiera  $B$   $N \times N$  se tiene que

$$|h^T B h| = \sum_{i,j=1}^N |B_{ij}| |h_i| |h_j| \leq \left( \sum_{i,j} |B_{ij}| \right) \|h\|^2$$

Por lo tanto

$$|h^T [f''(x_0 + \xi_h h) - f''(x_0)] h| \leq \left( \sum_{i,j} |f_{x_i x_j}(x_0 + \xi_h h) - f_{x_i x_j}(x_0)| \right) \|h\|^2. \quad (5.33)$$

Como la función

$$\psi(k) = \sum_{i,j} |f_{x_i x_j}(x_0 + k) - f_{x_i x_j}(x_0)|$$

es continua en  $k = 0$ , y  $\psi(0) = 0$ , tenemos que existe  $\delta > 0$  tal que para todo  $k$  con  $\|k\| < \delta$  se tiene que

$$\psi(k) \leq \frac{\beta}{2}$$

donde  $\beta > 0$  es el número en (5.32). Por lo tanto, como  $\xi_h \in ]0, 1[$ , se sigue que si  $\|h\| < \delta$ ,

$$\sum_{i,j} |f_{x_i x_j}(x_0 + \xi_h h) - f_{x_i x_j}(x_0)| \leq \frac{\beta}{2}.$$

De (5.33) obtenemos entonces que

$$|h^T [f''(x_0 + \xi_h h) - f''(x_0)] h| \leq \frac{\beta}{2} \|h\|^2.$$

De aquí, y las relaciones (5.31) y (5.32), se sigue que

$$f(x_0 + h) - f(x_0) \geq \frac{\beta}{4} \|h\|^2 \quad \forall h \in B(0, \delta).$$

Por lo tanto  $f$  tiene un mínimo local estricto en  $x_0$ . La afirmación para máximo se sigue a partir de aquella correspondiente a  $-f$ .  $\square$

**Ejemplo 5.1.** Consideremos la función  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  dada por

$$f(x, y) = y^3 + x^2 + 4y^2 - 4x + 5y + 10$$

Busquemos los puntos críticos de esta función. Tenemos

$$f_x(x, y) = 2x - 4, \quad f_y(x, y) = 3y^2 + 8y + 5 = (y + 1)(3y + 5)$$

De este modo,  $\nabla f(x, y) = (0, 0)$  si y solo si

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

Así, los puntos críticos de  $f$  son  $(2, -1)$  y  $(2, -\frac{5}{3})$ . Calculamos también

$$f_x(x, y) = 2, \quad f_{xy}(x, y) = 0, \quad f_{yy}(x, y) = 6y + 8$$

Tenemos entonces que

$$f''(x, y) = \begin{bmatrix} 2 & 0 \\ 0 & 6y + 8 \end{bmatrix}.$$

Por lo tanto,

$$f''(2, -1) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \quad f''(2, -\frac{5}{3}) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}.$$

Como  $f''(2, -1)$  es una matriz diagonal, sus valores propios son precisamente las entradas de esta:  $\lambda_1 = 2$ ,  $\lambda_2 = 2$ . Como ambos positivos, concluimos que el punto  $(2, -1)$  es un mínimo local de  $f$ . En cambio los valores propios de  $f''(2, -\frac{5}{3})$ ,  $\lambda_1 = 2$ ,  $\lambda_2 = -2$ , tienen signos opuestos, de modo que el punto crítico no es ni un mínimo ni un máximo. Precisemos un poco más el comportamiento de la función  $f$  cerca de este punto. Observemos que

$$\begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = (-2) \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

de modo que los vectores propios respectivamente asociados a 2 y  $-2$  son

$$e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{y} \quad e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Estas son las, así llamadas, *direcciones principales* de  $f$  en el punto crítico  $(2, -\frac{5}{3})$ . Observemos que si

$$\phi_1(t) = f((2, -\frac{5}{3}) + te_1), \quad \phi_2(t) = f((2, -\frac{5}{3}) + te_2),$$

entonces, evaluamos directamente

$$\phi_1'(t) = f_x((2, -\frac{5}{3}) + te_1) = 2t, \quad \phi_2'(t) = f_y((2, -\frac{5}{3}) + te_1) = (-2 + 3t)t,$$

$$\phi_1''(t) = 2, \quad \phi_2''(t) = -2 + 6t.$$

Vemos entonces que en la dirección de  $e_1$  la función  $f$  tiene segunda derivada positiva, esto es, es convexa, exhibiendo un mínimo en  $t = 0$ . En la dirección de  $e_2$  en cambio, vemos que la segunda derivada es negativa para todo  $t$  cercano a 0, siendo la función en esta dirección, maximizándose en  $t = 0$ . La forma del gráfico de  $f$  en torno a este punto se denomina *punto silla*, en referencia a la forma de una silla de montar.

Diremos en general, que para  $f : \Omega \subset \mathbb{R}^N \rightarrow \mathbb{R}$  de clase  $C^2$ , un punto crítico  $x_0$  es un punto silla, si todos los valores propios de  $f''(x_0)$  son distintos de cero, y hay presentes valores propios positivos y negativos. Los vectores propios asociados a valores propios negativos corresponden a direcciones en las cuales la función *baja* a partir de  $x_0$  (hacia ambos lados), mientras que en las direcciones complementarias, las de los vectores propios asociados a valores propios positivos, la función *sube*.

**Ejemplo 5.2.** Sea

$$f(x, y, z) = 3x^2 - 6x + 3y^2 - 6y - 2xy + (z^2 - 1)^2.$$

En este caso  $(x, y, z)$  es un punto crítico de  $f$  si y solo si tenemos

$$\begin{aligned} f_x(x, y, z) &= 6x - 6 - 2y = 0, & f_y(x, y, z) &= 6y - 6 - 2x = 0, \\ f_z(x, y, z) &= 4(z^2 - 1)z = 0. \end{aligned}$$

Este sistema tiene tres soluciones, lo que nos conduce a la presencia de tres puntos críticos:

$$(1, 1, 0), \quad (1, 1, 1), \quad (1, 1, -1),$$

La matriz Hessiana de  $f$  está dada por

$$f''(x, y, z) = \begin{bmatrix} 6 & -2 & 0 \\ -2 & 6 & 0 \\ 0 & 0 & 12z^2 - 4 \end{bmatrix}.$$

Calculemos los valores propios de  $f''(1, 1, 0)$ . Estos corresponden a las raíces del polinomio característico

$$\begin{vmatrix} 6 - \lambda & -2 & 0 \\ -2 & 6 - \lambda & 0 \\ 0 & 0 & -4 - \lambda \end{vmatrix} = -((\lambda - 6)^2 - 4)(\lambda + 4),$$

que son

$$\lambda_1 = 8, \quad \lambda_2 = 4, \quad \lambda_3 = -4.$$

Como tenemos valores positivos y negativos, concluimos que  $(1, 1, 0)$  es un punto silla.

Similarmente, encontramos que los valores propios de  $f''(1, 1, 1)$  y de  $f''(1, 1, -1)$  están dados, en ambos casos por

$$\lambda_1 = 8, \quad \lambda_2 = 4, \quad \lambda_3 = 8.$$

Siendo estos tres valores positivos, concluimos que  $(1, 1, 1)$  y  $(1, 1, -1)$  son mínimos locales.

La pregunta surge naturalmente, tanto en este ejemplo como en el anterior, de si los puntos de mínimo local encontrados son en realidad mínimos globales. Por cierto, la pregunta básica es la de si un mínimo global de la función  $f$  en realidad existe.

Para  $f(x, y, z) = 3x^2 - 6x + 3y^2 - 6y - 2xy + (z^2 - 1)^2$  tenemos que

$$f(x, y, z) = 2(x^2 + y^2) + (x - y)^2 - 6x - 6y + (z^2 - 1)^2.$$

Por la relación  $a^2 \pm 2ab + b^2 \geq 0$ , obtenemos las desigualdades

$$(z^2 - 1)^2 \geq 2(z^2 - 1) - 1, \quad -6x \geq -x^2 - 9, \quad -6y \geq -y^2 - 9,$$

de modo que

$$f(x, y, z) \geq x^2 + y^2 + z^2 - 21.$$

De este modo,

$$f(x, y, z) \rightarrow +\infty \quad \text{si} \quad \|(x, y, z)\| = \sqrt{x^2 + y^2 + z^2} \rightarrow +\infty.$$

Como  $f$  es obviamente continua, el teorema 3.2 nos garantiza que un punto de mínimo (global) de  $f$  en efecto existe. Este punto de mínimo debe ser un punto crítico, que a su vez debe ser un mínimo local, por ende la matriz Hessiana de  $f$  en este punto debe ser semidefinida positiva. La única posibilidad es entonces que los puntos Sin embargo, vemos que  $f(1, 1, 1) = f(1, 1, -1)$ , por ende *ambos* son puntos de mínimo global de  $f$ . Tenemos además que

$$\min_{\mathbb{R}^3} f = f(1, 1, 1) = -8.$$

Consideremos ahora la función  $f$  del ejemplo 5.1,  $f(x, y) = y^3 + x^2 + 4y^2 - 4x + 5y + 10$ . Aquí encontramos que el punto  $(2, -1)$  es un mínimo local. Nos preguntamos si se trata de un mínimo global. En este caso la respuesta es **no**. En efecto, no existe mínimo global de  $f$  pues, por ejemplo, a lo largo de la sucesión  $(0, -n)$  tenemos

$$f(0, -n) = -n^3 + 4n^2 - 5n + 10 \rightarrow -\infty \quad \text{si} \quad n \rightarrow \infty.$$

Esto implica que la función  $f$  no es acotada por abajo, y un valor mínimo absoluto por cierto no puede existir.

## 6. LOS TEOREMAS DE LA FUNCIÓN IMPLÍCITA E INVERSA

Consideremos una función  $F : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}$  de clase  $C^1$ . Un problema natural es aquél de entender la estructura del conjunto de puntos, en  $\mathbb{R}^2$  que satisfacen la ecuación  $F(x, y) = 0$ . Como ya hemos discutido antes, se espera muchas veces que esta relación defina una "curva", lo que hemos entendido como el hecho que *localmente*, esto es en una

vecindad de cada uno de sus puntos, la relación en realidad puede describirse como el gráfico de una función de una variable, ya sea  $y$  como función de  $x$ , o  $x$  como función de  $y$ . Supongamos que éste es el caso, en torno a un punto  $(x_0, y_0)$  de la relación, esto es, existe una función  $y = \phi(x)$  cuyo gráfico la describe en una vecindad de  $(x_0, y_0)$ , esto es,  $\phi(x_0) = y_0$  y existe  $\delta > 0$  tal que

$$F(x, \phi(x)) = 0 \quad \text{para todo } x \in ]x_0 - \delta, x_0 + \delta[.$$

Supongamos que  $\phi(x)$  es diferenciable. Entonces, por la regla de la cadena,

$$\frac{\partial F}{\partial x}(x, \phi(x)) + \frac{\partial F}{\partial y}(x, \phi(x))\phi'(x) = 0 \quad \text{para todo } x \in ]x_0 - \delta, x_0 + \delta[.$$

En particular en  $x = x_0$  obtenemos

$$\phi'(x_0) = - \left( \frac{\partial F}{\partial x}(x_0, y_0) \right)^{-1} \frac{\partial F}{\partial y}(x_0, y_0)$$

por cierto siempre que se tenga  $\frac{\partial F}{\partial x}(x_0, y_0) \neq 0$ . El Teorema de la Función implícita es una suerte de recíproca de esta afirmación: Si  $\frac{\partial F}{\partial y}(x_0, y_0) \neq 0$  entonces una función  $\phi(x)$  con las características antes mencionadas en efecto existe. Probaremos ésto en realidad para funciones de varias variables del tipo

$$F : \Omega \subset \mathbb{R}^N \times \mathbb{R}^m \rightarrow \mathbb{R}^m, \quad (x, y) \mapsto F(x, y)$$

de clase  $C^1$ , solo que en este caso la condición  $\frac{\partial F}{\partial y}(x_0, y_0) \neq 0$  debe reemplazarse por la *invertibilidad* de la matriz derivada de  $F$  tomada parcialmente respecto a  $y$ ,  $m \times m$ . En efecto, si  $F(x_0, y_0) = 0$  y hay una función  $C^1$   $y = \phi(x)$  tal que  $F(x, \phi(x)) = 0$  para todo  $x$  en una vecindad de  $x_0$  entonces la regla de la cadena nos dice que

$$F'(x, \phi(x)) \begin{bmatrix} I_{N \times N} \\ \phi'(x)_{m \times N} \end{bmatrix} = 0,$$

donde  $I$  es la matriz identidad. Obtenemos entonces que

$$F'(x_0, y_0) \begin{bmatrix} I_{N \times N} \\ \phi'(x_0)_{m \times N} \end{bmatrix} = 0.$$

Denotemos entonces

$$F_y(x_0, y_0) = \left[ \frac{\partial f}{\partial y_1}(x_0, y_0) \cdots \frac{\partial f}{\partial y_m}(x_0, y_0) \right]$$

y análogamente  $F_x(x_0, y_0)$  de modo que

$$F'(x_0, y_0) = [F_x(x_0, y_0) \quad F_y(x_0, y_0)]$$



y el producto anterior se convierte en la relación matricial

$$F_x(x_0, y_0) + F_y(x_0, y_0)\phi'(x_0) = 0,$$

y así, si la matriz  $F_y(x_0, y_0)$  es invertible, podemos despejar

$$\phi'(x_0) = F_y(x_0, y_0)^{-1} F_x(x_0, y_0) .$$

Antes de enunciar y demostrar este teorema, nos centraremos en un caso especial, el así llamado *Teorema de la función inversa*, del cual el caso general se sigue. Nos centramos en este caso en una función de la forma  $F(x, y) = f(x) - y$ , con  $f : \Omega \subset \mathbb{R}^N \rightarrow \mathbb{R}^N$ . Note que despejar  $x$  en función de  $y$  de la relación  $f(x) - y = 0$  cerca de un par dado  $(x_0, y_0)$  que la satisface, corresponde al problema de encontrar una *inversa local*  $x = f^{-1}(y)$ , la que resulta de hecho existir, y ser de clase  $C^1$  bajo el requerimiento que la matriz  $f'(x_0)$  sea invertible. El teorema enuncia como sigue.

**Teorema 6.1.** (*Teorema de la función inversa*) Sea  $f : \Omega \subset \mathbb{R}^N \rightarrow \mathbb{R}^N$ ,  $\Omega$  abierto, una función de clase  $C^1(\Omega)$  y  $x_0 \in \Omega$ . Supongamos que  $f'(x_0)^{-1}$  existe. Entonces existe  $\mathcal{U}$ , un abierto contenido en  $\Omega$  que contiene a  $x_0$ , tal que  $\mathcal{V} = f(\mathcal{U})$  es un abierto y

$$f : \mathcal{U} \rightarrow \mathcal{V}$$

es inyectiva. Entonces la función

$$f^{-1} : \mathcal{V} \rightarrow \mathcal{U}$$

es diferenciable, con derivada continua en  $\mathcal{V}$ . Se tiene además la fórmula

$$(f^{-1})'(y) = f'(f^{-1}(y))^{-1} \quad \forall y \in \mathcal{V} .$$

Para la demostración necesitamos el siguiente lema, consecuencia del teorema del punto fijo de Banach.

**Lema 6.1.** Sea  $\psi : \bar{B}(0, R) \subset \mathbb{R}^N \rightarrow \mathbb{R}^N$  con  $\psi(0) = 0$ , una función contractante, con constante  $0 < \alpha < 1$ , esto es

$$\|\psi(x_1) - \psi(x_2)\| \leq \alpha \|x_1 - x_2\| \quad \forall x_1, x_2 \in \bar{B}(0, R) ,$$

y definamos  $g(x) = x - \psi(x)$ . Entonces  $g$  es inyectiva en  $B(0, R)$ . Más precisamente, se tiene que

$$\|x_1 - x_2\| \leq \frac{1}{1 - \alpha} \|g(x_1) - g(x_2)\| \quad \forall x_1, x_2 \in B(0, R) .$$

Por otra parte, para todo  $y \in B(0, (1 - \alpha)R)$ , la ecuación

$$g(x) = y$$

posee una única solución  $x \in B(0, R)$ . Más aún,  $\mathcal{V} = g(B(0, R))$  es un conjunto abierto.

**Demostración.** Sea  $y \in B(0, R(1 - \alpha))$ , y consideremos la ecuación  $g(x) = y$ , que se escribe

$$\psi_y(x) := \psi(x) + y = x$$

La función  $\psi_y$  es claramente contractante en  $\bar{B}(0, R)$ . Además, si  $x \in \bar{B}(0, R)$  entonces

$$\begin{aligned} \|\psi_y(x)\| &= \|\psi(x) - \psi(0) + y\| \leq \|\psi(x) - \psi(0)\| + \|y\| \leq \\ &\alpha\|x\| + \|y\| < \alpha R + (1 - \alpha)R = R \end{aligned}$$

de modo que  $\psi_y(x) \in B(0, R)$ . De este modo, tenemos en particular que  $\psi_y$  aplica el cerrado  $\bar{B}(0, R)$  en si mismo. Por el teorema del punto fijo de Banach, teorema 3.3, tenemos que la ecuación  $x = \psi_y(x)$  posee una única solución  $x$  en  $\bar{B}(0, R)$ , que además está en realidad en  $B(0, R)$ . Así, la ecuación  $g(x) = y$  tiene una única solución en  $B(0, R)$  que denotamos Sean  $x_1 = g^{-1}(y_1)$ ,  $x_2 = g^{-1}(y_2)$ . Entonces

$$x_1 - x_2 = \psi(x_1) - \psi(x_2) + y_1 - y_2$$

y por lo tanto

$$\|x_1 - x_2\| \leq \|\psi(x_1) - \psi(x_2)\| + \|y_1 - y_2\| \leq \alpha\|x_1 - x_2\| + \|y_1 - y_2\|.$$

Deducimos que

$$\|g^{-1}(y_1) - g^{-1}(y_2)\| \leq \frac{1}{1 - \alpha} \|y_1 - y_2\|.$$

Solo resta demostrar que  $g(B(0, R))$  es abierto. Sea  $y_0 \in g(B(0, R))$ , de modo que  $y_0 = g(x_0)$  con  $x_0 \in B(0, R)$ . Existe entonces  $\rho > 0$  tal que  $B(x_0, \rho) \subset B(0, R)$ . Definamos

$$\tilde{\psi}(\tilde{x}) = \psi(\tilde{x} + x_0) - \psi(x_0) .$$

$\tilde{\psi}$  es claramente contractante de constante  $\alpha$  en  $\bar{B}(0, \rho)$  con  $\tilde{\psi}(0) = 0$ . Por lo tanto, si  $y \in B(y_0, (1 - \alpha)\rho)$ , se tiene que  $\tilde{y} = y - y_0 \in B(0, (1 - \alpha)\rho)$ , y de acuerdo a lo ya demostrado, existe una solución  $\tilde{x} \in B(0, \rho)$  de la ecuación

$$\tilde{x} - \tilde{\psi}(\tilde{x}) = \tilde{y},$$

esto es

$$\tilde{x} - \psi(x_0 + \tilde{x}) + \psi(x_0) = y - y_0.$$

Como  $x_0 - \psi(x_0) + \psi(x_0) = y_0$  Concluimos que

$$x_0 + \tilde{x} - \psi(x_0 + \tilde{x}) = y,$$

y por lo tanto  $x = x_0 + \tilde{x} \in B(x_0, \rho)$  satisface que  $g(x) = y$ , esto es,  $y \in g(B(x_0, \rho)) \subset g(B(0, R))$ . Hemos demostrado que

$$(\forall y_0 \in g(B(0, R))) (\exists \rho > 0) : B(y_0, \rho(1 - \alpha)) \subset g(B(0, R)) ,$$

y por lo tanto todo punto de  $g(B(0, R))$  es a su vez punto interior de este conjunto, esto es  $g(B(0, R))$  es abierto. Esto concluye la demostración.  $\square$

**Demostración del teorema 6.1.** Consideremos la ecuación  $f(x) = y$  para  $y$  cerca de  $y = y_0 = f(x_0)$ , y  $x$  cerca de  $x_0$ . Escribiendo  $x = x_0 + h$ , esta ecuación es equivalente a  $f(x_0 + h) = y$ , la que reenunciamos como

$$h - f'(x_0)^{-1}(f(x_0 + h) - f(x_0) - f'(x_0)h) = f'(x_0)^{-1}(y_0 - y). \quad (6.34)$$

Consideremos entonces la función

$$\psi(h) = f'(x_0)^{-1}(f(x_0 + h) - f(x_0) - f'(x_0)h).$$

Entonces  $\psi(0) = 0$ . Además,

$$\psi'(h) = f'(x_0)^{-1}(f'(x_0 + h) - f'(x_0)).$$

Tenemos que

$$\psi'(h) = \begin{bmatrix} \psi'_1(h) \\ \vdots \\ \psi'_N(h) \end{bmatrix}.$$

Por hipótesis  $\nabla \psi_i(h) = \psi'_i(h)^T$  es una función continua, y tenemos además  $\nabla \psi'_i(0) = 0$ . Fijemos un número  $\delta > 0$  tal que para todo  $i = 1, \dots, N$ ,

$$\|\nabla \psi_i(h)\| \leq \frac{1}{2N} \quad \forall h \in \bar{B}(0, \delta).$$

Sean  $k_1, k_2 \in \bar{B}(0, \delta)$ . Entonces  $k_1 + t(k_2 - k_1) \in \bar{B}(0, \delta)$  para todo  $t \in [0, 1]$ . Por el teorema del valor medio, existe  $\xi \in ]0, 1[$  tal que

$$\psi_i(k_1) - \psi_i(k_2) = \nabla \psi_i(k_1 + \xi(k_2 - k_1)) \cdot (k_2 - k_1),$$

de modo que por Cauchy-Schwartz,

$$|\psi_i(k_2) - \psi_i(k_1)| \leq \|\nabla \psi_i(k_1 + \xi(k_2 - k_1))\| \|k_2 - k_1\| \leq \frac{1}{2\sqrt{N}} \|k_2 - k_1\|.$$

Concluimos, sumando que

$$\|\psi(k_2) - \psi(k_1)\| = \sqrt{\sum_{i=1}^N |\psi_i(k_2) - \psi_i(k_1)|^2} \leq \frac{1}{2} \|k_2 - k_1\|$$

y entonces  $\psi$  es contractante de constante  $\alpha = \frac{1}{2}$  en  $\bar{B}(0, \delta)$ . Las hipótesis del lema 6.1 entonces se cumplen para este  $\psi$  y concluimos que la ecuación (6.34) posee una única solución  $h \in B(0, \delta)$  para cada  $y$  que satisfaga

$$\|f'(x_0)^{-1}(y_0 - y)\| < (1 - \alpha)\delta.$$

Más aún, la función

$$g(h) = f'(x_0)^{-1}(f(x_0 + h) - f(x_0))$$

es inyectiva,  $g(B(0, \delta))$  es abierto, y la inversa de  $g$  sobre este último conjunto es Lipschitz. Concluimos que  $f(x) = f'(x_0)g(x - x_0) + f(x_0)$  es también inyectiva, y también que  $\mathcal{V} = f(B(x_0, \delta))$  es abierto. La inversa de  $f$  es también Lipschitz, en particular continua, sobre todo  $\mathcal{V}$ . Se propone al lector la verificación en detalle de estos últimos hechos. Nos resta demostrar que la inversa de  $f$  es diferenciable.

Consideremos un  $y \in f(B(x_0, \delta))$  y  $y = f(x)$ . Suponemos, reduciendo  $\delta$  si es necesario, que la inversa  $f'(x)^{-1}$  existe para todo  $x \in B(x_0, \delta)$ . Consideremos  $k$  pequeño y denotemos

$$h(k) = f^{-1}(y + k) - x.$$

Se tiene entonces que

$$\begin{aligned} \theta(k) &= f^{-1}(y + k) - f^{-1}(y) - f'(x)^{-1}k = h - f'(x)^{-1}(f(x + h) - f(x)) = \\ &\quad -f'(x)^{-1}[f(x + h) - f(x) - f'(x)h] \end{aligned}$$

Notemos que  $h(0) = 0$  y que  $h$  es Lipschitz pues  $f^{-1}$  lo es. Entonces

$$\|h(k)\| = \|h(k) - h(0)\| \leq C\|k - 0\|,$$

para cierto  $C > 0$ , de modo que  $\frac{\|h\|}{\|k\|} \leq C$ . Tenemos,

$$\frac{\theta(k)}{\|k\|} = -\frac{\|h\|}{\|k\|} f'(x)^{-1} \left[ \frac{f(x + h) - f(x) - f'(x)h}{\|h\|} \right].$$

Como  $f$  es diferenciable en  $x$ ,  $h(k) \rightarrow 0$  cuando  $k \rightarrow 0$ ,  $f'(x)$  es una matriz constante y  $\frac{\|h\|}{\|k\|}$  es acotada, concluimos

$$\lim_{k \rightarrow 0} \frac{\theta(k)}{\|k\|} = 0$$

y entonces  $f^{-1}$  es diferenciable en  $y$  con  $f^{-1'}(y) = f'(f^{-1}(y))^{-1}$ . Esta última fórmula define además una función continua de  $y$ , pues  $f^{-1}$  es continua al ser Lipschitz, y la aplicación a valores matriciales  $x \mapsto f'(x)^{-1}$  tiene componentes continuas. Proponemos al lector la verificación en detalle de este hecho. Esto concluye la demostración.  $\square$

**Ejemplo 6.1.** La hipótesis de continuidad de la derivada es necesaria para la validez del teorema anterior. En efecto, consideremos la función de una variable

$$f(x) = \begin{cases} x + x^2 \sin \frac{1}{x} & \text{si } x \neq 0, \\ 0 & \text{si } x = 0. \end{cases}$$

Notemos que  $f$  es diferenciable en todo  $\mathbb{R}$ . En efecto, para  $x \neq 0$  esto es claro. Si  $x = 0$  se tiene

$$f'(0) = \lim_{h \rightarrow 0} \frac{f(h) - f(0)}{h} = \lim_{h \rightarrow 0} 1 + h \sin \frac{1}{h} = 1 \neq 0.$$

Por otra parte,  $f$  no es inyectiva en ningún intervalo abierto que contiene a 0. En efecto, si  $x \neq 0$ ,

$$f'(x) = 1 - \cos \frac{1}{x} + 2x \sin \frac{1}{x}.$$

Consideremos la sucesión que tiende a cero

$$x_k = \frac{1}{2k\pi},$$

y notemos que

$$f'(x_k) = 0, \quad f''(x_k) = \frac{1}{2k\pi} > 0.$$

Así,  $f$  tiene un mínimo local estricto en cada  $x_k$  y  $f$  por ende no es inyectiva en un entorno de  $x_k$ . Como todo intervalo que contiene a 0 contiene una infinidad de estos  $x_k$ , concluimos el resultado:  $f$  no es inyectiva en en ninguno de estos intervalos.

Como dijimos antes, la consecuencia principal del teorema de la función inversa es el teorema de la función implícita, que se refiere a la posibilidad de “despejar” la variable  $y$  en términos de  $x$  en un sistema de ecuaciones de la forma

$$f(x, y) = 0, \quad x \in \mathbb{R}^N, \quad y \in \mathbb{R}^m. \quad (6.35)$$

donde  $f : \mathbb{R}^N \times \mathbb{R}^m \rightarrow \mathbb{R}$  es una función de clase  $C^1$ . Consideremos las derivadas parciales matriciales

$$f_y(x_0, y_0) = \left[ \frac{\partial f}{\partial y_1}(x_0, y_0) \cdots \frac{\partial f}{\partial y_m}(x_0, y_0) \right]_{m \times m},$$

$$f_x(x_0, y_0) = \left[ \frac{\partial f}{\partial x_1}(x_0, y_0) \cdots \frac{\partial f}{\partial x_N}(x_0, y_0) \right]_{m \times N}.$$

Este resultado enuncia básicamente que si  $f(x_0, y_0) = 0$  y la matriz  $f_y(x_0, y_0)$  es invertible, entonces puede despejarse  $y$  en función de  $x$  en

una vecindad de  $x_0$  como una función de clase  $C^1$  cuyo valor en  $x_0$  es  $y_0$ .

**Teorema 6.2.** *(Teorema de la función implícita). Sean  $\Omega \subset \mathbb{R}^N$  y  $\Lambda \subset \mathbb{R}^m$  conjuntos abiertos y*

$$f : \Omega \times \Lambda \rightarrow \mathbb{R}^m, \quad (x, y) \mapsto f(x, y)$$

*una función de clase  $C^1(\Omega \times \Lambda)$ . Supongamos que  $(x_0, y_0) \in \Omega \times \Lambda$  es tal que  $f(x_0, y_0) = 0$  y que la matriz  $m \times m$   $f_y(x_0, y_0)$  es invertible. Entonces existe un abierto  $\mathcal{U}$  con  $x_0 \in \mathcal{U} \subset \Omega$  y una única función  $\phi : \mathcal{U} \rightarrow \Lambda$  de clase  $C^1(\mathcal{U})$  tal que*

$$f(x, \phi(x)) = 0 \quad \forall x \in \mathcal{U}.$$

**Demostración.** Consideremos la función

$$F : \Omega \times \Lambda \rightarrow \mathbb{R}^N \times \mathbb{R}^m,$$

definida por

$$F(x, y) = \begin{bmatrix} x \\ f(x, y) \end{bmatrix}.$$

$F$  es claramente de clase  $C^1(\Omega \times \Lambda)$  y su derivada en  $(x_0, y_0)$  es la matriz cuadrada  $(N + m) \times (N + m)$  dada por

$$F'(x_0, y_0) = \begin{bmatrix} I_{N \times N} & 0_{N \times m} \\ f_x(x_0, y_0) & f_y(x_0, y_0) \end{bmatrix}.$$

Aquí  $I_{N \times N}$  denota la matriz identidad  $N \times N$  y  $0_{N \times m}$  la matriz nula  $N \times m$ . Afirmamos que esta matriz es invertible. En efecto, si

$$F'(x_0, y_0)h = \begin{bmatrix} I_{N \times N} & 0_{N \times m} \\ f_x(x_0, y_0) & f_y(x_0, y_0) \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \in \mathbb{R}^N \times \mathbb{R}^m$$

entonces

$$h_1 = 0, \quad f_x(x_0, y_0)h_1 + f_y(x_0, y_0)h_2 = 0.$$

Por lo tanto  $f_y(x_0, y_0)h_2 = 0$ , y como esta matriz es invertible, se sigue que  $h_2 = 0$ . Entonces  $h = 0$ , lo que implica que la inversa  $F'(x_0, y_0)^{-1}$  existe. Por el teorema de la función inversa, concluimos que existe una vecindad del punto  $(x_0, y_0)$  cuya imagen a través de  $F$  es un abierto, donde  $F$  es inyectiva, y posee una inversa de clase  $C^1$ . Empequeñeciendo esta vecindad si es necesario, la podemos suponer de la forma  $\mathcal{W} \times \mathcal{V}$  con  $x_0 \in \mathcal{W}$ ,  $y_0 \in \mathcal{V}$ . Así, como

$$F(x_0, y_0) = \begin{bmatrix} x_0 \\ 0 \end{bmatrix}$$

entonces para todo punto  $(x, z) \in \mathcal{U} \times \mathcal{Z}$ , esta última una pequeña vecindad de  $(x_0, 0)$ , la ecuación

$$F(t, y) = \begin{bmatrix} x \\ z \end{bmatrix}$$

posee una única solución  $(t, y) \in \mathcal{W} \times \mathcal{V}$ , que define una función de clase  $C^1$

$$F^{-1}(x, z) = \begin{bmatrix} \phi_1(x, z) \\ \phi_2(x, z) \end{bmatrix}.$$

Las funciones  $\phi_1$  y  $\phi_2$  son entonces de clase  $C^1$  en  $\mathcal{U} \times \mathcal{Z}$ , con  $\phi_1(x_0, 0) = x_0$  y  $\phi_2(x_0, 0) = y_0$ . Tenemos entonces que

$$F(\phi_1(x, z), \phi_2(x, z)) = \begin{bmatrix} \phi_1(x, z) \\ f(\phi_1(x, z), \phi_2(x, z)) \end{bmatrix} = \begin{bmatrix} x \\ z \end{bmatrix}.$$

De modo que, en particular, para todo  $x \in \mathcal{U}$  se tiene que  $(x, 0) \in \mathcal{U} \times \mathcal{Z}$  y

$$\begin{bmatrix} \phi_1(x, 0) \\ f(\phi_1(x, 0), \phi_2(x, 0)) \end{bmatrix} = \begin{bmatrix} x \\ 0 \end{bmatrix},$$

esto es  $\phi_1(x, 0) = x$  y  $f(x, \phi_2(x, 0)) = 0$ . El resultado del teorema entonces se concluye, tomando la función  $\phi(x) =: \phi_2(x, 0)$ .  $\square$

**Ejemplo 6.2.** Consideremos el sistema de ecuaciones de dos ecuaciones y tres incógnitas

$$\begin{aligned} x^2y - x^5ty + 3(t^2 - 1) &= 0 \\ x^3y^2 - 6xt^2y + 3t^5 &= 0. \end{aligned}$$

Observemos que este sistema tiene a  $(x, y, t) = (1, 1, 1)$  como solución. Sea

$$F(x, y, t) = \begin{bmatrix} x^2y - x^5ty + 3(t^2 - 1) \\ x^3y^2 - 6xt^2y + 3t^5 \end{bmatrix}.$$

Vemos que

$$F'(x, y, t) = \begin{bmatrix} 2xy - 5x^4ty & x^2 - x^5t & -x^5y + 6t \\ 3x^2y^2 - 6t^2y & 2x^3y - 6xt^2 & -12txy + 15t^4 \end{bmatrix},$$

de modo que

$$F'(1, 1, 1) = \begin{bmatrix} -3 & 0 & 5 \\ -3 & -4 & 3 \end{bmatrix},$$

y entonces

$$F_{(x,y)}(1, 1, 1) = \begin{bmatrix} -3 & 0 \\ -3 & -4 \end{bmatrix}.$$

Esta matriz es invertible, de modo que por el teorema de la función implícita pueden despejarse  $x$  e  $y$  como funciones de  $t$  en modo  $C^1$  en una vecindad de  $t = 1$ . Más precisamente, existe  $\delta > 0$  y funciones

$$x : ] - \delta, \delta[ \rightarrow \mathbb{R}, \quad y : ] - \delta, \delta[ \rightarrow \mathbb{R}$$

de clase  $C^1$ , tales que  $x(1) = 1$ ,  $y(1) = 1$ , que satisfacen

$$\begin{aligned} x^2(t)y(t) - x^5(t)ty(t) + 3(t^2 - 1) &= 0 \\ x^3(t)y^2(t) - 6x(t)t^2y(t) + 3t^5 &= 0 \end{aligned} \quad \text{para todo } t \in ] - \delta, \delta[ .$$

Podemos además calcular las derivadas  $x'(1)$ ,  $y'(1)$ , por ejemplo mediante derivación implícita de estas relaciones. Obtenemos de ellas

$$\begin{aligned} 2xx'y + x^2y' - 5x^4x'ty - x^5y - x^5y' + 6t &= 0 \\ 3x^2y^2x' + 2x^3yy' - 6x't^2y - 12txy - 6t^2xy' + 15t^4 &= 0 \end{aligned} \quad \text{para todo } t \in ] - \delta, \delta[$$

y entonces, evaluando en  $t = 1$ ,

$$\begin{aligned} -3x'(1) + 5 &= 0 \\ -3x'(1) - 4y'(1) + 3 &= 0 \end{aligned}$$

sistema lineal que nos entrega los valores

$$x'(1) = \frac{5}{3}, \quad y'(1) = -\frac{11}{12}.$$

Alternativamente, podríamos haber calculado esta derivada a partir de la fórmula matricial

$$\begin{bmatrix} x'(1) \\ y'(1) \end{bmatrix} = -F_{(x,y)}(1, 1, 1)F_t(1, 1, 1) = \begin{bmatrix} 3 & 0 \\ 3 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 5 \\ 3 \end{bmatrix}^{-1}$$

fórmula que corresponde precisamente al sistema que resolvimos.

Una consecuencia interesante del Teorema de la función implícita es que provee una condición de primer orden necesaria para la resolución de problemas de minimización con restricciones de igualdad, esto es, el minimizar una función  $f(z)$  sujeta a  $m$  restricciones de la forma  $g_i(z)$ , donde  $m$  es estrictamente menor que la dimensión del espacio. Este es el contenido del resultado siguiente.

**Teorema 6.3.** (*Teorema de los multiplicadores de Lagrange*) Sean  $f : \mathbb{R}^{N+m} \rightarrow \mathbb{R}$ ,  $g : \mathbb{R}^{N+m} \rightarrow \mathbb{R}^m$ , funciones de clase  $C^1$ . Sea

$$A = \{z \in \mathbb{R}^{N+m} / g(z) = 0\}$$

y supongamos que

$$f(z_0) = \min_{z \in A} f(z) .$$



Supongamos además que la matriz  $(N + m) \times m$ ,  $g'(z_0)$  es de rango completo, esto es posee  $m$  columnas linealmente independientes. Entonces existen números  $\lambda_1, \lambda_2, \dots, \lambda_m$  tales que la siguiente igualdad se satisface.

$$\nabla f(z_0) = \sum_{i=1}^m \lambda_i \nabla g_i(z_0) .$$

**Demostración.** Supongamos que la variable  $z \in \mathbb{R}^{N+m}$  se descompone como

$$z = (x, y) \in \mathbb{R}^N \times \mathbb{R}^m$$

donde, luego de reordenar las variables en caso de ser necesario, podemos suponer que las últimas  $m$  columnas de  $g'(x_0, y_0)$  son linealmente independientes. Así, la matriz  $g_y(x_0, y_0)$  es invertible.

Supongamos que  $z_0 = (x_0, y_0)$  es tal que  $f$  se minimiza sobre  $\mathcal{D} = \{(x, y) / g(x, y) = 0\}$ . Como, por hipótesis, la matriz  $g_y(x_0, y_0)$  es invertible, el teorema de la función implícita nos garantiza la existencia de una función  $y = \phi(x)$ , definida en un abierto  $\mathcal{U}$  que contiene a  $x_0$ , de modo que  $\phi(x_0) = y_0$  y  $g(x, \phi(x)) = 0$  para todo  $x \in \mathcal{U}$ . Así, los puntos  $(x, \phi(x))$ ,  $x \in \mathcal{U}$  yacen todos en  $\mathcal{D}$  y tenemos entonces que

$$f(x, \phi(x)) \geq f(x_0, y_0) \quad \text{para todo } x \in \mathcal{U}.$$

Por lo tanto, la función  $\psi(x) := f(x, \phi(x))$  satisface que  $\psi(x) \geq \psi(x_0)$  para todo  $x \in \mathcal{U}$ . Se sigue que  $\psi'(x_0) = 0$ , lo que quiere decir, gracias a la regla de la cadena, que

$$0 = \psi'(x_0) = [f_x(x_0, y_0) \ f_y(x_0, y_0)] \begin{bmatrix} I_m \\ \phi'(x_0) \end{bmatrix} ,$$

esto es,

$$0 = f_x(x_0, y_0) + f_y(x_0, y_0)\phi'(x_0).$$

Por otra parte, Como la función  $\xi(x) := g(x, \phi(x))$  es constante en  $\mathcal{U}$ , se sigue en particular que  $\xi'(x_0) = 0$ , lo que quiere decir

$$0 = g_x(x_0, y_0) + g_y(x_0, y_0)\phi'(x_0),$$

de modo que

$$\phi'(x_0) = -g_y(x_0, y_0)^{-1}g_x(x_0, y_0),$$

y obtenemos entonces

$$f_x(x_0, y_0) = f_y(x_0, y_0)g_y(x_0, y_0)^{-1}g_x(x_0, y_0)$$

como así también,

$$f_y(x_0, y_0) = f_y(x_0, y_0)g_y(x_0, y_0)^{-1}g_y(x_0, y_0)$$

Sea

$$\Lambda_{1 \times m} = f_y(x_0, y_0)g_y(x_0, y_0)^{-1}.$$

Se sigue entonces que

$$[f_x(x_0, y_0) \ f_y(x_0, y_0)] = \Lambda[g_x(x_0, y_0) \ g_y(x_0, y_0)]$$

esto es,

$$f'(x_0, y_0) = \Lambda g'(x_0, y_0)$$

o sea, tomando transpuesta,

$$\nabla f(x_0, y_0) = g'(x_0, y_0)^T \Lambda^T.$$

Si  $\Lambda = [\lambda_1 \cdots \lambda_m]$ , la relación anterior se lee precisamente como

$$\nabla f(x_0, y_0) = \sum_{i=1}^m \lambda_i \nabla g_i(x_0, y_0),$$

y la demostración queda entonces concluida.  $\square$

**Ejemplo 6.3.** Consideremos el problema de minimizar la función

$$f(x, y, z) := x + y - z$$

bajo la restricción

$$g(x, y, z) := x^2 + y^2 + z^2 - 1 = 0.$$

La primera observación es que el conjunto de puntos que satisfacen la restricción es cerrado y acotado en  $\mathbb{R}^3$ , por lo tanto, siendo la función  $f$  continua, alcanza en efecto su valor mínimo. Consideremos entonces los puntos  $(x, y, z)$  que satisfacen la restricción y tales que para algún  $\lambda \in \mathbb{R}$

$$\nabla f(x, y, z) - \lambda \nabla g(x, y, z) = 0$$

Esto corresponde al sistema de ecuaciones  $4 \times 4$

$$\begin{cases} 1 - \lambda 2x = 0 \\ 1 - \lambda 2y = 0 \\ -1 - \lambda 2z = 0 \\ x^2 + y^2 + z^2 - 1 = 0 \end{cases}$$

Así, claramente para una solución de este sistema tenemos que

$$x = y = -z, \quad 3x^2 = 1$$

lo que nos entrega dos puntos posibles:

$$\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}\right), \quad \left(-\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right).$$

Notemos que

$$f\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}\right) = \sqrt{3}, \quad f\left(-\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right) = -\sqrt{3},$$

y por lo tanto el segundo de estos puntos es aquél que corresponde al valor mínimo.

La regla de los multiplicadores de Lagrange tiene una versión nemotécnica útil: Si  $x_0$  minimiza a  $f$  sujeto a las restricciones  $g_i(x) = 0$ ,  $i = 1, \dots, m$ , entonces existen  $m$  números  $\lambda_i^0$  tal que  $(x_0, \lambda_0^1, \dots, \lambda_0^m)$  es un punto crítico en  $\mathbb{R}^{N+m}$ , sin restricción, del *Lagrangiano* del problema, definido por

$$\mathcal{L}(x, \lambda_1, \dots, \lambda_m) := f(x) - \sum_{i=1}^m \lambda_i g_i(x).$$

**Ejemplo 6.4.** Una desigualdad clásica en los números reales enuncia que el *promedio geométrico* de  $m$  números positivos es menor que el *promedio aritmético*, a menos que todos estos números sean iguales. Así, afirmamos que si  $a_i > 0$  para  $i = 1, \dots, N$ , entonces se tiene la desigualdad

$$(a_1 a_2 \cdots a_N)^{\frac{1}{N}} \leq \frac{1}{N} (a_1 + \cdots + a_N),$$

la cual es estricta a menos que todos los números sean iguales. Para probar esto, consideremos la función

$$f(x_1, \dots, x_N) = \frac{1}{N} \sum_{i=1}^N x_i$$

sobre el conjunto definido por la restricción

$$g(x_1, \dots, x_N) := \prod_{i=1}^N x_i - 1 = 0$$

Consideramos a las funciones  $f$  y  $g$  a su vez definidas sobre el abierto de  $\mathbb{R}^N$  dado por el conjunto de puntos  $x$  con  $x_i > 0$  para todo  $i$ . Observemos primero que nada que si para algún  $i$  se tiene  $x_i > N$ , entonces

$$f(x_1, \dots, x_N) > 1 = f(1, \dots, 1).$$

Por lo tanto, el mínimo de  $f$  sobre la región cerrada y acotada

$$\{x \in \mathbb{R}^N / g(x) = 0, 0 \leq x_i \leq N \text{ para todo } i\},$$

que existe pues  $f$  es continua, debe ser necesariamente el el mínimo de  $f$  sujeto a la restricción completa  $g = 0$ . Así, definamos

$$\mathcal{L}(x, \lambda) = \frac{1}{N} \sum_{i=1}^N x_i - \lambda \left( \prod_{i=1}^N x_i - 1 \right).$$

Tenemos entonces que en el punto de mínimo, tenemos una solución  $(x, \lambda)$  del sistema

$$\frac{\partial}{\partial x_j} \mathcal{L}(x, \lambda) = 0 \quad \text{para todo } j, \quad \frac{\partial}{\partial \lambda} \mathcal{L}(x, \lambda) = 0,$$

esto es

$$\frac{1}{N} = \lambda \prod_{i \neq j} x_i = 0 \quad \text{para todo } j, \quad \prod_{i=1}^N x_i - 1 = 0.$$

Tenemos entonces que, en el mínimo

$$\frac{x_j}{N} = \lambda \prod_{i=1}^N x_i = \lambda \quad \text{para todo } j,$$

y por lo tanto,  $x_1 = x_2 = \cdots x_N = 1$ . esto se traduce en que

$$1 \leq \frac{1}{N} \sum_{i=1}^N x_i, \quad \text{si } \prod_{i=1}^N x_i = 1,$$

Con desigualdad estricta a menos que todos estos números sean iguales. Consideremos ahora  $a_1, \dots, a_N$  números positivos cualesquiera y definamos

$$x_j := \frac{a_j}{\left( \prod_{i=1}^N a_i \right)^{\frac{1}{N}}}.$$

Claramente  $\prod_{i=1}^N x_i = 1$ , y por lo tanto

$$1 \leq \frac{1}{N} \sum_{j=1}^N \frac{a_j}{\left( \prod_{i=1}^N a_i \right)^{\frac{1}{N}}},$$

y el resultado deseado se concluye.

## 7. INTEGRACIÓN DE FUNCIONES DE VARIAS VARIABLES

En este capítulo extenderemos la noción de integral en el sentido de Riemann a funciones definidas sobre subconjuntos del espacio  $\mathbb{R}^N$ . Consideremos una función  $f : \mathcal{D} \subset \mathbb{R}^N \rightarrow \mathbb{R}$ . Si  $N = 1$  y  $\mathcal{D} = [a, b]$  se definió, bajo ciertas condiciones, la cantidad  $\int_a^b f(x)dx$ , cuya interpretación geométrica, cuando  $f$  es positiva, es el *área* bajo la curva que define el gráfico de  $f$ ,  $\{(x, f(x)) / x \in [a, b]\}$ , por sobre el intervalo  $[a, b]$ , esto es, de la región

$$\{(x, y) / x \in [a, b], 0 \leq y \leq f(x)\}.$$

Para una función de dos variables, el gráfico típicamente se entiende como una *superficie* en  $\mathbb{R}^3$ , y deseamos definir la cantidad

$$\int \int_{\mathcal{D}} f(x, y) dx dy$$

como una noción apropiada de *volúmen* de la región en  $\mathbb{R}^3$  dada por

$$\{(x, y, z) / (x, y) \in \mathcal{D}, 0 \leq z \leq f(x, y)\}.$$

Consideremos el especial de la función  $f \equiv 1$ . Así, en particular

$$\int \int_{\mathcal{D}} 1 dx dy$$

debiese corresponder al volumen del cilindro de altura 1 que tiene a  $\mathcal{D}$  como sección transversal, esto es al *área de  $\mathcal{D}$* . Del mismo modo intentamos definir, para el caso de una función de tres variables la cantidad

$$I = \int \int \int_{\mathcal{D}} f(x, y, z) dx dy dz,$$

que si bien no tendrá interpretación geométrica intuitiva directa, debiese corresponder a que  $\int \int \int_{\mathcal{D}} 1 dx dy dz$  sea el *volumen* de la región  $\mathcal{D}$  del espacio tridimensional. Por otra parte, la cantidad  $I$  puede interpretarse físicamente como *masa total* de  $\mathcal{D}$ , suponiendo que su densidad de masa por unidad de volumen en el punto  $(x, y, z)$  está dada por la función  $f(x, y, z)$ , lo cual quiere decir que un "rectángulo infinitesimal de volumen  $dx dy dz$ ." en este punto tiene por masa  $f(x, y, z) dx dy dz$ .

**7.1. Definición de la Integral y propiedades básicas.** La definición que daremos es la extensión directa de la *integral de Riemann* para funciones de una variable hecha en el curso de Cálculo anterior. Consideremos primero el caso de los dominios  $\mathcal{D}$  más sencillos que llamamos

*rectángulos de lados paralelos a los ejes coordenados.* Un rectángulo  $\mathcal{R}$  de este tipo en  $\mathbb{R}^N$  es un conjunto de la forma

$$\mathcal{R} = \{x \in \mathbb{R}^N \mid a_i \leq x_i \leq b_i \text{ para todo } i = 1, \dots, N\} = [a_1, b_1] \times \cdots \times [a_N, b_N],$$

en otras palabras,

$$\mathcal{R} = \prod_{i=1}^N [a_i, b_i]. \quad (7.1)$$

Definimos el *volumen de  $\mathcal{R}$*  simplemente como la cantidad

$$V(\mathcal{R}) = \prod_{i=1}^N (b_i - a_i),$$

lo que en dimensiones 1, 2 y 3 corresponde a nuestras nociones habituales de largo, área y volumen respectivamente. Consideramos en lo que sigue un rectángulo  $\mathcal{R} \subset \mathbb{R}^N$  y una función  $f : \mathcal{R} \rightarrow \mathbb{R}$  la cual supondremos *acotada*, esto es tal que existe  $M > 0$  para el cual

$$|f(x)| \leq M \quad \text{para todo } x \in \mathcal{R}.$$

Un *reticulado  $\mathcal{S}$*  del rectángulo  $\mathcal{R}$  dado por (7.1) es una familia de rectángulos

$$\mathcal{S} = \{\mathcal{R}_i\}_{i \in I}$$

Donde  $I$  es un conjunto de índices, con cada  $\mathcal{R}_i$  un rectángulo de lados paralelos a los ejes coordenados tales que

$$\bigcup_{i \in I} \mathcal{R}_i = \mathcal{R}, \quad \text{int } \mathcal{R}_i \cup \text{int } \mathcal{R}_j = \emptyset \text{ para todo } i \neq j$$

$$\mathcal{R}_i = [x_{i_1}, x_{i_1+1}] \times [x_{i_2}, x_{i_2+1}] \times \cdots \times [x_{i_N}, x_{i_N+1}]$$

donde  $i_j = 0, \dots, k_j$ ,  $j = 1, \dots, N$  y

$$x_{i_0} = a_i < x_{i_1} < x_{i_2} < \cdots < x_{i_{k_i}} = b_i$$

Definimos la *suma inferior* para la función  $f$  asociada al reticulado  $\mathcal{S}$  como

$$I_{\mathcal{S}}(f) := \sum_{i \in I} m_{\mathcal{R}_i}(f) V(\mathcal{R}_i)$$

donde

$$m_i(f) = \inf_{x \in \mathcal{R}_i} f(x).$$

En modo similar, la *suma superior* asociada al reticulado  $\mathcal{S}$  está dada por

$$S_{\mathcal{S}}(f) := \sum_{i \in I} M_{\mathcal{R}_i}(f) V(\mathcal{R}_i)$$

donde

$$M_i(f) = \sup_{x \in \mathcal{R}_i} f(x).$$

Observemos, por ejemplo, que si la dimensión  $N = 2$ , el número  $m_{\mathcal{R}_i}(f)V(\mathcal{R}_i)$  corresponde, si  $f \geq 0$ , al volumen del paralelepípedo de base rectangular  $\mathcal{R}_i$  y altura  $m_{\mathcal{R}_i}(f)$ , de modo que si  $\mathcal{R}_i$  tiene lados pequeños y  $f$  es, por ejemplo, continua, este volumen debiese aproximar bien (y por abajo) el de la región del espacio comprendida entre el rectángulo  $\mathcal{R}_i$ , contenido en el plano  $x-y$  y el gráfico de la función  $f$ , esto es la "superficie"  $z = f(x, y)$ . Así, la suma inferior  $I_S(f)$  es una aproximación por abajo del volumen total comprendido entre el rectángulo  $R$  y el gráfico de  $f$ . En modo similar,  $S_S(f)$  es una aproximación por arriba de este volumen, aproximación que debiese mejorar y mejorar si los rectángulos del reticulado se hacen más y más pequeños. Básicamente, el volumen total en cuestión debiese considerarse por sí mismo bien definido, precisamente si las sumas inferiores y superiores de  $f$  aproximan un número común a medida que los reticulados se hacen *más finos*. Precisamente en este caso diremos que  $f$  es integrable sobre  $\mathcal{R}$ .

Sean  $\mathcal{S}_1$  y  $\mathcal{S}_2$  dos reticulados de  $\mathcal{R}$ . decimos que  $\mathcal{S}_2$  es *más fino que*  $\mathcal{S}_1$  si se tiene todo rectángulo en  $\mathcal{S}_2$  está contenido en algún rectángulo en  $\mathcal{S}_1$ . Claramente se tiene que

$$I_{\mathcal{S}_1} \leq I_{\mathcal{S}_2} \leq S_{\mathcal{S}_2} \leq S_{\mathcal{S}_1}$$

Por otra parte, observemos que dados dos reticulados cualesquiera  $\mathcal{S}_1$  y  $\mathcal{S}_2$  se asocia canónicamente un reticulado  $\mathcal{S}_3$  más fino tanto a  $\mathcal{S}_1$  como a  $\mathcal{S}_2$  simplemente mediante la colección de todas las intersecciones de rectángulos en  $\mathcal{S}_1$  con rectángulos en  $\mathcal{S}_2$ .

Así, concluimos que en realidad, para *todo* par de reticulados  $\mathcal{S}_1$  y  $\mathcal{S}_2$  (no necesariamente uno más fino que el otro) se tiene que

$$I_{\mathcal{S}_1} \leq S_{\mathcal{S}_2}.$$

Para la función  $f$  acotada en cuestión tiene entonces sentido definir su *integral inferior* en  $\mathcal{R}$  como

$$\int_{\mathcal{R}} f := \sup \{ I_S(f) / \mathcal{S} \text{ es un reticulado de } \mathcal{R} \}.$$

Definimos, en modo similar su *integral inferior sobre*  $\mathcal{R}$  como

$$\overline{\int}_{\mathcal{R}} f := \sup \{ I_S(f) / \mathcal{S} \text{ es un reticulado de } \mathcal{R} \}.$$

De este modo, para cualquier función acotada tenemos la validez de la desigualdad

$$\int_{\underline{\mathcal{R}}} f \leq \overline{\int_{\mathcal{R}}} f .$$

Decimos que la función  $f$  es *Riemann-integrable* si

$$\int_{\underline{\mathcal{R}}} f = \overline{\int_{\mathcal{R}}} f .$$

en cuyo caso llamamos a este valor común *la integral de  $f$  sobre  $\mathcal{R}$*  y lo denotamos como

$$\int_{\mathcal{R}} f \quad \text{o} \quad \int_{\mathcal{R}} f(x)dx \quad \text{o} \quad \int \cdots \int_{\mathcal{R}} f(x_1, \dots, x_N)dx_1 \cdots dx_N ,$$

donde, en la última notación el símbolo de integral se repite  $N$  veces.

Una caracterización útil de la condición de integrabilidad es la siguiente.

**Proposición 7.1.** *Si existe una sucesión de reticulados  $\{\mathcal{S}_n\}_{n \in \mathbb{N}}$  tal que*

$$\lim_{n \rightarrow \infty} S_{\mathcal{S}_n}(f) - I_{\mathcal{S}_n}(f) = 0$$

*entonces  $f$  es integrable. Además,*

$$\lim_{n \rightarrow \infty} S_{\mathcal{S}_n}(f) = \lim_{n \rightarrow \infty} I_{\mathcal{S}_n}(f) = \int_R f(x)dx$$

**Demostración.** Tenemos que

$$I_{\mathcal{S}_n}(f) \leq \int_{\underline{R}} f(x)dx \leq \overline{\int_R} f(x)dx \leq S_{\mathcal{S}_n}(f) .$$

Por ende, pasando al límite, por el teorema del sandwich, obtenemos que

$$\int_{\underline{R}} f(x)dx = \overline{\int_R} f(x)dx,$$

y entonces  $f$  es integrable.  $\square$

**Ejemplo 7.1.** Consideremos  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  dada por  $f(x, y) = x + y$  y  $R = [0, 1] \times [0, 1]$  Consideremos para  $n \in \mathbb{N}$  la partición  $\mathcal{S}_n$  que consta de los elementos

$$R_{jk}^n = \left[\frac{j}{n}, \frac{j+1}{n}\right] \times \left[\frac{k}{n}, \frac{k+1}{n}\right], \quad 0 \leq j, k \leq n-1.$$

Entonces

$$I_{\mathcal{S}_n}(f) = \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} V(R_{jk}^n) m_{R_{jk}^n}(f).$$



Ahora, claramente

$$m_{R_{jk}^n}(f) = \frac{j}{n} + \frac{k}{n},$$

de modo que

$$\begin{aligned} I_{\mathcal{S}_n}(f) &= \frac{1}{n^3} \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} (j+k) = \\ &= \frac{1}{n^3} \sum_{j=0}^{n-1} \left( \frac{n(n-1)}{2} + nj \right) = \frac{1}{n^3} n^2(n-1) = 1 - \frac{1}{n} \end{aligned}$$

En modo similar, obtenemos

$$M_{R_{jk}^n}(f) = \frac{j+1}{n} + \frac{k+1}{n},$$

y entonces

$$S_{\mathcal{S}_n}(f) = \frac{1}{n^3} \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} (j+k+2) = 1 + \frac{1}{n}.$$

Vemos entonces que

$$\lim_{n \rightarrow +\infty} S_{\mathcal{S}_n}(f) - I_{\mathcal{S}_n}(f) = 0$$

y en virtud de la proposición anterior,  $f$  es integrable en  $R$ , y tenemos

$$\int_{[0,1] \times [0,1]} (x+y) dx dy = 1.$$

A continuación probaremos el importante resultado que afirma que las funciones continuas sobre  $\mathcal{R}$  son en efecto integrables.

**Teorema 7.1.** *Sea  $f : \mathcal{R} \subset \mathbb{R}^N \rightarrow \mathbb{R}$  una función continua, donde  $\mathcal{R}$  es un rectángulo. Entonces  $f$  es integrable.*

Para probar este resultado necesitamos el siguiente resultado intermedio:

**Proposición 7.2.** *Sea  $f : K \subset \mathbb{R}^N \rightarrow \mathbb{R}$  una función continua, donde  $K$  es un conjunto cerrado y acotado. Entonces  $f$  es uniformemente continua, esto es:*

$$(\forall \varepsilon > 0)(\exists \delta > 0)(\forall x, y \in K) : \|x - y\| < \delta \implies |f(x) - f(y)| < \varepsilon.$$

Esto se traduce diciendo que el número  $\delta$ , dependiente de  $\varepsilon$  en la caracterización  $\varepsilon$ - $\delta$  de la continuidad en  $x_0 \in K$  puede escogerse *independiente* del punto particular  $x_0$  considerado en  $K$ .

**Demostración de la Proposición 7.2.** Supongamos que esta propiedad no es válida. Entonces se tiene que

$$(\exists \varepsilon_0 > 0)(\forall n \in \mathbb{N})(\exists x_n, y_n \in K) : \|x_n - y_n\| < \frac{1}{n} \text{ y } |f(x_n) - f(y_n)| \geq \varepsilon_0.$$

Como  $x_n$  está contenido en  $K$ , conjunto cerrado y acotado, se sigue que  $x_n$  posee una subsucesión convergente, la que denotamos  $x_{n_j}$ ,  $j \in \mathbb{N}$ , digamos

$$x_{n_j} \rightarrow \bar{x} \in K \quad \text{cuando } n \rightarrow \infty.$$

Así, tenemos también que  $y_{n_j} \rightarrow \bar{x}$ . Por otra parte, como  $f$  es continua tenemos que

$$\lim_{j \rightarrow \infty} f(x_{n_j}) = \lim_{j \rightarrow \infty} f(y_{n_j}) = f(\bar{x})$$

por ende, en particular,  $|f(x_{n_j}) - f(y_{n_j})| \rightarrow 0$ . Esto es claramente una contradicción con  $|f(x_{n_j}) - f(y_{n_j})| \geq \varepsilon_0$  y la demostración queda por tanto concluida.  $\square$

**Demostración del Teorema 7.2.** En virtud de la proposición anterior, basta demostrar que existe una sucesión de reticulados  $\mathcal{S}_n$ ,  $n \in \mathbb{N}$  con la propiedad que

$$S_{\mathcal{S}_n}(f) - I_{\mathcal{S}_n}(f) \rightarrow 0.$$

Probaremos a continuación que este es en efecto el caso para una función continua definida sobre un rectángulo. Consideremos para un  $n \geq 1$  dado, el reticulado uniforme dado por

$$\begin{aligned} & \left[ a_1 + \frac{k_1 - 1}{n}(b_1 - a_1), a_1 + \frac{k_1}{n}(b_1 - a_1) \right] \times \cdots \\ & \times \left[ a_N + \frac{k_N - 1}{n}(b_N - a_N), a_N + \frac{k_N}{n}(b_N - a_N) \right] \end{aligned} \quad (7.2)$$

donde los números  $k_j$  están comprendidos en el rango

$$0 \leq k_j \leq n, \quad \text{para todo } j = 1, \dots, N.$$

Consideremos una enumeración  $R_i$  de estos rectángulos, para  $i = 1, \dots, n^N$ . Entonces notemos que si  $x, y \in R_i$ , se tiene que

$$|x - y| < \frac{C}{\sqrt{n}}.$$

De este modo, si consideramos  $m > 1$ , podemos encontrar, gracias a la continuidad uniforme de  $f$ , un  $n = n_m$ , suficientemente grande tal que para todo  $i$  se tiene que

$$|f(x) - f(y)| < \frac{1}{m} \text{ para todo } x, y \in R_i.$$

En particular

$$M_{R_i}(f) \leq m_{R_i}(f) + \frac{1}{m}$$

y entonces

$$\begin{aligned} S_{S_m} &= \sum_i M_{R_i} V(R_i) \leq \sum_i i_{R_i} m_{R_i} V(R_i) \frac{1}{m} \sum_i i_{R_i} V(R_i) = \\ &I_{S_m} + \frac{1}{m} V(\mathcal{R}) . \end{aligned}$$

Concluimos de aquí la conclusión deseada, pues  $m$  es arbitrario.  $\square$ .

Si bien el resultado anterior es de gran importancia, no es suficiente en la práctica del cálculo de casos concretos. En efecto, nos interesan casos notables en que la función  $f$  no es necesariamente continua. Consideremos una región  $\mathcal{D} \subset \mathbb{R}^N$  acotada, no necesariamente rectangular, y una función acotada  $f : \mathcal{D} \rightarrow \mathbb{R}$ . Deseamos definir el número  $\int_{\mathcal{D}} f$  como aquél correspondiente al volumen de la región entre la base  $\mathcal{D}$  y el gráfico de  $f$ . Para ello, definamos

$$f_{\mathcal{D}}(x) = \begin{cases} f(x) & \text{si } x \in \mathcal{D} \\ 0 & \text{si } x \notin \mathcal{D}. \end{cases}$$

Sea  $\mathcal{R}$  un rectángulo tal que  $\mathcal{D} \subset \mathcal{R}$ . Decimos que  $f$  es integrable sobre  $\mathcal{D}$  si la función  $f_{\mathcal{D}}$  es integrable sobre  $\mathcal{R}$ , y en tal caso definimos

$$\int_{\mathcal{D}} f(x) dx := \int_{\mathcal{R}} f_{\mathcal{D}}(x) dx.$$

Esta definición es en realidad independiente del rectángulo  $\mathcal{R}$  que se escoja conteniendo a  $\mathcal{D}$ , pues las contribuciones a las sumas superiores e inferiores de cualquier región fuera de  $\mathcal{D}$  son nulas. Dejamos la verificación detallada de este hecho como un ejercicio.

La función  $f_{\mathcal{D}}$  puede ser ciertamente discontinua, aunque  $f$  lo sea, pero en la mayor parte de los casos que se enfrentan en la práctica sí será integrable.

Un caso fundamental es probablemente el de la función  $f \equiv 1$ . Si la integral  $\int_{\mathcal{D}} 1 dx$  está bien definida, le denominamos en general *volumen de  $\mathcal{D}$* :

$$V(\mathcal{D}) := \int_{\mathcal{D}} 1 dx.$$

En el caso  $N = 1$  le llamamos en realidad *longitud de  $\mathcal{D}$* , y si  $N = 2$  su *área*. Una pregunta natural por cierto es para qué tipo de regiones el volumen está bien definido, o más en general, sobre qué regiones pueden integrarse funciones, digamos, continuas. Discutiremos este tema en la subsección siguiente.

Antes de esto, probaremos algunas propiedades fundamentales de la integral de Riemann.

**Proposición 7.3.** *Supongamos que  $f$  y  $g$  son integrables sobre una región  $\mathcal{D}$ . Se tiene entonces que*

(a) *La función  $f + g$  también es integrable y*

$$\int_{\mathcal{D}} (f + g) = \int_{\mathcal{D}} f + \int_{\mathcal{D}} g.$$

(b) *Si  $\alpha \in \mathbb{R}$  entonces la función  $\alpha f$  es integrable y*

$$\int_{\mathcal{D}} \alpha f = \alpha \int_{\mathcal{D}} f.$$

(c) *Si  $f(x) \leq g(x)$  para todo  $x \in \mathcal{D}$  entonces*

$$\int_{\mathcal{D}} f \leq \int_{\mathcal{D}} g.$$

(d) *La función  $|f(x)|$  es integrable sobre  $\mathcal{D}$  y*

$$\left| \int_{\mathcal{D}} f \right| \leq \int_{\mathcal{D}} |f|.$$

(e) *Si  $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$ ,  $\mathcal{D}_1 \cap \mathcal{D}_2 = \emptyset$ , y  $f$  es integrable en  $\mathcal{D}_1$  y  $\mathcal{D}_2$ , entonces*

$$\int_{\mathcal{D}} f = \int_{\mathcal{D}_1} f + \int_{\mathcal{D}_2} f.$$

**Demostración.** Supondermos en las propiedades (a)-(d) que  $\mathcal{D} = \mathcal{R}$ , un rectángulo. Si no, basta aplicar los resultados obtenidos a las funciones  $f_{\mathcal{D}}$ ,  $g_{\mathcal{D}}$ . Sea  $\mathcal{S}$  un reticulado de  $\mathcal{R}$ . Veamos la propiedad (a). Tenemos que si  $R \in \mathcal{S}$  entonces

$$\inf_R f + \inf_R g \leq \inf_R (f + g) \leq \sup_R (f + g) \leq \sup_R f + \sup_R g.$$

Por lo tanto, deducimos que

$$I_{\mathcal{S}}(f) + I_{\mathcal{S}}(g) \leq I_{\mathcal{S}}(f + g) \leq S_{\mathcal{S}}(f + g) \leq S_{\mathcal{S}}(f) + S_{\mathcal{S}}(g). \quad (7.3)$$

Sean  $\mathcal{S}_n^1$  y  $\mathcal{S}_n^2$  sucesiones de reticulados tales que

$$S_{\mathcal{S}_n^1}(f) - I_{\mathcal{S}_n^1}(f) \rightarrow 0, \quad S_{\mathcal{S}_n^2}(g) - I_{\mathcal{S}_n^2}(g) \rightarrow 0.$$

Consideremos un reticulado  $\mathcal{S}_n$  más fino que, simultáneamente,  $\mathcal{S}_n^1$  y  $\mathcal{S}_n^2$ , por ejemplo aquél obtenido de las intersecciones de los elementos de ambos. Se tiene entonces también que

$$S_{\mathcal{S}_n}(f) - I_{\mathcal{S}_n}(f) \rightarrow 0, \quad S_{\mathcal{S}_n}(g) - I_{\mathcal{S}_n}(g) \rightarrow 0,$$

y por cierto que

$$\lim_{n \rightarrow \infty} S_{\mathcal{S}_n}(f) = \lim_{n \rightarrow \infty} I_{\mathcal{S}_n}(f) = \int_R f(x)dx, \quad \lim_{n \rightarrow \infty} S_{\mathcal{S}_n}(g) = \lim_{n \rightarrow \infty} I_{\mathcal{S}_n}(g) = \int_R g(x)dx.$$

Deducimos de las desigualdades (7.3) para  $\mathcal{S} = \mathcal{S}_n$  que

$$S_{\mathcal{S}_n}(f + g) - I_{\mathcal{S}_n}(f + g) \rightarrow 0,$$

y por lo tanto que  $f + g$  es integrable sobre  $\mathcal{R}$  y que

$$\int_R (f + g) = \lim_{n \rightarrow \infty} S_{\mathcal{S}_n}(f + g) = \lim_{n \rightarrow \infty} I_{\mathcal{S}_n}(f + g) = \int_R f, \quad + \int_R g.$$

Para probar la parte (b), observemos que si  $\alpha \geq 0$ , entonces

$$I_{\mathcal{S}}(\alpha f) = \alpha I_{\mathcal{S}}(f), \quad S_{\mathcal{S}}(\alpha f) = \alpha S_{\mathcal{S}}(f).$$

Por otra parte, si  $\alpha \leq 0$  tenemos que

$$I_{\mathcal{S}}(\alpha f) = \alpha S_{\mathcal{S}}(f), \quad S_{\mathcal{S}}(\alpha f) = \alpha I_{\mathcal{S}}(f).$$

De aquí la conclusión deseada se sigue en modo directo, en ambos casos.

Probemos ahora la parte (c). Gracias a las partes (a) y (b) tenemos que la función  $h(x) := g(x) - f(x)$  es integrable sobre  $\mathcal{R}$  y

$$\int_{\mathcal{R}} h(x)dx = \int_{\mathcal{R}} g(x)dx - \int_{\mathcal{R}} f(x)dx.$$

Por otra parte, si  $\mathcal{S}$  es cualquier reticulado, observamos de inmediato que como  $h(x) \geq 0$  para todo  $x \in \mathcal{R}$  entonces

$$0 \leq I_{\mathcal{S}}(h) \leq \int_{\mathcal{R}} h(x)dx.$$

Deducimos que  $\int_{\mathcal{R}} g(x)dx - \int_{\mathcal{R}} f(x)dx \geq 0$  y el resultado se concluye.

Probemos ahora (d). Consideremos un reticulado  $\mathcal{S}$  y  $R \in \mathcal{S}$ . Consideremos distintos casos. Si  $f \geq 0$ , obviamente

$$M_R(f) - m_R(f) = M_R(|f|) - m_R(|f|).$$

Si  $f \leq 0$ , entonces

$$M_R(|f|) - m_R(|f|) = M_R(-f) - m_R(-f)$$

Si  $f$  cambia de signo, tenemos entonces que

$$M_R(|f|) - m_R(|f|) \leq (M_R(f) - m_R(f)) + (M_R(-f) - m_R(-f))$$

Así, en cualquier caso, tenemos que

$$S_{\mathcal{S}}(|f|) - I_{\mathcal{S}}(|f|) \leq S_{\mathcal{S}}(f) - I_{\mathcal{S}}(f) + S_{\mathcal{S}}(-f) - I_{\mathcal{S}}(-f)$$

Por ende, si  $\mathcal{S}_n$  es una sucesión de reticulados tales que

$$S_{\mathcal{S}_n}(f) - I_{\mathcal{S}_n}(f) \rightarrow 0, \quad S_{\mathcal{S}_n}(-f) - I_{\mathcal{S}_n}(-f) \rightarrow 0,$$

se sigue que

$$S_{S_n}(|f|) - I_{S_n}(|f|) \rightarrow 0,$$

por lo tanto  $|f|$  es integrable. Además, tenemos que  $\pm f \leq |f|$ , por lo que a partir de las partes (b) y (c), se sigue que

$$\pm \int_{\mathcal{R}} f = \int_{\mathcal{R}} (\pm f) \leq \int_{\mathcal{R}} |f(x)| dx$$

y la demostración de (d) queda concluida. Finalmente, para la parte (e) nos basta observar que

$$f_{\mathcal{D}} = f_{\mathcal{D}_1} + f_{\mathcal{D}_2},$$

por lo cual el resultado deseado se sigue por linealidad.  $\square$

**7.2. Dónde integrar: Conjuntos Jordan-medibles.** Como dijimos anteriormente, una pregunta fundamental es qué tipo de regiones  $\mathcal{D}$  son apropiadas para calcular integrales, esto es, en particular bajo qué condiciones podemos calcular el volumen de  $\mathcal{D}$ ,  $V(\mathcal{D}) := \int_{\mathcal{D}} 1 dx$ . Podemos dar al menos una respuesta negativa: No todo conjunto es apropiado para esto. Consideremos por ejemplo el conjunto

$$\mathcal{D} = \{(x, y) \in \mathbb{Q} \times \mathbb{Q} / 0 \leq x, y \leq 1\}.$$

La función  $1_{\mathcal{D}}$  en efecto *no es integrable*, por ejemplo sobre  $\mathcal{R} = [0, 1] \times [0, 1]$  pues cualquier rectángulo de un reticulado  $\mathcal{S}$  de  $\mathcal{R}$  contendrá tanto puntos de  $\mathcal{D}$  como otros que no están en  $\mathcal{D}$  esto hace que para todo reticulado

$$I_{\mathcal{S}}(1_{\mathcal{D}}) = 0, \quad S_{\mathcal{S}}(1_{\mathcal{D}}) = 1,$$

y por ende no puede definirse (al menos mediante la integral de Riemann) el área de  $\mathcal{S}$ .

Necesitamos un concepto preliminar para encontrar una clase suficientemente amplia de conjuntos a los cuales se les puede asociar un volumen, el de *conjunto de medida nula*. Decimos que un conjunto  $A \subset \mathbb{R}^N$  acotado tiene medida nula, si para todo  $\varepsilon > 0$  existe una colección finita de rectángulos  $\{R_i\}_{i \in I}$  tal que

$$A \subset \bigcup_{i \in I} R_i, \quad \sum_{i \in I} V(R_i) < \varepsilon$$

A manera de ejemplo, consideremos una función  $g : \mathcal{R} \rightarrow \mathbb{R}$  continua, donde  $\mathcal{R}$  es un rectángulo en  $\mathbb{R}^N$  y su gráfico, esto es, el subconjunto de  $\mathbb{R}^{N+1}$  dado por

$$A = \{(x, y) / x \in \mathcal{R}, \quad y = g(x)\}.$$

Afirmamos que  $A$  tiene medida nula. En efecto, consideremos el reticulado uniforme dado por (7.2) de  $\mathcal{R}$ . Observemos entonces que

$$\mathcal{D} \subset \cup_i R_i \times [m_{R_i}(g), M_{R_i}(g)]$$

Como  $f$  es uniformemente continua sobre  $\mathcal{R}$  podemos escoger, dado  $\varepsilon > 0$ ,  $n$  suficientemente grande como para que

$$M_{R_i}(f) - m_{R_i}(f) < \frac{\varepsilon}{V(\mathcal{R})}$$

y de este modo,

$$\begin{aligned} \sum_i V(R_i \times [m_{R_i}(f), M_{R_i}(f)]) &= \sum_i V(R_i)(M_{R_i}(f) - m_{R_i}(f)) < \\ &\frac{\varepsilon}{V(\mathcal{R})} \sum_i V(R_i) = \varepsilon. \end{aligned}$$

Decimos que un conjunto  $\mathcal{D}$  en  $\mathbb{R}^N$  es *medible en el sentido de Jordan* o simplemente *Jordan-medible* si su frontera  $Fr(\mathcal{D})$  es de medida nula. Es fácil ver, y lo proponemos como un ejercicio, que toda unión de un número finito de conjuntos Jordan-medibles también lo es.

Tenemos la validez del siguiente resultado fundamental.

**Proposición 7.4.** *Sea  $\mathcal{D}$  un conjunto cerrado y acotado en  $\mathbb{R}^N$  Jordan-medible y  $f : \mathcal{D} \rightarrow \mathbb{R}$  una función continua. Entonces  $f$  es integrable sobre  $\mathcal{D}$ . En particular su volumen,  $V(\mathcal{D}) = \int_{\mathcal{D}} 1$  está bien definido.*

**Demostración.** Consideremos un rectángulo  $\mathcal{R}$  que contiene a  $\mathcal{D}$ , y dado  $\varepsilon > 0$ , una familia de rectángulos  $R_i$  cuya unión recubre a  $Fr(\mathcal{D})$  con  $\sum_i V(R_i) < \varepsilon$ . Completando esta familia de rectángulos a un reticulado del rectángulo  $\mathcal{R}$ , vemos que dado un reticulado  $\tilde{\mathcal{S}}$  arbitrario de  $\mathcal{R}$  podemos encontrar un reticulado  $\mathcal{S}$  más fino que  $\tilde{\mathcal{S}}$  tal que

$$\sum_{R \in \mathcal{S}, R \cap Fr(\mathcal{D}) \neq \emptyset} V(R) < \varepsilon.$$

Más aún, como  $f$  es uniformemente continua sobre  $\mathcal{D}$  podemos escoger el reticulado  $\mathcal{S}$  de modo que

$$\sum_{R \in \mathcal{S}, R \subset \text{int}(\mathcal{D})} [M_R(f) - m_R(f)]V(R) < \varepsilon V(\mathcal{R}).$$

Así, suponiendo que  $|f(x)| \leq M$  para todo  $x \in \mathcal{D}$ , tenemos que

$$I_{\mathcal{S}}(f_{\mathcal{D}}) \geq \sum_{R \subset \text{int}(\mathcal{D})} m_R(f)V(R) - M \sum_{R \in \mathcal{S}, R \cap Fr(\mathcal{D}) \neq \emptyset} V(R),$$

$$S_{\mathcal{S}}(f_{\mathcal{D}}) \leq \sum_{R \in \mathcal{S}, R \subset \text{int}(\mathcal{D})} M_R(f)V(R) + M \sum_{R \in \mathcal{S}, R \cap \text{Fr}(\mathcal{D}) \neq \emptyset} V(R),$$

de modo que

$$S_{\mathcal{S}}(f_{\mathcal{D}}) - I_{\mathcal{S}}(f_{\mathcal{D}}) \leq V(\mathcal{R})\varepsilon + 2M\varepsilon.$$

Escogiendo  $\varepsilon = \frac{1}{n}$ , podemos encontrar entonces una familia de reticulados  $\mathcal{S}_n$  tal que

$$S_{\mathcal{S}_n}(f_{\mathcal{D}}) - I_{\mathcal{S}_n}(f_{\mathcal{D}}) \rightarrow 0,$$

lo que demuestra que  $f_{\mathcal{D}}$  es integrable sobre  $\mathcal{R}$ . Esto concluye la demostración.  $\square$

A manera de ejemplo notable, consideremos dos funciones continuas  $g, h : [a, b] \rightarrow \mathbb{R}$  tales que  $g(x) \leq h(x)$  para todo  $x \in [a, b]$  y la región  $\mathcal{D} \subset \mathbb{R}^2$  definida por

$$\mathcal{D} = \{(x, y) / x \in [a, b], g(x) \leq y \leq h(x)\}$$

Observemos que la frontera de  $\mathcal{D}$  es la región

$$\text{Fr}(\mathcal{D}) = \{(x, y)/x \in [a, b], y = h(x)\} \cup$$

$$\{(x, y)/x \in [a, b], y = g(x)\} \cup \{(x, y)/x = a, b, h(x) \leq y = g(x)\}.$$

Los dos primeros conjuntos en la descomposición anterior son de medida nula, en virtud de lo ya demostrado. El tercero es la unión de dos segmentos de recta, conjuntos también de medida nula, de modo que la unión de todos estos lo es. Así  $\mathcal{D}$  arriba dado es Jordan-medible. Evidentemente, también es es Jordan-medible una región de la forma

$$\mathcal{D} = \{(x, y) / y \in [c, d], p(y) \leq x \leq q(y)\}$$

para funciones  $p$  y  $q$  continuas. En realidad, la casi totalidad de los ejemplos que consideraremos corresponden o a regiones de esta forma, o a regiones que pueden obtenerse como una unión finita de regiones de este tipo. Consideremos por ejemplo la región anular

$$\mathcal{D} = \{(x, y) / 1 \leq x^2 + y^2 \leq 4\}.$$

Podemos entonces escribir

$$\mathcal{D} = \{(x, y) / x \in [-1, 1], \sqrt{1-x^2} \leq y \leq \sqrt{4-x^2}\} \cup$$

$$\{(x, y) / x \in [-1, 1], -\sqrt{4-x^2} \leq y \leq -\sqrt{1-x^2}\} \cup$$

$$\{(x, y) / y \in [-\sqrt{3}\sqrt{3}], -\sqrt{4-y^2} \leq x \leq -1\} \cup$$

$$\{(x, y) / y \in [-\sqrt{3}\sqrt{3}], 1 \leq x \leq \sqrt{4-y^2}\}.$$



En modo inductivo, podemos obtener una amplia clase de regiones Jordan-medibles en dimensiones mayores: Sea  $A$  una región Jordan-medible cerrada y acotada en  $\mathbb{R}^N$ . Consideremos una región en  $\mathbb{R}^{N+1}$  de la forma

$$\mathcal{D} = \{(x, y) \in \mathbb{R}^{N+1} / x \in A, \quad h(x) \leq y \leq g(x)\} \quad (7.4)$$

donde  $h$  y  $g$  son funciones continuas con  $h \leq g$  en  $A$ . Proponemos como un ejercicio (no-trivial) demostrar que este conjunto es en efecto Jordan-medible en  $\mathbb{R}^{N+1}$ . Por cierto una región construida a partir de una unión finita de tales regiones será también Jordan-medible, donde las funciones consideradas pueden ser de cualquier selección de  $N - 1$  variables (no necesariamente las primeras). Consideremos por ejemplo la bola en  $\mathbb{R}^3$

$$\mathcal{D} = \{(x, y, z) / x^2 + y^2 + z^2 < 1\}.$$

Podemos escribir este conjunto en la forma (7.4) pues

$$\mathcal{D} = \{(x, y, z) \in \mathbb{R}^3 / (x, y) \in A, \quad -\sqrt{1 - x^2 - y^2} \leq z \leq \sqrt{1 - x^2 - y^2}\},$$

donde  $A$  es un disco en  $\mathbb{R}^2$ , el que a su vez puede describirse como

$$A = \{(x, y) \in \mathbb{R}^2 / x \in [-1, 1], \quad -\sqrt{1 - x^2} \leq y \leq \sqrt{1 - x^2}\}.$$

$A$  es entonces medible en  $\mathbb{R}^2$  de acuerdo a lo ya demostrado, y por ende  $\mathcal{D}$  también lo es.

**7.3. Cálculo de integrales: El Teorema de Fubini.** El problema que queremos abordar a continuación es el del cálculo de integrales múltiples. La sola definición, por cierto es de difícil aplicación en el cálculo explícito. Afortunadamente, en casos concretos este cálculo se reduce al de *integrales iteradas* en el modo que explicamos a continuación. Consideremos el rectángulo  $R = [a, b] \times [c, d] \subset \mathbb{R}^2$  y una función  $f(x, y)$  continua en  $R$ . Consideremos para números naturales  $n, m$ , el reticulado de  $R$  de elementos

$$R_{kj} = [a + \frac{k-1}{n}(b-a), a + \frac{k}{n}(b-a)] \times [c + \frac{j-1}{m}(d-c), c + \frac{j}{m}(d-c)],$$

para  $k = 1, \dots, n$ ,  $j = 1, \dots, m$ . Como  $f$  es uniformemente continua sobre  $R$ , tenemos que, dado  $\varepsilon > 0$ , existe  $n_0$  tal que para todo  $n, m \geq n_0$ ,

$$M_{R_{kj}}(f) - m_{R_{kj}}(f) \leq \frac{\varepsilon}{V(R)}.$$

Por ende tenemos que

$$-\varepsilon + \int_R f \leq I_S(f) \leq \frac{(b-a)(d-c)}{nm} \sum_{k=1}^n \sum_{j=1}^m f(a + \frac{k}{n}(b-a), c + \frac{j}{m}(d-c)) \leq$$

$$S_S(f) \leq \varepsilon + \int_R f,$$

y por lo tanto

$$\left| \int_R f - \frac{b-a}{n} \sum_{k=1}^n \frac{d-c}{m} \sum_{j=1}^m f\left(a + \frac{k}{n}(b-a), c + \frac{j}{m}(d-c)\right) \right| < \varepsilon$$

para todo  $m, n > n_0(\varepsilon)$ . Haciendo tender  $m$  a infinito en la desigualdad anterior, obtenemos que

$$\left| \int_R f - \frac{b-a}{n} \sum_{k=1}^n \int_c^d f\left(a + \frac{k}{n}(b-a), y\right) dy \right| \leq \varepsilon,$$

y, en modo simétrico, haciendo tender  $n$  a infinito,

$$\left| \int_R f - \frac{d-c}{m} \sum_{j=1}^m \int_a^b f\left(x, c + \frac{j}{m}(d-c)\right) dx \right| \leq \varepsilon$$

Así, haciendo tender respectivamente  $m$  y  $n$  a infinito obtenemos que

$$\begin{aligned} \left| \int_R f - \int_a^b \left( \int_c^d f(x, y) dx \right) dy \right| &\leq \varepsilon, \\ \left| \int_R f - \int_c^d \left( \int_a^b f(x, y) dy \right) dx \right| &\leq \varepsilon \end{aligned}$$

y como  $\varepsilon$  es arbitrario,

$$\int_R f = \int_a^b \left( \int_c^d f(x, y) dx \right) dy = \int_c^d \left( \int_a^b f(x, y) dy \right) dx.$$

Este resultado se conoce como *Teorema de Fubini* y es la herramienta principal en el cálculo de integrales de funciones de varias variables, pues reduce su cálculo al de varias integrales iteradas de funciones de una variable. Este resultado es en realidad válido en total generalidad, en el modo siguiente.

**Teorema 7.2.** (*Teorema de Fubini*) Sean  $R_1 \subset \mathbb{R}^N$ ,  $R_2 \subset \mathbb{R}^m$ ,  $R = R_1 \times R_2 \subset \mathbb{R}^{N+m}$  y  $f : R \rightarrow \mathbb{R}$ , una función integrable, y tal que las funciones

$$x \in R_1 \mapsto \int_{R_2} f(x, y) dy, \quad y \in R_2 \mapsto \int_{R_1} f(x, y) dx,$$

están bien definidas y son integrables. Entonces se tiene la validez de las igualdades

$$\int_R f = \int_{R_1} \left( \int_{R_2} f(x, y) dy \right) dx = \int_{R_2} \left( \int_{R_1} f(x, y) dx \right) dy.$$

Este resultado puede aplicarse en modo iterado para deducir que si  $R = [a_1, b_1] \times \cdots \times [a_N, b_N]$  y  $f : R \rightarrow \mathbb{R}$ , entonces si todas las integrales que siguen están bien definidas se tiene que

$$\int_R f = \int_{a_1}^{b_1} \left( \int_{a_2}^{b_2} \cdots \left( \int_{a_N}^{b_N} f(x_1, \dots, x_N) dx_N \right) dx_{N-1} \right) \cdots dx_1 .$$

donde en realidad el orden en las integraciones sucesivas puede alterarse como se desee. Cuando se quiera enfatizar el orden de integración es conveniente escribir

$$\int_R f = \int_{a_1}^{b_1} dx_1 \int_{a_2}^{b_2} \cdots dx_{N-1} \int_{a_N}^{b_N} f(x_1, \dots, x_N) dx_N,$$

y si no, escribimos

$$\int_R f = \int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_N}^{b_N} f(x_1, \dots, x_N) dx_1 \cdots dx_N.$$

**Ejemplo 7.2.** Consideremos por ejemplo la integral

$$I = \int \int_{[0,1] \times [0,2]} xy^2 dx dy.$$

De acuerdo al teorema anterior,

$$\begin{aligned} I &= \int_0^1 dx \int_0^2 xy^2 dy = \int_0^1 x \frac{y^3}{3} \Big|_{y=0}^{y=2} dx = \\ &= \frac{8}{3} \int_0^1 x dx = \frac{4}{3}. \end{aligned}$$

Podemos también calcular  $I$  como

$$\begin{aligned} I &= \int_0^2 dy \int_0^1 xy^2 dx = \int_0^2 y^2 \frac{x^2}{2} \Big|_{x=0}^{x=1} dy = \\ &= \frac{1}{2} \int_0^2 y^2 dy = \frac{4}{3}. \end{aligned}$$

**Ejemplo 7.3.** El teorema de Fubini también es útil para calcular integrales sobre regiones que no son necesariamente rectángulos. Por ejemplo consideremos la región triangular

$$\mathcal{D} = \{(x, y) / 0 \leq x \leq 1, 0 \leq y \leq x\}$$

Se pide calcular la integral

$$I = \int \int_{\mathcal{D}} f(x, y) dx dy.$$

Tenemos que para un rectángulo  $\mathcal{R}$  que contiene a  $\mathcal{D}$ , digamos  $\mathcal{R} = [0, 1] \times [0, 1]$ ,

$$I = \int \int_{\mathcal{R}} f_{\mathcal{D}}(x, y) dx dy = \int_0^1 dx \int_0^1 f_{\mathcal{D}}(x, y) dy.$$

Recordemos que  $f_{\mathcal{D}}(x, y) = f(x, y)$  si  $(x, y) \in \mathcal{D}$ ,  $f_{\mathcal{D}}(x, y) = 0$  si no. Entonces, para  $x$  dado en  $[0, 1]$ ,  $f_{\mathcal{D}}(x, y) = f(x, y)$  si  $0 \leq y \leq x$ ,  $= 0$  si  $x < y \leq 1$ . Así,

$$\int_0^1 f_{\mathcal{D}}(x, y) dy = \int_0^x f(x, y) dy,$$

y por lo tanto

$$I = \int_0^1 dx \int_0^x f(x, y) dy.$$

Notemos que, calculando en el orden inverso,

$$I = \int_0^1 dy \int_0^1 f_{\mathcal{D}}(x, y) dx,$$

y tenemos que para  $y$  dado en  $[0, 1]$ ,  $f_{\mathcal{D}}(x, y) = f(x, y)$  si  $y \leq x \leq 1$ ,  $= 0$  en caso contrario. Por lo tanto, podemos también expresar

$$I = \int_0^1 dy \int_y^1 f(x, y) dx.$$

Consideremos por ejemplo  $f(x, y) = x^2 + y^2$ . Entonces

$$I = \int_0^1 dx \int_0^x (x^2 + y^2) dy = \frac{4}{3} \int_0^1 x^3 dx = \frac{1}{3}.$$

En el orden inverso,

$$\begin{aligned} I &= \int_0^1 dy \int_y^1 (x^2 + y^2) dx = \int_0^1 dy \left( \frac{x^3}{3} + xy^2 \right) \Big|_{x=y}^{x=1} = \\ &= \int_0^1 \left( \frac{1-y^3}{3} + (1-y)y^2 \right) dy = \\ &= \frac{1}{3} - \frac{1}{12} + \frac{1}{3} - \frac{1}{4} = \frac{1}{3}. \end{aligned}$$

Geométricamente, el número calculado corresponde al volumen comprendido entre el triángulo  $\mathcal{D}$  y el paraboloide  $z = x^2 + y^2$ .

Notemos, que más en general, si  $A \subset \mathbb{R}^2$  y  $f : A \rightarrow \mathbb{R}^+$  definimos

$$\mathcal{D} = \{(x, y, z) / (x, y) \in A, 0 \leq z \leq f(x, y)\}$$

entonces, bajo las hipótesis necesarias,

$$v(\mathcal{D}) = \int_{\mathcal{R}} 1_{\mathcal{D}}$$

donde  $\mathcal{R} = [a_1, b_1] \times [a_2, b_2] \times [a_3, b_3]$  contiene a  $\mathcal{D}$ . Así, por Fubini obtenemos

$$\begin{aligned} v(\mathcal{D}) &= \int_{[a_1, b_1] \times [a_2, b_2]} dx dy \int_{a_3}^{b_3} 1_{\mathcal{D}}(x, y, z) dz = \\ &= \int \int_A dx dy \int_0^{f(x, y)} dz = \int \int_A f(x, y) dx dy. \end{aligned}$$

**Ejemplo 7.4.** Se pide calcular el volumen del tetraedro

$$\mathcal{D} = \{(x, y, z) / x, y, z \geq 0, x + y + z \leq 1\}$$

Podemos describir esta región en el modo siguiente:

$$\mathcal{D} = \{(x, y, z) / (x, y) \in A, 0 \leq z \leq 1 - x - y\}.$$

donde  $A$  es el triángulo

$$A = \{(x, y) / x \in [0, 1], 0 \leq y \leq 1 - x\}.$$

Entonces

$$\begin{aligned} V(\mathcal{D}) &= \int \int_A (1 - x - y) dx dy = \int_0^1 dx \int_0^{1-x} (1 - x - y) dy = \\ &= \frac{1}{2} \int_0^1 (1 - x)^2 dx = \frac{1}{6} \end{aligned}$$

**Ejemplo 7.5.** Calculemos el volumen de la porción de la bola  $B(0, 1)$  en  $\mathbb{R}^3$  comprendida dentro del primer octante, esto es

$$\mathcal{D} = \{(x, y, z) / x^2 + y^2 + z^2 \leq 1, x, y, z \geq 0\}.$$

Para ello, describimos  $\mathcal{D}$  razonando en el modo siguiente: El rango de variación de la coordenada  $x$  para los puntos  $(x, y, z) \in \mathcal{D}$  es  $0 \leq x \leq 1$ . Ahora, para  $x$  dado en este rango, la coordenada  $y$  tiene por rango total de variación  $0 \leq y \leq \sqrt{1 - x^2}$ , y finalmente para  $(x, y)$  dado en estos rangos,  $z$  puede variar en  $0 \leq z \leq \sqrt{1 - x^2 - y^2}$ . Así, tenemos la descripción

$$\mathcal{D} = \{(x, y, z) / 0 \leq x \leq 1, 0 \leq y \leq \sqrt{1 - x^2}, 0 \leq z \leq \sqrt{1 - x^2 - y^2}\}$$

Entonces

$$\begin{aligned} v(\mathcal{D}) &= \int \int \int_{\mathcal{D}} 1 dx dy dz = \\ &= \int_0^1 dx \int_0^{\sqrt{1-x^2}} dy \int_0^{\sqrt{1-x^2-y^2}} dz = \\ &= \int_0^1 dx \int_0^{\sqrt{1-x^2}} \sqrt{1-x^2-y^2} dy \end{aligned}$$

Calculemos la integral interior mediante el cambio de variables  $y = t\sqrt{1-x^2}$ , de modo que  $dy = dt\sqrt{1-x^2}$  y entonces

$$\int_0^{\sqrt{1-x^2}} \sqrt{1-x^2-y^2} dy = (1-x^2) \int_0^1 \sqrt{1-t^2} dt$$

Haciendo  $t = \sin u$  obtenemos

$$\int_0^1 \sqrt{1-t^2} dt = \int_0^{\frac{\pi}{2}} \cos^2 t dt = \frac{\pi}{4},$$

y entonces

$$v(\mathcal{D}) = \frac{\pi}{4} \int_0^1 (1-x^2) dx = \frac{\pi}{6}.$$

#### 7.4. Cálculo de integrales: El Teorema del cambio de variables.

El cálculo del volumen de una porción esférica en el ejemplo anterior, si bien fue posible, resultó relativamente complicado. La razón es que no es del todo natural intentar describir una región de esta clase directamente en coordenadas cartesianas  $xyz$ . El teorema del cambio de variables nos entrega una herramienta útil de cálculo en caso que la región de integración y/o la función involucrada pueden ser expresadas en modo más simple mediante la introducción de coordenadas alternativas.

Consideremos a manera de ejemplo, las coordenadas polares en  $\mathbb{R}^2$ ,  $(r, \theta)$ ,  $0 < r < \infty$ ,  $\theta \in ]0, 2\pi[$ . Consideremos una región  $\mathcal{D}$  en  $\mathbb{R}^2$  y su representación en coordenadas polares,

$$\tilde{\mathcal{D}} = \{(r, \theta) / (r \cos \theta, r \sin \theta) \in \mathcal{D}\}.$$

Notemos que si a partir de un punto de coordenadas  $(r_0, \theta_0)$  incrementamos la variable  $r$  en magnitudes pequeñas  $\Delta r$  y  $\theta$  en  $\Delta \theta$  entonces el volumen de la célula correspondiente a

$$(r, \theta) \in [r_0, r_0 + \Delta r] \times [\theta_0, \theta_0 + \Delta \theta],$$

corresponde aproximadamente al de un rectángulo de lados  $\Delta r$  y  $r_0 \Delta \theta$ . Así, si llenamos la región  $\mathcal{D}$  con un reticulado de pequeñas celdas de este tipo, debiésemos tener que

$$\int_{\mathcal{D}} f \sim \sum_{r_0, \theta_0} f(r_0 \cos \theta_0, r_0 \sin \theta_0) r_0 \Delta r \Delta \theta,$$

en otras palabras, es dable esperar que bajo hipótesis razonables,

$$\int_{\mathcal{D}} f(x, y) dx dy = \int_{\mathcal{D}'} f(r \cos \theta, r \sin \theta) r dr d\theta.$$

Este es efectivamente el caso, y veremos a continuación como enfocar esto en modo más sistemático. Consideremos dos regiones abiertas y acotadas  $\mathcal{D}$  y  $\mathcal{D}'$  y una función biyectiva y de clase  $C^1$   $T : \mathcal{D}' \rightarrow \mathcal{D}$ . En otras palabras,  $T$  es inyectiva y  $\mathcal{D} = T(\mathcal{D}')$ .

Notemos primero que si  $R = [u_0, u_0 + h_1] \times [v_0, v_0 + h_2]$  y  $T$  es una aplicación lineal afín de la forma

$$T(u, v) = T(u_0, v_0) + A \begin{bmatrix} u - u_0 \\ v - v_0 \end{bmatrix},$$

con  $A$  una matriz invertible  $2 \times 2$ , que escribimos

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = [a_{.1} \ a_{.2}].$$

El conjunto  $T(R)$  está dado por

$$T(R) = \{T(u_0, v_0) + t a_{.1} s a_{.2} \mid 0 \leq t \leq h_1, 0 \leq s \leq h_2\}.$$

Esto es, por un paralelogramo que tiene por lados los vectores  $h_1 a_{.1}$  y  $h_2 a_{.2}$ . El área de este paralelogramo, recordemos, es la norma del producto cruz de estos dos vectores, pensados como vectores de  $\mathbb{R}^3$  con tercera coordenada 0. Vemos entonces de inmediato que el área de  $T(R)$  está dada por Vemos de inmediato que

$$V(T(R)) = |\det(A)| V(R).$$

Supongamos ahora que  $T$  no es afín sino una aplicación continuamente diferenciable. Entonces si los lados  $h_1$  y  $h_2$  del rectángulo  $R$  son pequeños, podemos aproximar  $T(x, y)$  en  $R$  por una aplicación afín

$$T(u, v) \sim T(u_0, v_0) + T'(u_0, v_0) \begin{bmatrix} u - u_0 \\ v - v_0 \end{bmatrix},$$

De modo que si  $f$  es una función continua, tenemos

$$\int \int_{T(R)} f(x, y) dx dy \sim f(T(u_0, v_0)) V(T(R)) \sim$$

$$f(T(u_0, v_0))|\det(T'(u_0, v_0))|V(R) \sim \int \int_R f(T(u, v))|\det(T'(u, v))|dudv.$$

De este modo, suponiendo ahora que  $\mathcal{D} = T(\mathcal{D}')$  y que  $\mathcal{D}'$  se aproxima por un reticulado formado por una colección de rectángulos pequeños  $\{R_i\}_{i \in I}$ , tendremos entonces que  $\mathcal{D}$  se aproxima por un reticulado de "casi paralelógramos"  $\{T(R_i)\}_{i \in I}$ , de modo que, suponiendo que  $T$  es inyectiva, las imágenes de estos rectángulos no se intersecten en sus interiores,

$$\begin{aligned} \int \int_{\mathcal{D}} f(x, y) dx dy &\sim \sum_{i \in I} \int \int_{T(R_i)} f(x, y) dx dy \sim \\ &\int \int_{R_i} f(T(u, v))|\det(T'(u, v))|dudv \sim \\ &\int \int_{\mathcal{D}'} f(T(u, v))|\det(T'(u, v))|dudv \end{aligned}$$

La igualdad de las cantidades primera y última se denomina el *Teorema del cambio de variables*, y es válida en realidad para una integral múltiple en cualquier dimensión. Vale la pena recordar, su versión ya conocida por el lector en funciones de una variable: si  $T : [a, b] \rightarrow \mathbb{R}$  es continuamente diferenciable y  $T'(u) > 0$ , entonces para  $f$  continua se tiene que

$$\int_a^b f(T(u))T'(u)du = \int_{T(a)}^{T(b)} f(x)dx$$

Si  $T'(u) < 0$  la misma fórmula es válida, y puede escribirse como

$$\int_a^b f(T(u))(-T'(u))du = \int_{T(b)}^{T(a)} f(x)dx .$$

así, siempre que  $T'(u) \neq 0$  para todo  $u \in [a, b]$ , tenemos que

$$\int_{]a, b[} f(T(u))|T'(u)|du = \int_{T([a, b])} f(x)dx .$$

Notemos que la inclusión estricta de los extremos del intervalo  $[a, b]$  o no, no altera el valor de la integral. Enunciemos ahora el teorema en su versión general.

**Teorema 7.3.** (*Teorema del Cambio de Variables*). Sea  $\Omega \subset \mathbb{R}^N$  un abierto y  $T : \Omega \rightarrow \mathbb{R}^N$  una función de clase  $C^1$ . Sea  $\mathcal{D}'$  una región abierta y acotada con  $\text{Adh}(\mathcal{D}') \subset \Omega$ , y supongamos además que  $T$  es inyectiva en  $\mathcal{D}'$ , que la matriz  $T'(u)$  es invertible para todo  $u \in \mathcal{D}'$  y



que  $\mathcal{D} = T(\mathcal{D}')$  es un abierto. Sea  $f : \bar{\mathcal{D}} \rightarrow \mathbb{R}$  una función continua. Se tiene entonces la validez de la igualdad

$$\int_{\mathcal{D}} f(x) dx = \int_{\mathcal{D}'} f(T(u)) |\det(T'(u))| du.$$

**Ejemplo 7.6.** Calcular la integral

$$\int \int_{\mathcal{D}} (x + y) dx dy$$

donde

$$\mathcal{D} = \{(x, y) / 1 - x \leq y \leq 2 - x, y + 1 \geq 2x \leq 5 + y\}.$$

Escribamos

$$u = x + y, \quad v = 2x - y$$

y definimos entonces la transformación de modo que

$$T(u, v) = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \frac{1}{3}(u + v) \\ \frac{1}{2}(u - 2v) \end{bmatrix}.$$

De este modo, tenemos que

$$T'(u, v) = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & -1 \end{bmatrix}.$$

Así

$$|\det(T'(u, v))| = \frac{1}{6}.$$

La región  $\mathcal{D}$  queda entonces descrita en términos de las coordenadas  $(u, v)$  como

$$\mathcal{D}' = \{(u, v) / 1 \leq u \leq 2, 1 \leq v \leq 5\}$$

y de acuerdo al teorema del cambio de variables,

$$\begin{aligned} \int \int_{\mathcal{D}} (x + y) dx dy &= \int \int_{\mathcal{D}'} u \frac{1}{6} du dv = \\ &= \frac{1}{6} \int_1^5 dv \int_1^2 u du = \frac{4}{6} \frac{3}{2} = 1. \end{aligned}$$

**Ejemplo 7.7.** Apliquemos este resultado para realizar el cálculo del volumen del primer octante de la esfera  $B(0, R)$  en  $\mathbb{R}^3$ , esto es aquél de la región  $\{x^2 + y^2 + z^2 < R^2, x, y, z \geq 0\}$ . Este volumen es aquel bajo el gráfico de la función  $z = f(x, y) = \sqrt{R^2 - x^2 - y^2}$  sobre la región del plano  $x$ - $y$ ,

$$\mathcal{D} = \{(x, y) / x, y \geq 0, x^2 + y^2 < R^2\}.$$

En otras palabras, queremos calcular la cantidad

$$I = \int \int_{\mathcal{D}} f(x, y) \, dx dy.$$

Consideremos coordenadas polares

$$T(r, \theta) = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} r \cos \theta \\ r \sin \theta \end{bmatrix}.$$

De este modo tenemos que (salvo por una zona de medida nula: la periferia y el origen),  $\mathcal{D} = T(\mathcal{D}')$  donde  $\mathcal{D}'$  es simplemente

$$\mathcal{D}' = \{(r, \theta) / 0 < r < R, 0 < \theta < \frac{\pi}{2}\}$$

Notemos que

$$T'(r, \theta) = \begin{bmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{bmatrix}.$$

De modo que  $\det(T'(r, \theta)) = r$ , y entonces, de acuerdo al teorema del cambio de variables,

$$\begin{aligned} \int \int_{\mathcal{D}} f(x, y) \, dx dy &= \int \int_{\mathcal{D}'} f(r \cos \theta, r \sin \theta) r dr d\theta = \\ &= \int_0^{\frac{\pi}{2}} \int_0^R \sqrt{R^2 - r^2 \cos^2 \theta - r^2 \sin^2 \theta} r dr d\theta = \\ &= \frac{\pi}{2} \int_0^R \sqrt{R^2 - r^2} r dr = \frac{\pi}{6} R^3. \end{aligned}$$

Notemos que esto está en acuerdo, multiplicando por 8, con la fórmula familiar

$$V(B(0, R)) = \frac{4}{3} \pi R^3.$$

Viene al caso mencionar que es lenguaje común decir que en coordenadas polares el *elemento de área* está dado por  $r dr d\theta$ , mientras que en coordenadas cartesianas lo está por  $dx dy dz$ .

**Ejemplo 7.8.** Consideremos el círculo  $(x - a)^2 + y^2 < a^2$ . Se pide calcular el área de la región  $\mathcal{D}$  interior al círculo, comprendida entre las rectas  $y = x$  e  $y = -x$ . Para resolver este problema, expresaremos la región  $\mathcal{D}$  en términos de coordenadas polares *relativas al origen*. La primera observación es que la periferia del círculo  $(x - a)^2 + y^2 = a^2$  queda expresada en coordenadas polares como

$$(r \cos \theta - a)^2 + (r \sin \theta)^2 = a^2,$$

esto es,  $r^2 - 2ar \cos \theta = 0$  o sea la curva  $r = 2a \cos \theta$ . La región en cuestión queda entonces descrita como

$$\mathcal{D}' = \{(r, \theta) \mid -\frac{\pi}{4} < \theta < \frac{\pi}{4}, 0 < r < 2a \cos \theta\}$$

y entonces su área es

$$\begin{aligned} V(\mathcal{D}) &= \int \int_{\mathcal{D}} 1 dx dy = \int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} \int_0^{2a \cos \theta} 1 r dr d\theta = \\ &= \int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} 2a^2 \cos^2 \theta d\theta = \\ &= a^2 \int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} (1 + \cos 2\theta) d\theta = a^2 \left( \frac{\pi}{2} + 1 \right). \end{aligned}$$

**Ejemplo 7.9.** Consideremos un cono de revolución de altura  $h$  y radio  $R$  con vértice en el origen cuya densidad de masa está dada por  $\rho(x, y, z) = z$ . Se pide calcular la masa total del cono. para este problema es conveniente introducir un sistema de coordenadas que extiende las coordenadas polares del plano con la coordenada  $z$ , las así llamadas, *coordenadas cilíndricas*, definidas en el modo siguiente.

$$T(r, \theta, z) = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} r \cos \theta \\ r \sin \theta \\ z \end{bmatrix}.$$

Notemos que, ahora,

$$T'(r, \theta, z) = \begin{bmatrix} \cos \theta & -r \sin \theta & 0 \\ \sin \theta & r \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

De modo que  $\det(T'(r, \theta, z)) = r$ . Decimos entonces, que para coordenadas cilíndricas, el elemento de volumen está dado por  $r dr d\theta dz$ . Esto tiene una interpretación geométrica simple nuevamente, pues si a partir de un punto de coordenadas  $(r, \theta, z)$  se hacen variar estas variables respectivamente en magnitudes  $dr$ ,  $d\theta$  y  $dz$  se obtiene un rectángulo infinitesimal de lados:  $dr$ ,  $r d\theta$  y  $dz$ .

Usando estas coordenadas, podemos describir el cono,  $\mathcal{D}$ , salvo por un conjunto de medida nula, como la región

$$\mathcal{D}' = \{(r, \theta, z) \mid r \in ]0, R[, \theta \in ]0, 2\pi[, \frac{h}{R}r < z < h\}.$$

No necesitamos trabajar una representación explícita del dominio  $\mathcal{D}$  en coordenadas cartesianas. Tenemos que

$$\begin{aligned} \int \int \int_{\mathcal{D}} z dx dy dz &= \int \int \int_{\mathcal{D}'} z r dr d\theta dz = \\ &= \int_0^{2\pi} d\theta \int_0^R r dr \int_{\frac{h}{R}r}^h z dz = \\ &= \pi h^2 \int_0^R \left(1 - \frac{r^2}{R^2}\right) r dr = \frac{\pi}{4} h^2 R^2. \end{aligned}$$

Notemos también que el volumen del cono está dado por

$$\begin{aligned} V(\mathcal{D}) &= \int \int \int_{\mathcal{D}'} 1 r dr d\theta dz = \int_0^{2\pi} d\theta \int_0^R r dr \int_{\frac{h}{R}r}^h dz = \\ &= 2\pi h \int_0^R r \left(1 - \frac{r}{R}\right) dr = \frac{\pi}{3} h R^2, \end{aligned}$$

otra fórmula familiar.

En el ejemplo siguiente introducimos otro sistema de coordenadas útil que extiende las polares a  $\mathbb{R}^3$ , las *coordenadas esféricas*.

**Ejemplo 7.10.** Sea  $\mathcal{D} = \{x^2 + y^2 + z^2 < R^2\}$  Consideremos a continuación el problema de calcular la integral

$$I = \int_{\mathcal{D}} (x^2 + y^2) dx dy dz$$

Consideremos a continuación las *coordenadas esféricas* definidas como

$$T(r, \theta, \phi) = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} r \cos \theta \sin \phi \\ r \sin \theta \sin \phi \\ r \cos \phi \end{bmatrix}.$$

Con esta definición,  $r$  representa la distancia del punto  $(x, y, z)$  al origen:  $r = \sqrt{x^2 + y^2 + z^2}$  mientras que  $\theta$  es el ángulo polar en el plano  $x$ - $y$  como antes,  $\theta \in [0, 2\pi]$  y ahora  $\phi$  es el ángulo medido desde el eje  $z$  al vector  $(x, y, z)$ ,  $\phi \in [0, \pi]$ , de modo que el  $r$  polar antiguo, el largo de  $(x, y)$  está dado ahora por  $r \sin \phi$ .

Tenemos que, ahora,

$$T'(r, \theta, \phi) = \begin{bmatrix} \cos \theta \sin \phi & -r \sin \theta \sin \phi & r \cos \theta \cos \phi \\ \sin \theta \sin \phi & r \cos \theta \sin \phi & r \sin \theta \cos \phi \\ \cos \phi & 0 & -r \sin \phi \end{bmatrix}.$$

De modo que

$$|\det(T'(r, \theta, \phi))| = r^2 \sin \phi.$$

El elemento de volumen en coordenadas esféricas está entonces dado por  $r^2 \sin \phi dr d\theta d\phi$ . Esto tiene una interpretación geométrica simple nuevamente, pues si a partir de un punto de coordenadas  $(r, \theta, \phi)$  se hacen variar estas variables respectivamente en magnitudes  $dr$ ,  $d\theta$  y  $d\phi$  se obtiene un rectángulo infinitesimal de lados:  $dr$ ,  $r \sin \phi d\theta$  y  $rd\phi$ .

Usando estas coordenadas, podemos describir la esfera  $\mathcal{D}$ , salvo por un conjunto de medida nula, como la región

$$\mathcal{D}' = \{(r, \theta, \phi) / r \in ]0, R[, \theta \in ]0, 2\pi[, \phi \in ]0, \pi[ \},$$

De modo que

$$\begin{aligned} I &= \int \int \int_{\mathcal{D}} (x^2 + y^2) dx dy dz = \\ &= \int_0^R \int_0^{2\pi} \int_0^\pi (r^2 \sin^2 \phi) r^2 \sin \phi dr d\theta d\phi = \\ &= \frac{2}{5} \pi R^4 \int_0^\pi \sin^3 \phi d\phi = \frac{2}{5} \pi R^4 \int_0^\pi (1 - \cos^2 \phi) \sin \phi d\phi = \frac{8}{15} \pi R^4. \end{aligned}$$

**Ejemplo 7.11.** Consideremos el toro de revolución  $\mathcal{D}$ , constituido por el sólido obtenido al rotar el disco de centro  $(b, 0, 0)$  y de radio  $a$  con  $b > a$ . Las siguientes *coordenadas toroidales* describen apropiadamente a este sólido:

$$T(r, \theta, \phi) = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} (b + r \cos \phi) \cos \theta \\ (b + r \cos \phi) \sin \theta \\ r \sin \phi \end{bmatrix}.$$

Tenemos ahora que

$$T'(r, \theta, \phi) = \begin{bmatrix} \cos \phi \cos \theta & -(b + r \cos \phi) \sin \theta & -r \sin \phi \cos \theta \\ \cos \phi \sin \theta & (b + r \cos \phi) \cos \theta & -r \sin \phi \sin \theta \\ \sin \phi & 0 & r \cos \phi \end{bmatrix}.$$

De modo que

$$|\det(T'(r, \theta, \phi))| = r(b + r \cos \phi)$$

y, por ejemplo, el volumen del toro está dado por

$$V(\mathcal{D}) = \int_0^{2\pi} \int_0^{2\pi} \int_0^a r(b + r \cos \phi) d\theta d\phi dr = \pi^2 b a^2 = (2\pi b)(\pi a^2),$$

que corresponde al área del disco multiplicada por la longitud del círculo central.

**Ejemplo 7.12.** Una aplicación interesante del teorema del cambio de variables es el cálculo de una integral clásica, especialmente relevante en Probabilidades. Proponemos el cálculo de la integral impropia

$$I = \int_{-\infty}^{\infty} e^{-x^2} dx .$$

Esto no es tan sencillo puesto que la función  $e^{-x^2}$  no posee una primitiva expresable en términos de funciones elementales. El truco es expresar la cantidad  $I$  de la siguiente manera

$$I^2 = \left( \int_{-\infty}^{\infty} e^{-x^2} dx \right) \left( \int_{-\infty}^{\infty} e^{-y^2} dy \right) = \lim_{R \rightarrow +\infty} J_R$$

donde

$$J_R = \left( \int_{-R}^R e^{-x^2} dx \right) \left( \int_{-R}^R e^{-y^2} dy \right) .$$

Este producto corresponde exactamente a la integral iterada

$$J_R = \int_{-R}^R dx \int_{-R}^R e^{-x^2-y^2} dy$$

la cual es, gracias al Teorema de Fubini igual a la integral doble

$$J_R = \int \int_{C_R} e^{-x^2-y^2} dx dy, \quad C_R = [-R, R] \times [-R, R].$$

Notemos que se tiene la validez de las inclusiones

$$B(0, R) \subset C_R \subset B(0, 2R).$$

Por ende tenemos que

$$\int \int_{B(0, R)} e^{-x^2-y^2} dx dy \leq J_R \leq \int \int_{B(0, 2R)} e^{-x^2-y^2} dx dy.$$

Por otra parte, en coordenadas polares, la región  $B(0, R)$  corresponde simplemente a  $0 < r < R$ ,  $0 < \theta < 2\pi$ , por lo tanto del teorema del cambio de variables obtenemos

$$\begin{aligned} \int \int_{B(0, R)} e^{-x^2-y^2} dx dy &= \int_0^{2\pi} \int_0^R e^{-r^2} r dr = \\ &= \pi(1 - e^{-R^2}). \end{aligned}$$

la conclusión que obtenemos es entonces que

$$\pi(1 - e^{-R^2}) \leq J_R \leq \pi(1 - e^{-4R^2})$$

y por lo tanto

$$\lim_{R \rightarrow \infty} J_R = \pi,$$

lo que implica la hermosa fórmula explícita

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi} .$$