Funciones diferenciables de $\mathbb{R}^n$ en $\mathbb{R}$

Cálculo II (2004) *

En este capítulo estudiamos funciones definidas en subconjuntos de $\mathbb{R}^n$ que toman valores reales, que llamamos *funciones (reales) de varias variables*. Veamos ejemplos sencillos de este tipo de funciones.

Ejemplo 1. Dado un vector $A = (A_1, \dots, A_n) \in \mathbb{R}^n$, llamamos *transformación lineal* de $\mathbb{R}^n$ en $\mathbb{R}$ a la función $T\colon \mathbb{R}^n \to \mathbb{R}$ definida mediante

$$T(v) = \langle A, v \rangle . \tag{1}$$

Se puede ver también que toda función $T\colon \mathbb{R}^n \to \mathbb{R}$ que sea lineal (es decir, que verifique $T(\alpha v + \beta w) = \alpha T(v) + \beta T(w)$ para α, β escalares y v, w vectores de $\mathbb{R}^n$, todos arbitrarios) tiene la forma (1) para algún vector A.

Ejemplo 2. Dada una matriz $n \times n$ simétrica Q, llamamos *forma cuadrática* a la función $Q\colon \mathbb{R}^n \to \mathbb{R}$ definida mediante

$$Q(v) = \langle v, Qv \rangle ,$$

lo que en notación matricial es $Q(v) = vQv^t$, asumiendo que v es un vector fila, y v^t su traspuesto.

Ejemplo 3. Casos interesantes de formas cuadráticas son las funciones dadas por

$$f(x, y) = x^2 + y^2, \qquad g(x, y) = x^2 - y^2,$$

definidas para todo $(x, y) \in \mathbb{R}^2$. Decimos que el gráfico de f es un *paraboloide*, y que el gráfico de g es un *hiperboloide*.

Dado que no es posible graficar las funciones de $\mathbb{R}^n$ en $\mathbb{R}$ como podemos hacer en el caso de funciones reales de variable real, la siguiente noción ayuda a formarse una idea sobre el comportamiento de una función en algunos casos.

---

* Notas para el curso de la Licenciatura en Matemática, Facultad de Ciencias, preparadas por Ernesto Mordecki en base a notas manuscritas de Fernando Peláez.

Definición 1 (Conjuntos de nivel).
Consideremos una función $f\colon U \to \mathbb{R}$*, donde* U *es abierto en* $\mathbb{R}^n$*, y un real* c*. Llamamos* conjunto de nivel c *de* f *al conjunto* M *de los puntos* $x \in U$ *tales que* $f(x) = c$*, es decir* $M = f^{-1}(c)$*. Si* $n = 2, 3$ *llamamos respectivamente* curva de nivel, superficie de nivel *a los conjuntos de nivel recién definidos.*

Es sencillo ver que el conjunto de nivel c de una transformación lineal como la del ejemplo 1 es una recta de ecuación $\langle A, v\rangle = c$, y que, cuando $n = 2$ las curvas de nivel c de una forma cuadrática como la del ejemplo 2 son cónicas. En particular, las curvas de nivel de un paraboloide de ecuación $f(x, y) = x^2 + y^2$ son circunferencias, y las de un hiperboloide de ecuación $f(x, y) = x^2 - y^2$ son hipérbolas.

1. Derivadas parciales y direccionales

Nos interesa en primer lugar extender el concepto de *derivada* de una función de una variable, una de cuyas finalidades es determinar los extremos absolutos de una función real y continua, definida en un conjunto compacto, cuya existencia asegura el teorema de Weierstrass.

Definición 2 (Derivadas parciales). *Sea* $f\colon U \to \mathbb{R}$ *donde* U *es un subconjunto abierto de* $\mathbb{R}^n$*, y consideremos un punto* $a \in U$*. Para* $i = 1, \dots, n$*, la* i-ésima derivada parcial *de* f *en el punto* a*, que designamos indistintamente mediante*

$$\partial_i f(a) \quad \textit{o también} \quad \frac{\partial f}{\partial x_i}(a).$$

es el valor del límite

$$\lim_{t\to 0} \frac{f(a + te_i) - f(a)}{t}$$

cuando éste existe.

En el caso particular en el que $n = 2$, decimos *derivada parcial respecto de* x *o de* y. En este caso, si $a = (x, y)$, tenemos:

$$\partial_1 f(a) = \frac{\partial f}{\partial x}(a) = \lim_{h\to 0} \frac{f(x + h, y) - f(x, y)}{h},$$
$$\partial_2 f(a) = \frac{\partial f}{\partial y}(a) = \lim_{k\to 0} \frac{f(x, y + k) - f(x, y)}{k},$$

que también designamos mediante $f_x(x,y)$ y $f_y(x,y)$. Una situación análoga obtenemos cuando $n=3$: en este caso $a=(x,y,z)$, y designamos la tercer derivada parcial, o derivada parcial respecto de z, mediante $\partial_3 f(a) = f_z(a) = f_z(x,y,z)$.

En lo que respecta al cálculo de las derivadas parciales, como la definición indica que la derivada parcial i-ésima es la derivada de una función real, con respecto de una variable real privilegiada x_i, cuando las otras permanecen constantes, se aplican las reglas de derivacion de funciones reales de variable real. Por ejemplo, si $f\colon \mathbb{R}^2 \to \mathbb{R}$ viene dada por

$$f(x,y) = ax^2 + 2bxy + cy^2,$$

sus derivadas parciales valen

$$\frac{\partial f}{\partial x}(x,y) = 2ax + 2by, \qquad \frac{\partial f}{\partial y}(x,y) = 2bx + 2cy.$$

Consideremos $f\colon U \to \mathbb{R}$, con U abierto en $\mathbb{R}^n$ y $a \in U$. Supongamos que existe $\partial_i f(a)$ para algún $i = 1, \dots, n$. Consideremos la recta (en $\mathbb{R}^n$) que pasa por el punto a y es paralela al vector e_i, esto es, la función $\lambda\colon \mathbb{R} \to \mathbb{R}^n$ definida mediante

$$\lambda(t) = a + te_i.$$

Como U es abierto y λ es continua, existe $\varepsilon > 0$ tal que $\lambda(t) \in U$ cuando $-\varepsilon < t < \varepsilon$. La función $\phi(t)$ resultante de la composición de f con λ, con dominio en el intervalo $(-\varepsilon, \varepsilon)$, a valores reales, es decir, la función

$$\phi(t) = (f \circ \lambda)(t) = f\big(\lambda(t)\big) = f(a + te_i), \quad t \in (-\varepsilon, \varepsilon),$$

es una función de una variable real, cuya derivada en el origen vale

$$\phi'(0) = \lim_{t \to 0} \frac{(f \circ \lambda)(t) - (f \circ \lambda)(0)}{t} = \lim_{t \to 0} \frac{f(a + te_i) - f(a)}{t} = \frac{\partial f}{\partial x_i}(a).$$

Esto prueba que $\phi(t)$ es una función derivable en $t=0$, y que el valor de su derivada es el de la derivada parcial de la función f en el punto a.

Consideremos el caso $n=2$, y representemos gráficamente a la función $z = f(x,y)$ un un sistema de ejes coordenados $Oxyz$. Para la derivada parcial respecto de x (es decir, $i=1$), el recorrido de la recta λ es el intervalo $(a - \varepsilon e_1, a + \varepsilon e_1)$, por lo que la función compuesta $f \circ \lambda$ se puede representar gráficamente tomando como eje el paralelo a Ox con origen en el punto a, obteniendo el gráfico de la función compuesta como intersección del gráfico de la función original $f(x, y)$ con el plano paralelo a Oxz por el punto a. La derivada parcial con respecto a x es entonces el valor de la la pendiente de la recta tangente a este gráfico, obtenido como intersección.

Sin hipótesis adicionales de regularidad sobre la función f, las derivadas parciales apenas dan información sobre el comportamiento de la función en las direcciones de los ejes coordenados, y no permiten obtener conclusiones acerca del comportamiento global de la función en un punto. Esto es una diferencia muy importante con el caso $n = 1$, de las funciones reales de variable real.

Ejemplo 4. Consideremos $f: \mathbb{R}^2 \to \mathbb{R}$ definida mediante

$$f(x, y) = \begin{cases} 1, & \text{si } xy = 0, \\ 0, & \text{en otro caso.} \end{cases}$$

Es claro que existen ambas derivadas parciales: $f_x(0, 0) = f_y(0, 0) = 0$, pero la función no es continua en el punto $(0, 0)$.

Ejemplo 5. Consideremos $f: \mathbb{R}^2 \to \mathbb{R}$ definida mediante $f(0, 0) = 0$ y

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

En todo punto distinto de $(0, 0)$ la función tiene derivadas parciales, dadas por

$$\frac{\partial f}{\partial x}(x, y) = \frac{y^3 - x^2 y}{x^2 + y^2}, \qquad \frac{\partial f}{\partial y}(x, y) = \frac{x^3 - x^2 y}{x^2 + y^2}.$$

En el origen $(0, 0)$ tenemos

$$\frac{\partial f}{\partial x}(0, 0) = \lim_{t \to 0} \frac{f(0 + t, 0) - f(0, 0)}{t} = 0,$$

y análogamente obtenemos $f_y(0, 0) = 0$. Sin embargo, y aunque existen ambas derivadas parciales en todos los puntos, la función $f(x, y)$ no es continua $(0, 0)$ como resulta de observar, que el límite

$$\lim_{(x,y) \to (0,0)} \frac{xy}{x^2 + y^2}$$

no existe: tenemos $f(x, x) = 1/2$ (si $x \neq 0$), pero $f(x, 0) = 0$.

La siguiente noción generaliza la de derivada parcial.

Definición 3 (Derivadas direccionales). *Sean $f\colon U \to \mathbb{R}$, donde U es un subconjunto abierto de $\mathbb{R}^n$, y un punto $a \in U$. Consideremos un vector v de $\mathbb{R}^n$, no nulo. La* derivada direccional *de f con respecto de v en el punto a, que designamos mediante $\frac{\partial f}{\partial v}(a)$, es el valor del límite*

$$\lim_{t\to 0} \frac{f(a+tv)-f(a)}{t}$$

cuando éste existe.

Observemos primero, que si $v = e_i$ para algún $i = 1, \dots, n$, la derivada direccional es la derivada parcial de la definición anterior. Estamos efectivamente generalizando la definción de derivada parcial.

En segundo lugar, definimos la derivada direccional para todos los vectores no nulos de $\mathbb{R}^n$ (y no sólo cuando $\|v\| = 1$, como en la mayoría de los libros de Cálculo), para obtener la propiedad de linealidad de $\big(\partial f/\partial v\big)(a)$ con respecto de v. En particular, si $w = \alpha v$, con α real, tenemos

$$\frac{\partial f}{\partial w}(a) = \lim_{t\to 0} \frac{f(a+t\alpha v)-f(a)}{t} = \alpha \lim_{t\to 0} \frac{f(a+t\alpha v)-f(a)}{\alpha t} = \alpha \frac{\partial f}{\partial v}(a). \quad (2)$$

Como en el caso de las derivadas parciales, obtenemos que la derivada direccional es la derivada de la función real $\phi(t)$, que resulta de componer f con la recta de ecuación $\lambda(t) = a + tv$, definida para $t \in (-\varepsilon, \varepsilon)$, para algún $\varepsilon > 0$. Tenemos $\phi(t) = (f \circ \lambda)(t) = f(a + tv)$, y

$$\phi'(0) = \lim_{t\to 0} \frac{\phi(t)-\phi(0)}{t} = \lim_{t\to 0} \frac{f(a+tv)-f(a)}{t} = \frac{\partial f}{\partial v}(a).$$

La existencia de todas las derivadas direccionales de una función en un punto, si bien es una condición mas fuerte que la existencia de las derivadas parciales, tampoco asegura la continuidad de la función, como vemos a continuación.

Ejemplo 6. Consideremos $f\colon \mathbb{R}^2 \to \mathbb{R}$, definida mediante $f(0,0) = 0$ y

$$f(x,y) = \frac{x^3 y}{x^6 + y^2}, \quad \text{si } (x,y) \neq (0,0).$$

Consideremos un vector $v = (h,k)$ no nulo, y estudiemos la existencia de $\partial f/\partial v$ en el origen $(0,0)$:

$$\frac{\partial f}{\partial v}(0,0) = \lim_{t\to 0} \frac{f(th,tk)-f(0,0)}{t} = \lim_{t\to 0} \frac{th^3 k}{t^4 h^6 + k^2} = 0$$

Obtenemos entonces que existen todas las derivadas direccionales de $f(x,y)$ en el punto $(0,0)$, y toman el valor 0, sin embargo, la función no es continua en $(0,0)$: tenemos $f(x,x^3) = 1/2$ (si $x \neq 0$), pero $f(x,0) = 0$.

Nos proponemos ahora generalizar el teorema de Lagrange para una función de n variables.

Teorema 1 (Teorema del valor medio). *Sea $f\colon U \to \mathbb{R}$ con U abierto en $\mathbb{R}^n$. Consideremos un punto $a \in U$ y un vector $v \in \mathbb{R}^n$, tal que el intervalo $[a, a+v]$ esté contenido en U. Supongamos que f restringida a $[a, a+v]$ es una función continua, y que existe la derivada direccional $(\partial f/\partial v)(x)$ para todo $x \in (a, a+v)$. Entonces, existe $\theta \in (0,1)$ tal que*

$$f(a+v) - f(a) = \frac{\partial f}{\partial v}(a + \theta v).$$

Demostración. La demostración se basa en el teorema del valor medio (de Lagrange) para funciones reales. Consideremos la curva $\lambda\colon [0,1] \to \mathbb{R}^n$ de ecuación $\lambda(t) = a + tv$ y la función compuesta $\phi = f \circ \lambda$, es decir $\phi(t) = f(a+tv)$. La función ϕ es continua en el intervalo $[0,1]$. Veamos que es derivable, calculando su derivada. En efecto

$$\phi'(\theta) = \lim_{t\to 0} \frac{\phi(\theta+t) - \phi(\theta)}{t} = \lim_{t\to 0} \frac{f(a+\theta v + tv) - f(a+\theta v)}{t} = \frac{\partial f}{\partial v}(a+\theta v). \tag{3}$$

Aplicando el teorema de Lagrange a la función ϕ, tenemos

$$\frac{\partial f}{\partial v}(a+\theta v) = \phi'(\theta) = \phi(1) - \phi(0) = f(a+v) - f(a),$$

para algún $\theta \in (0,1)$, concluyendo la demostración. □

Observación. En muchas situaciones sabemos que $\frac{\partial f}{\partial v}(x)$ existe para todo x en un cierto intervalo $[a, a+v] \subset U$. En este caso la restricción de f a dicho intervalo es una función continua. En efecto, la función auxiliar $\phi(t) = f(a+tv)$ con $t \in [0,1]$ es derivable en $(0,1)$ y tiene derivadas laterales en $t=0$ y en $t=1$. Esto implica que ϕ es continua en $[0,1]$ y de allí obtenemos la continuidad de f restringida a $[a, a+v]$.

En vista de la observación anterior, dada $f(x,y)$ si sabemos que existe f_x en un intervalo $[x_0, x_0+h]$, aplicando el teorema del valor medio con $v = he_1$, obtenemos que existe $\theta \in (0,1)$ tal que

$$\begin{aligned} f(x_0+h,y_0)-f(x_0,y_0) &= \frac{\partial f}{\partial v}(x_0+\theta h,y_0) = h\frac{\partial f}{\partial e_1}(x_0+\theta h,y_0) \\ &= h\frac{\partial f}{\partial x}(x_0+\theta h,y_0), \end{aligned} \tag{4}$$

donde utilizamos (2).

Generalizamos ahora el teorema que, para funciones reales de variable real, afirma que si la derivada es nula en un intervalo, la función es constante.

Teorema 2. *Sea $f\colon U \to \mathbb{R}$, donde U es un conjunto abierto y conexo de $\mathbb{R}^n$. Supongamos que $(\partial f/\partial v)(x) = 0$ para todo vector $v \in \mathbb{R}^n$, y para todo punto $x \in U$. Entonces, f es constante en U.*

Demostración. El conjunto U es poligonalmente conexo, por ser abierto y conexo (Teorema 25 del capítulo 1). Supongamos primero que U es convexo.

Tomemos un punto $a \in U$ de referencia, y $x = a+v \in U$ arbitrario. En primer lugar $[a,a+v] \subset U$, dado que U es convexo. Luego, de la existencia de la derivada direccional $(\partial f/\partial v)(y)$ en todos los puntos $y \in [a,a+v]$ obtenemos que es aplicable el teorema 1 del valor medio. Combinado con nuestra hipótesis, obtenemos que existe $\theta \in (0,1)$ tal que

$$f(x)-f(a) = \frac{\partial f}{\partial v}(a+\theta v) = 0,$$

concluyendo la demostración en el caso convexo. En el caso poligonalmente convexo repetimos el argumento anterior en cada intervalo de la poligonal que une dos puntos arbitrarios. $\square$

2. Diferenciabilidad

Si bien la noción de derivada direccional nos permitió demostrar el teorema del valor medio, no es una generalización suficiente de la noción de derivada de funciones reales, en particular, porque existen funciones que no son continuas en un punto, pero poseen todas las derivadas direccionales en ese punto.

Nos proponemos definir la *diferenciabilidad* de una función en un punto, generalizando la noción de derivabilidad de funciones reales, como sigue.

Recordemos que dada $f\colon I \to \mathbb{R}$, donde I es un intervalo abierto de $\mathbb{R}$ y $a \in I$, definimos la derivada de f en el punto a, que designamos indistintamente

$$f'(a), \quad \text{o también} \quad \frac{df}{dt}(a)$$

como el valor del límite

$$\lim_{t\to 0} \frac{f(a+t) - f(a)}{t},$$

cuando existe y es finito. Supongamos que f es derivable en el punto a, y definamos la función $p(t)$, en un entorno reducido $B^*(0,\varepsilon)$, suficientemente pequeño, mediante

$$p(t) = \frac{f(a+t) - f(a)}{t} - f'(a). \tag{5}$$

Es claro que $\lim_{t\to 0} p(t) = 0$, y podemos escribir, despejando,

$$f(a+t) - f(a) = f'(a)t + tp(t), \qquad \lim_{t\to 0} p(t) = 0.$$

Mas aún, podemos dar la siguiente definición.

Definición 4 (Función real diferenciable).
Decimos que $f\colon I \to \mathbb{R}$ es diferenciable *en un punto a de un intervalo abierto $I \subset \mathbb{R}$, cuando existen una constante A y una función $p(t)\colon B^*(0,\varepsilon) \to \mathbb{R}$, con $\varepsilon > 0$ suficientemente pequeño, tales que se verifica*

$$f(a+t) = f(a) + At + tp(t), \qquad \lim_{t\to 0} p(t) = 0. \tag{6}$$

Es sencillo verificar que esta definición es equivalente a la de *función derivable* en a: basta despejar $p(t)$ de (6) y compararlo con (5), de donde obtenemos que $f'(a) = A$.

La siguiente definición generaliza, para funciones de n variables, la definición (4).

Definición 5 (Función diferenciable). *Sea $f\colon U \to \mathbb{R}$, donde U es abierto en $\mathbb{R}^n$. La función f es* diferenciable *en un punto $a \in U$, cuando existen un vector $A = (A_1, \dots, A_n)$ y una función $p\colon B^*(0,\varepsilon) \to \mathbb{R}$, con $\varepsilon > 0$*

suficientemente pequeño, tales que para todo $v \in \mathbb{R}^n$ *con* $a+v \in B^*(a,\varepsilon)$*, se verifica*

$$f(a+v) - f(a) = \langle A, v\rangle + \|v\|\, p(v), \quad \lim_{v\to 0} p(v) = 0. \tag{7}$$

Decimos además que f *es* diferenciable en U *cuando es diferenciable en todo* $a \in U$*.*

Introduciendo la función $r(v) = \|v\| p(v)$, que llamamos *resto*, las fórmulas en (7) pueden escribirse como

$$f(a+v) - f(a) = \langle A, v\rangle + r(v), \qquad \lim_{v\to 0} \frac{r(v)}{\|v\|} = 0. \tag{8}$$

Estamos entonces definiendo que una función es diferenciable en un punto a cuando su incremento se puede aproximar por una transformación lineal de la forma $T(v) = \langle A, v\rangle$.

Una función $f(x,y)$ de dos variables, definida en un conjunto U abierto en $\mathbb{R}^2$, es entonces diferenciable en un punto $a = (x,y)$ cuando existen dos constantes A, B y una función $r(h,k)\colon B^*(0,\varepsilon) \to \mathbb{R}$, con $\varepsilon > 0$ suficientemente pequeño, tales que si $v = (h,k)$, $a+v \in B^*(a,\varepsilon)$, entonces

$$f(x+h, y+k) - f(x,y) = Ah + Bk + r(h,k), \quad \lim_{v\to 0} \frac{r(h,k)}{\sqrt{h^2+k^2}} = 0. \tag{9}$$

Observemos finalmente que en la definición utilizamos la norma euclideana usual. Despejando $r(v)$ en (8), obtenemos que f es diferenciable en $a \in U$ si y sólo si existe un vector A tal que

$$\lim_{v\to 0} \frac{1}{\|v\|}\big(f(a+v) - f(a) - \langle A, v\rangle\big) = 0.$$

Como este límite no depende de la norma en $\mathbb{R}^n$ (porque todas las normas son equivalentes), la definición de diferenciabilidad no depende de la norma elegida. A continuación, el resultado que estábamos buscando.

Teorema 3. *Sean* $f\colon U \to \mathbb{R}$*, con* U *abierto en* $\mathbb{R}^n$*,* $a \in U$*. Si* f *es diferenciable en el punto* a*, entonces es continua en* a*, y, dado un vector* v *cualquiera, existe la derivada direccional* $(\partial f/\partial v)(a)$*, que verifica*

$$\frac{\partial f}{\partial v}(a) = \langle A, v\rangle\,, \tag{10}$$

donde $A = (A_1, \dots, A_n)$ es el vector de la definición de diferenciabilidad. En particular, existen las derivadas parciales, y se verifica

$$\frac{\partial f}{\partial x_1}(a) = A_1, \dots, \frac{\partial f}{\partial x_n}(a) = A_n.$$

Demostración. La continuidad es inmediata, dado que

$$\lim_{v\to 0}\big(f(a+v) - f(a)\big) = \lim_{v\to 0}\big(\langle A, v\rangle + \|v\| p(v)\big) = 0.$$

Respecto de la derivada direccional, dado un vector v, si ponemos tv en (7), tenemos

$$f(a+tv) - f(a) = t\,\langle A, v\rangle + |t|\|v\|p(tv),$$

Dividiendo por t y tomando límite, obtenemos

$$\lim_{t\to 0}\frac{f(a+tv) - f(a)}{t} = \langle A, v\rangle + \lim_{t\to 0}\frac{|t|}{t}\|v\|p(tv) = \langle A, v\rangle\,,$$

porque $p(tv) \to 0$ ($t \to 0$), lo que demuestra (10). En particular, si $v = e_i$, obtenemos

$$\partial_i f(a) = \frac{\partial f}{\partial x_i}(a) = \langle A, e_i\rangle = A_i,$$

concluyendo la demostración. □

La fórmula (10) muestra que la derivada direccional $(\partial f/\partial v)(a)$ es una función lineal de v, cuando la función es diferenciable en a. Además, el teorema muestra que las constantes en la definición de diferenciabilidad son únicas, por coincidir con las derivadas parciales. Estos hechos motivan las siguientes definiciones.

Definición 6 (Diferencial y gradiente). *Supongamos que $f\colon U \to \mathbb{R}$ es diferenciable en $a \in U$, con U abierto de $\mathbb{R}^n$.*
(a) *Llamamos* diferencial *de f en el punto a, y designamos df_a, a la transformación lineal*

$$df_a\colon \mathbb{R}^n \to \mathbb{R}, \qquad df_a(v) = \frac{\partial f}{\partial v}(a) = \sum_{i=1}^{n}\frac{\partial f}{\partial x_i}(a)v_i,$$

para todo $v = (v_1, \dots, v_n) \in \mathbb{R}^n$.
(b) *Llamamos* gradiente *de f en el punto a, y designamos $\nabla f(a)$, al vector*

$$\nabla f(a) = \Big(\frac{\partial f}{\partial x_1}(a), \dots, \frac{\partial f}{\partial x_n}(a)\Big).$$

El diferencial es el término lineal en la definición de función diferenciable. Si $v \in \mathbb{R}^n$ es tal que $a + v \in U$, la fórmula (7) se escribe

$$f(a+v) - f(a) = df_a(v) + \|v\|\, p(v), \quad \lim_{v \to 0} p(v) = 0.$$

A su vez, el gradiente es el vector A de la definición de función diferenciable, y, según el teorema 3, la derivada direccional con respecto de v en el punto a no es otra cosa que el diferencial de f en a evaluado en v. Es decir

$$df_a(v) = \langle \nabla f(a), v \rangle = \frac{\partial f}{\partial v}(a).$$

Veamos una propiedad del gradiente. Supongamos que $\nabla f(a) \neq 0$, y consideremos v arbitrario, que verifique $\|v\| = \|\nabla f(a)\|$. Aplicando la desigualdad de Cauchy-Schwarz, tenemos

$$\begin{aligned}\Big|\frac{\partial f}{\partial v}(a)\Big| &= \big|\, \langle \nabla f(a), v \rangle \,\big| \leq \|\nabla f(a)\|\, \|v\| = \|\nabla f(a)\|^2 \\ &= \langle \nabla f(a), \nabla f(a) \rangle = \frac{\partial f}{\partial \nabla f(a)}(a).\end{aligned}$$

Como los vectores v tienen norma constante, esto indica que la dirección de mayor crecimiento de la función en el dominio U (indicada por el mayor valor absoluto de la derivada direccional), está dada por el gradiente $\nabla f(a)$.

Estudiemos ahora la diferenciabilidad de algunas funciones sencillas a partir de la definición.

Ejemplo 7. Consideremos la función $s\colon \mathbb{R}^2 \to \mathbb{R}$ dada por $s(x,y) = x + y$. Si $a = (x,y)$ y $v = (h,k)$, tenemos

$$s(x+h, y+k) - s(x,y) = h + k,$$

y se verifica (9) con $r(h,k) = 0$. Las derivadas parciales verifican $s_x(x,y) = s_y(x,y) = 1$ en todos los puntos $a = (x,y)$.

Consideremos ahora la función $p\colon \mathbb{R}^2 \to \mathbb{R}$ dada por $p(x,y) = xy$. Con a, v como antes, tenemos

$$p(a+v) - p(a) = (x+h)(y+k) - xy = yh + xk + hk.$$

Como

$$\frac{hk}{\sqrt{h^2+k^2}} = \frac{h}{\sqrt{h^2+k^2}} \times k \to 0, \quad \text{si } (h,k) \to (0,0),$$

porque $|h| \leq \sqrt{h^2+k^2}$, se verifica (9) con $r(h,k) = hk$, y las derivadas parciales valen $p_x(x,y) = y$, $p_y(x,y) = x$.

Consideremos por último la función $q\colon \mathbb{R} \times (\mathbb{R} \setminus \{0\}) \to \mathbb{R}$ dada por $q(x,y) = x/y$, definida si $y \neq 0$. Tenemos

$$q(a+v) - q(a) = \frac{x+h}{y+k} - \frac{x}{y} = \frac{hy - xk}{y^2 + yk} = \frac{1}{y}h - \frac{x}{y^2}k + \frac{k(xk-yh)}{y^2(y+k)}.$$

En este caso

$$\frac{k}{\sqrt{h^2+k^2}} \frac{(xk-yh)}{y^2(y+k)} \to 0, \quad \text{si } (h,k) \to (0,0),$$

porque $|k| \leq \sqrt{h^2+k^2}$ e $y \neq 0$. Se verifica la definición de diferenciabilidad, y las derivadas parciales valen $q_x(x,y) = 1/y$, $q_y(x,y) = -x/y^2$.

Consideremos $U \subset \mathbb{R}^2$, y una función $f\colon U \to \mathbb{R}$, diferenciable en $a \in U$. Si ponemos $a = (x_0, y_0)$ y $v = (x - x_0, y - y_0)$ en la definición de diferenciabilidad, podemos escribir

$$f(x,y) = f(x_0,y_0) + \frac{\partial f}{\partial x}(x_0,y_0)(x-x_0) + \frac{\partial f}{\partial y}(x_0,y_0)(y-y_0) + r(v)$$

Llamamos *plano tangente* al gráfico de la función $f(x,y)$ en el punto (x_0,y_0), al plano de ecuación

$$z = f(x_0,y_0) + f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)(y-y_0), \quad (x,y) \in \mathbb{R}^2$$

Este plano verifica la propiedad de ser, entre todos los planos que pasan por el punto $\big(x_0, y_0, f(x_0,y_0)\big)$, el que mejor aproxima localmente a $f(x,y)$, en el siguiente sentido: si

$$f(x,y) - \big(f(x_0,y_0) + A(x-x_0) + B(y-y_0)\big),$$

es un infinitésimo de orden superior a $\|v\| = \sqrt{(x-x_0)^2 + (y-y_0)^2}$, entonces $A = f_x(x_0,y_0)$ y $B = f_y(x_0,y_0)$, como se puede verificar a partir de la definición de diferenciabilidad. Observemos que este plano es perpendicular al vector $\big(f_x(x_0,y_0), f_y(x_0,y_0), -1\big)$ que llamamos *vector normal* al gráfico de la función $f(x,y)$ en el punto (x_0,y_0).

Teorema 4 (Condición suficiente de diferenciabilidad).
Supongamos que una función $f\colon U \to \mathbb{R}$, donde U es abierto en $\mathbb{R}^n$, tiene $n-1$ derivadas parciales definidas en alguna bola $B(a,\varepsilon) \subset U$, continuas en a, y que la restante derivada parcial existe en a. Entonces f es diferenciable en a.

Demostración. Para simplificar la notación consideramos $n = 2$. Supongamos entonces que $a = (x_0, y_0)$, que $f_x(x, y)$ existe en $B(a, \varepsilon) \subset U$, es continua en a, y que existe $f_y(x_0, y_0)$. Para $v = (h, k)$ tal que $a + v \in B(a, \varepsilon)$, definimos

$$r(h,k) = f(x_0 + h, y_0 + k) - f(x_0, y_0) - \frac{\partial f}{\partial x}(x_0, y_0)h - \frac{\partial f}{\partial y}(x_0, y_0)k.$$

Según (9) tenemos que demostrar que

$$\lim_{(h,k)\to(0,0)} \frac{r(h,k)}{\sqrt{h^2 + k^2}} = 0. \tag{11}$$

Podemos escribir

$$\begin{aligned} r(h,k) = f(x_0 + h, y_0 + k) - f(x_0, y_0 + k) - \frac{\partial f}{\partial x}(x_0, y_0)h \\ + f(x_0, y_0 + k) - f(x_0, y_0) - \frac{\partial f}{\partial y}(x_0, y_0)k. \end{aligned} \tag{12}$$

Aplicando el teorema del valor medio como en (4) sabemos que existe $\theta \in (0, 1)$ tal que

$$f(x_0 + h, y_0 + k) - f(x_0, y_0 + k) = \frac{\partial f}{\partial x}(x_0 + \theta h, y_0 + k)h.$$

Sustituyendo esta expresión en (12) y dividiendo por $\sqrt{h^2 + k^2}$, obtenemos

$$\begin{aligned} \frac{r(h,k)}{\sqrt{h^2 + k^2}} = \Big[\frac{\partial f}{\partial x}(x_0 + \theta h, y_0 + k) - \frac{\partial f}{\partial x}(x_0, y_0)\Big]\frac{h}{\sqrt{h^2 + k^2}} \\ + \Big[\frac{f(x_0, y_0 + k) - f(x_0, y_0)}{k} - \frac{\partial f}{\partial y}(x_0, y_0)\Big]\frac{k}{\sqrt{h^2 + k^2}} \end{aligned}$$

Examinemos el límite de esta expresión cuando $(h, k) \to (0, 0)$. En primer lugar los factores $h/\sqrt{h^2 + k^2}$ y $k/\sqrt{h^2 + k^2}$ permanecen acotados. Luego, ambos sumandos a la derecha de la igualdad tienden a cero: el primero por ser f_x continua en el punto (x_0, y_0); el segundo por existir la derivada $f_y(x_0, y_0)$. Verificamos entonces (11), concluyendo la demostración. $\square$

Del punto de vista práctico este teorema nos da un criterio para obtener la diferenciabilidad de funciones en todo su dominio. Decimos que una función $f\colon U \to \mathbb{R}$ es de *clase* C^1 cuando todas sus derivadas parciales son funciones continuas. Del teorema anterior obtenemos entonces, como corolario, que las funciones de clase C^1 en un dominio abierto son diferenciables en ese dominio.

Ejemplo 8. Sea $f\colon \mathbb{R} \to \mathbb{R}$ definida mediante $f(0) = 0$ y $f(x) = x^2 \operatorname{sen}(1/x)$, si $x \neq 0$. Es sencillo ver que f verifica la definición de diferenciabilidad en $a = 0$ con $A = 0$, pero que no es de clase C^1 en $\mathbb{R}$. Este ejemplo muestra que para una función, ser diferenciable no equivale a ser de clase C^1.

Teorema 5 (Regla de la Cadena).
Consideremos las funciones $f = (f_1, \dots, f_m)\colon U \to \mathbb{R}^m$ *donde* $U \subset \mathbb{R}^n$, $g\colon V \to \mathbb{R}$, *donde* $f(U) \subset V \subset \mathbb{R}^m$, *con* U, V *abiertos. Supongamos que cada función coordenada* f_j $(j = 1, \dots, m)$ *es diferenciable en un punto* $a \in U$, *y que* g *es diferenciable en* $b = f(a)$. *Entonces, la función compuesta* $h = g \circ f\colon U \to \mathbb{R}$ *es diferenciable en el punto* a, *y sus derivadas parciales verifican*

$$\frac{\partial(g \circ f)}{\partial x_i}(a) = \frac{\partial g}{\partial y_1}(b)\frac{\partial f_1}{\partial x_i}(a) + \cdots + \frac{\partial g}{\partial y_m}(b)\frac{\partial f_m}{\partial x_i}(a), \tag{13}$$

para cada $i = 1, \dots, n$.

Observación. Si $m = n = 1$, en (13) tenemos un único sumando, obteniendo la regla de la cadena de funciones reales de variable real.

Demostración. Veamos que $h = g \circ f$ verifica la definición 5 de diferenciabilidad; al calcular el vector A en la definición, obtendremos (13).

Como f_j es diferenciable en a, existe $p_j\colon B^*(0, \varepsilon_j) \to \mathbb{R}$ tal que si $a + v \in B^*(a, \varepsilon_j)$, tenemos

$$f_j(a + v) - f_j(a) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_j}(a) v_i + \|v\|\, p_j(v), \quad \lim_{v \to 0} p_j(v) = 0. \tag{14}$$

Designemos $w_1 = f_1(a + v) - f_1(a), \dots, w_m = f_m(a + v) - f_m(a)$, y w tal que

$$w = (w_1, \dots, w_m) = f(a + v) - f(a).$$

Como $w + b = w + f(a) = f(a + v) \in V$, aplicando la defininción de diferenciabilidad, ahora a la función g, existe $q\colon B^*(0, \delta) \to \mathbb{R}$ tal que

$$g(b + w) - g(b) = \sum_{j=1}^{m} \frac{\partial g}{\partial y_j}(b) w_j + \|w\| q(w), \quad \lim_{w \to 0} q(w) = 0. \tag{15}$$

Sustituyendo en (15) el factor $w_j = f_j(a+v) - f_j(a)$ ($j = 1, \dots, m$) calculado en (14), tenemos

$$\begin{aligned} h(a+v) &= g\big(f(a+v)\big) = g(w+b) \\ &= g(b) + \sum_{j=1}^{m} \frac{\partial g}{\partial y_j}(b)\Big[\sum_{i=1}^{n} \frac{\partial f_j}{\partial x_i}(a)v_i + \|v\|p_j(v)\Big] + \|w\|q(w) \end{aligned}$$

Como $g(b) = g\big(f(a)\big) = h(a)$, cambiando el orden en la suma doble e introduciendo la función auxiliar $P(v)$, podemos escribir

$$h(a+v) = h(a) + \sum_{i=1}^{n}\Big[\sum_{j=1}^{m} \frac{\partial g}{\partial y_j}(b)\frac{\partial f_j}{\partial x_i}(a)\Big]v_i + \|v\|P(v), \tag{16}$$

$$P(v) = \sum_{j=1}^{m} \frac{\partial g}{\partial y_j}(b)p_j(v) + \frac{\|w\|}{\|v\|}q(w). \tag{17}$$

Las constantes en (16) cumplen la fórmula (13). Para obtener la diferenciabilidad resta verificar que $\lim_{v\to 0} P(v) = 0$. Veamos primero que $\|w\|/\|v\|$ está acotado. Tenemos

$$\frac{|w_j|}{\|v\|} = \Big|\sum_{i=1}^{n} \frac{\partial f_j}{\partial x_i}(a)\frac{v_i}{\|v\|} + p_j(v)\Big| \le \sum_{i=1}^{n}\Big|\frac{\partial f_j}{\partial x_i}(a)\Big|\frac{|v_i|}{\|v\|} + |p_j(v)|.$$

Como $|v_i|/\|v\| \le 1$ y $p_j(v) \to 0$, resulta que $\|w\|/\|v\|$ está acotado. Finalmente, si $v \to 0$, $w \to 0$, y de la definción (17) obtenemos que $P(v) \to 0$. Esto concluye la demostración. □

En la práctica, es frecuente que todas las funciones $f_1, \dots, f_n, g$ tengan derivadas parciales continuas (es decir, son de clase C^1) en U. Sabemos entonces que son diferenciables, y se aplica la regla de la cadena en todos los puntos de U, obteniendo

$$\frac{\partial(g\circ f)}{\partial x_i}(x) = \frac{\partial g}{\partial y_1}\big(f(x)\big)\frac{\partial f_1}{\partial x_i}(x) + \cdots + \frac{\partial g}{\partial y_m}\big(f(x)\big)\frac{\partial f_m}{\partial x_i}(x),$$

para cada $i = 1, \dots, n$, lo que muestra que las derivadas parciales de $g \circ f$ son también funciones continuas, por ser composición, producto y suma de funciones continuas. En otras palabras, la composición de funciones de clase C^1 es una función de clase C^1. Veamos ahora el caso particular en que $m = 1$.

Corolario 1. *Sean* $f\colon U \to \mathbb{R}$, *con* $U \subset \mathbb{R}^n$ *abierto,* $g\colon I \to \mathbb{R}$ *con* I *intervalo abierto en* $\mathbb{R}$, *y* $f(U) \subset I$. *Si* f *es diferenciable en* $a \in U$ *y* g *diferenciable en* $b = f(a)$, *entonces* $g \circ f$ *es diferenciable en el punto* a, *y se verifica*

$$\frac{\partial (g \circ f)}{\partial x_i}(a) = g'(b)\frac{\partial f}{\partial x_i}(a) \qquad (i = 1, \dots, n)$$

Cuando $n = 1$ en la regla de la cadena, la primer función es una curva. Pongamos $\lambda = (\lambda_1, \dots, \lambda_n)\colon I \to \mathbb{R}^n$, donde I es un intervalo abierto en la recta. Decimos que la curva λ es *diferenciable* en un punto $a \in I$, cuando existe el límite vectorial

$$\lim_{h \to 0} \frac{\lambda(t+h) - \lambda(t)}{h}$$

que designamos $\lambda'(a)$, y llamamos *derivada* de la curva, *vector tangente* a la curva, y también *velocidad* de la curva. Es claro que

$$\lambda'(a) = \big(\lambda_1'(a), \dots, \lambda_n'(a)\big),$$

donde $\lambda_i'(a)$ son las derivadas usuales de funciones reales $\lambda_1, \dots, \lambda_n$.

Ejemplo 9. La curva $\lambda\colon \mathbb{R} \to \mathbb{R}^n$, dada por $\lambda(t) = a + tv$, donde a, v son vectores de $\mathbb{R}^n$ representa el *movimiento rectilíneo uniforme* en $\mathbb{R}^n$, que parte del punto a ($\lambda(0) = a$) y tiene velocidad constante v, dado que $\lambda' = v$.

Ejemplo 10. Es importante notar que la curva es la función, y no su imagen en $\mathbb{R}^n$ (esta imagen se llama la *traza* de la curva). Las curvas $\lambda, \mu\colon \mathbb{R} \to \mathbb{R}^2$ definidas mediante

$$\lambda(t) = (r\cos t, r\operatorname{sen} t), \quad \mu(t) = \big(r\cos(2t), r\operatorname{sen}(2t)\big),$$

son curvas distintas con la misma traza. Sus vectores velocidad valen

$$\lambda'(t) = (-r\operatorname{sen} t, r\cos t), \quad \mu'(t) = 2\big(-r\operatorname{sen}(2t), r\cos(2t)\big),$$

y se verifica que el punto cuyo movimento describe μ va más rápido que el que describe λ. Ambas curvas representan el *movimiento circular uniforme.* La frecuencia del segundo es el doble que la del primero. Es fácil ver que $\langle \lambda(t), \lambda'(t) \rangle = 0$, es decir, la velocidad es perpendicular a la posición, dada por el segmento $O\lambda(t)$, y que $\|\lambda'\| = r$, es decir, el vector velocidad tiene norma constante.

Corolario 2. *Consideremos la curva* $\lambda = (\lambda_1, \dots, \lambda_n)\colon I = (a - \varepsilon, a + \varepsilon) \to \mathbb{R}^n$*,* $f\colon U \to \mathbb{R}$*, con* U *abierto en* $\mathbb{R}^n$*, y* $f(I) \subset U$*. Si* λ *es diferenciable en un punto* $a \in I$*, y* f *es diferenciable en* $b = \lambda(a)$*, entonces* $\phi = f \circ \lambda\colon I \to \mathbb{R}$ *es diferenciable en el punto* a*, y se verifica*

$$\phi'(a) = \frac{d(f \circ \lambda)}{dt}(a) = \frac{\partial f}{\partial x_1}(b)\lambda_1'(a) + \cdots + \frac{\partial f}{\partial x_n}(b)\lambda_n'(a) = \frac{\partial f}{\partial v}(a),$$

donde $v = \lambda'(a)$*.*

Observación. En el caso particular en el que la curva $\lambda\colon (a - \varepsilon, a + \varepsilon) \to \mathbb{R}^n$ es una recta dada por $\lambda(t) = b + (t - a)v$, como el vector tangente $\lambda'(a) = v$, el corolario anterior nos da

$$\phi'(a) = \frac{\partial f}{\partial v}(b).$$

Concluimos que, para calcular la derivada direccional de f en el punto b con respecto de v, podemos componer f con *cualquier* curva λ que verifique $\lambda(a) = b$, $\lambda'(0) = v$.

Teorema 6 (Operaciones con funciones diferenciables).
Si las funciones $f, g\colon U \to \mathbb{R}$ *son diferenciables en el punto* $a \in U$ *abierto en* $\mathbb{R}^n$*, también son diferenciables las funciones* $f + g$*,* fg *y* f/g*, ésta última cuando* $g(a) \neq 0$*, y sus diferenciales verifican*

(a) $d\big(f + g\big)_a = df_a + dg_a$*,*

(b) $d\big(fg\big)_a = g(a)df_a + f(a)dg_a$*,*

(c) $d\big(f/g\big)_a = \frac{1}{g(a)}df_a - \frac{f(a)}{g(a)^2}dg_a$*.*

Demostración. Consideremos $F\colon U \to \mathbb{R}^2$ tal que $F(x) = \big(f(x), g(x)\big)$, y las funciones $s\colon \mathbb{R}^2 \to \mathbb{R}$ dada por $s(u, v) = u + v$, $p\colon \mathbb{R}^2 \to \mathbb{R}$ dada por $p(u, v) = uv$, y $q\colon \mathbb{R} \times (\mathbb{R} \setminus \{0\}) \to \mathbb{R}$ dada por $q(u, v) = u/v$, que son diferenciables, como vimos en el ejemplo 7.

Como la función F tiene funciones coordenadas diferenciables, aplicando la regla de la cadena, obtenemos que $(s \circ F)(x, y) = f(x, y) + g(x, y)$, $(p \circ F)(x, y) = f(x, y)g(x, y)$, y $(q \circ F)(x, y) = f(x, y)/g(x, y)$, (ésta última cuando $g(x,y) \neq 0$), son funciones diferenciables. El cálculo de los diferenciales se hace a través del cálculo de las derivadas parciales. Por ejemplo para la suma, tenemos

$$\frac{\partial(f+g)}{\partial x_i}(a) = \frac{\partial s}{\partial u}\frac{\partial f}{\partial x_i} + \frac{\partial s}{\partial v}\frac{\partial g}{\partial x_i} = \frac{\partial f}{\partial x_i} + \frac{\partial g}{\partial x_i}, \quad \text{para cada } i = 1, \dots, n,$$

mostrando (a). Las partes (b) y (c) son análogas. □

3. Derivadas de orden superior

Consideremos $f \colon U \to \mathbb{R}$ diferenciable en cada punto del abierto $U \subset \mathbb{R}^n$. Para cada natural $i = 1, \dots, n$ tenemos definida la función

$$\frac{\partial f}{\partial x_i} \colon U \to \mathbb{R}.$$

Nos planteamos estudiar las derivadas parciales de estas funciones. Es así que definimos, cuando existe el límite, la j-ésima derivada parcial de $\partial_i f$ en un punto $a \in U$:

$$\frac{\partial}{\partial x_j}\Big(\frac{\partial f}{\partial x_i}\Big)(a) = \frac{\partial^2 f}{\partial x_i \partial x_j}(a) = \partial_{ij} f,$$

que llamamos *derivada segunda* de f con respecto de x_i y de x_j en el punto a. Si $n = 2$ también escribimos $\partial_{11} f = f_{xx}$, $\partial_{12} f = f_{xy}$, $\partial_{21} f = f_{yx}$ y $\partial_{22} f = f_{yy}$, y procedemos análogamente si $n = 3$ escribiendo, por ejemplo, $\partial_{13} f(x,y,z) = f_{xz}(x,y,z)$. Además, cuando derivamos por segunda vez con respecto de la misma variable, escribimos indistintamente

$$\frac{\partial^2 f}{\partial x_i^2} = \frac{\partial^2 f}{\partial x_i \partial x_i}.$$

En forma análoga se definen las derivadas terceras, cuartas, etc. que llamamos *derivadas de orden superior*.

Definición 7. *Consideremos $f \colon U \to \mathbb{R}$ con U abierto en $\mathbb{R}^n$. Decimos que f es de clase C^0 en U, y escribimos $f \in C^0(U)$, cuando f es continua en U. Dado $k = 1, 2, \dots$ decimos que f es de clase C^k, y escribimos $f \in C^k(U)$, cuando todas las derivadas parciales de f de orden k son funciones continuas en U.*

Finalmente, decimos que f es de clase C^∞ en U cuando f es de clase C^k en U para todo k.

Como cualquier función de clase C^1 es diferenciable, y por lo tanto continua, obtenemos que si $f \in C^k(U)$ entonces las derivadas de $f \in C^p$ para todo $p = 0, 1, \dots, k-1$. En otros términos, tenemos

$$C^0(U) \supset C^1(U) \supset \cdots \supset C^\infty(U).$$

El siguiente resultado muestra que la composición de funciones de clase C^k es de clase C^k.

Corolario 3 (Regla de la Cadena en C^k).
Supongamos que las funciones $f = (f_1, \dots, f_m) \colon U \to \mathbb{R}^m$ donde $U \subset \mathbb{R}^n$, $g \colon V \to \mathbb{R}$, donde $f(U) \subset V \subset \mathbb{R}^m$, con U, V abiertos, son de clase C^k en U. Entonces, la función compuesta $g \circ f \colon U \to \mathbb{R}$ es de clase C^k en U.

Demostración. Veamos la demostración para $k = 2$. Sabemos que, por ser $f_1, \dots, f_n, g$ de clase C^1, la función $g \circ f$ es de clase C^1. Sus derivadas verifican

$$\frac{\partial (g \circ f)}{\partial x_i}(x) = \frac{\partial g}{\partial y_1}\big(f(x)\big)\frac{\partial f_1}{\partial x_i}(x) + \cdots + \frac{\partial g}{\partial y_m}\big(f(x)\big)\frac{\partial f_m}{\partial x_i}(x), \tag{18}$$

por lo que, aplicando la regla de la cadena a las funciones que aparecen a la derecha de la igualdad en (18) (que son de clase C^1), obtenemos que las derivadas de la función compuesta son de clase C^1 en U, es decir, la función compuesta es de clase C^2 en U. $\square$

Las derivadas segundas del tipo $\partial_{ij} f, \partial_{ji} f$ se llaman *derivadas cruzadas*, y pueden ser distintas. Vemos ahora dos versiones de un teorema que, bajo distintas hipótesis de *regularidad* (existencia y continuidad de algunas derivadas), nos asegura la igualdad de las derivadas cruzadas.

Teorema 7 (Teorema de Schwarz I).
Sean U abierto de $\mathbb{R}^n$, $f \colon U \to \mathbb{R}$. Supongamos que las derivadas cruzadas $\partial_{ij} f, \partial_{ji} f$ están definidas en una bola $B(a, \varepsilon) \subset U$, y que son continuas en el punto a. Entonces ambas derivadas coniciden en a.

Demostración. Por simplicidad en la notación, y sin pérdida de generalidad, suponemos que $n = 2$, $a = (x_0, y_0)$. Para h suficientemente pequeño, tal que $(x_0 + h, y_0 + h) \in B(a, \varepsilon)$, definimos

$$\psi(x) = f(x, y_0 + h) - f(x, y_0), \tag{19}$$

de forma que

$$\begin{aligned}\varphi(h) &= f(x_0+h, y_0+h) - f(x_0+h, y_0) - f(x_0, y_0+h) + f(x_0, y_0) \\ &= \psi(x_0+h) - \psi(x_0).\end{aligned}$$

Llamamos a $\varphi(h)$ el *incremento doble* de la función f en el rectángulo $[x_0, x_0+h] \times [y_0, y_0+h]$. La función ψ es derivable en el intervalo $[x_0, x_0+h]$, y por el teorema del valor medio, existe $\theta_1 \in (0,1)$ tal que

$$\psi(x_0+h) - \psi(x_0) = \psi'(x_0+\theta_1 h)h,$$

por lo que, derivando en (19), obtenemos

$$\varphi(h) = \Big[\frac{\partial f}{\partial x}(x_0+\theta_1 h, y_0+h) - \frac{\partial f}{\partial x}(x_0+\theta_1 h, y_0)\Big]h.$$

Para este último incremento de la función f_x, aplicamos nuevamente el teorema del valor medio (la existencia de f_{xy} asegura la derivabilidad de f_x con respecto de y), para obtener que existe $\theta_2 \in (0,1)$, tal que

$$\varphi(h) = \frac{\partial^2 f}{\partial x \partial y}(x_0+\theta_1 h, y_0+\theta_2 h)h^2.$$

De la continuidad de f_{xy} en (x_0, y_0) obtenemos que

$$\lim_{h\to 0}\frac{\varphi(h)}{h^2} = \frac{\partial^2 f}{\partial x \partial y}(x_0, y_0).$$

El resto de la demostración consiste en ver, que cambiando el rol de x por el de y, este mismo límite es igual a la otra derivada cruzada. Definimos entonces

$$\phi(y) = f(x_0+h, y) - f(x_0, y),$$

de forma que el mismo incremento doble verifica

$$\varphi(h) = \phi(y_0+h) - \phi(y_0).$$

Los mismos argumentos nos permiten obtener

$$\lim_{h\to 0}\frac{\varphi(h)}{h^2} = \frac{\partial^2 f}{\partial y \partial x}(x_0, y_0),$$

lo que concluye la demostración. □

El siguiente ejemplo muestra que algún tipo de condición de regularidad es necesario para obtener la igualdad de las derivadas cruzadas.

Ejemplo 11. Veamos que la función $f\colon \mathbb{R}^2 \to \mathbb{R}$ definida mediante $f(0,0) = 0$, y

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

tiene derivadas cruzadas distintas en el origen. En efecto, si $y = 0$ tenemos

$$\frac{\partial f}{\partial x}(0,0) = \lim_{x\to 0} \frac{f(x,0) - f(0,0)}{x} = 0.$$

Si $y \neq 0$, tenemos

$$\frac{\partial f}{\partial x}(0,y) = \lim_{x\to 0} \frac{f(x,y) - f(0,y)}{x} = \lim_{x\to 0} \frac{y(x^2-y^2)}{x^2+y^2} = -y.$$

Por esto,

$$\frac{\partial}{\partial y}\Big(\frac{\partial f}{\partial x}\Big)(0,0) = \lim_{y\to 0} \frac{f_x(0,y) - f_x(0,0)}{y} = -1.$$

Cálculos similares muestran que

$$\frac{\partial}{\partial x}\Big(\frac{\partial f}{\partial y}\Big)(0,0) = 1.$$

El siguiente teorema permite derivar bajo el signo de integración. Si bien se utiliza para dar otra demostración del Teorema de Schwarz, tiene relevancia por sí mismo.

Teorema 8 (Regla de Leibnitz). *Sea $f\colon U \times [a,b] \to \mathbb{R}$, donde U es abierto en $\mathbb{R}^n$, y se verifica:*

(i) *Para cada $x \in U$ fijo, la función de una variable $f(x,t)$ es integrable en el intervalo $[a,b]$.*

(ii) *$\partial_i f$ existe y es una función continua en $U \times [a,b]$.*

Entonces la función $\varphi\colon U \to \mathbb{R}$ definida mediante

$$\varphi(x) = \int_a^b f(x,t)dt$$

tiene derivada parcial i-ésima en U, que verifica

$$\frac{\partial \varphi}{\partial x_i}(x) = \int_a^b \frac{\partial f}{\partial x_i}(x,t)dt.$$

Demostración. Tenemos que ver que, para un x arbitrario en U, se verifica

$$\lim_{h\to 0} \frac{\varphi(x+he_i)-\varphi(x)}{h} = \int_a^b \frac{\partial f}{\partial x_i}(x,t)dt.$$

Sea $\varepsilon > 0$. Tomemos h suficientemente pequeño tal que $x + he_i \in U$. En primer lugar,

$$\frac{\varphi(x+he_i)-\varphi(x)}{h} - \int_a^b \frac{\partial f}{\partial x_i}(x,t)dt$$
$$= \int_a^b \Big[\frac{f(x+he_i,t)-f(x,t)}{h} - \frac{\partial f}{\partial x_i}(x,t)\Big]dt.$$

Por el teorema del valor medio, existe $\theta \in (0,1)$ tal que $f(x+he_i,t)-f(x,t) = h\partial_i f(x+\theta he_i,t)$. Por otra parte, al ser $\partial_i f$ una función continua en el conjunto $[x, x+h\varepsilon_i] \times [a,b]$ compacto, obtenemos la continuidad uniforme, de donde existe $\delta > 0$ tal que si $|h| < \delta$ tenemos

$$\Big|\frac{\partial f}{\partial x_i}(y,t) - \frac{\partial f}{\partial x_i}(x,t)\Big| < \frac{\varepsilon}{b-a}, \text{ para todo } (y,t) \in [x, x+h\varepsilon_i] \times [a,b].$$

Sustituyendo, obtenemos

$$\Big|\int_a^b \Big[\frac{f(x+he_i,t)-f(x,t)}{h} - \frac{\partial f}{\partial x_i}(x,t)\Big]dt\Big|$$
$$\leq \int_a^b \Big|\frac{\partial f}{\partial x_i}(x+\theta he_i,t) - \frac{\partial f}{\partial x_i}(x,t)\Big|dt < \varepsilon,$$

completando la demostración. □

Teorema 9 (Teorema de Schwarz II).
Sea $f\colon U \to \mathbb{R}$, con U abierto de $\mathbb{R}^n$. Supongamos que existe la derivada $\partial_j f$ y que existen y son continuas las derivadas $\partial_i f$ y $\partial_{ij} f$ en todo U. Entonces, existe $\partial_{ji} f$ y verifica

$$\frac{\partial}{\partial x_j}\Big(\frac{\partial f}{\partial x_i}\Big) = \frac{\partial}{\partial x_j}\Big(\frac{\partial f}{\partial x_j}\Big)$$

en todos los puntos de U.

Demostración. Por simplicidad en la notación, y sin pérdida de generalidad, suponemos que $n = 2$, $a = (x_0, y_0)$. Como U es abierto, existen intervalos I, J centrados en x_0 e y_0 respectivamente, y tales que $I \times J \subset U$. Si $x \in I$, $y \in J$, podemos escribir

$$f(x, y) = f(x_0, y) + \int_{x_0}^{x} \frac{\partial f}{\partial x}(t, y)dt.$$

La continuidad de f_{xy} nos permite aplicar la regla de Leibnitz, obteniendo:

$$\frac{\partial f}{\partial y}(x, y) = \frac{\partial f}{\partial y}(x_0, y) + \int_{x_0}^{x} \frac{\partial^2 f}{\partial x \partial y}(t, y)dt.$$

Derivando con respecto de x a la derecha de la igualdad mediante el teorema fundamental del cálculo, obtenemos la tesis. $\square$

Observación. El teorema de Schwarz se aplica también para derivadas de orden superior, obteniendose, por ejemplo, para las derivadas terceras de $f(x, y)$, que

$$\frac{\partial^3 f}{\partial x \partial x \partial y} = \frac{\partial^3 f}{\partial x \partial y \partial x} = \frac{\partial^3 f}{\partial y \partial x \partial x},$$

cuando U es abierto en $\mathbb{R}^2$, y $f \colon U \to \mathbb{R}$ verifica las hipótesis correspondientes.

4. Fórmula de Taylor

Definición 8 (Diferencial segundo). *Sea $f\colon U \to \mathbb{R}$. Si existen todas las derivadas de segundo orden de f en un punto a del abierto U de $\mathbb{R}^n$, llamamos* diferencial segundo *de f en a, y lo designamos $d^2 f_a$, a la función*

$$d^2 f_a\colon \mathbb{R}^n \to \mathbb{R}, \qquad d^2 f_a(v) = \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(a) v_i v_j,$$

para todo $v = (v_1, \dots, v_n) \in \mathbb{R}^n$.

Observemos que el diferencial segundo es un polinomio homogéneo de segundo grado de las variables $v_1, \dots, v_n$. Podemos escribir, en forma matricial

$$d^2 f_a(v) = vHv^t,$$

donde v^t es el vector traspuesto de v, y la matriz H, que llamamos *matriz Hessiana*, o *Hessiano*, es

$$H = \begin{bmatrix} \partial_{11} f(a) & \partial_{12} f(a) & \cdots & \partial_{1n} f(a) \\ \partial_{21} f(a) & \partial_{22} f(a) & \cdots & \partial_{2n} f(a) \\ \vdots & \vdots & & \vdots \\ \partial_{n1} f(a) & \partial_{n2} f(a) & \cdots & \partial_{nn} f(a) \end{bmatrix}. \tag{20}$$

Cuando estamos en las hipótesis del teorema de Schwarz, las derivadas cruzadas coinciden, es decir $\partial_{ij} f(a) = \partial_{ji} f(a)$. En este caso el Hessiano es una matriz simétrica, y por lo tanto, la diferencial segunda es una *forma cuadrática.*

Por ejemplo, si $n = 2$, $v = (h, k)$, y las derivadas cruzadas coinciden en un punto $a \in U$, tenemos

$$d^2 f_a(h, k) = \frac{\partial^2 f}{\partial x^2}(a) h^2 + 2\frac{\partial^2 f}{\partial x \partial y}(a) hk + \frac{\partial^2 f}{\partial y^2}(a) k^2.$$

Dado p natural, si existen todas las derivadas parciales de orden p de $f\colon U \to \mathbb{R}$ en $a \in U$, definimos el *diferencial de orden* p de f en el punto a, que designamos $d^p f_a$, a la función

$$d^p f_a\colon \mathbb{R}^n \to \mathbb{R}, \qquad d^p f_a(v) = \sum_{i_1,\dots,i_p=1}^{n} \frac{\partial^p f}{\partial x_{i_1} \dots x_{i_p}}(a + tv) v_{i_1} \dots v_{i_p}$$

para todo $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$. El diferencial de orden p resulta ser un polinomio homogéneo de grado p, por lo que se verifica

$$d^p f_a(tv) = t^p d^p f_a(v).$$

Teorema 10 (Fórmula de Taylor).
Sean $f\colon U \to \mathbb{R}$ y $a \in U$, donde U es abierto en $\mathbb{R}^n$.
(a) *Supongamos que f es de clase C^p en $B(a, \varepsilon) \subset U$ y que las derivadas parciales de orden p son funciones diferenciables en la misma bola. Entonces, para cada $v \in \mathbb{R}^n$ con $a + v \in B(a, \varepsilon)$, existe $\theta \in (0, 1)$ tal que*

$$f(a+v) = f(a) + \sum_{i=1}^{p} \frac{1}{i!} d^i f_a(v) + \frac{1}{(p+1)!} d^{p+1} f_{a+\theta v}(v). \tag{21}$$

(b) *Si además se verifica que las derivadas parciales de orden $p+1$ son funciones continuas en a, entonces vale*

$$f(a+v) = f(a) + \sum_{i=1}^{p+1} \frac{1}{i!} d^i f_a(v) + r(v), \qquad \lim_{v \to 0} \frac{r(v)}{\|v\|^{p+1}} = 0.$$

Demostración. Sea $v = (v_1, \ldots, v_n)$ tal que $a + v \in B(a, \varepsilon)$ y $\lambda\colon [0, 1] \to \mathbb{R}^n$ dada por $\lambda(t) = a + tv$. La demostración consiste obtener el desarrollo de Taylor de la función auxiliar $\phi\colon [0, 1] \to \mathbb{R}$, definida mediante

$$\phi(t) = (f \circ \lambda)(t) = f(a + tv) \tag{22}$$

en el intervalo $[0, 1]$ (tenemos $\phi(0) = f(a)$, $\phi(1) = f(a + v)$).

Como f es de clase C^p y λ es de clase C^∞ la regla de la cadena (teorema 3) nos asegura que ϕ es de clase C^p. Calculemos las derivadas de ϕ. Aplicando la regla de la cadena en (22), obtenemos

$$\phi'(t) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a + tv) v_i = df_{a+tv}(a), \quad \phi'(0) = df_a(v).$$

Derivando nuevamente

$$\phi''(t) = \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(a + tv) v_i v_j = d^2 f_{a+tv}(a), \quad \phi''(0) = d^2 f_a(v).$$

Análogamente obtenemos

$$\phi^{(p)}(t) = \sum_{i_1,\dots,i_p=1}^{n} \frac{\partial^p f}{\partial x_{i_1}\dots x_{i_p}}(a+tv)v_{i_1}\dots v_{i_p} = d^p f_{a+tv}(v)$$
$$\phi^{(p)}(0) = d^p f_a(v).$$

Como las derivadas de orden p son funciones diferenciables, la fórmula anterior nos permite ver, aplicando nuevamente la regla de la cadena, que la funcón $\phi^{(p)}$ es derivable en el intervalo $(0,1)$, es decir, la función ϕ tiene derivada de orden $p+1$ en ese intervalo, que vale $\phi^{(p+1)}(t) = d^{p+1}f_{a+tv}(v)$.

Estamos entonces en condiciones de aplicar la fórmula de Taylor con resto de Lagrange a la función ϕ en el intervalo $[0,1]$ hasta el orden p: existe $\theta \in (0,1)$ tal que

$$\phi(1) = \phi(0) + \sum_{i=1}^{p} \frac{1}{i!}\phi^{(i)}(0) + \frac{1}{(p+1)!}\phi^{(p+1)}(\theta). \tag{23}$$

La fórmula (21) a obtener es esta misma ecuación, escrita en términos de f, dado que $\phi(1) = f(a+v)$, $\phi(0) = f(a)$, $\phi'(0) = df_a(v)$, $\dots$ $\phi^{(p)}(0) = d^p f_a(v)$ $\phi^{(p+1)}(\theta) = d^{p+1} f_{a+\theta v}(v)$.

Veamos ahora (b). En efecto

$$\begin{aligned}(p+1)!\Big|\frac{r(v)}{\|v\|^{p+1}}\Big| &= \Big|d^{p+1}f_{a+\theta v}(v) - d^{p+1}f_a(v)\Big| \\ &\leq \sum_{i_1,\dots,i_{p+1}=1}^{n} \Big|\frac{\partial^{p+1}}{\partial x_{i_1}\dots\partial x_{i_n}}f(a+\theta v) - \frac{\partial^{p+1}}{\partial x_{i_1}\dots\partial x_{i_n}}f(a)\Big|\frac{|v_{i_1}\cdots v_{i_{p+1}}|}{\|v\|^{p+1}} \\ &\leq \sum_{i_1,\dots,i_{p+1}=1}^{n} \Big|\frac{\partial^{p+1}}{\partial x_{i_1}\dots\partial x_{i_n}}f(a+\theta v) - \frac{\partial^{p+1}}{\partial x_{i_1}\dots\partial x_{i_n}}f(a)\Big| \to 0\end{aligned}$$

si $v \to 0$, debido a la continuidad de las derivadas de orden $p+1$ en a, completando la demostración. $\square$

Dada $f(x,y)$ con derivadas primeras continuas y diferenciables en alguna bola centrada en $a=(x,y)$, y derivadas segundas continuas en a, del teorema de Taylor con $p=1$, obtenemos que

$$\begin{aligned}f(x+h,y+k) - f(x,y) &= f_x(x,y)h + f_y(x,y)k \\ &\quad + \frac{1}{2}f_{xx}(x,y)h^2 + f_{xy}(x,y)hk + \frac{1}{2}f_{yy}f(x,y)k^2 + r(h,k)\end{aligned}$$

donde

$$\frac{r(h,k)}{h^2+k^2} \to 0 \quad \text{cuando} \quad (h,k) \to (0,0).$$

5. Extremos de funciones de varias variables

Definición 9 (Extremos relativos y absolutos). *Sean $f\colon X \to \mathbb{R}$ y $a \in X$, subconjunto de $\mathbb{R}^n$.*
(a) *Decimos que $f(a)$ es un* máximo relativo *de f, o que f* presenta un máximo relativo *en a, si existe $\varepsilon > 0$ tal que*

$$f(a) \geq f(x) \quad \textit{para todo } x \in B(a,\varepsilon) \cap X. \tag{24}$$

La definición de mínimo relativo *es análoga, con un signo $\leq$ en lugar de $\geq$ en* (24). *Si f presenta máximo o mínimo relativo, decimos que f presenta* extremo relativo *en a.*
(b) *Si en $a \in X$ se verifica*

$$f(a) \geq f(x) \quad \textit{para todo } x \in X,$$

decimos que $f(a)$ es el máximo absoluto *de f en X, o que f* presenta su máximo absoluto *en a, con el valor $f(a)$. (Análogamente se define el* mínimo absoluto *en X.)*

Teorema 11 (Condición necesaria de extremo).
Si $f\colon U \to \mathbb{R}$ presenta extremo relativo en un punto a del abierto $U \subset \mathbb{R}^n$, en el que existe la i-ésima derivada parcial, entonces esta derivada es nula.

Demostración. La función $\varphi(t) = f(a_1,\ldots,a_{i-1},t,a_{i+1},\ldots,a_n)$ de una variable real presenta un máximo relativo en $t = a_i$. Dado que $\varphi'(a_i) = f_{x_i}(a)$. sabemos que φ es derivable en $t = a_i$, y de la condición necesaria de extremo para funciones de una variable obtenemos

$$0 = \varphi'(a_i) = \frac{\partial f}{\partial x_i}(a),$$

lo que concluye la demostración. □

El teorema anterior nos da una herramienta para la búsqueda de extremos de una función. Si sabemos que existe alguna derivada parcial, y que no se anula en un punto a, sabemos que en a no hay extremo relativo. De otra forma, si existen todas las derivadas parciales en un dominio U, los extremos se presentan únicamente en los puntos donde se anulan todas las derivadas parciales. Estos puntos se denominan *puntos críticos*, y son los "candidatos" a ser extremos relativos, y absolutos. Llamamos *punto de ensilladura*, o más brevemente *punto silla* a un punto crítico que no es ni máximo ni mínimo relativo. Veamos ejemplos sencillos de búsqueda de extremos.

Ejemplo 12. Determinemos los extremos relativos y absolutos de la función dada por

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

definida para todo $(x,y) \in \mathbb{R}^2$. Como existen las derivadas parciales en todos los puntos del plano, los posibles extremos se presentan en puntos en donde ambas derivadas parciales son nulas simultáneamente. Tenemos que determinar entonces los puntos del plano que verifiquen

$$\begin{cases} f_x = 2x = 0 \\ f_y = 2y = 0 \end{cases}$$

Obtenemos que $(0,0)$ es el único punto donde f puede presentar extremo relativo. (Decimos que $(0,0)$ es un "candidato" a extremo.) Como $f(x,y) \geq 0$ para todo (x,y) y $f(0,0) = 0$, concluímos que f presenta mínimo absoluto en $(0,0)$, y que este mínimo vale 0. Concluímos además que f no presenta máximo absoluto, dado que de presentarlo, sería relativo, y se anularían las derivadas, cosa que no ocurre en ningún otro punto del plano.

Ejemplo 13. Determinemos ahora los extremos relativos y absolutos de la función dada por

$$g(x,y) = x^2 - y^2,$$

definida para todo $(x,y) \in \mathbb{R}^2$. Como existen las derivadas parciales en todos los puntos del plano, tenemos que determinar los puntos del plano que verifiquen

$$\begin{cases} g_x & = 2x = 0 \\ g_y & = -2y = 0 \end{cases}$$

Obtenemos que $(0,0)$ es el único punto donde f puede presentar extremo relativo. (Decimos que $(0,0)$ es un "candidato" a extremo.) Tenemos $g(0,0) = 0$. Como $g(x,0) = x^2 > 0$ si $x \neq 0$, sabemos que g no presenta máximo en $(0,0)$. Como $g(0,y) = -y^2 < 0$ si $y \neq 0$, sabemos que g no presenta mínimo en $(0,0)$. Concluímos que $(0,0)$ es un punto silla, y que g no presenta máximo ni mínimo absoluto en $\mathbb{R}^2$.

La clasificación de puntos críticos no siempre es sencilla como en los ejemplos anteriores. Estudiamos entonces cómo reconocer si un punto crítico es máximo relativo, mínimo relativo, o punto silla.

Teorema 12 (Criterio de clasificación de puntos críticos).
Sean $f\colon U \to \mathbb{R}$ y $a \in U$ un punto crítico de f. Supongamos que f es de clase C^2 en alguna bola $B(a,\varepsilon) \subset U$. Entonces:

(a) *Si todos los valores propios de la matriz Hessiana* (20) *son positivos (es decir, si la forma cuadrática $d^2 f_a$ es definida positiva), f presenta mínimo relativo en a.*

(b) *Si todos los valores propios de la matriz Hessiana son negativos (es decir, si $d^2 f_a$ es una forma cuadrática definida negativa), f presenta máximo relativo en a.*

(c) *Si existen valores propios de la matriz Hessiana postivos y negativos (es decir, $d^2 f_a$ es una forma cuadrática indefinida), f presenta punto silla en a.*

Observación. El teorema no analiza los casos en que existe algún valor propio nulo del Hessiano, y todos los demás tienen signo constante, (correspondiente a los casos en que $d^2 f_a$ es una forma cuadratica semidefinida positiva o negativa). En estos casos decimos que el criterio *no clasifica* al punto crítico.

Ejemplo 14. Consideremos $f\colon \mathbb{R}^2 \to \mathbb{R}$ dada por $f(x,y) = x^2 + y^3$. Es fácil ver que f presenta un punto silla en $(0,0)$. Su matriz Hessiana en el punto $(0,0)$ tiene la forma

$$H = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$$

con vectores propios 2 y 0. Por su parte, la función $g\colon \mathbb{R}^2 \to \mathbb{R}$ dada por $g(x,y) = x^2 + y^4$ presenta mínimo relativo (y absoluto) en $(0,0)$, y tiene la misma matriz Hessiana en el punto $(0,0)$. Esto muestra que con el análisis de la matriz Hessiana no es suficiente para clasificar todos los puntos críticos.

Demostración. Veamos (a). Estamos en las hipótesis de la parte (b) del teorema de Taylor con $p = 1$. Existe $\varepsilon > 0$ tal que si $a + v \in B(a,\varepsilon) \subset U$ tenemos

$$f(a+v) - f(a) = df_a(v) + \frac{1}{2} d^2 f_a(v) + r(v),$$

donde $r(v)/\|v\|^2 \to 0$ $(v \to 0)$. Como a es punto crítico $df_a(v) = 0$, y si $v \neq 0$, podemos escribir

$$f(a+v) - f(a) = \|v\|^2 \Big[\frac{1}{2} d^2 f_a\Big(\frac{v}{\|v\|}\Big) + \frac{r(v)}{\|v\|^2}\Big], \tag{25}$$

Estudiemos primero el signo del diferencial segundo. Como H es una matriz real y simétrica, existe una base ortonormal de vectores propios de H, que designamos $\{v_1, \dots, v_n\}$, con valores propios $\lambda_1, \dots, \lambda_n$. Dado entonces $x \neq 0 \in \mathbb{R}^n$, podemos escribir $x = a_1 v_1 + \cdots + a_n v_n$, y tenemos

$$\begin{aligned} d^2 f_a(x) = xHx^t &= \Big(\sum_{i=1}^n a_i v_i\Big) H \Big(\sum_{j=1}^n a_j v_j^t\Big) = \sum_{i,j=1}^n a_i a_j v_i H v_j^t \\ &= \sum_{i,j=1}^n a_i a_j \lambda_j v_i v_j^t = \sum_{i=1}^n \lambda_i a_i^2, \end{aligned} \tag{26}$$

y $d^2 f_a(x) > 0$, porque todos los valores propios son positivos y algún a_i es no nulo.

Consideremos el conjunto compacto $S = \{x \in \mathbb{R}^n \colon \|x\| = 1\}$. Tenemos $d_a^2 f(x) > 0$ para todo $x \in S$ y, por ser S compacto y $d^2 f_a(x)$ una función continua de x, f alcanza su mínimo absoluto, que es entonces un número m positivo. En conclusión $d^2 f_a(x) \geq m > 0$ para todo $x \in S$. Si $v \neq 0$, $x = v/\|v\| \in S$, y

$$\frac{1}{2} d^2 f_a\Big(\frac{v}{\|v\|}\Big) \geq \frac{m}{2}.$$

Por otra parte, como $m/2 > 0$, existe $\delta > 0$ $(\delta < \varepsilon)$ tal que si $\|v\| < \delta$

$$\frac{|r(v)|}{\|v\|^2} < \frac{m}{2}.$$

En vista de (25) y las dos últimas desigualdades, tenemos una bola $B(a, \delta) \subset U$ tal que si $v \in B^*(a, \delta)$ se verifica $f(a+v) > f(a)$, demostrando que f presenta un mínimo relativo en a.

La demostración de (b) es análoga. Para demostrar (c) observemos que la forma cuadrática $d^2 f_a$ restringida a S alcanza su mínimo m y su máximo M en versores x e y, y se cumple $m < 0 < M$, como resulta de (26), dado que $d^2 f_a(v_i) = \lambda_i$, y que existen valores propios positivos y negativos. En todos los vectores de la forma $v = tx$ $(t \neq 0)$ el primer sumando a la derecha en (25) verifica

$$\frac{1}{2}d^2 f_a\Big(\frac{v}{\|v\|}\Big) = \frac{m}{2} < 0,$$

y como el segundo sumando tiende a cero, en todo entorno de a existen puntos tales que $f(a+v) < f(a)$ (tomando t suficientemente pequeño). Análogamente, en cualquier entorno de a existen puntos de forma $v = ty$ en donde $f(a+v) > f(a)$, por lo que f no presenta ni máximo ni mínimo en a, es decir, presenta un punto de ensilladura. □

El teorema anterior reduce la clasificación de puntos críticos, en una gran cantidad de casos, al de la determinación del signo de los valores propios de la matriz H. Veamos que ocurre cuando $n = 2$.

Corolario 4 (Clasificación de puntos críticos en $\mathbb{R}^2$).
Sean $f\colon U \to \mathbb{R}$, $a \in U \subset \mathbb{R}^2$ un punto crítico de f. Supongamos que f es de clase C^2 en alguna bola $B(a,\varepsilon) \subset U$. Sea

$$H = \begin{bmatrix} f_{xx}(a) & f_{xy}(a) \\ f_{xy}(a) & f_{yy}(a) \end{bmatrix}.$$

Entonces:

(a) *Si* $\det(H) > 0$ *y* $f_{xx}(a) > 0$*, f presenta mínimo relativo en a.*

(b) *Si* $\det(H) > 0$ *y* $f_{xx}(a) < 0$*, f presenta máximo relativo en a.*

(c) *Si* $\det(H) < 0$*, f presenta punto silla en a.*

Demostración. Como H es una matriz simétrica y real, tiene valores propios reales λ_1 y λ_2, como se puede verificar directamente, calculando las raíces de $\det(H - \lambda I) = 0$. Sabemos que

$$\det(H) = f_{xx}(a)f_{yy}(a) - f_{xy}^2(a) = \lambda_1\lambda_2, \tag{27}$$

$$\operatorname{tr}(H) = f_{xx}(a) + f_{yy}(a) = \lambda_1 + \lambda_2. \tag{28}$$

Si el determinante $\det(H) > 0$ de (27) obtenemos que los valores propios tienen el mismo signo. Además $f_{xx}(a)f_{yy}(a) > f_{xy}(a)^2 \geq 0$, y las derivadas segundas $f_{xx}(a)$, $f_{yy}(a)$ tienen igual signo. Si $f_{xx}(a) > 0$, en vista de (28) la traza $\operatorname{tr}(H)$ resulta positiva, y ambos valores propios son positivos. Se aplica entonces (a) en el teorema 12, para obtener (a) en el corolario. El caso (b) es análogo. Si $\det(H) < 0$ los valores propios son de signo opuesto, y se aplica (c) en el teorema 12 para completar la demostración del corolario. □

6. Funciones implícitas y aplicaciones

Consideremos un abierto U en $\mathbb{R}^2$, una función $F\colon U \to \mathbb{R}$, y un punto (x_0, y_0) tal que $F(x_0, y_0) = 0$. Queremos averiguar bajo que condiciones esta raíz de F se "extiende" al gráfico de una curva en el conjunto U, es decir, la ecuación $F(x, y) = 0$ tiene soluciones de la forma $\big(x, f(x)\big)$ para $x \in I = (x_0 - \varepsilon, x_0 + \varepsilon)$, con algún $\varepsilon > 0$. En este caso pensamos que "despejamos" la función $y = f(x)$ de la condición $F(x, y) = 0$, y decimos que f es la *función implícita* definida en I por la ecuación $F(x, y) = 0$.

Alternativamente, si sabemos que $F(x_0, y_0) = 0$, nos preguntamos si el conjunto $F^{-1}(0)$ es el gráfico de alguna función f en un entorno de x_0.

Teorema 13 (Teorema de la función implícita).
Sean $(x_0, y_0) \in U$ abierto en $\mathbb{R}^2$, y $F\colon U \to \mathbb{R}$ de clase C^1, tales que $F(x_0, y_0) = 0$. Supongamos que $F_y(x_0, y_0) \neq 0$. Entonces:
(a) *Existen intervalos $I = (x_0 - \delta, x_0 + \delta)$, $J = (y_0 - \varepsilon, y_0 + \varepsilon)$ tales que $I \times J \subset U$, y una función $f\colon I \to J$ de clase C^1 tal que*

$$F(x, y) = 0 \Leftrightarrow y = f(x) \text{ para todo } (x, y) \in I \times J.$$

(b) *Si $x \in I$ tenemos*

$$f'(x) = -\frac{F_x\big(x, f(x)\big)}{F_y\big(x, f(x)\big)}. \tag{29}$$

Demostración. Supongamos, para fijar ideas, que $F_y(x_0, y_0) > 0$. Como F_y es continua existen intervalos $I = (x_0 - \delta, x_0 + \delta)$, $J = (y_0 - \varepsilon, y_0 + \varepsilon)$ tales que $I \times J \subset U$ y $F_y(x, y) > 0$ si $(x, y) \in [x_0 - \delta, x_0 + \delta] \times [y_0 - \varepsilon, y_0 + \varepsilon]$.

Entonces, para $x_0 \in I$ la función de una variable $F(x_0, y)$ es estrictamente creciente en el intervalo $[y_0 - \varepsilon, y_0 + \varepsilon]$. Como $F(x_0, y_0) = 0$, tenemos $F(x_0, y_0 - \varepsilon) < 0 < F(x_0, y_0 + \varepsilon)$.

La conservación del signo de la función F en un entorno de los puntos $(x_0, y_0 - \varepsilon)$ y $(x_0, y_0 + \varepsilon)$ nos asegura (tomando otro δ menor que el anterior, si es necesario) que, para cada $x \in [x_0 - \delta, x_0 + \delta]$, tenemos

$$F(x, y_0 - \varepsilon) < 0 < F(x, y_0 + \varepsilon).$$

Como además cada función $\phi^x(y) = F(x, y)$ es continua en J, por la propiedad de Bolzano-Darboux, existe (para cada x) una raíz $y \in J$ tal que

$\phi^x(y) = F(x,y) = 0$. Definimos $f\colon I \to J$ mediante $f(x) = y$, donde y es la raíz hallada. Esto demuestra (a).

Veamos (b). Si $h \neq 0$ verifica $x+h \in I$ tenemos $F\big(x+h, f(x+h)\big) = 0$. Designamos $k = f(x+h) - f(x)$ y, por el teorema del valor medio aplicado a F en el intervalo de extremos $\big(x, f(x)\big)$ y $\big(x+k, f(x)+k\big)$, sabemos que existe $\theta \in (0,1)$ tal que

$$\begin{aligned} 0 &= F\big(x+h, f(x)+k\big) - F\big(x, f(x)\big) \\ &= \frac{\partial F}{\partial x}\big(x+\theta h, f(x)+\theta k\big)h + \frac{\partial F}{\partial y}\big(x+\theta h, f(x)+\theta k\big)k. \end{aligned}$$

Entonces

$$\frac{f(x+h)-f(x)}{h} = \frac{k}{h} = -\frac{F_x\big(x+\theta h, f(x)+\theta k\big)}{F_y\big(x+\theta h, f(x)+\theta k\big)}. \tag{30}$$

La continuidad de las derivadas parciales F_x y F_y nos permite asegurar que existen constantes M y $H > 0$ tales que

$$\Big|F_x(x,y)\Big| \leq M, \quad \Big|F_y(x,y)\Big| \geq H,$$

para todo $(x,y) \in [x_0-\delta, x_0+\delta] \times [y_0-\varepsilon, y_0+\varepsilon]$. Esto nos permite obtener que

$$\big|f(x+h)-f(x)\big| \leq \frac{M}{H}h,$$

de donde resulta la continuidad de la función f en el punto $x \in I$. Es decir, si $h \to 0$, tenemos $k \to 0$. Tomando ahora límite en (30) si $h \to 0$, obtenemos

$$\begin{aligned} f'(x) &= \lim_{h\to 0}\frac{f(x+h)-f(x)}{h} = -\lim_{h\to 0}\frac{F_x\big(x+\theta h, f(x)+\theta k\big)}{F_y\big(x+\theta h, f(x)+\theta k\big)} \\ &= -\frac{F_x\big(x, f(x)\big)}{F_y\big(x, f(x)\big)}, \end{aligned}$$

por la continuidad de las derivadas parciales, obteniendo (29). La demostración concluye observando que la derivada de f es una función continua por ser composición de funciones continuas. □

Observemos que la función f del teorema anterior es única, en el siguiente sentido. Supongamos que existe $f_0\colon I_0 \to J_0$, con I_0, J_0 intervalos centrados en x_0, y_0, tales que

$$F(x,y) = 0 \Leftrightarrow y = f_0(x) \qquad \text{para todo } (x,y) \in I_0 \times J_0.$$

Sean $x \in I \cap I_0$, $y = f(x)$, $y_0 = f_0(x)$. Si $y \neq y_0$, por ejemplo $y < y_0$, como $F(x,y) = F(x,y_0) = 0$ resulta, por el teorema de Rolle aplicado a $\phi^x(y) = F(x,y)$ en el intervalo $[y - y_0]$, que existe $\bar{y}$ tal que $F_y(x,\bar{y}) = 0$, contradiciendo $F_y > 0$ en $I \times J$. Esto demuestra la unicidad.

La fórmula (29) permite obtener que si F es de clase C^k, entonces f es de clase C^k, por la regla de la cadena. Veamos como obtener las derivadas segundas de f, suponiendo que F es de clase C^2. Podemos derivar (29), pero es mas sencillo derivar dos veces la igualdad

$$F\big(x, f(x)\big) = 0 \quad \text{para } x \in I.$$

Derivando una vez, obtenemos

$$F_x\big(x, f(x)\big) + F_y\big(x, f(x)\big) f'(x) = 0,$$

que equivale a (29). Derivando nuevamente, obtenemos

$$\begin{aligned} F_{xx}\big(x, f(x)\big) &+ 2F_{xy}\big(x, f(x)\big) f'(x) \\ &+ F_{yy}\big(x, f(x)\big) f'(x)^2 + F_y\big(x, f(x)\big) f''(x) = 0, \end{aligned} \tag{31}$$

de donde, como $F_y\big(x, f(x)\big) \neq 0$, se obtiene el valor de $f''(x)$.

Ejemplo 15. Consideremos $F \colon \mathbb{R}^2 \to \mathbb{R}$ dada por

$$F(x,y) = y^3 + 3xy + x + y.$$

Apliquemos el teorema de la función implícita en $(0,0)$, dado que $F(0,0) = 0$. La función F es de clase C^1. Además

$$F_x = 3y + 1, \quad F_y = 3y^2 + 3x + 1.$$

por lo que $F_y(0,0) = 1$ y se puede aplicar el teorema. Luego existe un entorno $I = (-\delta, \delta)$ y una función $y = f(x)$ definida en I tal que $F\big(x, f(x)\big) = 0$ para $x \in I$. En particular $f(0) = 0$. Calculemos la derivada primera:

$$f'(0) = -\frac{F_x(0,0)}{F_y(0,0)} = -1.$$

Calculemos ahora la derivada segunda. Tenemos

$$F_{xx} = 0, \quad F_{xy} = 3, \quad F_{yy} = 6y,$$

de donde $F_{xx}(0,0) = F_{yy}(0,0) = 0$, $F_{xy}(0,0) = 3$, por lo que sustituyendo en (31) tenemos $f''(0) = 6$. Como sabemos que f es de clase C^2, y conocemos el valor de sus derivadas, tenemos que su desarrollo de Taylor en $x = 0$ es

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2}x^2 + r(x) = -x + 3x^2 + r(x),$$

donde $r(x)/x^2 \to 0$ $(x \to 0)$.

Con una demostración análoga a la del teorema 13 se obtiene el siguiente resultado.

Teorema 14 (Teorema de la función implícita en $\mathbb{R}^{n+1}$).
Sean $(x_{10}, \ldots, x_{n0}, y_0) \in U$ abierto de $\mathbb{R}^{n+1}$, $F\colon U \to \mathbb{R}$ de clase C^1. Supongamos que $F(x_{10}, \ldots, x_{n0}, y_0) = 0$, y que $\partial_{n+1}F(x_{10}, \ldots, x_{n0}, y_0) \neq 0$. Entonces:
(a) *Existe una bola $B = B(x_0, \delta)$, donde $x_0 = (x_{10}, \ldots, x_{n0})$, un intervalo $J = (y_0 - \varepsilon, y_0 + \varepsilon)$, $B \times J \subset U$, y una función $f\colon B \to J$ de clase C^1, tales que*

$$F(x_1, \ldots, x_n, y) = 0 \Leftrightarrow y = f(x_1, \ldots, x_n) \quad \forall (x_1, \ldots, x_n, y) \in B \times J.$$

(b) *Si $(x_1, \ldots, x_n) \in B$ tenemos*

$$\frac{\partial f}{\partial x_i}(x_1, \ldots, x_n) = -\frac{\partial_i F\big(x_1, \ldots, x_n, f(x_1, \ldots, x_n)\big)}{\partial_{n+1} F\big(x_1, \ldots, x_n, f(x_1, \ldots, x_n)\big)}.$$

Es importante observar que el teorema de la función implícita se aplica indistintamente a cualquier variable. Elejimos la última para simplificar los enunciados de los teoremas. Por ejemplo, si $F(x_0, y_0, z_0) = 0$ para una función de tres variables de clase C^1, y $F_x(x_0, y_0, z_0) \neq 0$, obtenemos que existe una función f de dos variables tales que $F(x, y, z) = 0 \Leftrightarrow x = f(y, z)$. En lo que respecta a las derivadas parciales, obtenemos

$$\frac{\partial f}{\partial y}(y, z) = -\frac{F_y\big(f(y,z), y, z\big)}{F_x\big(f(y,z), y, z\big)},$$
$$\frac{\partial f}{\partial z}(y, z) = -\frac{F_z\big(f(y,z), y, z\big)}{F_x\big(f(y,z), y, z\big)}.$$

Veamos ahora una aplicación geométrica del teorema de la función implícita. Dado un conjunto $M \subset \mathbb{R}^n$ y un punto $p \in M$ definimos el *espacio tangente* a M en p, que designamos T_pM, como el conjunto de los vectores tangentes en el punto p a curvas contenidas en M. En otros términos

$$T_pM = \big\{v \in \mathbb{R}^n \colon \exists\ \lambda\colon (-\varepsilon,\varepsilon) \to M,\ \lambda(0) = p,\ \lambda'(0) = v\big\}.$$

Recordemos que el gradiente $\nabla F(a)$ en un punto a de una función F de varias variables indica la dirección de mayor crecimiento de F. El siguiente resultado completa esta afirmación, formalizando el hecho de que el gradiente es "ortogonal" a los conjuntos en donde F es constante. Es decir, el vector $\nabla F(a)$ es ortogonal al conjunto de nivel $c = F(a)$.

Teorema 15. *Sea $F\colon U \to \mathbb{R}$ de clase C^1 en U, abierto de $\mathbb{R}^n$. Dado c consideremos $M = F^{-1}(c)$, el conjunto de nivel c de F, y un punto $p \in M$, tal que $\nabla F(p) \neq 0$. Entonces, el espacio tangente a M en p es el subespacio ortogonal a $\nabla F(p)$, es decir $T_pM = [\nabla F(p)]^\perp$. En particular, T_pM es un subsespacio vectorial de $\mathbb{R}^n$ de dimensión $n-1$.*

Demostración. Consideremos primero $v \in T_pM$. Existe $\lambda(t)\colon (-\varepsilon,\varepsilon) \to M$ tal que $\lambda(0) = p$ y $\lambda'(0) = v$. Como $F\big(\lambda(t)\big) = c$, si derivamos, obtenemos

$$0 = \big(F \circ \lambda\big)'(0) = \langle \nabla F(p), \lambda'(0)\rangle = \langle \nabla F(p), v\rangle\,,$$

de donde $v \perp \nabla F(p)$, es decir, $v \in [\nabla F(p)]^\perp$, y concluimos que $T_pM \subset [\nabla F(p)]^\perp$

Consideremos ahora $v = (v_1,\dots,v_n) \in [\nabla F(p)]^\perp$, es decir

$$\alpha_1\partial_1 F(p) + \cdots + \alpha_n\partial_n F(p) = 0. \tag{32}$$

Como $\nabla F(p) \neq 0$, alguna derivada parcial de F no es nula, supongamos entonces que $\partial_n F(p) \neq 0$. Designemos $p = (p_1,\dots,p_n)$. Del teorema de la función implícita obtenemos un bola $B = B(a,\delta)$, con $a = (p_1,\dots,p_{n-1})$, un intervalo $J = (p_n - \varepsilon, p_n + \varepsilon)$, y una función $f\colon B \to J$ tales que

$$x_n = f(x_1,\dots,x_{n-1}) \Leftrightarrow F(x_1,\dots,x_n) = 0,\ \forall (x_1,\dots,x_n) \in B \times J.$$

Como además se verifica $\partial_i f(a) = -\partial_i F(p)/\partial_n F(p)$, para $i = 1,\dots,n-1$, en vista de (32), resulta que

$$\alpha_n = \partial_1 f(a)\alpha_1 + \cdots + \partial_{n-1} f(a)\alpha_{n-1}.$$

Por eso, la curva

$$\lambda(t) = \big(p_1 + t\alpha_1, \dots, p_{n-1} + t\alpha_{n_1}, f(p_1 + t\alpha_1, \dots, p_{n-1} + t\alpha_{n-1})\big)$$

verifica $\lambda(0) = p$, $\lambda'(0) = v$, obteniendo que $v \in T_pM$. Hemos probado entonces que $[\nabla F(p)]^\perp \subset T_pM$, concluyendo la demostración. $\square$

Dada $f\colon U \to \mathbb{R}$ de clase C^1 en U, abierto de $\mathbb{R}^n$, llamamos *gráfico* de f al conjunto $G \subset \mathbb{R}^{n+1}$ definido mediante

$$G = \{(x_1, \dots, x_n, x_{n+1})\colon (x_1, \dots, x_n) \in U, x_{n+1} = f(x_1, \dots, x_n)\}. \qquad (33)$$

El gráfico de f es entonces un conjunto que está en correspondencia biyectiva con el dominio U de f: cada punto $a = (x_1, \dots, x_n) \in U$ se corresponde con un punto $p = \big(x_1, \dots, x_n, f(x_1, \dots, x_n)\big) \in G$. En vista de que la noción de espacio tangente está definida para subconjuntos arbitrarios de $\mathbb{R}^{n+1}$, tiene sentido considerar el espacio tangente al gráfico G de la función f en un punto p, que designamos T_pG. El teorema anterior nos permite obtener una caracterización de este espacio tangente.

Corolario 5. *Sea $f\colon U \to \mathbb{R}$ de clase C^1 en U, abierto de $\mathbb{R}^n$. Entonces, dado $a = (x_1, \dots, x_n) \in U$, el espacio tangente al gráfico G de f en el punto $p = \big(x_1, \dots, x_n, f(x_1, \dots, x_n)\big)$ es el subespacio de $\mathbb{R}^{n+1}$ ortogonal al vector $\big(\partial_1 f(a), \dots, \partial_n f(a), -1\big)$. En otros términos*

$$T_pG = \big[\big(\partial_1 f(a), \dots, \partial_n f(a), -1\big)\big]^\perp.$$

Demostración. Consideremos la función auxiliar $F\colon U \times \mathbb{R} \to \mathbb{R}$ definida mediante

$$F(x_1, \dots, x_n, x_{n+1}) = f(x_1, \dots, x_n) - x_{n+1},$$

que es de clase C^1. Según la definición (33) de G, el gráfico de f es el conjunto de nivel 0 de F. Dado un punto $a = (x_1, \dots, x_n) \in U$ tenemos $p = \big(x_1, \dots, x_n, f(x_1, \dots, x_n)\big)$, y

$$\nabla F(p) = \big(\partial_1 f(a), \dots, \partial_n f(a), -1\big) \neq 0.$$

Aplicando entonces el teorema 15, obtenemos

$$T_pG = [\nabla F(p)]^\perp = \big[\big(\partial_1 f(a), \dots, \partial_n f(a), -1\big)\big]^\perp,$$

concluyendo la demostración. $\square$

Estudiemos ahora el caso $n = 2$, considerando $f\colon U \to \mathbb{R}$ de clase C^1 en U, abierto de $\mathbb{R}^2$. Consideremos $a = (x_0, y_0) \in U$, $p = \big(x_0, y_0, f(x_0, y_0)\big) \in G$, y veamos la relación que existe entre el espacio tangente al gráfico de f en p y el plano tangente al gráfico de la función f en a. Según vimos en la página 12, el plano tangente es el plano por el punto p ortogonal al vector $\big(f_x(a), f_y(a), -1\big)$. Por otra parte, el espacio tangente al gráfico de f en p es el subespacio vectorial $T_pG = \big[\big(f_x(a), f_y(a), -1\big)\big]^{\perp}$, en $\mathbb{R}^3$. En conclusión, el plano tangente está formado por los puntos del espacio afín que se obtiene sumando el punto p a los vectores de T_pG. Esto explica la denominación de *tangente* para ambos objetos matemáticos.

7. Extremos condicionados

Supongamos que queremos determinar los extremos de la suma de las tres coordenadas de los puntos de $\mathbb{R}^3$ que pertenecen a la esfera de centro en el origen y radio unidad. En otros términos, queremos determinar los extremos de $f(x, y, z) = x + y + z$ en el dominio $M = \{(x, y, z)\colon x^2 + y^2 + z^2 = 1\}$.

Más en general, si consideramos U abierto en $\mathbb{R}^n$, y dos funciones $f\colon U \to \mathbb{R}$ y $g\colon U \to \mathbb{R}$, el problema que estudiamos en esta sección es el de determinar los extremos de la función f en $M = g^{-1}(0)$, el conjunto de nivel 0 de g.

Definición 10 (Extremos condicionados).
Consideremos U abierto en $\mathbb{R}^n$, $f\colon U \to \mathbb{R}$, $g\colon U \to \mathbb{R}$, el conjunto de nivel $M = g^{-1}(0)$, y el punto $p \in M$.
(a) *Decimos que f presenta un* máximo relativo condicionado *a g en el punto p, si existe $\varepsilon > 0$ tal que*

$$f(p) \geq f(x) \quad \textit{para todo } x \in B(p, \varepsilon) \cap M.$$

La definición de mínimo relativo condicionado *es análoga, con un signo $\leq$ en lugar de $\geq$ en* (24). *Si f presenta máximo o mínimo relativo condicionado, decimos que f presenta* extremo relativo condicionado *en p.*
(b) *Si en $p \in M$ se verifica*

$$f(p) \geq f(x) \quad \textit{para todo } p \in M,$$

decimos que $f(p)$ es el máximo absoluto *de f condicionado a g, o que f* presenta su máximo absoluto condicionado a *g en p, con el valor $f(p)$. (Análogamente se define el* mínimo absoluto condicionado a *g.)*

Teorema 16 (Multiplicador de Lagrange).
Consideremos U abierto en $\mathbb{R}^3$, $f\colon U \to \mathbb{R}$, $g\colon U \to \mathbb{R}$, ambas de clase C^1, $p \in M = g^{-1}(0)$. Si f presenta un extremo relativo condicionado a g en p, y se verifica $\nabla g(p) \neq 0$, entonces existe un real λ tal que

$$\nabla f(p) = \lambda \nabla g(p).$$

El número real λ del teorema anterior se llama multiplicador de Lagrange*.*

Demostración. Como $\nabla g(p) \neq 0$, sabemos que el espacio tangente $T_pM = [\nabla g(p)]^{\perp}$. Para ver entonces que los gradientes de f y g son colineales, es suficiente ver que que $\nabla f(p) \perp T_pM$.

Sea entonces $v \in T_pM$ y $\lambda\colon (-\varepsilon, \varepsilon) \to M$ una curva diferenciable en $t = 0$ con $\lambda(0) = p$, $\lambda'(0) = v$. Como f presenta extremo en p restringido a M, la función $f \circ \lambda$ presenta extremo en $t = 0$, de donde

$$0 = (f \circ \lambda)'(0) = \langle \nabla f(p), \lambda'(0) \rangle = \langle \nabla f(p), v \rangle,$$

y obtenemos que $\nabla f(p) \perp v$. Como $v \in T_pM$ es arbitrario, resulta que $\nabla f(p) \perp T_pM$, es decir, los gradientes de f y g en p son colineales, completando la demostración. $\square$

Al igual que en la búsquda de los extermos de $f\colon U \to \mathbb{R}$, que llamamos *extremos libres*, en contraposición a los extremos condicionados, el teorema del multiplicador de Lagrange da una condición necesaria para la existencia de extremos relativos condicionados. Una vez descartados los puntos de $M = g^{-1}(0)$ en los que f, g no son de clase C^1, y aquellos en los que $\nabla g = 0$, los extremos condicionados se presentan únicamente en aquellos puntos en los que existe λ real tal que $\nabla f(p) = \lambda \nabla g(p)$. Para buscar estos puntos consideramos la función auxiliar

$$L(x, y, z, \lambda) = f(x, y, z) - \lambda g(x, y, z),$$

y observamos, que como

$$\nabla L = (f_x - \lambda g_x, f_y - \lambda g_y, f_z - \lambda g_z, g),$$

los extremos condicionados de f a g son los puntos críticos de L. Veamos dos situaciones sencillas, comenzando por el problema expuesto al principio de esta sección.

Ejemplo 16. Determinar los extremos absolutos de $f(x, y, z) = x + y + z$ en el dominio $M = \{(x, y, z) \in \mathbb{R}^3 \colon x^2 + y^2 + z^2 = 1\}$. Observemos en primer lugar que f, $g(x, y, z) = x^2 + y^2 + z^2 - 1$ son funciones de clase C^1, y que $\nabla g = (2x, 2y, 2z) \neq 0$ si $(x, y, z) \in M$. Por este motivo, todos los posibles extremos condicionados son extremos libres de la función

$$L(x, y, z, \lambda) = x + y + z + \lambda(x^2 + y^2 + z^2 - 1).$$

Para determinarlos debemos resolver el sistema de ecuaciones

$$\begin{cases} L_x & = 1 + 2\lambda x = 0 \\ L_y & = 1 + 2\lambda y = 0 \\ L_z & = 1 + 2\lambda z = 0 \\ L_\lambda & = x^2 + y^2 + z^2 - 1 = 0 \end{cases}$$

Como $\lambda = 0$ no es solución, de las tres primeras ecuaciones obtenemos $x = y = z = -1/(2\lambda)$, por lo que, en vista de la cuarta ecuación tenemos dos soluciones:

$$a = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right) \text{ para } \lambda = \frac{\sqrt{3}}{2}$$
$$b = \left(\frac{-1}{\sqrt{3}}, \frac{-1}{\sqrt{3}}, \frac{-1}{\sqrt{3}}\right) \text{ para } \lambda = \frac{-\sqrt{3}}{2}$$

Ademas $f(a) = \sqrt{3}$, y $f(b) = -\sqrt{3}$.

El teorema de Weierstrass afirma, dado que M es compacto y f continua, que existen dos puntos en donde se alcanzan el máximo y el mínimo de la función, que, necesariamente son los puntos hallados, dado que son los únicos extremos relativos.

En conclusion, f alcanza su máximo absoluto $\sqrt{3}$ en $(1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$ y su mínimo absoluto $-\sqrt{3}$ en $(-1/\sqrt{3}, -1/\sqrt{3}, -1/\sqrt{3})$.