Análisis II

Artemio González López

Madrid, enero de 1999

Índice general

Capítulo 1

Topología de $\mathbb{R}^n$. Límites y continuidad

1.1. El espacio euclídeo $\mathbb{R}^n$. Espacios euclídeos, normados y métricos.

El conjunto

$$\mathbb{R}^n = \big\{(x_1, \ldots, x_n) : x_1 \in \mathbb{R},\ 1 \le i \le n\big\}$$

es un espacio vectorial real n-dimensional, con las operaciones

$$\begin{aligned}(x_1, \ldots, x_n) + (y_1, \ldots, y_n) &= (x_1 + y_1, \ldots, x_n + y_n)\\ a(x_1, \ldots, x_n) &= (a\,x_1, \ldots, a\,x_n), \qquad \forall a \in \mathbb{R}.\end{aligned}$$

Las propiedades fundamentales de $\mathbb{R}^n$ como espacio vectorial constituyen una parte esencial del curso de Álgebra Lineal. (En efecto, recuérdese que todo espacio vectorial real n-dimensional es isomorfo a $\mathbb{R}^n$.) En Análisis lo que nos interesará es desarrollar sus propiedades *topológicas* (nociones de límite, continuidad,...) y *métricas* (distancia, ángulos,...), que son fundamentales para definir el concepto de *derivabilidad*.

Definición 1.1. Dados dos vectores $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$, su *producto escalar euclídeo* se define mediante

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

El producto escalar euclídeo en $\mathbb{R}^n$ goza de las siguientes propiedades cuya comprobación es elemental:

I) $x \cdot y = y \cdot x, \qquad x, y \in \mathbb{R}^n$

ii) $x \cdot (y + z) = x \cdot y = x \cdot z, \qquad x, y, z \in \mathbb{R}^n$

iii) $(a\, x) \cdot y = a(x \cdot y), \qquad a \in \mathbb{R},\ x, y \in \mathbb{R}^n$

iv) $\forall x \in \mathbb{R}^n,\ x \cdot x \geq 0,\ \text{y } x \cdot x = 0 \iff x = 0$

En otras palabras, el producto escalar euclídeo es una forma bilineal simétrica definida positiva. En general, un espacio vectorial *real* (aunque no necesariamente de dimensión finita) en el que hay definida una forma bilineal simétrica definida positiva recibe el nombre de **espacio euclídeo** (real). (Hay una generalización del concepto de espacio euclídeo al caso complejo, el llamado *espacio pre-Hilbertiano*, que se estudiará en la segunda parte de la asignatura de Métodos Matemáticos I.)

Ejemplo 1.2. Sea

$$E = C[a, b] = \big\{ f : [a, b] \to \mathbb{R} \,\big|\, f \text{ continua} \big\},$$

con el producto escalar definido por

$$(f, g) = \int_a^b f(x)\, g(x)\, dx. \tag{1.1}$$

Es inmediato comprobar que $(\cdot\,,\cdot) : E \times E \to \mathbb{R}$ es una forma bilineal simétrica y definida positiva. Por tanto, $\big(E, (\cdot\,,\cdot)\big)$ es un espacio euclídeo, de dimensión infinita.

Si $\big(E, (\cdot\,,\cdot)\big)$ es un espacio euclídeo, se define la *norma* de un vector $x \in E$ mediante

$$\|x\| = \sqrt{(x, x)}, \tag{1.2}$$

donde se ha de tomar la raiz cuadrada positiva. La norma cumple las siguientes dos desigualdades fundamentales:

i) $|(x, y)| \leq \|x\|\, \|y\|$ (desigualdad de *Cauchy–Schwarz*)

ii) $\|x + y\| \leq \|x\| + \|y\|$ (*desigualdad triangular*)

En efecto, veamos en primer lugar que ii) es consecuencia de i):

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

de donde se sigue ii). Para probar i), podemos suponer que $y \neq 0$, ya que si $y = 0$ i) se cumple trivialmente. Desarrollando entonces $\|x + t\, y\|^2$ obtenemos

$$\|x + t\, y\|^2 = \|x\|^2 + 2t(x, y) + t^2\, \|y\|^2 \geq 0, \qquad \forall t \in \mathbb{R}.$$

Completando el cuadrado:

$$\left(t\,\|y\| + \frac{(x,y)}{\|y\|}\right)^2 + \left(\|x\|^2 - \frac{|(x,y)|^2}{\|y\|^2}\right) \geq 0, \qquad \forall t \in \mathbb{R}.$$

Tomando $t = -\frac{(x,y)}{\|y\|^2}$ en la desigualdad anterior el primer paréntesis se anula, y queda la desigualdad de Cauchy–Schwarz.

Definición 1.3. Un *espacio normado* es un par $(E, \|\cdot\|)$, siendo E un espacio vectorial real y $\|\cdot\| : E \to \mathbb{R}$ una aplicación con las siguientes propiedades:

I) $\|x\| \geq 0 \; \forall x \in E, \quad \text{y} \quad \|x\| = 0 \Leftrightarrow x = 0$

II) $\|\lambda x\| = |\lambda|\;\|x\|, \qquad \forall \lambda \in \mathbb{R},\; \forall x \in E$

III) $\|x + y\| \leq \|x\| + \|y\|, \qquad \forall x, y \in E$

Por lo visto anteriormente, todo espacio euclídeo $\big(E, (\cdot\,,\cdot)\big)$ es un espacio normado con la *norma euclídea* (1.2) asociada al producto escalar. En particular, $\mathbb{R}^n$ es un espacio normado con la *norma usual*

$$|x| = \sqrt{x \cdot x} = \left(\sum_{i=1}^n x_i^2\right)^{1/2}$$

asociada al producto escalar euclideo, que denotaremos con el símbolo $|\cdot|$ porque es la norma que utilizaremos por defecto en $\mathbb{R}^n$. Por supuesto, se pueden definir muchas otras normas en $\mathbb{R}^n$. Por ejemplo, las fórmulas

$$\|x\| = \sum_{i=1}^n |x_i| \tag{1.3}$$

ó

$$\|x\| = \text{máx}\big\{|x_i| : 1 \leq i \leq n\big\} \tag{1.4}$$

definen dos normas en $\mathbb{R}^n$ distintas de la usual. Cualquiera de estas dos normas tiene la propiedad de no provenir de un producto escalar; en otras palabras, no hay ningún producto escalar en $\mathbb{R}^n$ cuya norma euclídea sea una de las dos normas definidas anteriormente. En particular, esto demuestra que hay espacios normados que *no* son euclídeos, es decir cuya norma no es la norma asociado a ningún producto escalar.

Ejercicio. Probar que en un espacio euclídeo $\big(E, (\cdot\,,\cdot)\big)$ el producto escalar se puede "reconstruir" a partir de la norma mediante la *identidad de polarización*

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

Utilizando esta fórmula, probar que las normas (1.3) y (1.4) no están asociadas a ningún producto escalar.

En un espacio euclídeo arbitrario E se define el *ángulo* entre dos vectores no nulos $x, y \in E$ como el único número $\theta \in [0, \pi]$ que cumple

$$\cos\theta = \frac{(x, y)}{\|x\| \, \|y\|}.$$

En efecto, el miembro derecho es un número real en el intervalo $[-1, 1]$ en virtud de la desigualdad de Cauchy–Schwarz, y la función $\cos : [0, \pi] \to [-1, 1]$ es biyectiva.

Ejemplo 1.4. Si $E = C[a, b]$, con la norma euclídea asociada al producto escalar (1.1):

$$\|f\| = \sqrt{\int_a^b f(x)^2 \, dx}. \tag{1.5}$$

Las desigualdades de Cauchy–Schwarz y triangular aplicadas a este caso proporcionan las desigualdades entre integrales siguientes:

$$\left| \int_a^b f(x)\, g(x)\, dx \right| \le \left(\int_a^b f(x)^2 \, dx \right)^{1/2} \left(\int_a^b g(x)^2 \, dx \right)^{1/2}$$

$$\left(\int_a^b [f(x) + g(x)]^2 \, dx \right)^{1/2} \le \left(\int_a^b f(x)^2 \, dx \right)^{1/2} + \left(\int_a^b g(x)^2 \, dx \right)^{1/2}$$

En un espacio normado $(E, \|\cdot\|)$ podemos definir la *distancia* entre dos puntos $x, y \in E$ mediante la fórmula

$$d(x, y) = \|x - y\| \,. \tag{1.6}$$

De las propiedades de la norma se deducen las siguientes propiedades de la distancia:

i) $\forall x, y \in E,\ d(x, y) \ge 0, \quad$ y $\quad d(x, y) = 0 \Leftrightarrow x = y$

ii) $d(x, y) = d(y, x)$

iii) $d(x, z) \le d(x, y) + d(y, z) \quad$ (*desigualdad triangular*)

La desigualdad triangular de la distancia es una consecuencia inmediata de la desigualdad análoga para la norma, ya que

$$\begin{aligned} d(x, z) &= \|x - z\| = \|(x - y) + (y - z)\| \\ &\le \|x - y\| + \|y - z\| = d(x, y) + d(y, z). \end{aligned}$$

Recibe el nombre de desigualdad triangular porque expresa la propiedad geométrica bien conocida de que la longitud del lado xz del triángulo de vertices xyz es menor que la suma de las longitudes de los dos lados restantes.

Definición 1.5. Sea M un conjunto cualquiera. Una *distancia* en M es una aplicación $d : M \times M \to \mathbb{R}$ con las propiedades i)–iii) que acabamos de enunciar. Un **espacio métrico** es un par (M, d), donde M es un conjunto y d es una distancia en M.

Por lo que acabamos de ver, los espacios normados (y por tanto los euclídeos) son espacios métricos, con la distancia (1.6) asociada a la norma. En particular, $\mathbb{R}^n$ es un espacio métrico, con la distancia

$$d(x, y) = |x - y| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

asociada a la norma euclídea. Si no se dice nada en contrario, esta es la estructura de espacio métrico que se considera usualmente en $\mathbb{R}^n$. Sin embargo, está claro que el concepto de espacio métrico es mucho más general que el de espacio normado. En efecto, a diferencia de los espacios euclídeos y normados, los espacios métricos no tienen por qué ser ni siquiera espacios vectoriales. Esto no es arbitrario, sino que es debido a que para definir los axiomas del producto escalar y de la norma hay que recurrir a las operaciones de espacio vectorial, mientras que en el enunciado de los axiomas i)–iii) de la distancia no aparecen ni la suma de puntos de M ni el producto de números reales por elementos de M. Incluso en el caso en que M es un espacio vectorial, puede probarse que hay distancias en M que no están asociadas a ninguna norma. (Por ejemplo, una de ellas es la distancia trivial que veremos a continuación.)

Ejemplo 1.6. En el espacio normado $C[a, b]$ con la norma (1.5), la distancia asociada a la norma está dada por:

$$d(f, g) = \sqrt{\int_a^b [(f(x) - g(x)]^2 \, dx}.$$

Por ejemplo, se calcula fácilmente que la distancia entre las funciones $\operatorname{sen} x$ y $\cos x$ en $C[0, 2\pi]$ es $\sqrt{2\pi} \approx 2{,}507$, mientras que la distancia entre x y x^2 en $C[0, 1]$ es $1/\sqrt{30} \approx 0{,}183$.

1.2. Interior, cierre, frontera. Abiertos y cerrados.

Vamos a estudiar en el resto de este capítulo la topología definida en un espacio métrico por la distancia. Para ello, es fundamental el concepto de bola abierta, que definimos a continuación:

Definición 1.7. Sea (M, d) un espacio métrico, a un punto de M y $r > 0$. La **bola abierta** de centro a y radio r es el conjunto $B_r(a)$ definido por:

$$B_r(a) = \{x \in M : d(x, a) < r\}$$

Análogamente, si $r \geq 0$ se define la *bola cerrada* de centro a y radio r mediante

$$\overline{B}_r(a) = \{x \in M : d(x, a) \leq r\}.$$

En un espacio normado,

$$B_r(a) = \{x \in M : \|x - a\| < r\};$$

en particular, en $\mathbb{R}^n$ se tiene

$$B_r(a) = \{x \in M : |x - a| < r\}.$$

Evidentemente, se obtienen fórmulas análogas para $\overline{B}_r(a)$ sin más que reemplazar $<$ por $\leq$.

En $\mathbb{R}$ las bolas abiertas son intervalos abiertos, en $\mathbb{R}^2$ son círculos (excluída la circunferencia), y en $\mathbb{R}^3$ son esferas sólidas (excluída la superficie esférica). En general, es evidente que la forma de las bolas en un espacio métrico depende no sólo del conjunto M sino también de la distancia d. Así, si en $\mathbb{R}^2$ consideramos la distancia asociada a la norma (1.3) entonces las bolas abiertas son cuadrados cuyos lados forman un ángulo de 45° con los ejes de coordenadas, mientras que si la distancia es la asociada a la norma (1.4) las bolas son cuadrados con lados paralelos a los ejes coordenados. Un ejemplo todavía más exótico es el siguiente:

Ejemplo 1.8. Sea M un conjunto no vacío arbitrario, y definamos la aplicación $d : M \times M \to \mathbb{R}$ mediante:

$$d(x, y) = \begin{cases} 0, & x = y \\ 1, & x \neq y. \end{cases} \tag{1.7}$$

Es fácil probar que d es una distancia en M. El único axioma de la distancia cuya verificación no es trivial es la desigualdad triangular. Pero si $x = z$ entonces $d(x, z) = 0$ es trivialmente menor que $d(x, y) + d(y, z)$, mientras que si $x \neq z$ entonces $d(x, z) = 1$, y $d(x, y) + d(y, z) \geq 1$, porque en caso contrario $d(x, y) = d(y, z) = 0 \Longrightarrow x = y, x = z \Longrightarrow x = z$. En este espacio métrico, es inmediato comprobar que

$$B_r(a) = \begin{cases} \{a\}, & r \leq 1 \\ M, & r > 1. \end{cases}$$

En el espacio métrico del ejemplo anterior, las bolas de radio ≤ 1 son conjuntos finitos. Esto no puede ocurrir en un espacio normado:

Proposición 1.9. *En un espacio normado distinto de $\{0\}$ las bolas abiertas (y por tanto las cerradas) son conjuntos infinitos.*

Demostración. En efecto, sea $(E, \|\cdot\|)$ un espacio normado, y escojamos un vector cualquiera v de norma 1. Un vector de la forma $a + t\,v$ pertenece a la bola abierta $B_r(a)$ si y sólo si $t \in (-r, r)$, ya que

$$d(a, a + t\,v) = \|a + t\,v - a\| = \|t\,v\| = |t|\ \|v\| = |t|\,.$$

Por tanto, $B_r(a)$ contiene al segmento $\big\{a + t\,v : t \in (-r, r)\big\}$, y dicho segmento es un conjunto infinito, ya que se puede poner en correspondencia biunívoca con el intervalo $(-r, r)$, que es no vacío al ser $r > 0$. *Q.E.D.*

Ejemplo 1.10. Sea $C_0(\mathbb{R})$ el conjunto formado por las funciones continuas y acotadas de $\mathbb{R}$ en $\mathbb{R}$. Es fácil ver que $C_0(\mathbb{R})$ es un espacio vectorial real. En $C_0(\mathbb{R})$ definimos

$$\|f\| = \sup\big\{|f(x)| : x \in \mathbb{R}\big\}.$$

Es fácil ver que $\|f\|$ está bien definida para toda $f \in C_0(\mathbb{R})$, precisamente por ser acotadas las funciones en dicho espacio. También es inmediato verificar que $\|\cdot\|$ cumple los axiomas de la norma. La distancia asociada a esta norma está dada por:

$$d(f, g) = \sup\big\{|f(x) - g(x)| : x \in \mathbb{R}\big\}.$$

Veamos cómo son las bolas cerradas en este espacio métrico. Si $f \in C_0(\mathbb{R})$ y $r \geq 0$, Una función continua y acotada g pertenece a la bola cerrada $\overline{B}_r(f)$ si y sólo si

$$\sup\big\{|f(x) - g(x)| : x \in \mathbb{R}\big\} \leq r.$$

Es inmediato comprobar (utilizando la definición y las propiedades del supremo de un conjunto de números reales) que lo anterior es equivalente a la desigualdad

$$|f(x) - g(x)| \leq r, \qquad \forall x \in \mathbb{R},$$

o lo que es lo mismo

$$f(x) - r \leq g(x) \leq g(x) + r, \qquad \forall x \in \mathbb{R}.$$

Por tanto, la bola cerrada $\overline{B}_r(f)$ está formada por todas las funciones continuas y acotadas g cuya gráfica está comprendida en la “bandaçerrada limitada por las gráficas de las funciones $f-r$ y $f+r$. No es cierto, sin embargo, que la bola abierta $B_r(f)$ esté formada por todas las funciones $g \in C_0(\mathbb{R})$ cuya gráfica esté comprendida en la banda *abierta* limitada por las gráficas de las funciones $f - r$ y $f + r$. Por ejemplo, si $f = 0$ y $g(x) = x^2/(x^2 + 1) \in C_0(\mathbb{R})$, la gráfica de g está dentro de la banda abierta limitada por las funciones $f - 1 = -1$ y $f + 1 = 1$, pero

$$d(g, f) = d(g, 0) = \sup\big\{|g(x)| : x \in \mathbb{R}\big\} = 1,$$

luego $g \notin B_1(f)$.

Ejercicio. Probar que $g \in B_r(f)$ si y sólo si la gráfica de g está en la banda abierta limitada por las funciones $f - \rho$ y $f + \rho$, para algún $0 < \rho < r$.

Definición 1.11. Sea (M, d) un espacio métrico, y sea $A \subset M$.

i) $x \in M$ es un *punto interior* a $A \iff \exists r > 0$ tal que $B_r(x) \subset A$.

ii) $x \in M$ es un *punto exterior* a $A \iff \exists r > 0$ tal que $B_r(x) \subset A^c$, siendo $A^c = M - A$ el complementario de A:

$$A^c = \{y \in M : y \notin A\}.$$

iii) $x \in M$ es un *punto frontera* de A si x no es punto interior ni exterior a A. Equivalentemente, para todo $r > 0$ $B_r(x) \cap A \neq \emptyset$ y $B_r(x) \cap A^c \neq \emptyset$.

Utilizaremos a partir de ahora la siguiente notación:

$$\overset{\circ}{A} = \{x \in M : x \text{ es punto interior a } A\} \equiv \textit{interior} \text{ de} A$$

$$\operatorname{ext}(A) = \{x \in M : x \text{ es punto exterior a } A\} \equiv \textit{exterior} \text{ de} A$$

$$\operatorname{fr}(A) = \{x \in M : x \text{ es punto frontera de } A\} \equiv \textit{frontera} \text{ de} A$$

Las siguientes propiedades de los conjuntos $\overset{\circ}{A}$, $\operatorname{fr}(A)$ y $\operatorname{ext}(A)$ son consecuencia inmediata de sus definiciones:

i) $M = \overset{\circ}{A} \cup \operatorname{fr}(A) \cup \operatorname{ext}(A)$, siendo los conjuntos $\overset{\circ}{A}$, $\operatorname{fr}(A)$ y $\operatorname{ext}(A)$ disjuntos dos a dos

ii) $\operatorname{ext}(A) = (A^c)^\circ \quad (\Longrightarrow \overset{\circ}{A} = \operatorname{ext}(A^c))$

iii) $\operatorname{fr}(A) = \operatorname{fr}(A^c)$

iv) $\overset{\circ}{A} \subset A \quad (\Longrightarrow \operatorname{ext}(A) \subset A^c)$

Definición 1.12. Dado un subconjunto A de un espacio métrico (M, d), el *cierre* de A es el conjunto $\overline{A}$ definido por

$$\overline{A} = \overset{\circ}{A} \cup \operatorname{fr}(A).$$

En otras palabras, $x \in \overline{A}$ si y sólo si para todo $r > 0$, $B_r(x) \cap A \neq \emptyset$. Por lo tanto,

$$\operatorname{fr}(A) \subset \overline{A}, \qquad A \subset \overline{A}.$$

A los puntos del cierre de A se les llama a veces *puntos adherentes* a A, y al cierre de A se le llama también *adherencia* o *clausura* de A.

Los puntos del cierre de A se dividen en dos clases disjuntas: puntos aislados y puntos de acumulación. Por definición, $x \in M$ es un *punto aislado* de A si existe $r > 0$ tal que $B_r(x) \cap A = \{x\}$. En particular, los puntos aislados de A pertenecen a A y por tanto a $\overline{A}$. Los puntos de $\overline{A}$ que no son aislados se denominan *puntos de acumulación* de A. Equivalentemente, x es un punto de acumulación de A si y sólo si para todo $r > 0$ $(B_r(x) - \{x\}) \cap A \neq \emptyset$. Por construcción, $\overline{A}$ es la unión disjunta

$$\overline{A} = A' \cup \big\{x \in M : x \text{ es punto aislado de } A\big\}, \tag{1.8}$$

siendo A' el conjunto formado por los puntos de acumulación de A (*conjunto derivado* de A). Si además M es un espacio normado (con la distancia asociada a la norma) se verifican las dos propiedades siguientes:

i) x punto aislado de $A \Longrightarrow x \in \mathrm{fr}(A)$

ii) $\mathring{A} \subset A'$

Demostración. La demostración se basa en que las bolas abiertas en un espacio normado no nulo son conjuntos infinitos (Proposición 1.9).
i) Si x es un punto aislado de A, $x \in A$, y toda bola centrada en x corta a A por lo menos en x. Basta por tanto probar que para todo $r > 0$ $B_r(x) \cap A^c \neq \emptyset$. Como x es punto aislado de A, existe un $\epsilon > 0$ tal que $B_\epsilon(x) \cap A = \{x\}$. Dado $r > 0$, sea $\rho = \min\{r, \epsilon\}$. Entonces $0 < \rho \leq \epsilon$ y $B_\rho(x) \subset B_\epsilon(x) \subset A$, por lo que $B_\rho(x) \cap A = \{x\}$. Al ser $B_\rho(x) \neq \{x\}$, existe $z \in B_\rho(x)$ tal que $z \neq x$. Entonces $z \in A^c$, pues $B_\rho(x) \cap A = \{x\}$, y $z \in B_r(x)$, pues $B_r(x) \supset B_\rho(x)$.
ii) Si $x \in \mathring{A}$, existe un $\epsilon > 0$ tal que $B_\epsilon(x) \subset A$. Dado $r > 0$, sea $\rho = \min\{r, \epsilon\}$. Entonces $0 < \rho \leq \epsilon$, por lo que $B_\rho(x) \subset B_\epsilon(x) \subset A$, y por tanto

$$(B_r(x) - \{x\}) \cap A \subset (B_\rho(x) - \{x\}) \cap A = B_\rho(x) - \{x\} \neq \emptyset,$$

porque en caso contrario $B_\rho(x)$ tendría un sólo punto. *Q.E.D.*

Ejemplo 1.13. Sea $M = \mathbb{R}^2$ con la distancia usual, y sea

$$A = [0,1) \times [0,1) \cup \left\{\left(2 - \frac{1}{n}, 0\right) : n \in \mathbb{N}\right\} \equiv [0,1) \times [0,1) \cup S.$$

(Nótese que $(1,0) \in S \subset A$, mientras que $(2,0) \notin A$.) Intuitivamente, es evidente que el interior de A es el cuadrado abierto

$$\mathring{A} = (0,1) \times (0,1) \equiv C.$$

La frontera de A está formada por los lados de C, los puntos de la sucesión S y el límite de dicha sucesión $(2,0)$:

$$\mathrm{fr}(A) = \mathrm{fr}(C) \cup S \cup \{(2,0)\},$$

siendo

$$\operatorname{fr}(C) = [0,1] \times \{0\} \cup [0,1] \times \{1\} \cup \{0\} \times [0,1] \cup \{1\} \times [0,1].$$

Por tanto el cierre de A está dado por

$$\overline{A} = \mathring{A} \cup \operatorname{fr}(A) = [0,1] \times [0,1] \cup S \cup \{(2,0)\},$$

y el exterior es

$$\operatorname{ext}(A) = \mathbb{R}^2 - \overline{A}.$$

Los puntos aislados de A son los puntos de S menos el punto $(1,0)$, es decir

$$\left\{ \left(1 - \frac{1}{n}, 0\right) : n \in \mathbb{N}, n \geq 2 \right\};$$

por tanto, el conjunto derivado de A está dado por

$$A' = \overline{A} - (S - \{(1,0)\}) = [0,1] \times [0,1] \cup \{(2,0)\}.$$

(Probaremos más adelante teoremas que permiten justificar rigurosamente el cálculo de $\mathring{A}$, $\operatorname{fr}(A)$, $\operatorname{ext}(A)$ y A' en este ejemplo.)

Definición 1.14. Sea (M,d) un espacio métrico, y sea $A \subset M$.

I) A es **abierto** $\iff$ $A = \mathring{A}$

II) A es **cerrado** $\iff$ $\overline{A}$ es abierto

- Como siempre se cumple que $\mathring{A} \subset A$,

$$A \text{ es abierto} \iff A \subset \mathring{A}.$$

- En virtud de la anterior, A es abierto si y sólo si A contiene una bola abierta centrada en cada uno de sus puntos. En otras palabras,

$$A \text{ abierto} \iff \forall x \in A, \exists r > 0 \text{ tal que } B_r(x) \subset A.$$

- Un conjunto puede ser a la vez abierto y cerrado (por ejemplo, veremos a continuación que M y $\emptyset$ lo son), o no ser ni abierto ni cerrado (como, por ejemplo, un intervalo acotado semicerrado $[a,b)$ en $\mathbb{R}$).

Proposición 1.15. *En un espacio métrico (M,d), M y $\emptyset$ son abiertos y cerrados.*

Demostración. En efecto, $\mathring{M} = M$ (pues M contiene una bola abierta centrada en cada uno de sus puntos), y $\mathring{\emptyset} = \emptyset$ (pues $\mathring{\emptyset} \subset \emptyset$). Por tanto, M y $\emptyset$ son abiertos. Luego también son cerrados, ya que $M = \emptyset^c$ y $\emptyset = M^c$. *Q.E.D.*

Proposición 1.16. *Las bolas abiertas en un espacio métrico son conjuntos abiertos.*

Demostración. Sea (M, d) espacio métrico, sea $a \in M$, y sea $r > 0$. Para ver que $B_r(a)$ es abierto, hay que probar que para todo $y \in B_r(a)$ existe $\delta > 0$ tal que $B_\delta(y) \subset B_r(a)$. Pero si $z \in B_\delta(y)$ entonces se tiene

$$d(z, a) \leq d(z, y) + d(y, a) < \delta + d(y, a) = r$$

si tomamos $\delta = r - d(y, a) > 0$ (nótese que $r > 0$ por ser $y \in B_r(a)$). *Q.E.D.*

Ejercicio. Probar que en un espacio métrico arbitrario las bolas cerradas son conjuntos cerrados.

Proposición 1.17. *Sea (M, d) un espacio métrico, $A \subset M$. Entonces A es cerrado si y sólo si $\overline{A} = A$.*

Demostración.

$$\begin{aligned} A \text{ cerrado} &\iff A^c \text{ abierto} \iff (A^c)^\circ = A^c \\ &\iff \operatorname{ext}(A) = A^c = \left[\mathring{A} \cup \operatorname{fr}(A)\right]^c = (\overline{A})^c \\ &\iff A = \overline{A}. \end{aligned}$$

Q.E.D.

Como siempre se cumple que $A \subset \overline{A}$,

$$A \text{ cerrado} \iff \overline{A} \subset A.$$

Por otra parte, al ser $\overline{A} = \mathring{A} \cup \operatorname{fr}(A)$ y $\mathring{A} \subset A$,

$$A \text{ cerrado} \iff \operatorname{fr}(A) \subset A.$$

En otras palabras,*un conjunto es cerrado si y sólo si contiene todos sus puntos frontera.* Del mismo modo, de (1.8) y de que los puntos aislados de A están en A se deduce que

$$A \text{ cerrado} \iff A' \subset A.$$

Luego *un conjunto es cerrado si y sólo si contiene todos sus puntos de acumulación.*

Ejemplo 1.18. Sea (M, d) el espacio métrico trivial, en que la distancia entre dos puntos está dada por (1.7). Veamos que todo subconjunto de M es a la vez abierto y cerrado, y con frontera vacía. En efecto, sea $A \subset M$, y sea $a \in A$. Como $B_{\frac{1}{2}}(a) = \{a\} \subset A$, A contiene un entorno de cada uno de sus puntos, y es por tanto abierto. Esto prueba que todo subconjunto de M

es abierto, lo que a su vez implica (¿por qué?) que todo subconjunto de M es también cerrado. De esto se sigue que $\mathrm{fr}(A)$ es vacío para todo $A \subset M$, ya que

$$\mathrm{fr}(A) = M - \big(\mathring{A} \cup (A^c)^\circ\big) = M - (A \cup A^c) = M - M = \emptyset.$$

Una consecuencia interesante de todo esto es que en este espacio métrico hay una bola abierta cuyo cierre no coincide con la bola cerrada del mismo radio. En efecto, las bolas cerradas en este espacio métrico son

$$\overline{B}_r(a) = \begin{cases} \{a\}, & r < 1 \\ M, & r \geq 1. \end{cases}$$

Por tanto

$$\overline{B}_1(a) = M, \qquad \overline{B_1(a)} = B_1(a) = \{a\}$$

y $\overline{B}_1(a) \neq \overline{B_1(a)}$ si M tiene más de un elemento.

Ejercicio. Probar que en un espacio normado siempre se cumple que $\overline{B}_r(a) = \overline{B_r(a)}$, y

$$\mathrm{fr}\,(B_r(a)) = \big\{x \in M : d(x,a) = r\big\}.$$

¿Es esta última propiedad válida en cualquier espacio métrico?

1.3. Límites y continuidad

Sean (M, d) y (M_1, d_1) dos espacios métricos. Una función $f : M \to M_1$ es cualquier regla que a cada x perteneciente a un cierto subconjunto $D \subset M$ le asigna un *único* elemento $f(x) \in M_1$. A $f(x)$ se le llama la *imagen* de x bajo f. El *dominio* de f es el conjunto $\mathrm{dom}(f) = D \subset M$ en que la actuación de f está bien definida, es decir

$$\mathrm{dom}(f) = \{x \in M : \exists f(x)\}.$$

La *imagen* de f es el conjunto $\mathrm{Im}(f) \subset M_1$ al que pertenecen los valores de $f(x)$ cuando x recorre $\mathrm{dom}(f)$, es decir

$$\mathrm{Im}(f) = \{f(x) : x \in M\}.$$

Nótese que, aunque escribamos "$f : M \to M_1$", en general $\mathrm{dom}(f)$ e $\mathrm{Im}(f)$ son subconjuntos propios de M y M_1, respectivamente. Si $A \subset M$, utilizaremos a veces la notación $f(A)$ para denotar el conjunto de las imágenes bajo f de los elementos de A en que f esté bien definida, es decir

$$f(A) = \{f(x) : x \in A \cap \mathrm{dom}(f)\} \subset M_1.$$

Por tanto, $f(A) = f\big(A \cap \operatorname{dom}(f)\big)$; en particular,

$$f(M) = f\big(\operatorname{dom}(f)\big) = \operatorname{Im}(f).$$

Si $B \subset M_1$, el conjunto $f^{-1}(B)$ es el conjunto de los elementos de $\operatorname{dom}(f)$ cuyas imágenes caen en B, es decir

$$f^{-1}(B) = \{x \in \operatorname{dom}(f) : f(x) \in B\} \subset M.$$

Nótese que $f^{-1}(B) = f^{-1}(B \cap \operatorname{Im}(f))$, y

$$\operatorname{dom}(f) = f^{-1}(\operatorname{Im}(f)) = f^{-1}(M_1).$$

Ejemplo 1.19. Sea $f : \mathbb{R}^2 \to \mathbb{R}$ la función definida por la fórmula

$$f(x, y) = \sin x + \sqrt{x - y^2}.$$

Claramente, $\operatorname{dom}(f)$ es el subconjunto de los puntos $(x, y) \in \mathbb{R}^2$ que satisfacen la desigualdad $x \geq y^2$, es decir la región a la derecha de la parábola $x = y^2$ (incluida la propia parábola). En cuanto a la imagen, un sencillo cálculo muestra que es el intervalo $[-1, \infty)$, pues $\sin x + \sqrt{x - y^2} \geq -1$ para todo $(x, y) \in \operatorname{dom}(f)$, y si $t \geq -1$ la ecuación $\sin x + \sqrt{x - y^2} = t$ siempre puede satisfacerse. (Si $t \in [-1, 0]$, tomamos un $x \geq 0$ tal que $\sin x = t$, e $y = \sqrt{x}$. Si $t \in (0, \infty)$, tomamos $x = k\pi \geq t^2$, con $k \in \mathbb{N}$, e $y = \sqrt{x - t^2}$.)

Definición 1.20. Sea $f : M \to M_1$, sea $b \in M_1$, y sea $a \in \operatorname{dom}(f)'$. Se dice que b es el límite de f en, y se escribe

$$\lim_{x \to a} f(x) = b,$$

si para todo $\epsilon > 0$ existe $\delta > 0$ tal que

$$f\left(B_\delta(a) - \{a\}\right) \subset B_\epsilon(b).$$

Esta definición, una de las más importantes del curso, merece varios comentarios:

I) f no tiene por qué estar definida en el punto a, y si lo está el valor de f en a no interviene para nada en la definición de $\lim_{x \to a} f(x)$. En particular, ni la existencia de $\lim_{x \to a} f(x)$ ni su valor dependen de la existencia ó el valor de f en a.

II) Para que exista el límite de f cuando $x \to a$ es imprescindible que a sea un punto de acumulación de $\operatorname{dom}(f)$. Esta condición no es arbitraria, sino que se impone para que $f\left(B_r(a) - \{a\}\right)$ sea no vacío para todo $r > 0$. En caso contrario, la condición de límite se cumpliría trivialmente tomando cualquier δ cualquier $r > 0$ tal que $(B_r(a) - \{a\}) \cap \operatorname{dom}(f) = \emptyset$.

III) La condición (1.20) se puede expresar como sigue:

$$\forall \epsilon > 0,\ \exists \delta > 0 \text{ t.q. } \big(x \in \operatorname{dom}(f),\ 0 < d(x,a) < \delta\big) \Longrightarrow d_1\big(f(x), b\big) < \epsilon.$$

De esto se sigue que

$$\lim_{x\to a} f(x) = b \qquad \Longleftrightarrow \qquad \lim_{x\to a} d_1\big(f(x), b\big) = 0.$$

El segundo límite es el límite de la función numérica (a valores reales) $d\big(f(\cdot), b\big)$.

Proposición 1.21. *Sea $f : M \to M'$. Entonces $\exists \lim_{x\to a} f(x) = b$ si y sólo si para toda sucesión $\big(x_n\big)_{n=1}^{\infty}$ tal que $x_n \neq a$ para n suficientemente grande y $\lim_{n\to\infty} x_n = a$ se tiene $\lim_{n\to\infty} f(x_n) = b$.*

Demostración.

$\Longrightarrow$) Dado $\epsilon > 0$, $\exists \delta > 0$ tal que $0 < d(x,a) < \delta \Longrightarrow d(f(x), b) < \epsilon$. Escojamos $N \in \mathbb{N}$ tal que $n \geq N \Longrightarrow 0 < d(x_n, a) < \delta$; entonces $\forall n \geq N$ se cumple $d(f(x_n), b) < \epsilon \Longrightarrow \lim_{n\to\infty} f(x_n) = b$.

$\Longleftarrow$) Supongamos que $\lim_{x\to a} f(x) \neq b$ (lo cual ocurrirá, en particular, si no existe dicho límite). Entonces, $\exists \epsilon > 0$ tal que $\forall \delta > 0$ $f\left(B_\delta(a) - \{a\}\right) \not\subset B_\epsilon(b)$. Apliquemos lo anterior a $\delta = \frac{1}{n}$: $\forall n \in \mathbb{N}, \exists x_n$ con $0 < d(x_n, a) < \frac{1}{n}$ tal que $d(f(x_n), b) \geq \epsilon$. Consideremos la sucesión $\big(x_n\big)_{n=1}^{\infty}$ construída de esta forma. Por un lado, se tiene claramente que $x_n \neq a$ para todo n, y $\lim_{n\to\infty} x_n = a$, pues $d(x_n, a) < \frac{1}{n}@ >> n \to \infty > 0$. Por otra parte, $\lim_{n\to\infty} f(x_n) \neq b$, pues $d(f(x_n), b) \geq \epsilon$ $\forall n \in \mathbb{N}$. *Q.E.D.*

Una generalización de lo anterior consiste en considerar el comportamiento de $f : M \to M'$ a lo largo de curvas continuas que pasen por el punto a. Concretamente, una *curva continua por a* es una función continua $\gamma : I \to M$, donde I es un intervalo abierto de la recta real que contiene a 0, tal que $\gamma(0) = a$. Por definición, el *límite de f a lo largo de γ* en el punto $a \in M$ es el $\lim_{t\to 0} f\big(\gamma(t)\big)$. Se tiene entonces el siguiente resultado:

Proposición 1.22. *Si $f : M \to M'$, $\lim_{x\to a} f(x) = b$ y γ es cualquier curva continua por a, entonces el límite de f a lo largo de γ en el punto a es igual a b.*

Demostración. Basta considerar la función $f \circ \gamma : I \to M'$. Como γ es continua en 0, $\lim_{t\to 0} \gamma(t) = \gamma(0) = a$, y por el teorema sobre el límite de una composición de funciones y la hipótesis acerca de la existencia del límite de f cuando $x \to a$ se obtiene el resultado anterior. *Q.E.D.*

Comentarios:

I) Por lo tanto, una condición *necesaria* para que el $\lim_{x\to a} f(x)$ exista es que exista el límite de f a lo largo de cualquier curva continua por a en el punto a, y que todos estos límites sean iguales. (Se puede probar que esta condición es también suficiente, aunque esto último no es de utilidad práctica para probar la existencia del límite de una función en un punto.)

II) En particular, si se encuentran dos curvas por a tal que los límites de f en a a lo largo de dichas curvas son distintos, no puede existir el límite de f en a.

Ejemplo: Sea

$$f(x,y) = \begin{cases} \dfrac{x^3+2y^2}{x^2+y^4}, & (x,y) \neq (0,0) \\ 0, & (x,y) = (0,0). \end{cases}$$

Entonces f no es continua en el origen, pues no existe el límite de f en $(0,0)$. En efecto, si tendemos al origen a lo largo de una recta $\gamma(t) = (t, \lambda t)$ de pendiente λ entonces se tiene:

$$\lim_{t\to 0} f\bigl(\gamma(t)\bigr) = \lim_{t\to 0} \frac{2\lambda^2+t}{1+\lambda^4 t^2} = 2\lambda^2.$$

Este límite existe, pero depende de la pendiente λ de la recta por la que tendamos al origen. Por tanto, no existe $\lim_{(x,y)\to(0,0)} f(x,y)$.

III) Sin embargo, si el límite de f a lo largo de una familia de curvas por a existe y es igual a b, ¡no se sigue necesariamente que exista el límite de f en a!

Ejemplo: Sea ahora

$$f(x,y) = \begin{cases} \dfrac{xy^2}{x^2+y^4}, & (x,y) \neq (0,0) \\ 0, & (x,y) = (0,0). \end{cases}$$

Si de nuevo tendemos a $(0,0)$ a lo largo de la recta $\gamma(t) = (t, \lambda t)$ obtenemos

$$\lim_{t\to 0} f\bigl(\gamma(t)\bigr) = \lim_{t\to 0} \frac{\lambda^2 t}{1+\lambda^4 t^2} = 0.$$

Por tanto, el límite de f a lo largo de cualquier recta por el origen existe y es igual a 0. Pero sería incorrecto deducir de esto que $\lim_{(x,y)\to(0,0)} f(x,y) = 0$. En efecto, si tendemos al origen a lo largo de la parábola $\gamma(t) = (t^2, t)$ obtenemos

$$\lim_{t\to 0} f\bigl(\gamma(t)\bigr) = \lim_{t\to 0} \frac{t^4}{2t^4} = \frac{1}{2}.$$

Este límite es distinto del límite de f en el origen a lo largo de una recta cualquiera, por lo que no existe el $\lim_{(x,y)\to(0,0)} f(x,y)$. $\square$

Veamos a continuación algunos lemas que son útiles en el cálculo de límites de funciones escalares.

Lema 1.23. *Si* $f : M \to \mathbb{R}$ *y* $\lim_{x\to a} f(x) > 0$*, entonces* $f > 0$ *en un entorno reducido de* a*.*

Demostración. Sea $b = \lim_{x\to a} f(x) > 0$. Por definición de límite, $\forall\epsilon > 0$ $\exists\delta(\epsilon) > 0$ tal que $|f(x) - b| < \epsilon$ si $x \in B_{\delta(\epsilon)}(a) - \{a\}$. En particular, $x \in B_{\delta(\epsilon)}(a) - \{a\} \Longrightarrow f(x) > b - \epsilon$, y por tanto basta tomar $\epsilon = b/2 > 0$ para concluir que $f > 0$ en $B_{\delta(\epsilon)}(a) - \{a\}$. *Q.E.D.*

Corolario 1.24. *Si* $f : M \to \mathbb{R}$ *es continua en* a *y* $f(a) > 0$*, entonces* $f > 0$ *en un entorno de* a*.*

Nota. Aplicando el lema anterior a $-f$ se obtiene un resultado análogo cuando $\lim_{x\to a} f(x) < 0$.

Lema 1.25. *Sean* $f, g, h : M \to \mathbb{R}$*. Si* $g(x) \le f(x) \le h(x)$ $\forall x$ *en un entorno reducido de* a*, y* $\lim_{x\to a} g(x) = \lim_{x\to a} h(x) = b$*, entonces* $\exists \lim_{x\to a} f(x) = b$*.*

Demostración. Dado $\epsilon > 0$, $\exists\delta_1 > 0$ tal que $|g(x) - b| < \epsilon$ si $0 < d(x, a) < \delta_1$, y análogamente $\exists\delta_2 > 0$ tal que $|h(x) - b| < \epsilon$ si $0 < d(x, a) < \delta_2$. Tomando $\delta = \min(\delta_1, \delta_2, R)$ (siendo R el radio del entorno reducido del enunciado), para todo x tal que $0 < d(x, a) < \delta$ se tiene entonces

$$b - \epsilon < g(x) \le f(x) \le h(x) < b + \epsilon,$$

es decir $|f(x) - b| < \epsilon$. *Q.E.D.*

Lema 1.26. *Sean* $f, g : M \to \mathbb{R}$ *tales que* f *está acotada en un entorno reducido de* $a \in M$*, y* $\lim_{x\to a} g(x) = 0$*. Entonces* $\lim_{x\to a} f(x)g(x) = 0$*.*

Demostración. Sea $|f(x)| < K$ en $B_r(a) - \{a\}$. Entonces $\forall x \in B_r(a) - \{a\}$ se tiene

$$0 \le |f(x)g(x)| = |f(x)|\,|g(x)| < K\,|g(x)|\,,$$

y basta aplicar el lema anterior. *Q.E.D.*

Definición 1.27. Si $f : M \to M'$ y $A \subset M'$, la *imagen inversa* de A es el subconjunto

$$f^{-1}(A) = \big\{x \in M : f(x) \in A\big\} \subset \operatorname{dom}(f).$$

El resultado más importante de esta sección afirma que *la imagen inversa de un abierto bajo una función continua es un abierto*:

Proposición 1.28. *Si* $f : M \to M'$ *es continua en* M *y* $V \subset M'$ *es abierto, entonces* $f^{-1}(V)$ *es abierto en* M*.*

Demostración. Sea $a \in f^{-1}(V)$, y por tanto $f(a) \in V$. Por ser V abierto, $\exists \epsilon > 0$ tal que $B_\epsilon\big(f(a)\big) \subset V$. Al ser f continua en a, $\exists \delta > 0$ tal que $f\big(B_\delta(a)\big) \subset B_\epsilon\big(f(a)\big) \implies f\big(B_\delta(a)\big) \subset V \implies B_\delta(a) \cap \mathrm{dom}(f) = B_\delta(a) \subset f^{-1}(V)$. *Q.E.D.*

Corolario 1.29. *Si $f : M \to M'$ es continua en M y $A \subset M'$ es cerrado, entonces $f^{-1}(A)$ es cerrado en M.*

Demostración. A^c es abierto, y por tanto $f^{-1}(A^c)$ es abierto. Pero

$$f^{-1}(A^c) = \big\{x \in M : f(x) \in A^c\big\} = \big\{x \in M : f(x) \notin A\big\} = M - f^{-1}(A),$$

y por tanto $f^{-1}(A)$ es cerrado. *Q.E.D.*

Nota. Si $f : M \to M'$ es continua y $S \subset M$ es abierto (cerrado), $f(S)$ no tiene por qué ser abierto (cerrado).

Ejemplo 1.30. Sea $f : \mathbb{R} \to \mathbb{R}$ definida por $f(x) = x^2$, $\forall x \in \mathbb{R}$, y $S = (-1, 1)$. Entonces $f(S) = [0, 1)$, que no es abierto. Por otra parte, si $f(x) = e^{-x}$ $\forall x \in \mathbb{R}$ se tiene $f([0, \infty]) = (0, 1]$, que no es cerrado.

Nota. Se dice que $f : M \to \mathbb{R}$ es una aplicación *abierta* si transforma abiertos en abiertos. El ejemplo anterior demuestra que las aplicaciones continuas no tienen por qué ser abiertas.

Ejemplo 1.31. De las proposiciones 1.28 y 1.29 se deduce el siguiente resultado. Supongamos que $f_1, \dots, f_m : \mathbb{R}^n \to \mathbb{R}$ son funciones continuas, y consideremos los subconjunto de $\mathbb{R}^n$

$$\begin{aligned} A_0 &= \big\{x \in \mathbb{R}^n : f_i(x) = 0,\ 1 \le i \le m\big\} \\ A_1 &= \big\{x \in \mathbb{R}^n : f_i(x) > 0,\ 1 \le i \le m\big\} \\ A_2 &= \big\{x \in \mathbb{R}^n : f_i(x) \ge 0,\ 1 \le i \le m\big\} \\ A_3 &= \big\{x \in \mathbb{R}^n : f_i(x) < 0,\ 1 \le i \le m\big\} \\ A_4 &= \big\{x \in \mathbb{R}^n : f_i(x) \le 0,\ 1 \le i \le m\big\}. \end{aligned}$$

Es inmediato comprobar que

$$\begin{aligned} A_0 &= \bigcap_{i=1}^m f_i^{-1}(\{0\}), \quad A_1 = \bigcap_{i=1}^m f_i^{-1}\big((0,\infty)\big), \quad A_2 = \bigcap_{i=1}^m f_i^{-1}\big([0,\infty)\big), \\ A_3 &= \bigcap_{i=1}^m f_i^{-1}\big((-\infty,0)\big), \quad A_4 = \bigcap_{i=1}^m f_i^{-1}\big((-\infty,0]\big). \end{aligned}$$

De aquí se deduce que los conjuntos A_0, A_2 y A_4 son cerrados, y A_1 y A_3 son abiertos. De hecho, este resultado proporciona uno de los métodos más utilizados para probar que un subconjunto de $\mathbb{R}^n$ es abierto o cerrado.

Ejercicio. Utilizar el ejemplo anterior para probar que los hiperparalelepípedos $(a_1, b_1) \times \cdots \times (a_m, b_m)$ y $[a_1, b_1] \times \cdots \times [a_m, b_m]$ (donde se admite que a_i ó b_i puedan ser iguales a $-\infty$ ó $+\infty$ para algún i) son abierto y cerrado, respectivamente, en $\mathbb{R}^m$. Deducir de esto que el producto cartesiano de m abiertos o cerrados de $\mathbb{R}$ es abierto o cerrado, respectivamente, en $\mathbb{R}^m$.

El que las imágenes inversas de conjuntos abiertos (cerrados) sean abiertos (cerrados) caracteriza a las funciones continuas:

Proposición 1.32. *Sea $f : M \to M'$ con $\operatorname{dom}(f) = M$. Si $f^{-1}(V)$ es abierto para todo abierto $V \subset M'$ entonces f es continua en M.*

Demostración. Dado $a \in M$ y $\epsilon > 0$, el conjunto $V = B_\epsilon\big(f(a)\big)$ es abierto en M'. Por tanto, $f^{-1}(V)$ es un abierto de M que contiene a $a \in M$, lo cual implica que existe $\delta > 0$ tal que $B_\delta(a) \subset f^{-1}(V)$, de donde $f\left(B_\delta(a)\right) \subset V = B_\epsilon\big(f(a)\big)$. Esto equivale a decir que $\lim_{x \to a} f(x) = f(a)$, y por tanto f es continua en a. *Q.E.D.*

Nota. El resultado anterior vale cambiando la palabra "abierto"por la palabra "cerrado". En efecto, basta aplicar que $f^{-1}(A^c) = M - f^{-1}(A)$.

1.4. Subespacios y topología relativa

Cuando el dominio de f no es todo M, se pueden probar resultados análogos a los del final de la sección anterior. Sin embargo, para ello hay que estudiar con más detalle el concepto de subespacio de un espacio métrico introducido anteriormente. Recordemos que si (M, d) es un espacio métrico y $S \subset M$, automáticamente (S, d) es un espacio métrico, en que la distancia entre dos puntos es la misma que en M. Se dice que (S, d) es un *subespacio métrico* de S.

Nota. No confundir este concepto con el de subespacio vectorial. M no tiene por qué ser un espacio vectorial, y en cualquier caso al subconjunto S no se le exige absolutamente nada.

Queremos ver qué relación existe entre las topologías de (S, d) y (M, d). Veamos para ello, en primer lugar, como son las bolas abiertas de centro $a \in S$ y radio r en el espacio métrico (S, d), que denotaremos por $B^S_r(a)$. Por definición,

$$B^S_r(a) = \big\{x \in S : d(x, a) < r\big\} = B_r(a) \cap S.$$

Por tanto, las bolas abiertas en (S, d) son simplemente la intersección de S con bolas abiertas de M. Nótese, en particular, que $B^S_r(a)$ no tiene por qué ser una bola abierta, ni siquiera un subconjunto abierto, de M. (Por ejemplo, considérese el caso $M = \mathbb{R}^2$, $S = \overline{B}_1(0)$ y a un punto de la circunferencia $S_1(0)$.)

Definición 1.33. Si $a \in S$, un *entorno relativo de a en S* es cualquier bola abierta $B_r^S(a)$ centrada en a. Un subconjunto $A \subset S$ es *abierto respecto de S* (o abierto en S) si A es un conjunto abierto del espacio métrico (S, d).

Como puede verse por el ejemplo anterior, un abierto respecto de S no tiene por qué ser abierto en M. Teniendo en cuenta que un subconjunto de un espacio métrico es abierto si y sólo si contiene un entorno de cada uno de sus puntos, es evidente que $A \in S$ es abierto respecto de $S \iff \forall a \in A$ $\exists r > 0$ tal que $B_r(a) \cap S \subset A$. La topología de (S, d), es decir la familia de conjuntos

$$\mathcal{T}_S = \big\{A \in S : A \text{ abierto respecto de } S\big\}$$

se denomina *topología relativa de S* (respecto de M).

Veamos a continuación qué relación hay entre los abiertos de S y los de M:

Proposición 1.34. *Un subconjunto $A \subset S$ es abierto respecto de S si y sólo si $\exists V \subset M$ abierto tal que $A = V \cap S$.*

Demostración.

$\Longrightarrow$) Sea A abierto en S. Entonces $\forall a \in A$ $\exists r(a) > 0$ tal que $B_{r(a)}(a) \cap S \subset A$. Entonces $V = \bigcup_{a \in A} B_{r(a)}(a)$ es un abierto en M (unión de abiertos), y

$$V \cap S = \Big(\bigcup_{a \in A} B_{r(a)}(a)\Big) \cap S = \bigcup_{a \in A} \big[B_{r(a)}(a) \cap S\big] \subset A.$$

Por otra parte, si $a \in A$ entonces $a \in S \cap B_{r(a)}(a)$, de donde $a \in \bigcup_{a \in A} \big[B_{r(a)}(a) \cap S\big] = V \cap S$.

$\Longleftarrow$) Sea $A = V \cap S$, con $V \subset M$ abierto. Entonces $A \subset S$, y dado $a \in A$ se tiene $a \in V$, por lo que $\exists r > 0$ tal que $B_r(a) \subset V$, de donde $B_r^S(a) = B_r(a) \cap S \subset V \cap S = A$. *Q.E.D.*

Corolario 1.35. *$A \subset S$ es cerrado en S—es decir, cerrado en el espacio métrico (S, d)—si y sólo si $A = C \cap S$, con C cerrado en M.*

Demostración. A cerrado en $S \iff S - A$ abierto en $S \iff \exists V \subset M$ abierto tal que $S - A = V \cap S \iff A = S - (V \cap S) = S \cap (M - V)$, con $M - V$ cerrado en M. *Q.E.D.*

Nota. Por supuesto, $A \subset S$ abierto (cerrado) en S no implica que A sea abierto (cerrado) en M. Por ejemplo, tómese $M = \mathbb{R}$, $S = [-1, 1]$ y $A = (0, 1]$. Entonces $A = (0, 2) \cap S$ es abierto en S pero no en M. Sin embargo, de la proposición anterior se sigue inmediatamente el siguiente corolario:

Corolario 1.36. *Si S es abierto (cerrado) en M, entonces $A \subset S$ es abierto (cerrado) en S si y sólo si lo es en M.*

Volviendo al estudio de las funciones continuas, si $f : M \to M'$ definimos la *restricción* $f|_S : S \to M'$ mediante $f|_S(x) = f(x)$, $\forall x \in S$. Nótese que $f|_S$ se considera una función del espacio métrico (S, d) en el espacio métrico M'.

Proposición 1.37. *$f : M \to M'$ continua en $S \implies f|_S : S \to M'$ continua en S.*

Demostración. En efecto, dado $a \in S$ f es continua en a, y por tanto $\forall \epsilon > 0$ $\exists \delta > 0$ tal que $f\left(B_\delta(a)\right) \subset B_\epsilon\bigl(f(a)\bigr)$, de donde $f\left(B_\delta^S(a)\right) = f\left(B_\delta(a) \cap S\right) \subset B_\epsilon\bigl(f(a)\bigr)$. *Q.E.D.*

Comentarios:

I) La proposición anterior sigue siendo válida si reemplazamos M' por $f(S)$ en el enunciado, pues

$$f\left(B_\delta^S(a)\right) = f\left(B_\delta(a) \cap S\right) \subset B_\epsilon\bigl(f(a)\bigr) \cap f(S) = B_\epsilon^{f(S)}\bigl(f(a)\bigr).$$

II) El recíproco no es cierto. Por ejemplo, sea $M = \mathbb{R}^2$, $S = \bigl\{(x, x) : x \in \mathbb{R}\bigr\}$ y defínase

$$f(x, y) = \begin{cases} \dfrac{x\,y}{x^2 + y^2}, & (x, y) \neq (0, 0) \\ \frac{1}{2}, & (x, y) = (0, 0). \end{cases}$$

Entonces $f : M \to \mathbb{R}$ no es continua en $(0, 0) \in S$ (el límite a lo largo de una recta de pendiente λ vale $\lambda/(1+\lambda^2)$). Sin embargo, $f|_S : S \to \mathbb{R}$ es continua, ya que $f|_S(x, x) = 1/2$, $\forall x \in \mathbb{R}$. □

1.5. Conjuntos conexos

Definición 1.38. Un espacio métrico (M, d) es *conexo* si no existen dos abiertos disjuntos no vacíos $A, B \subset M$ tales que $M = A \cup B$. Un subconjunto $S \subset M$ es conexo si el espacio métrico (S, d) es conexo. Por último, S ó M es *disconexo* si y sólo si no es conexo.

Comentarios:

I) $S \subset M$ es conexo si y sólo si no existen $A, B \subset S$ abiertos respecto de S y no vacíos tales que $A \cap B = \emptyset$ y $S = A \cup B$.

II) Un espacio métrico M es conexo si y sólo si los únicos subconjuntos de M que son a la vez abiertos y cerrados son M y $\emptyset$. □

Intuitivamente, $S \subset M$ es conexo si y sólo si consta "de una sóla pieza". Por ejemplo, si $M = \mathbb{R}$ y $S = (-1, 0) \cup (1, 2]$ entonces es intuitivamente evidente que S no es conexo. Para probarlo rigurosamente, basta notar que $S = A \cup B$, con $A = (-1, 0)$ y $B = (1, 2]$. Estos subconjuntos de S son claramente no vacíos y disjuntos, y ambos son abiertos en S, ya que $A = S \cap (-1, 0)$ y $B = S \cap (1, 3)$.

Ejemplo 1.39. $\mathbb{Q} \subset \mathbb{R}$ no es conexo. En efecto, si x es un número irracional culaquiera entonces $\mathbb{Q} = \big((-\infty, x) \cap \mathbb{Q}\big) \cup \big((x, \infty) \cap \mathbb{Q}\big)$.

También resulta intuitivamente claro que un intervalo cualquiera de la recta real es conexo. (Formalmente, un *intervalo* de $\mathbb{R}$ se define como un subconjunto $I \subset \mathbb{R}$ que goza de la siguiente propiedad: si $x, y \in I$ con $x < y$, entonces $[x, y] \subset I$.) De hecho, se puede probar el siguiente resultado (cf. Fleming, Prop. 2.11):

Proposición 1.40. *Los únicos subconjuntos conexos de $\mathbb{R}$ son los intervalos.*

Demostración. En primer lugar, veamos que si $S \subset \mathbb{R}$ no es un intervalo entonces S es disconexo. En efecto, si S no es un intervalo $\exists x < z < y$ con $x, y \in S$ y $z \notin S$. Pero entonces $S = \big((-\infty, z) \cap S\big) \cup \big((z, \infty) \cap S\big)$ es una unión disjunta de dos subconjuntos no vacíos y abiertos en S.

Probemos ahora que cualquier intervalo $I \subset \mathbb{R}$ es conexo. En efecto, si I no fuera conexo entonces existirían dos subconjuntos disjuntos no vacíos $A, B \subset I$ abiertos respecto de I tales que $I = A \cup B$. Sean entonces $x_1 \in A$, $x_2 \in B$, y supongamos (sin pérdida de generalidad) que $x_1 < x_2$. Como $x_1, x_2 \in I$, se tiene $[x_1, x_2] \subset I$. Llamemos $B_1 = \big\{x \in B : x > x_1\big\}$; por ser B_1 no vacío ($x_2 \in B_1$) y acotado inferiormente, existe $y = \operatorname{ínf} B_1$, y claramente $y \in [x_1, x_2] \subset I$, por ser x_1 cota inferior de B_1 y ser $x_2 \in B_1$. Como $y \in I$, o bien $y \in A$ o bien $y \in B$. Supongamos que $y \in A$. Por ser A abierto en I, existe $\delta > 0$ tal que $(y - \delta, y + \delta) \cap I \subset A$. Como $[x_1, x_2] \subset I$ e $y \neq x_2$ (pues $y \notin B$), tomando $\delta_1 = \operatorname{mín}(\delta, x_2 - y) > 0$ se tiene $[y, y + \delta_1) \subset A$. Pero esto es imposible, ya que al ser $y = \operatorname{ínf} B_1$ en todo intervalo $(y, y + \delta_1)$ con $\delta_1 > 0$ debe haber puntos de B_1, y por tanto de B. Del mismo modo, si $y \in B$ entonces $y \neq x_1$, y $\exists \delta > 0$ tal que $(y - \delta, y + \delta) \cap I \subset B$. Tomando ahora $\delta_2 = \operatorname{mín}(\delta, y - x_1)$ se tiene $(y - \delta_2, y] \subset B_1$, luego y no es cota inferior de B_1. Por tanto y no puede pertenecer ni a A ni a B, en contradicción con la igualdad $I = A \cup B$. *Q.E.D.*

El resultado principal de esta sección afirma que la imagen de un subconjunto conexo de un espacio métrico bajo una función continua es conexo:

Proposición 1.41. *Si $f : M \to M'$ es continua en $S \subset M$ y S es conexo, entonces $f(S) \subset M'$ es conexo.*

Nota. Como se verá por la demostración que sigue, de hecho basta con exigir que $f|_S$ sea continua en S.

Demostración. Sea $g = f|_S : S \to f(S)$; entonces g es continua en S por la proposición 1.37, y $g(S) = f(S)$. Si $f(S)$ no fuera conexo, existirían $A_1, B_1 \subset f(S)$ no vacíos, disjuntos y abiertos en $f(S)$ tales que $f(S) = A_1 \cup B_1$. Si llamamos $A = g^{-1}(A_1)$ y $B = g^{-1}(B_1)$, entonces es claro que A y B son ambos no vacíos (al ser $g(S) = f(S)$), disjuntos (pues $x \in A \cap B \Longrightarrow g(x) \in A_1 \cap B_1 = \emptyset$), abiertos en S (por ser imágenes inversas de abiertos bajo una función continua), y claramente $S = A \cup B$. Por tanto S no sería conexo, en contra de la hipótesis. *Q.E.D.*

Teorema (de los valores intermedios)**.** *Si $f : M \to \mathbb{R}$ es continua en un subconjunto conexo $S \subset M$, entonces $f(S)$ es un intervalo.*

Demostración. $f(S) \subset \mathbb{R}$ ha de ser conexo por la proposición anterior, y los únicos subconjuntos conexos de $\mathbb{R}$ son los intervalos. *Q.E.D.*

En otras palabras, si $f : M \to \mathbb{R}$ es continua en un conjunto conexo $S \subset M$, $x, y \in S$ y (por ejemplo) $f(x) = a < f(y) = b$, entonces para todo $c \in [a, b]$ $\exists z \in M$ tal que $f(z) = c$.

1.5.1. Conexión por arcos

Definición 1.42. Si $p, q \in M$, un *arco en M por p y q* es una aplicación continua $\gamma : [0,1] \to M$ con $p = \gamma(0)$ y $q = \gamma(1)$. Un espacio métrico M es *conexo por arcos* (ó arco-conexo) si $\forall p, q \in M$ existe un arco en M por p y q. Un subconjunto $S \subset M$ es conexo por arcos si S, considerado como espacio métrico (subespacio de M), es conexo por arcos.

En otras palabras, $S \subset M$ es conexo por arcos si y sólo si $\exists \gamma : [0,1] \to S$ continua tal que $\gamma(0) = p$ y $\gamma(1) = q$. (Nótese que $\gamma : [0,1] \to S$ es continua $\Longleftrightarrow$ $\gamma : [0,1] \to M$ continua.)

Ejemplo 1.43. $\mathbb{R}^n$ es conexo por arcos. En efecto, si $p, q \in \mathbb{R}^n$ un arco en $\mathbb{R}^n$ por p y q es por ejemplo el segmento de extremos p y q. Formalmente, basta definir $\gamma : [0,1] \to \mathbb{R}^n$ mediante $\gamma(t) = p + t(q - p)$. (Nótese que γ es continua, por ser la suma de una función constante y una función lineal.)

Nota. De hecho, la demostración anterior prueba que *todo espacio vectorial normado es arco-conexo.*

El concepto de conexión por arcos es más fuerte que el de conexión. En efecto, se tiene el siguiente resultado:

Proposición 1.44. *Si un subconjunto S de un espacio métrico M es conexo por arcos, entonces S es conexo.*

Demostración. Supongamos, por reducción al absurdo, que S no fuera conexo. Existirían entonces dos subconjuntos no vacíos $A, B \subset S$, disjuntos y abiertos en S, tales que $S = A \cup B$. Sean entonces $p \in A$ y $q \in B$; por ser S

arco-conexo, existe $\gamma : [0,1] \to S$ continua con $\gamma(0) = p$, $\gamma(1) = q$. Entonces $[0,1] = \gamma^{-1}(A) \cup \gamma^{-1}(B)$ (por ser $S = A \cup B$), con $\gamma^{-1}(A)$ y $\gamma^{-1}(B)$ abiertos en $[0,1]$ (imágenes inversas de abiertos en S), disjuntos (por serlo A y B) y no vacíos ($0 \in \gamma^{-1}(A)$, $1 \in \gamma^{-1}(B)$). Por tanto $[0,1]$ sería disconexo, en contra de la hipótesis. *Q.E.D.*

Corolario 1.45. $\mathbb{R}^n$ (*y, en general, cualquier espacio vectorial normado*) *es conexo.*

En general, la conexión por arcos es más fuerte que la conexión; en otras palabras, hay conjuntos conexos que no son conexos por arcos. El ejemplo típico es el conjunto

$$S = \big\{(0,y) : -1 \le y \le 1\big\} \cup \left\{(x, \operatorname{sen}\frac{1}{x}) : x > 0\right\}$$

en $\mathbb{R}^2$. Sin embargo (cf. Fleming, Proposición 1.25) puede probarse que un abierto de $\mathbb{R}^n$ es conexo si y sólo si es arco-conexo.

1.6. Conjuntos compactos

Definición 1.46. Sea S un subconjunto de un espacio métrico M. Un *recubrimiento* de S es una familia $\mathcal{A} = \{A_j : j \in J\}$ de subconjuntos de M tal que $S \subset \cup_{j\in J} A_j$. Un recubrimiento $\mathcal{A}$ de S es *abierto* si todos los elementos de $\mathcal{A}$ son conjuntos abiertos, y es *finito* si $\mathcal{A}$ es una familia finita. Un *subrecubrimiento* de $\mathcal{A}$ es una subfamilia $\mathcal{A}' \subset \mathcal{A}$ que es recubrimiento de S.

Ejemplo 1.47. Sea $M = \mathbb{R}$, $S = (0,1]$. Si $A_n = (1/n, 2)$ $\forall n \in \mathbb{N}$, entonces $\mathcal{A} = \{A_n : n \in \mathbb{N}\}$ es un recubrimiento abierto de S, pero no es un recubrimiento finito. Un subrecubrimiento de $\mathcal{A}$ es $\{A_n : n \in \mathbb{N}, n \ge 2\}$. La familia $\mathcal{B} = \{B_n : n \in \mathbb{N}\}$ con $(1/n, 1]$ es otro recubrimiento de S, pero no es un subrecubrimiento de $\mathcal{A}$ (aunque $B_n \subset A_n$ $\forall n$). Por último, la familia $\mathcal{C} = \{(-1,1), (0,3]\}$ es un recubrimiento finito de S.

Definición 1.48. Un subconjunto S de un espacio métrico M es *compacto* si todo recubrimiento abierto de S contiene un subrecubrimiento finito.

Nota. Es inmediato ver que $S \subset M$ es compacto si y sólo si S, considerado como espacio métrico (subespacio de M) es compacto.

Ejemplo 1.49. Los conjuntos finitos son compactos. En efecto, si $\mathcal{A} = \{A_j : j \in J\}$ es un recubrimiento abierto de $S = \{s_1, \dots, s_m\}$ entonces todo $s_i \in S$ pertenece a algún $A_{j_i} \in \mathcal{A} \Longrightarrow \{A_{j_1}, \dots, A_{j_m}\}$ es un subrecubrimiento finito de $\mathcal{A}$.

Ejemplo 1.50. $\mathbb{R}$ no es compacto. En efecto, si $A_n = (n, n+2)$ entonces $\mathcal{A} = \{A_n : n \in \mathbb{Z}\}$ es un recubrimiento abierto de $\mathbb{R}$. Cualquier subfamilia finita $\mathcal{A}' = \{A_{n_1}, \dots, A_{n_k}\} \subset \mathcal{A}$ no puede recubrir $\mathbb{R}$, ya que si $n_1 < n_2 < \dots < n_k$ entonces $\bigcup_{i=1}^k A_{n_i} \subset (n_1, n_k + 2)$.

Ejemplo 1.51. El intervalo $(0, 1] \subset \mathbb{R}$ no es compacto. En efecto, si $A_n = (1/n, 2)$ entonces $\mathcal{A} = \{A_n : n \in \mathbb{N}\}$ es un recubrimiento abierto de $(0, 1]$ que no contiene ningún subrecubrimiento finito. En efecto, si $\mathcal{A}' = \{A_{n_1}, \dots, A_{n_k}\} \subset \mathcal{A}$ con $n_1 < \dots < n_k$ entonces se tiene $0 \notin \bigcup_{i=1}^k A_{n_i} = (1/n_k, 2)$.

Definición 1.52. Un subconjunto S de un espacio métrico M es *acotado* si $\exists R > 0$ tal que $S \subset B_R(0)$.

Si $S \subset M$ es un conjunto acotado, definimos su *diámetro* mediante

$$\operatorname{diam}(S) = \sup\big\{d(x, y) : x, y \in S\big\}.$$

Nótese que este supremo existe (es finito), ya que por ser S acotado se tiene

$$d(x, y) \leq d(x, 0) + d(y, 0) \leq 2R,$$

y por tanto $\big\{d(x, y) : x, y \in S\big\}$ está acotado superiormente por $2R$. En consecuencia, todo conjunto acotado tiene un diámetro finito.

Ejercicio. Probar que $\operatorname{diam} B_r(a) = 2r$, y que el diámetro de un hipercubo de lado L en $\mathbb{R}^n$ es $L\sqrt{n}$.

El objetivo principal de esta sección es el de obtener la siguiente caracterización de los conjuntos compactos de $\mathbb{R}^n$: *un subconjunto de $\mathbb{R}^n$ es compacto si y sólo si es cerrado y acotado.* Empezaremos probando que, en cualquier espacio métrico, los conjuntos compactos son necesariamente cerrados y acotados.

Proposición 1.53. *Si S es un subconjunto compacto de un espacio métrico M, entonces S es acotado.*

Demostración. La familia $\{B_r(0) : r > 0\}$ recubre S por lo que, aplicando la definición de compacidad, existirán $r_1 < r_2 < \dots < r_m$ tales que $S \subset \cup_{i=1}^m B_{r_i}(0) = B_{r_m}(0)$. *Q.E.D.*

Para probar que los subconjuntos compactos de un espacio métrico son cerrados, utilizaremos el siguiente lema:

Lema 1.54. *Si $x, y \in M$, existen sendos entornos U de x y V de y tales que $U \cap V = \emptyset$.*

Demostración. Basta tomar $U = B_r(x)$, $V = B_r(y)$ con $2r \leq d(x, y)$. *Q.E.D.*

El lema anterior, que parece totalmente trivial, no es cierto sin embargo en espacios topológicos arbitrarios (entendiendo por entorno de un punto un abierto que contiene a dicho punto). Por definición, se dice que un espacio topológico es *de Hausdorff* (o separado) si en dicho espacio el lema anterior es cierto. Por tanto, hemos probado que los espacios métricos son de Hausdorff. Con ayuda de esta propiedad de los espacios métricos, se demuestra el siguiente resultado:

Proposición 1.55. *Un subconjunto compacto de un espacio métrico es necesariamente cerrado.*

Demostración. Si $S \subset M$ es compacto, probaremos que S^c es abierto. En efecto, sea $x \in S^c$. Por el lema anterior, $\forall z \in S\ \exists U_z$ entorno de x y V_z entorno de z tales que $U_z \cap V_z = \emptyset$. La familia $\{V_z : z \in S\}$ es entonces un recubrimiento abierto de S; al ser este conjunto compacto, han de existir $z_1, \dots, z_m \in S$ tales que $S \subset V_{z_1} \cup \dots \cup V_{z_m}$. El conjunto $U = U_{z_1} \cap \dots \cap U_{z_m}$ es un entorno de x, por ser intersección finita de bolas abiertas centradas en x. Por otra parte,

$$U \cap S \subset U \cap \bigcup_{i=1}^{m} V_{z_i} = \bigcup_{i=1}^{m} U \cap V_{z_i} \subset \bigcup_{i=1}^{m} U_{z_i} \cap V_{z_i} = \emptyset.$$

Por tanto, dado un punto cualquiera $x \in S^c$ existe un entorno de x contenido en S^c. Esto significa que S^c es abierto, y por tanto S es cerrado. *Q.E.D.*

Con las dos proposiciones anteriores hemos establecido que los subconjuntos compactos de cualquier espacio métrico M son cerrados y acotados. El recíproco de este resultado es cierto en $\mathbb{R}^n$ (de hecho, puede probarse que no es cierto más que si M es isomorfo a $\mathbb{R}^n$):

Teorema (Heine–Borel–Lebesgue)**.** *Un subconjunto cerrado y acotado de $\mathbb{R}^n$ es compacto.*

1.6.1. Sucesiones de Cauchy. Completitud

En la demostración del teorema de Heine–Borel–Lebesgue es fundamental el concepto de sucesión de Cauchy:

Definición 1.56. Una sucesión $(x_m)_{m=1}^{\infty}$ en un espacio métrico (M, d) es *de Cauchy* (ó fundamental) si $\forall \epsilon > 0\ \exists N(\epsilon) \in \mathbb{N}$ tal que $d(x_l, x_m) < \epsilon$ si $l, m \geq N(\epsilon)$.

¿Qué relación hay entre el concepto de “sucesión de Cauchyτ el de “sucesión convergente¿ (Una sucesión es *convergente* si su límite existe, *divergente* en caso contrario.)

Proposición 1.57. *Toda sucesión convergente es de Cauchy.*

Demostración. Si $\lim_{m\to\infty} x_m = x$, dado $\epsilon > 0$ $\exists N \in \mathbb{N}$ tal que $d(x_m, x) < \frac{\epsilon}{2}$ si $m \geq N$. Entonces si $l, m \geq N$ se tiene $d(x_l, x_m) \leq d(x_l, x) + d(x_m, x) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$. *Q.E.D.*

Sin embargo, el recíproco de la proposición anterior no es cierto en espacios métricos arbitrarios:

Ejemplo 1.58. Sea $M = (0, 1]$ con la distancia euclídea. La sucesión $(1/m)_{m=1}^{\infty}$ es claramente de Cauchy, pues dado $\epsilon > 0$ si $l, m \geq N$ se tiene

$$d(x_l, x_m) = \left|\frac{1}{l} - \frac{1}{m}\right| \leq \frac{2}{N} < \epsilon$$

sin más que tomar $N > 2/\epsilon$. Sin embargo, es fácil ver que dicha sucesión no es convergente en M. Intuitivamente, esto es debido a que la sucesión converge a 0, y $0 \notin M$. Más rigurosamente, hay que probar que $\forall b \in (0, 1]$ se tiene $\lim_{m\to\infty} 1/m \neq b$. En efecto, como $b > 0$ podemos tomar $p \in \mathbb{N}$ tal que $b - 1/p > b/2$. Entonces $\forall m \geq p$ se tendrá

$$\left|\frac{1}{m} - b\right| = b - \frac{1}{m} \geq b - \frac{1}{p} > \frac{b}{2},$$

y la definición de límite falla tomando $\epsilon = b/2$.

El ejemplo anterior sugiere que la razón esencial por la cual hay sucesiones de Cauchy que no son convergentes es porque al espacio métrico "le faltan puntos". Esto motiva la siguiente terminología:

Definición 1.59. Un espacio métrico es *completo* si en él toda sucesión de Cauchy es convergente.

Una de las propiedades básicas de los números reales es que $\mathbb{R}$ *es un espacio métrico completo.* A partir de esto se demuestra sin dificultad que $\mathbb{R}^n$ es también completo:

Proposición 1.60. $\mathbb{R}^n$ *es completo.*

Demostración. Sea $\left(x_m\right)_{m=1}^{\infty}$ una sucesión de Cauchy en $\mathbb{R}^n$, y denotemos $x_m = \left(x_m^1, \ldots, x_m^n\right)$. Como $\left|x_l^i - x_m^i\right| \leq |x_l - x_m|$, la sucesión de números reales $\left(x_m^i\right)_{m=1}^{\infty}$ es de Cauchy, y por tanto convergente (pues $\mathbb{R}$ es completo). Al ser las sucesiones $\left(x_m^i\right)_{m=1}^{\infty}$ convergentes para $i = 1, 2, \ldots n$, por las propiedades de las sucesiones $\left(x_m\right)_{m=1}^{\infty}$ es convergente. *Q.E.D.*

Proposición 1.61. *Si M es un espacio métrico completo, entonces $S \subset M$ es completo si y sólo si es cerrado en M.*

Demostración.

$\Longrightarrow$) Sea S completo, y sea $x \in \overline{S}$. Entonces $\forall m \in \mathbb{N}$ $\exists x_m \in B_{\frac{1}{m}}(x) \cap S$. La sucesión $\left(x_m\right)_{m=1}^{\infty} \subset S$ es de Cauchy en S, pues al ser convergente a x en M es de Cauchy en M, y la distancia en S es la misma que en M. Por ser S completo, $\left(x_m\right)_{m=1}^{\infty}$ converge a un punto $z \in S$. Como $\lim_{m\to\infty} x_m = z$ también en M, por la unicidad del límite de una sucesión $z = x$, y por tanto $x \in S$. Luego $\overline{S} \subset S$, y S es cerrado.

$\Longleftarrow$) Supongamos ahora que S es cerrado en M, y sea $\left(x_m\right)_{m=1}^{\infty} \subset S$ una sucesión de Cauchy. Como M es completo, existe $\lim_{m\to\infty} x_m = x \in M$. Por definición de límite de una sucesión, $x \in \overline{S} = S$, por ser S cerrado. Luego toda sucesión de Cauchy en S es convergente en S, y S es completo. *Q.E.D.*

1.6.2. Teorema de Heine–Borel–Lebesgue

Proposición 1.62. *Si M es un espacio métrico compacto, un subconjunto de M es compacto si y sólo si es cerrado.*

Demostración. En primer lugar, todo subconjunto compacto de un espacio métrico es cerrado, por la proposición 1.55. Sea, por otra parte, $C \subset M$ cerrado, y sea $\mathcal{A} = \{A_j : j \in J\}$ un recubrimiento abierto de C. Como C es cerrado, C^c es abierto, y $\mathcal{A}' = \mathcal{A} \cup \{C^c\}$ es un recubrimiento abierto de M. Al ser M compacto, $\mathcal{A}'$ contiene un subrecubrimiento finito $\mathcal{A}''$ de M. Entonces $\mathcal{A}'' - \{C^c\}$ es un recubrimiento finito de M contenido en $\mathcal{A}$. *Q.E.D.*

Por la proposición anterior, para probar el teorema de Heine–Borel–Lebesgue basta demostrar que el hipercubo $[-R, R]^n$ es compacto en $\mathbb{R}^n$. En efecto, si $K \subset \mathbb{R}^n$ es cerrado y acotado, entonces $\exists R > 0$ tal que $K \subset B_R(0)$, y como $B_R(0) \subset [-R, R]^n$ (pruébese ésto analíticamente) entonces K es un subconjunto cerrado de un compacto, y por tanto es compacto. La demostración de la compacidad de $[-R, R]^n$ se basa en el siguiente lema, debido a Cantor:

Lema (Cantor)**.** *Sea A_m cerrado en $\mathbb{R}^n$ $\forall m \in \mathbb{N}$, y supongamos que $A_1 \supset A_2 \supset \cdots \supset A_m \supset A_{m+1} \supset \ldots$. Si $\lim_{m\to\infty} \operatorname{diam}(A_m) = 0$, entonces $\cap_{m=1}^{\infty} A_m$ consta de un sólo elemento.*

Demostración. Tomemos un punto x_m en cada cerrado A_m, y construyamos la sucesión $\left(x_m\right)_{m=1}^{\infty}$. Dicha sucesión es de Cauchy: en efecto, si $l, m \geq N$ entonces $A_l, A_m \subset A_N$, y por tanto $|x_m - x_l| \leq \operatorname{diam}(A_N) < \epsilon$ tomando N suficientemente grande. Al ser $\mathbb{R}^n$ completo, $\exists x = \lim_{m\to\infty} x_m$. Veamos, en primer lugar, que $x \in \cap_{m=1}^{\infty} A_m$. En efecto, para un m fijo cualquier bola $B_r(x)$ contiene un punto de A_m, ya que por definición de límite $\exists N \in \mathbb{N}$ tal que $x_k \in B_r(x)$ para $k \geq N$, y basta entonces tomar $k \geq \max(N, m)$ para concluir que $x_k \in A_m \cap B_r(x)$. Por tanto, $x \in \overline{A}_m$ para todo m, y como $\overline{A}_m = A_m$ por ser A_m cerrado, de aquí se sigue que $x \in \cap_{m=1}^{\infty} A_m$.

Finalmente, si $y \in \cap_{m=1}^{\infty} A_m$ entonces $0 \leq d(x,y) \leq \operatorname{diam}(A_m)\ \forall m \in \mathbb{N}$; como $\lim_{m\to\infty} \operatorname{diam}(A_m) = 0$, $d(x,y) = 0$ y, por tanto, ha de ser $y = x$. *Q.E.D.*

Demostremos, finalmente, que $[-R,R]^n$ es compacto en $\mathbb{R}^n$, con lo cual habremos completado la demostración del teorema de Heine–Borel–Lebesgue:

Teorema 1.63. *$[-R,R]^n$ es compacto en $\mathbb{R}^n$.*

Demostración. En efecto, sea $\mathcal{A}$ un recubrimiento abierto de $C_0 = [-R,R]^n$, y supongamos que *no* se puede extraer de $\mathcal{A}$ ningún subrecubrimiento finito de C_0. Dividiendo cada un de los lados del hipercubo cerrado C_0 en dos partes, obtenemos 2^n hipercubos cerrados más pequeños de lado R y diámetro $R\sqrt{n} = \frac{1}{2}\operatorname{diam}(C_0)$. Si todos estos hipercubos se pudieran recubrir por una subfamilia finita de $\mathcal{A}$, la unión de estas subfamilias sería un subrecubrimiento finito de C_0. Por tanto, al menos uno de estos hipercubos no puede ser recubierto por un número finito de elementos de $\mathcal{A}$. Sea $C_1 \subset C_0$ uno de dichos hipercubos; nótese que $\operatorname{diam}(C_1) = \frac{1}{2}\operatorname{diam}(C_0)$. Dividiendo C_1 en 2^n hipercubos cerrados iguales, por el mismo razonamiento que antes existe otro hipercubo cerrado $C_2 \subset C_1$ con $\operatorname{diam}(C_2) = \frac{1}{2}\operatorname{diam}(C_1) = 2^{-2}\operatorname{diam}(C_0)$ que no puede ser recubierto por una subfamilia finita de $\mathcal{A}$, y así sucesivamente. De esta forma construimos una serie de hipercubos cerrados tales que $C_0 \supset C_1 \supset \cdots \supset C_m \supset C_{m+1} \supset \dots$, con $\operatorname{diam}(C_m) = 2^{-m}\operatorname{diam}(C_0)@ >> m \to \infty > 0$ y con la propiedad de que ningún C_m puede ser recubierto por una subfamilia finita de $\mathcal{A}$. Por el lema de Cantor, $\cap_{m=0}^{\infty} C_m = \{x\}$. Como $x \in C_0$, $\exists A \in \mathcal{A}$ tal que $x \in A$, y al ser A abierto $\exists r > 0$ tal que $B_r(x) \subset A$. Por ser $\lim_{m\to\infty}\operatorname{diam}(C_m) = 0$, $\exists k \in \mathbb{N}$ tal que $\operatorname{diam}(C_k) < r$, y como $x \in C_k$ se tiene $C_k \subset B_r(x) \subset A$. Por tanto $\{A\} \subset \mathcal{A}$ es un subrecubrimiento finito de C_k, en contra de su construcción. *Q.E.D.*

Proposición 1.64. *Si S es un subconjunto infinito de un espacio métrico compacto M, entonces S tiene al menos un punto de acumulación.*

Demostración. Si S no tiene puntos de acumulación, entonces S es cerrado y todos sus puntos son aislados: $\forall x \in S$, $\exists U_x$ entorno de x tal que $U_x \cap S = \{x\}$. Al ser S cerrado y M compacto, S es compacto por la proposición 1.62. Consideremos entonces el recubrimiento abierto $\mathcal{A} = \big\{U_x : x \in S\big\}$ de S. Si $\mathcal{A}_1 \subset \mathcal{A}$ es una subfamilia finita de S, entonces $\mathcal{A}_1 = \big\{U_x : x \in S_1\big\}$, siendo S_1 un subconjunto finito de S. Nótese que $x \neq y \Rightarrow U_x \neq U_y$, pues $x \in U_x - U_y$, $y \in U_y - U_x$. Como S es un conjunto infinito y S_1 es finito, $S_1 \neq S$, y por tanto $\mathcal{A}_1$ no puede recubrir S, ya que sólo recubre S_1. Por tanto, $\mathcal{A}$ es un recubrimiento abierto de S que no contiene ningún subrecubrimiento finito, en contra de la compacidad de S. *Q.E.D.*

Corolario (Bolzano–Weierstrass). *Todo subconjunto infinito y acotado de $\mathbb{R}^n$ posee al menos un punto de acumulación.*

Demostración. En efecto, sea $S \subset \mathbb{R}^n$ acotado e infinito. Entonces $\overline{S}$ es también acotado ($\overline{S} \subset \overline{B}_R(0) \subset B_\rho(0)$ si $\rho > R$), y al ser además cerrado es compacto en $\mathbb{R}^n$ (teorema de Heine–Borel–Lebesgue). Aplicando la proposición anterior con $K = \overline{S}$ obtenemos que $S \subset \overline{S}$ posee algún punto de acumulación. *Q.E.D.*

Para finalizar este apartado, señalaremos sin demostración algunas propiedades que son equivalentes a la compacidad en un espacio métrico arbitrario:

Teorema 1.65. *Sea M un espacio métrico, y sea $K \subset M$. Las siguientes afirmaciones son equivalentes:*

i) *K es compacto*

ii) *Toda sucesión $\left(x_m\right)_{m=1}^{\infty} \subset K$ posee una subsucesión $\left(x_{m_i}\right)_{i=1}^{\infty}$ convergente en K*

iii) *Todo subconjunto infinito de K posee un punto de acumulación en K*

1.7. Continuidad y compacidad

Al igual que la conexión, la compacidad es una propiedad que se preserva al tomar imágenes bajo funciones continuas:

Teorema 1.66. *Si $f : M \to M'$ es continua en $K \subset M$ y K es compacto, entonces $f(K) \subset M'$ es compacto.*

Demostración. Sea $g = f|_K : K \to f(K)$; entonces g es continua en K, y $g(K) = f(K)$. Si $\mathcal{B} = \left\{B_j : j \in J\right\}$ es un recubrimiento abierto de $f(K)$, entonces es claro que $\mathcal{A} = \left\{g^{-1}(B_j) : j \in J\right\}$ es un recubrimiento abierto de K; nótese que $g^{-1}(B_j)$ es abierto en K, por ser la imagen inversa del abierto B_j de $f(K)$ bajo la función continua g. Por ser K compacto, este recubrimiento de K contiene un subrecubrimiento finito $\left\{g^{-1}(B_{j_k}) : j_k \in J, 1 \le k \le m\right\}$, de donde se sigue que $f(K) = g(K) = B_{j_1} \cup \cdots \cup B_{j_m}$, y por tanto $\{B_{j_k} : j_k \in J, 1 \le k \le m\}$ es un subrecubrimiento finito de $f(K)$. *Q.E.D.*

Nota. Como se ve de la demostración, el teorema es válido bajo la hipótesis más débil de que $f|_K : K \to M'$ sea continua.

En el caso de funciones escalares, el teorema anterior tiene el siguiente corolario:

Corolario 1.67. *Si $f : M \to \mathbb{R}$ es continua en un compacto $K \subset M$, entonces $\exists x_0, y_0 \in K$ tales que $f(x_0) \le f(x) \le f(y_0)$, $\forall x \in K$.*

Demostración. En efecto, por el teorema anterior $f(K) \subset \mathbb{R}$ es compacto. Por ser $f(K)$ acotado, $\exists a = \inf f(K)$, $\exists b = \sup f(K)$, y $a \leq f(x) \leq b$, $\forall x \in K$. Por definición de ínfimo y supremo, $a, b \in \overline{f(K)} = f(K)$, donde la última igualdad se debe a que $f(K)$ es cerrado por ser compacto. Por tanto, $\exists x_0, y_0 \in K$ tales que $a = f(x_0)$, $b = f(y_0)$, y el corolario está demostrado. *Q.E.D.*

En otras palabras, *una función escalar continua sobre un conjunto compacto alcanza sus valores máximo y mínimo en dicho conjunto.* En particular, una función escalar continua en un conjunto compacto está acotada en dicho conjunto.

1.7.1. Homeomorfismos e isometrías

Definición 1.68. Sean M_1 y M_2 espacios métricos. Un *homeomorfismo* de M_1 en M_2 es una aplicación biyectiva $f : M_1 \to M_2$ tal que f y $f^{-1} : M_2 \to M_1$ son continuas. Dos espacios métricos son *homeomorfos* si existe un homeomorfismo $f : M_1 \to M_2$.

Nótese que, si $f : M_1 \to M_2$ es un homeomorfismo, entonces f transforma abiertos de M_1 en abiertos de M_2 (pues $f(A) = \left(f^{-1}\right)^{-1}(A)$), y f^{-1} transforma abiertos de M_2 en abiertos de M_1. De aquí se sigue que cualquier propiedad topológica de M_1 (es decir, cualquier propiedad que se pueda expresar en términos de la topología de M_1 exclusivamente) se transforma en una propiedad topológica equivalente de M_2 al aplicar f (y viceversa al aplicar f^{-1}). Por tanto, *desde el punto de vista topológico dos espacios homeomorfos son equivalentes.* (De hecho, es inmediato probar que la relación "M_1 homeomorfo a M_2."es una relación de equivalencia.) Por ejemplo, si A es abierto, cerrado o compacto en M_1 entonces $f(A)$ es abierto, cerrado o compacto en M_2, etc. Sin embargo, un homeomorfismo $f : M_1 \to M_2$ no tiene por qué respetar propiedades *métricas*: por ejemplo, si A es acotado en M_1 $f(A)$ no necesariamente es acotado en M_2.

Definición 1.69. Una *isometría* entre dos espacios métricos (M_1, d_1) y (M_2, d_2) es una aplicación biyectiva $f : M_1 \to M_2$ que respeta la distancia: $\forall x, y \in M_1$ se tiene $d_1(x_1, x_2) = d_2\big(f(x_1), f(x_2)\big)$. Dos espacios métricos son *isomorfos* si existe una isometría entre ellos.

Nótese que una isometría es automáticamente continua en ambos sentidos; por tanto, toda isometría es un homeomorfismo. De nuevo, la relación "M_1 isomorfo a M_2."es una relación de equivalencia. Dos espacios isomorfos son totalmente equivalentes, tanto desde el punto de vista de las propiedades topológicas como desde el punto de vista métrico. Por ejemplo, $f\big(B_r(x)\big) = B_r\big(f(x)\big)$, si A es acotado en M_1 $f(A)$ es acotado en M_2, $\operatorname{diam}(A) = \operatorname{diam}\big(f(A)\big)$, etc.

Ejemplo 1.70. $M_1 = (-\frac{\pi}{2}, \frac{\pi}{2})$ y $M_2 = \mathbb{R}$ (cada uno de ellos con la distancia euclídea) son homeomorfos pero no isomorfos. En efecto, la aplicación $\operatorname{tg} : (-\frac{\pi}{2}, \frac{\pi}{2}) \to \mathbb{R}$ es claramente continua y biyectiva (tg es derivable en $(-\frac{\pi}{2}, \frac{\pi}{2})$, y su derivada $\operatorname{tg}'(x) = \sec^2 x$ no se anula en ningún punto). Su inversa, $\operatorname{arc\,tg} : \mathbb{R} \to (-\frac{\pi}{2}, \frac{\pi}{2})$ es también continua, y de hecho es derivable, por ser la función inversa de una función con derivada no nula en su dominio. En consecuencia, $(-\frac{\pi}{2}, \frac{\pi}{2})$ y $\mathbb{R}$ son homeomorfos, y tienen por tanto las mismas propiedades topológicas. Sin embargo, es claro que no hay ninguna isometría entre ambos conjuntos, porque el primero es acotado y el segundo no.

Ejemplo 1.71. $\mathbb{R}$ no es homeomorfo a $(0, 1]$. En efecto, $(0, 1] - \{1\}$ es conexo, mientras que $\mathbb{R} - \{x\}$ no es conexo para ningún $x \in \mathbb{R}$.

Nótese que para que una aplicación $f : M_1 \to M_2$ sea un homeomorfismo no sólo basta con que sea biyectiva y continua, sino que además es necesario que f^{-1} sea también continua, lo que no es en general consecuencia de las condiciones anteriores. Sin embargo, si M_1 es *compacto* entonces f^{-1} es automáticamente continua si f es biyectiva y continua:

Proposición 1.72. *Si M_1 es compacto y $f : M_1 \to M_2$ es biyectiva y continua, entonces f es un homeomorfismo.*

Demostración. Basta probar que $g = f^{-1} : M_2 \to M_1$ es continua, es decir que $g^{-1}(A)$ es cerrado para todo A cerrado en M_1. Pero, al ser M_1 compacto y A cerrado en M_1, A es compacto, y de la continuidad de f se sigue que $g^{-1}(A) = f(A)$ es compacto en M_2, y por tanto cerrado. *Q.E.D.*

1.7.2. Continuidad uniforme

Definición 1.73. Una función $f : M \to M'$ es *uniformemente continua* en $A \subset \operatorname{dom}(f)$ si $\forall \epsilon > 0\ \exists \delta(\epsilon)$ tal que $\forall x, y \in A$ con $d(x, y) < \delta(\epsilon)$ se tiene $d\big(f(x), f(y)\big) < \epsilon$.

Es inmediato probar que si f es uniformemente continua en un conjunto A automáticamente $f|_A$ es continua en dicho conjunto. En efecto, si fijamos $a \in A$ y damos $\epsilon > 0$, entonces $x \in A$ y $d(x, a) < \delta(\epsilon) \Longrightarrow d\big(f(x), f(a)\big) < \epsilon$, y por lo tanto $\lim_{x \to a} f(x) = f(a)$. Sin embargo, es claro que la continuidad uniforme es mucho más fuerte que la continuidad, y de hecho es fácil dar ejemplos de funciones continuas que no son uniformemente continuas:

Ejemplo 1.74. La función $f : \mathbb{R} \to \mathbb{R}$ definida por $f(x) = x^2$, $\forall x \in \mathbb{R}$, es continua pero no uniformemente continua en $\mathbb{R}$. En efecto, hay que probar que $\exists \epsilon > 0$ tal que $\forall \delta > 0\ \exists x, y \in \mathbb{R}$ tales que $|x - y| < \delta$ pero $|f(x) - f(y)| = \left|x^2 - y^2\right| \geq \epsilon$. Dado $\delta > 0$ arbitrario, si tomamos (por ejemplo) $x = R > 0, y = R + \frac{\delta}{2}$ se tiene $|x - y| = \frac{\delta}{2} < \delta$ pero $\left|x^2 - y^2\right| = |x + y|\,|x - y| = (2R + \frac{\delta}{2})\,\frac{\delta}{2} > R\delta \geq 1$ tomando $R \geq \frac{1}{\delta}$.

Sin embargo, una función continua sobre un conjunto compacto es automáticamente uniformemente continua en dicho conjunto:

Teorema 1.75. *Si $f : M \to M'$ es continua en un compacto $K \subset M$, entonces f es uniformemente continua en K.*

Demostración. Sea $\epsilon > 0$. Por ser f continua en K, $\forall z \in K$ $\exists \delta(z)$ tal que $p \in B_{\delta(z)}(z) \cap K \implies d\big(f(p), f(z)\big) < \epsilon/2$. La familia $\big\{B_{\delta(z)/2}(z) : z \in K\big\}$ es un recubrimiento abierto de K; al ser K compacto, $\exists z_1, \dots, z_m \in K$ tales que $K \subset B_{\delta(z_1)/2}(z_1) \cup \cdots \cup B_{\delta(z_m)/2}(z_m)$. Sea entonces $\delta = \frac{1}{2} \min\big\{\delta(z_1), \dots, \delta(z_m)\big\}$, y sean $x, y \in K$ con $d(x, y) < \delta$. Como $x \in B_{\delta(z_i)/2}(z_i)$ para algún $i \in \{1, \dots, m\}$ y $d(y, z_i) \leq d(x, y) + d(x, z_i) < \delta + \frac{\delta(z_i)}{2} \leq \frac{\delta(z_i)}{2} + \frac{\delta(z_i)}{2} = \delta(z_i)$, entonces $x, y \in B_{\delta(z_i)}(z_i) \cap K$, de donde

$$d\big(f(x), f(y)\big) \leq d\big(f(x), f(z_i)\big) + d\big(f(z_i), f(y)\big) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Q.E.D.

1.8. Espacios de funciones continuas y convergencia uniforme

Si $\big(f_n\big)_{n=1}^{\infty}$ es una sucesión de funciones $f_n : M \to M'$ con dominio M, se dice que $\big(f_n\big)_{n=1}^{\infty}$ *converge puntualmente* a una función $f : M \to M'$ si $\lim_{n\to\infty} f_n(x) = f(x)$, $\forall x \in M$. En otras palabras, $\forall x \in M, \forall \epsilon > 0$ $\exists N(\epsilon, x)$ tal que $d\big(f_n(x), f(x)\big) < \epsilon$ si $n \geq N(\epsilon, x)$. Aunque este concepto de convergencia es muy sencillo, adolece de serios defectos. Por ejemplo, puede ocurrir que una sucesión de funciones continuas converja puntualmente a una función no continua:

Ejemplo 1.76. Sea $M = [0, 1]$, $M' = \mathbb{R}$, $f_n(x) = x^n$, $\forall x \in [0, 1]$. Claramente, las funciones f_n son continuas (polinómicas), y

$$\lim_{n\to\infty} f_n(x) = \begin{cases} 0, & x \in [0, 1) \\ 1, & x = 1. \end{cases}$$

Por tanto, la sucesión $\big(f_n\big)_{n=1}^{\infty}$ converge puntualmente a una función discontinua en $x = 1$.

Para evitar este tipo de anomalías, se introduce el concepto de *convergencia uniforme*. La idea es exigir que el entero $N(\epsilon, x)$ en la definición de convergencia puntual se pueda tomar *independiente de x*. En otras palabras, se exige que $d\big(f_n(x), f(x)\big) < \epsilon$ simultáneamente (= uniformemente) para todo $x \in M$, si n es suficientemente grande:

Definición 1.77. Si $f_n : M \to M'$ $\forall n \in \mathbb{N}$, se dice que la sucesión de funciones $\big(f_n\big)_{n=1}^\infty$ *converge uniformemente* a la función $f : M \to M'$ si $\forall \epsilon > 0$ $\exists N(\epsilon) \in \mathbb{N}$ tal que $\forall x \in M$ y $\forall n \geq N(\epsilon)$ se cumple $d\big(f_n(x), f(x)\big) < \epsilon$.

Nótese que *la convergencia uniforme implica la convergencia puntual*: si $\big(f_n\big)_{n=1}^\infty$ converge uniformemente a f, entonces $\big(f_n\big)_{n=1}^\infty$ converge puntualmente a la misma función. Gráficamente (si $M' = \mathbb{R}$), si $\big(f_n\big)_{n=1}^\infty$ converge uniformemente a f entonces $\forall \epsilon > 0$ la gráfica de f_n está comprendida en la banda limitada por las gráficas de las funciones $f - \epsilon$ y $f + \epsilon$ para n suficientemente grande.

Si M' es un espacio vectorial normado, el concepto de convergencia uniforme se puede formular de manera más concisa introduciendo una norma en el espacio $\mathcal{B}(M, M')$ de todas las funciones acotadas $f : M \to M'$. Por definición, $f : M \to M'$ es *acotada* si $f(M)$ es acotado, i.e. si $\exists K > 0$ tal que $\|f(x)\| < K$ $\forall x \in M$. Si f y g son acotadas, entonces $\|\lambda f(x) + \mu g(x)\| \leq |\lambda|\,\|f(x)\| + |\mu|\,\|g(x)\|$, y por tanto $\lambda f + \mu g$ es también acotada. Por tanto, $\mathcal{B}(M, M')$ es un espacio vectorial. Introducimos entonces la *norma uniforme* (o norma del supremo) en $\mathcal{B}(M, M')$, definiendo

$$\|f\| = \sup\{\|f(x)\| : x \in M\}\,;$$

nótese que el supremo existe (es $< \infty$) por ser f acotada. Es inmediato probar que la norma uniforme cumple todas las propiedades de una norma. Por tanto, $\mathcal{B}(M, M')$ es un espacio vectorial normado. En particular, la norma uniforme induce una distancia en $\mathcal{B}(M, M')$ mediante

$$d(f, g) = \|f - g\| = \sup\{\|f(x) - g(x)\| : x \in M\}\,.$$

Con esta distancia (que llamaremos uniforme) es inmediato comprobar que una sucesión de funciones $f_n : M \to M'$ converge uniformemente a una función f si y sólo si $\lim_{n\to\infty} d(f_n, f) = \lim_{n\to\infty} \|f_n - f\| = 0$. En particular, si $f_n \in \mathcal{B}(M, M')$ $\forall n \in \mathbb{N}$ entonces la convergencia uniforme de $\big(f_n\big)_{n=1}^\infty$ a f es equivalente a la convergencia en $\mathcal{B}(M, M')$ con la norma uniforme de la sucesión $\big(f_n\big)_{n=1}^\infty \subset \mathcal{B}(M, M')$. Nótese que el límite uniforme de una sucesión de funciones acotadas es una función acotada.

Ejemplo 1.78. La sucesión del ejemplo 1.76 no converge uniformemente a ninguna función $f \in \mathcal{B}([0,1], \mathbb{R})$. En efecto, como la convergencia uniforme implica la convergencia puntual, si $\big(f_n\big)_{n=1}^\infty$ convergiera uniformemente su límite habría de ser la función f tal que $f(1) = 1$ y $f(x) = 0$ $\forall x \in [0, 1)$. Sin embargo, es evidente que $\big(f_n\big)_{n=1}^\infty$ no converge uniformemente a esta función, ya que

$$\|f_n - f\| = \sup\{x^n : x \in [0,1)\} = 1, \quad \forall n \in \mathbb{N}.$$

A diferencia de la convergencia puntual, la convergencia uniforme tiene la siguiente propiedad fundamental: *el límite uniforme de funciones continuas es una función continua*:

Teorema 1.79. *Si $\big(f_n\big)_{n=1}^{\infty}$ es una sucesión de funciones continuas $M \to M'$ uniformemente convergente a $f : M \to M'$, entonces f es continua.*

Demostración. La demostración es un típico "argumento $\epsilon/3$". En efecto, dado $a \in M$ y $\epsilon > 0$, se tiene

$$d\big(f(x), f(a)\big) \leq d\big(f(x), f_n(x)\big) + d\big(f_n(x), f_n(a)\big) + d\big(f_n(a), f(a)\big).$$

Por la convergencia uniforme de $\big(f_n\big)_{n=1}^{\infty}$ a f, podemos escoger n tal que $\|f_n - f\| < \epsilon/3$; por tanto, para este n el primer y el tercer sumando de la desigualdad anterior son menores que $\epsilon/3$. Por otra parte, por ser f_n continua en M $\exists\delta > 0$ tal que el segundo término de dicha desigualdad también es menor que $\epsilon/3$ si $d(x, a) < \delta$. En consecuencia, si $d(x, a) < \delta$ se tiene $d\big(f(x), f(a)\big) < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon$, y f es continua en a. *Q.E.D.*

Sea $\mathcal{C}(M, M')$ el espacio de las funciones continuas y acotadas de M en M' (siendo M' un espacio vectorial normado). Entonces $\mathcal{C}(M, M')$ es un espacio vectorial (subespacio de $\mathcal{B}(M, M')$), en el que se puede considerar la norma uniforme. Se tiene entonces el siguiente resultado:

Proposición 1.80. *Si M' es completo, $\mathcal{C}(M, M')$ es completo con la norma uniforme.*

Demostración. Sea $\big(f_n\big)_{n=1}^{\infty}$ una sucesión de funciones continuas acotadas de M en M' que es de Cauchy respecto de la distancia uniforme. Como $\|f_l(x) - f_m(x)\| \leq \|f_l - f_m\|$ $\forall x \in M$, para cada $x \in M$ la sucesión $\big(f_n(x)\big)_{n=1}^{\infty}$ es de Cauchy en M'. Al ser M' completo, $\exists f(x) = \lim_{n\to\infty} f_n(x)$ para todo $x \in M$. Probaremos a continuación que $f \in \mathcal{C}(M, M')$, y $\lim_{n\to\infty} \|f_n - f\| = 0$.

Empecemos por probar esta última afirmación. Por ser $\big(f_n\big)_{n=1}^{\infty}$ de Cauchy, dado $\epsilon > 0$ $\exists N \in \mathbb{N}$ tal que $\|f_l - f_m\| < \epsilon/3$ si $l, m \geq N$. Fijando un $x \in M$ y un $l \geq N$ se tiene entonces

$$\begin{aligned}\|f_l(x) - f(x)\| &\leq \|f_l(x) - f_m(x)\| + \|f_m(x) - f(x)\| \\ &< \frac{\epsilon}{3} + \|f_m(x) - f(x)\| < \frac{\epsilon}{3} + \frac{\epsilon}{3} = \frac{2\epsilon}{3},\end{aligned}$$

tomando $m \geq N$ tal que $\|f_m(x) - f(x)\| < \epsilon/3$. (Nótese que este m dependerá en general de x.) Como la igualdad anterior es cierta para cada $l \geq N$ fijo y para todo $x \in M$, se tiene $\|f_l - f\| \leq 2\epsilon/3 < \epsilon$, $\forall l \geq N$; en consecuencia, $\lim_{n\to\infty} \|f_n - f\| = 0$. Por el teorema anterior, f es continua, y sólo resta probar que es también acotada. Pero esto es sencillo, ya que tomando $n \in \mathbb{N}$ tal que (por ejemplo) $\|f_n - f\| < 1$ se obtiene $\|f(x)\| \leq \|f(x) - f_n(x)\| + \|f_n(x)\| < 1 + \|f_n\|$, $\forall x \in M$. *Q.E.D.*

Capítulo 2

Diferenciación de funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$

2.1. Derivadas direccionales y parciales

En el caso de una función $f : \mathbb{R} \to \mathbb{R}$, se define la derivada de f en un punto x_0 como el límite (si existe)

$$f'(x_0) = \lim_{h\to 0} \frac{f(x_0+h) - f(x_0)}{h};$$

se supone que f está definida en un entorno de x_0, para que $f(x_0+h)$ esté definido para $|h|$ suficientemente pequeño. La definición anterior se generaliza sin dificultad a funciones de $\mathbb{R}$ en $\mathbb{R}^m$ (es decir a curvas parametrizadas en $\mathbb{R}^m$):

Definición 2.1. Si $f : \mathbb{R} \to \mathbb{R}^m$ está definida en un entorno de $t_0 \in \mathbb{R}$, definimos la *derivada de f en t_0* como el límite

$$f'(t_0) \equiv \frac{df}{dt}(t_0) = \lim_{h\to 0} \frac{f(t_0+h) - f(t_0)}{h} \in \mathbb{R}^m,$$

cuando dicho límite existe.

Geométricamente, $f'(t_0)$ es un vector tangente a la curva f en el punto $f(t_0)$. Por las propiedades de los límites, si $f = \big(f_1, \dots, f_m\big)$ entonces $f'(t_0)$ existe si y sólo si existen las derivadas de las funciones componentes (funciones de $\mathbb{R}$ en $\mathbb{R}$) f_i en t_0, $1 \le i \le m$, y se tiene $f'(t_0) = \big(f_1'(t_0), \dots, f_m'(t_0)\big)$.

Sin embargo, si $f : \mathbb{R}^n \to \mathbb{R}^m$ con $n > 1$ es evidente que no puede utilizarse la definición anterior, pues en tal caso $h \in \mathbb{R}^n$ es un *vector*, y no tiene por tanto sentido dividir por h. ¿Cómo definir entonces la derivada de una función $f : \mathbb{R}^n \to \mathbb{R}^m$?

Un primer intento de definir la derivada de una función de $\mathbb{R}^n$ en $\mathbb{R}^m$ conduce al concepto de derivada direccional. La idea (suponiendo de nuevo que $f : \mathbb{R}^n \to \mathbb{R}^m$ está definida en un entorno $B_r(a)$ de un punto $a \in \mathbb{R}^n$) consiste en considerar la restricción de f a lo largo de rectas que pasen por a. En otras palabras, para cada vector $v \in \mathbb{R}^n$ consideramos la recta $\{a + t\,v : t \in \mathbb{R}\}$, y definimos la derivada direccional de f a lo largo del vector v en $a \in \mathbb{R}^n$ como la derivada en $t = 0$ de la función $f_v : \mathbb{R} \to \mathbb{R}^m$ definida por $f_v(t) = f(a + t\,v)$:

Definición 2.2. Si $f : \mathbb{R}^n \to \mathbb{R}^m$ está definida en un entorno de $a \in \mathbb{R}^n$, la *derivada direccional de f a lo largo de v en el punto a* es igual a

$$D_v f(a) = \left.\frac{d}{dt}\right|_{t=0} f(a + tv) = \lim_{t\to 0} \frac{f(a + t\,v) - f(a)}{t},$$

cuando este límite existe.

La derivada direccional tiene las siguientes propiedades, que se siguen fácilmente de su definición:

I) Si $f = (f_1, \dots, f_m)$, entonces $D_v f(a)$ existe si y sólo si existe $D_v f_i(a)$ para todo $i = 1, 2, \dots, m$, y $D_v f(a) = \big(D_v f_1(a), \dots, D_v f_m(a)\big)$

II) Si $D_v f(a)$ existe para algún $v \in \mathbb{R}^n$, entonces existe $D_{\lambda v} f(a)$ para todo $\lambda \in \mathbb{R}$, y se tiene $D_{\lambda v} f(a) = \lambda D_v f(a)$

Debido a esta última propiedad, y dado que si $v \neq 0$ entonces los vectores v y λv definen la misma recta para todo $\lambda \neq 0$, es habitual considerar derivadas direccionales sólo a lo largo de vectores v normalizados, i.e. con $|v| = 1$. Nosotros, sin embargo, no haremos esto. Nótese que, aunque exista $D_v f(a)$ para todo $v \in \mathbb{R}^n$, esto no implica en general que $D_v f(a)$ sea una función lineal de v: $D_{\lambda v + \mu w} f(a) \neq \lambda\, D_v f(a) + \mu\, D_w f(a)$, en general (véase el ejemplo siguiente).

Un caso particular de las derivadas direccionales son las *derivadas parciales*. Por definición, la i-ésima derivada parcial de $f : \mathbb{R}^n \to \mathbb{R}^m$ no es más que la derivada direccional de f a lo largo de e_i, el i-ésimo vector de la base canónica de $\mathbb{R}^n$:

$$D_i f(a) \equiv \frac{\partial f}{\partial x_i}(a) = D_{e_i} f(a).$$

Otras formas equivalentes de expresar la i-ésima derivada parcial de f en a, que se obtienen sin más que aplicar la definición de derivada direccional, son:

$$D_i f(a) = \left.\frac{d}{dt}\right|_{t=0} f(a + te_i) = \left.\frac{d}{dx_i}\right|_{x_i = a_i} f(a_1, \dots, a_{i-1}, x_i, a_{i+1}, \dots, a_n).$$

Por la propiedad i) de las derivadas direccionales, se tiene

$$\frac{\partial f}{\partial x_i}(a) = \left(\frac{\partial f_1}{\partial x_i}(a), \dots, \frac{\partial f_m}{\partial x_i}(a)\right),$$

donde el miembro izquierdo existe si y sólo si existen cada una de las componentes del miembro derecho.

Ejemplo 2.3. Sea $f : \mathbb{R}^2 \to \mathbb{R}$ definida por

$$f(x,y) = \begin{cases} \dfrac{x\,y^2}{x^2+y^4}, & (x,y) \neq (0,0) \\ 0, & (x,y) = (0,0). \end{cases}$$

Vimos en el capítulo anterior que f no es continua en el origen (el límite a lo largo de rectas por el origen no coincide con el límite a lo largo de la parábola $x = y^2$). Sin embargo, veamos que f tiene derivada direccional en el origen a lo largo de cualquier dirección $v = (v_1, v_2)$. En efecto,

$$D_v f(0,0) = \lim_{t\to 0} \frac{f(tv_1, tv_2)}{t} = \lim_{t\to 0} \frac{v_1 v_2^2}{v_1^2 + t^2 v_2^2} = \begin{cases} \frac{v_2^2}{v_1}, & v_1 \neq 0 \\ 0, & v_1 = 0. \end{cases}$$

En particular, $D_1 f(0,0) = D_2 f(0,0) = 0$. Nótese por último que $D_v f(0,0)$, aunque está definido para todo $v \in \mathbb{R}^2$, no es lineal en v.

Ejemplo 2.4. Sea $f : \mathbb{R}^2 \to \mathbb{R}$ definida por

$$f(x,y) = \begin{cases} \dfrac{xy}{x^2+y^2}, & (x,y) \neq (0,0) \\ 0, & (x,y) = (0,0). \end{cases}$$

Esta función no es continua en el origen (pues el límite de f en el origen a lo largo de rectas por el origen depende de la pendiente). Sin embargo, veamos que existen las dos derivadas parciales de f en $(0,0)$. En efecto,

$$D_1 f(0,0) = \lim_{t\to 0} \frac{f(t,0)}{t} = 0,$$

y análogamente $D_2 f(0,0) = 0$. Por otra parte, no existe la derivada direccional de f a lo largo de ninguna otra dirección $v \neq \lambda e_1$ ó λe_2, ya que

$$D_{(v_1,v_2)} f(0,0) = \lim_{t\to 0} \frac{f(tv_1, tv_2)}{t} = \lim_{t\to 0} \frac{1}{t} \frac{v_1 v_2}{v_1^2 + v_2^2}$$

es infinito.

2.2. Derivada

Como se ve en los ejemplos anteriores, la existencia de derivadas parciales no es suficiente para garantizar la continuidad de una función de $\mathbb{R}^n$ en $\mathbb{R}^m$. (Recuérdese que, para funciones $f : \mathbb{R} \to \mathbb{R}$, f diferenciable en un punto implica f continua en dicho punto.) Es por esto por lo que la existencia de derivadas parciales (o incluso de derivadas direccionales en cualquier dirección) no juega un papel importante en el caso de funciones de $\mathbb{R}^n \to \mathbb{R}^m$.

Necesitamos, por tanto, una generalización del concepto de diferenciabilidad al caso de funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$ que (entre otras propiedades) implique la continuidad. Examinemos con más detalle, para ello, la definición de la derivada de una función real de variable real en un punto. Si $f'(x_0)$ existe, se puede aproximar $f(x_0+h)-f(x_0)$ para $|h|$ suficientemente pequeño por $f'(x_0)\,h$, que es una función lineal del incremento h. Si llamamos $\lambda(h)$ a esta función, la afirmación anterior se expresa de forma más precisa mediante la igualdad

$$\lim_{h\to 0}\frac{1}{h}\big(f(x_0+h)-f(x_0)-\lambda(h)\big)=0.$$

En otras palabras, si f es derivable en x_0 la diferencia entre el incremento de la función en x_0, $f(x_0+h)-f(x_0)$, y la aproximación lineal $\lambda(h)=f'(x_0)\,h$ es $o(h)$. Recíprocamente, si se cumple la igualdad anterior para alguna función lineal $\lambda:\mathbb{R}\to\mathbb{R}$ entonces $\lambda(h)=a\,h$ para algún $a\in\mathbb{R}$, y $f'(x_0)=a$. En otras palabras, la existencia de una función lineal $\lambda:\mathbb{R}\to\mathbb{R}$ tal que se cumple la igualdad anterior es equivalente a la derivabilidad de f en x_0. Motivados por esta observación, hacemos la siguiente definición:

Definición 2.5. Si $f:\mathbb{R}^n\to\mathbb{R}^m$ está definida en un entorno de $a\in\mathbb{R}^n$, diremos que f es *derivable* o diferenciable *en a* si existe una aplicación lineal $\lambda_a:\mathbb{R}^n\to\mathbb{R}^m$ tal que

$$\lim_{h\to 0}\frac{|f(a+h)-f(a)-\lambda_a(h)|}{|h|}=0. \tag{2.1}$$

A la aplicación lineal $\lambda_a:\mathbb{R}^n\to\mathbb{R}^m$ se la denomina *derivada de f en el punto a.*

Nota. Nótese que (2.1) es equivalente a

$$\lim_{h\to 0}\frac{f(a+h)-f(a)-\lambda_a(h)}{|h|}=0. \tag{2.2}$$

La definición anterior de derivada tiene sentido en virtud de la siguiente proposición, que establece la unicidad de la derivada de una aplicación derivable:

Proposición 2.6. *Si $f:\mathbb{R}^n\to\mathbb{R}^m$ es derivable en a, su derivada es única.*

Demostración. Supongamos que existiera otra aplicación lineal $\mu:\mathbb{R}^n\to\mathbb{R}^m$ tal que

$$\lim_{h\to 0}\frac{f(a+h)-f(a)-\mu(h)}{|h|}=0.$$

Restando esta ecuación de (2.2) se obtiene

$$\lim_{h\to 0}\frac{\lambda_a(h)-\mu(h)}{|h|}=0.$$

En particular, haciendo h tender a cero a lo largo de una recta $h = tv$ con $v \neq 0$ arbitrario se obtiene, por linealidad,

$$0 = \lim_{t \to 0} \frac{|\lambda_a(tv) - \mu(tv)|}{|t|\,|v|} = \frac{|\lambda_a(v) - \mu(v)|}{|v|},$$

de donde se sigue que $\lambda_a(v) = \mu(v)$ para todo $v \neq 0$, lo cual implica que $\lambda_a = \mu$. *Q.E.D.*

Para denotar la derivada de una función $f : \mathbb{R}^n \to \mathbb{R}^m$ en un punto a, utilizaremos la notación $Df(a)$. Por definición de derivada, $Df(a) : \mathbb{R}^n \to \mathbb{R}^m$ es una aplicación *lineal*, que (como indica la notación) depende del punto a. Además, como es costumbre en las aplicaciones lineales, la imagen de $h \in \mathbb{R}^n$ bajo $Df(a)$ la denotaremos por $Df(a) \cdot h$. Por definición de derivada, $Df(a)$ (si existe) satisface

$$\lim_{h \to 0} \frac{|f(a+h) - f(a) - Df(a) \cdot h|}{|h|} = 0. \tag{2.3}$$

Otra forma de expresar la existencia de derivada de una función $f : \mathbb{R}^n \to \mathbb{R}^m$ en un punto $a \in \mathbb{R}^n$ es la siguiente:

$$f(a+h) - f(a) = Df(a) \cdot h + \eta_a(h), \tag{2.4}$$

con $\eta_a = o(|h|)$, es decir

$$\lim_{h \to 0} \frac{|\eta_a(h)|}{|h|} = 0. \tag{2.5}$$

La derivada de una función tiene las propiedades elementales siguientes, que se prueban fácilmente a partir de la definición:

i) $f = (f_1, \dots, f_m) : \mathbb{R}^n \to \mathbb{R}^m$ es derivable en a si y sólo si $f_i : \mathbb{R}^n \to \mathbb{R}$ es derivable en dicho punto para todo $i = 1, 2, \dots, m$

ii) Si $f : \mathbb{R}^n \to \mathbb{R}^m$ y $g : \mathbb{R}^n \to \mathbb{R}^m$ son ambas derivables en a, entonces $\lambda\, f + \mu\, g$ es derivable en a $\forall \lambda, \mu \in \mathbb{R}$, y se tiene

$$D(\lambda\, f + \mu\, g)(a) = \lambda\, Df(a) + \mu\, Dg(a)$$

iii) Si $L : \mathbb{R}^n \to \mathbb{R}^m$ es lineal, entonces L es derivable en cualquier punto, con derivada

$$DL(a) = L, \qquad \forall a \in \mathbb{R}^n$$

La definición de derivabilidad que acabamos de ver garantiza la continuidad:

Proposición 2.7. *Si* $f : \mathbb{R}^n \to \mathbb{R}^m$ *es derivable en* a*, entonces* f *es continua en* a*.*

Demostración. Si f es derivable en a, por hipótesis f está definida en un entorno de a, y por tanto $a \in \operatorname{dom}(f) \cap \operatorname{dom}(f)'$. Además, utilizando (2.4) y observando que $\lim_{h\to 0} \eta_a(h) = \lim_{h\to 0} |h| \, \frac{\eta_a(h)}{|h|} = 0 \cdot 0 = 0$, recordando que las aplicaciones lineales de $\mathbb{R}^n$ en $\mathbb{R}^m$ son continuas obtenemos

$$\lim_{x\to a} f(x) = \lim_{h\to 0} f(a+h) = f(a) + Df(a) \cdot 0 + 0 = f(a).$$

Q.E.D.

Veamos a continuación qué relación hay entre la derivada y las derivadas direccionales y parciales. Si $f : \mathbb{R}^n \to \mathbb{R}^m$ es derivable en a, haciendo h en (2.3) tender a cero a lo largo de una recta $h = t\,v$ $(v \neq 0)$ obtenemos

$$\begin{aligned} 0 &= \lim_{t\to 0} \frac{|f(a+tv) - f(a) - tDf(a) \cdot v|}{|t| \; |v|} \\ &= \lim_{t\to 0} \left| \frac{f(a+tv) - f(a) - tDf(a) \cdot v}{t} \right| \\ &= \lim_{t\to 0} \left| \frac{f(a+tv) - f(a)}{t} - Df(a) \cdot v \right|, \end{aligned}$$

lo que equivale a

$$\lim_{t\to 0} \frac{f(a+tv) - f(a)}{t} = Df(a) \cdot v.$$

Por tanto, hemos probado la siguiente proposición:

Proposición 2.8. *Si f es derivable en a, entonces existe la derivada direccional de f en a a lo largo de cualquier dirección v, y se tiene*

$$D_v f(a) = Df(a) \cdot v. \tag{2.6}$$

Como corolario de la proposición anterior, nótese que si f es derivable no sólo existen las derivadas parciales de f a lo largo de todas las direcciones v, sino que además $D_v f(a)$ ha de ser *lineal* en v. De la proposición anterior se sigue que si f es derivable en a entonces existen todas las derivadas parciales de f en a, y se tiene

$$D_i f(a) = Df(a) \cdot e_i, \tag{2.7}$$

de donde se obtiene por linealidad que si $v = \big(v_1, \dots, v_n\big)$ entonces

$$D_v f(a) = \sum_{i=1}^{n} v_i Df(a) \cdot e_i = \sum_{i=1}^{n} D_i f(a) \, v_i. \tag{2.8}$$

Nótese que esta fórmula no tiene por qué ser válida si f no es derivable en a (cf. el ejemplo 2.3). En otras palabras, si f es derivable en a entonces el *incremento de f en a*

$$\Delta f(a; h) = f(a+h) - f(a)$$

se puede expresar de la siguiente forma en términos de las derivadas parciales:

$$\Delta f(a;h) = \sum_{i=1}^{n} D_i f(a)\, h_i + o(|h|).$$

Por último, obsérvese que si $f = \big(f_1, \dots, f_m\big)$ entonces (2.7) afirma que en las bases canónicas de $\mathbb{R}^n$ y $\mathbb{R}^m$ la i-ésima columna de $Df(a)$ es el vector columna formado por las componentes de $D_i f(a)$, es decir el vector columna $\Big(\frac{\partial f_1}{\partial x_i}(a) \cdots \frac{\partial f_m}{\partial x_i}(a)\Big)^{\top}$. Por tanto, la matriz de $Df(a)$ en las bases canónicas de $\mathbb{R}^n$ y $\mathbb{R}^m$, llamada *matriz jacobiana de f en a*, está dada por

$$Jf(a) = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1}(a) & \cdots & \dfrac{\partial f_1}{\partial x_n}(a) \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1}(a) & \cdots & \dfrac{\partial f_m}{\partial x_n}(a) \end{pmatrix}. \tag{2.9}$$

Veamos a continuación algunos casos particulares de lo anterior.

i) Curvas
Si $f : \mathbb{R} \to \mathbb{R}^m$, entonces f es derivable en $a \in \mathbb{R}$ si y sólo si existe $f'(a)$, en cuyo caso $Df(a) : \mathbb{R} \to \mathbb{R}^m$ es la aplicación lineal definida por $Df(a) \cdot h = f'(a)\, h$, $\forall h \in \mathbb{R}$. En efecto, si f es derivable en a, por (2.9) su matriz jacobiana es la matriz columna

$$Jf(a) = \big(f_1'(a) \cdots f_m'(a)\big)^{\top},$$

de donde se sigue la fórmula para $Df(a) \cdot h$. Recíprocamente, si existe $f'(a)$ entonces

$$\lim_{h\to 0} \frac{1}{h}\big(f(a+h) - f(a) - h\, f'(a)\big) = \lim_{h\to 0} \left(\frac{f(a+h) - f(a)}{h} - f'(a) \right) = 0$$

por definición de $f'(a)$.

ii) Funciones escalares
El caso particular más importante es el de una función escalar $f : \mathbb{R}^n \to \mathbb{R}$. Si f es derivable en a, $Df(a)$ es una aplicación lineal $\mathbb{R}^n \to \mathbb{R}$, es decir un elemento de $(\mathbb{R}^n)^*$. La matriz jacobiana $Jf(a)$ es la matriz fila

$$Jf(a) = \Big(\frac{\partial f}{\partial x_1}(a) \cdots \frac{\partial f}{\partial x_n}(a)\Big);$$

por tanto, si $h = (h_1, \dots, h_n) \in \mathbb{R}^n$ se tiene

$$Df(a) \cdot h = Jf(a) \cdot (h_1 \dots h_n)^{\top}.$$

Es conveniente, en vista de lo anterior, definir el *gradiente* de f en a como el vector $\nabla f(a) \in \mathbb{R}^n$ cuyas componentes son las derivadas parciales de f en a, es decir

$$\nabla f(a) = \Big(\frac{\partial f}{\partial x_1}(a), \cdots, \frac{\partial f}{\partial x_n}(a)\Big); \tag{2.10}$$

con esta definición, si f es derivable en a entonces se tiene

$$Df(a) \cdot h = \nabla f(a) \cdot h,$$

donde el punto en el miembro derecho denota producto escalar.

Vimos anteriormente que la existencia de derivadas direccionales o parciales de una función en un punto no implica su derivabilidad en dicho punto. Sin embargo, probaremos a continuación que si las derivadas parciales de $f : \mathbb{R}^n \to \mathbb{R}^m$ existen y son *continuas* en $a \in \mathbb{R}^n$, entonces f es derivable en dicho punto. Para ello, recordemos el *teorema del valor medio* para funciones reales de una variable real:

Teorema (del valor medio)**.** *Si $f : \mathbb{R} \to \mathbb{R}$ es continua en $[a, b]$ y derivable en (a, b), entonces $\exists c \in (a, b)$ tal que $f(b) = f(a) + (b - a)f'(c)$.*

Teorema 2.9. *Supongamos que una de las derivadas parciales de $f : \mathbb{R}^n \to \mathbb{R}^m$ existe en $a \in \mathbb{R}^n$, y las restantes $n - 1$ derivadas parciales existen en un entorno de a y son continuas en dicho punto. Entonces f es derivable en a.*

Demostración. En primer lugar, el caso general se obtiene aplicando el caso $m = 1$ a cada una de las componentes de f (recuérdese que f es diferenciable en a si y sólo si lo son cada una de sus componentes). Por tanto, supondremos a partir de ahora que f es una función escalar. Sin pérdida de generalidad, podemos suponer que $D_1 f(a)$ existe, y que las derivadas parciales $D_i f$ con $2 \le i \le n$ existen en un entorno de a y son continuas en dicho punto. Nótese que si f es derivable en a, necesariamente su derivada Df tiene que venir dada por

$$Df(a) \cdot (h_1, \dots, h_n) = \sum_{i=1}^{n} D_i f(a)\, h_i;$$

por tanto, para probar que f es derivable hay que probar que

$$\lim_{h\to 0} \frac{|f(a + h) - f(a) - \sum_{i=1}^{n} D_i f(a)\, h_i|}{|h|} = 0.$$

Para ello, escribimos el numerador de esta última expresión como sigue:

$$\begin{aligned}
f(a+h) - f(a) - \sum_{i=1}^{n} D_i f(a)\, h_i \\
&= \sum_{i=1}^{n} [f(a_1 + h_1, \dots, a_i + h_i, a_{i+1}, \dots, a_n) \\
&\qquad - f(a_1 + h_1, \dots, a_{i-1} + h_{i-1}, a_i, \dots, a_n) - D_i f(a)\, h_i] \\
&= f(a_1 + h_1, a_2, \dots, a_n) - f(a_1, \dots, a_n) - D_1 f(a)\, h_1 \\
&\quad + \sum_{i=2}^{n} [f(a_1 + h_1, \dots, a_i + h_i, a_{i+1}, \dots, a_n) \\
&\qquad - f(a_1 + h_1, \dots, a_{i-1} + h_{i-1}, a_i, \dots, a_n) - D_i f(a)\, h_i]\,.
\end{aligned}$$

Por el teorema del valor medio aplicado a cada una de las funciones reales de variable real $f(a_1 + h_1, \dots, a_{i-1} + h_{i-1}, \cdot\,, a_{i+1}, \dots, a_n)$, $2 \le i \le n$, podemos escribir

$$\begin{aligned}
&f(a+h) - f(a) - \sum_{i=1}^{n} D_i f(a)\, h_i = \\
&\quad f(a_1 + h_1, a_2, \dots, a_n) - f(a_1, \dots, a_n) - D_1 f(a)\, h_1 \\
&\quad + \sum_{i=2}^{n} [D_i f(a_1 + h_1, \dots, a_{i-1} + h_{i-1}, a_i + \theta_i\, h_i, a_{i+1}, \dots, a_n) - D_i f(a)]\ h_i,
\end{aligned}$$

con $\theta_i \in (0,1)$, $2 \le i \le n$. Al existir $D_1 f(a)$, el primer término de esta ecuación se puede escribir

$$f(a_1 + h_1, a_2, \dots, a_n) - f(a) - D_1 f(a)\, h_1 = \mu(h_1)\, h_1,$$

donde $\lim_{h_1 \to 0} \mu(h_1) = 0$, y se puede tomar $\mu(0) = 0$ (pues el valor de μ en 0 no afecta la igualdad anterior). Por tanto se tiene, finalmente

$$\begin{aligned}
&\frac{|f(a+h) - f(a) - \sum_{i=1}^{n} D_i f(a)\, h_i|}{|h|} \le \frac{h_1}{|h|} \mu(h_1) \\
&\quad + \sum_{i=2}^{n} |D_i f(a_1 + h_1, \dots, a_{i-1} + h_{i-1}, a_i + \theta_i\, h_i, a_{i+1}, \dots, a_n) - D_i f(a)|\ \frac{h_i}{|h|}.
\end{aligned}$$

Como $|h_i| \,/\, |h| \le 1$ para todo i, el primero de estos sumando tiende a cero por la existencia de $D_1 f(a)$, mientras que el segundo tiende también a cero por la continuidad de $D_2 f, \dots, D_n f$ en a. *Q.E.D.*

Nota. El teorema anterior proporciona una condición *suficiente*, pero no necesaria, de diferenciabilidad. En otras palabras, hay funciones diferenciables que tienen alguna derivada parcial discontinua. (Por ejemplo, piénsese en el caso de funciones de $\mathbb{R}$ en $\mathbb{R}$.)

Ejemplo 2.10. Un polinomio $P : \mathbb{R}^n \to \mathbb{R}$ es una función diferenciable en cualquier punto. En efecto, las derivadas parciales de un polinomio respecto de cualquier coordenada son polinomios, y por tanto son continuas en todo punto.

Definición 2.11. Sea U un abierto de $\mathbb{R}^n$. Se dice que $f : \mathbb{R}^n \to \mathbb{R}^m$ es *de clase 0 en U*, y se escribe $f \in \mathbb{C}^0(U)$, si f es continua en U. Si todas las derivadas parciales de f existen y son continuas en U, se dice que f es *de clase 1 en U*, lo que se denota por $f \in C^1(U)$.

Si $f \in C^1(U)$, al ser U abierto puede aplicarse el teorema anterior, y f es diferenciable en U. Por la proposición 2.7, f es continua en U; en particualr, $C^1(U) \subset C^0(U)$.

2.3. Regla de Leibniz y de la cadena

En la sección anterior hemos visto algunas propiedades fundamentales de la derivada de una función de $\mathbb{R}^n$ en $\mathbb{R}^m$. Además, hemos desarrollado una condición suficiente para averiguar si una función es diferenciable en un punto (Teorema 2.9). En esta sección completaremos dichas propiedades con otros dos resultados fundamentales, que permiten muchas veces probar que una función es diferenciable sin necesidad de calcular sus derivadas parciales. Comenzaremos probando que el producto de dos funciones escalares diferenciables es diferenciable:

Teorema 2.12. *Si $f, g : \mathbb{R}^n \to \mathbb{R}$ son funciones diferenciables en $a \in \mathbb{R}^n$, el producto fg es diferenciable, y se tiene*

$$D(fg)(a) = f(a)\, Dg(a) + g(a)Df(a)$$

(regla de Leibniz).

Demostración. La demostración es un cálculo sencillo. El incremento de fg en el punto a es

$$\begin{aligned} &f(a+h)g(a+h) - f(a)g(a) \\ &\qquad = \big(f(a) + Df(a)\cdot h + \eta(h)\big)\big(g(a) + Dg(a)\cdot h + \mu(h)\big) - f(a)g(a) \\ &\qquad = f(a)\, Dg(a)\cdot h + g(a)\, Df(a)\cdot h + \rho(h), \end{aligned}$$

con

$$\begin{aligned} \rho(h) = f(a)\mu(h) + g(a)\eta(h) + [Df(a)\cdot h]\,[Dg(a)\cdot h] \\ + \mu(h)\, Df(a)\cdot h + \eta(h)\, Dg(a)\cdot h + \eta(h)\mu(h), \end{aligned}$$

y $\eta(h) = o(|h|)$, $\mu(h) = o(|h|)$. Por tanto, basta probar que también $\rho(h) = o(|h|)$. Todos los sumandos de que consta ρ son patentemente $o(|h|)$ menos el tercero. Para ver que este sumando es también $o(|h|)$, escribimos

$$\frac{|Df(a)\cdot h|\;|Dg(a)\cdot h|}{|h|} = \left|Df(a)\cdot\frac{h}{|h|}\right| |Dg(a)\cdot h|\,. \tag{2.11}$$

Como $Df(a)$ es continua en el compacto $\overline{B}_1(0)$, existe una constante M tal que

$$\left|Df(a)\cdot\frac{h}{|h|}\right| \leq M, \quad \forall h \neq 0,$$

y por tanto (2.11) está acotado por $M\,Dg(a)\cdot h$, que tiende a cero cuando $h \to 0$ por continuidad de $Dg(a)$. *Q.E.D.*

Nota. La regla de Leibniz se suele expresar también en términos del gradiente, en lugar de la derivada:

$$\nabla(fg)(a) = g(a)\,\nabla g(a) + g(a)\,\nabla f(a).$$

El siguiente teorema demuestra que, al igual que el producto, la composición de funciones diferenciables es diferenciable:

Teorema 2.13. *Si $g : \mathbb{R}^n \to \mathbb{R}^m$ es diferenciable en $a \in \mathbb{R}^n$, y $f : \mathbb{R}^m \to \mathbb{R}^p$ es diferenciable en $g(a) \in \mathbb{R}^m$, entonces la función compuesta $f \circ g : \mathbb{R}^n \to \mathbb{R}^p$ es diferenciable en a, con derivada*

$$D(f \circ g)(a) = Df(g(a)) \cdot Dg(a)$$

(regla de la cadena).

Demostración. (Nótese que, por definición de producto de aplicaciones lineales, $Df(g(a)) \cdot Dg(a) \equiv Df(g(a)) \circ Dg(a)$.) Sea $\Delta(h) = g(a+h) - g(a)$; entonces, por ser f diferenciable en $g(a)$ se tiene

$$f(g(a+h)) - f(g(a)) = Df(g(a)) \cdot \Delta(h) + \eta(\Delta(h)),$$

siendo $\eta(k) = o(|k|)$. A su vez, por ser g diferenciable en a el incremento $\Delta(h)$ puede expresarse de la forma siguiente:

$$\Delta(h) = Dg(a) \cdot h + \mu(h),$$

con $\mu(h) = o(|h|)$. Sustituyendo en la ecuación del incremento de $f \circ g$ queda

$$f(g(a+h)) - f(g(a)) - \big(Df(g(a)) \cdot Dg(a)\big) \cdot h = Df(g(a)) \cdot \mu(h) + \eta(\Delta(h)).$$

Basta probar, por tanto, que el miembro derecho de la ecuación anterior es $o(|h|)$. El primer término lo es claramente, pues

$$\frac{Df(g(a)) \cdot \mu(h)}{|h|} = Df(g(a)) \cdot \frac{\mu(h)}{|h|} @>> h \to 0 > 0,$$

por la continuidad de la función lineal $Df(g(a))$. En cuanto al segundo término, procediendo como en la demostración de la regla de Leibniz se demuestra que $\Delta(h)/|h|$ está acotada cuando h tiende a cero, es decir existe $M > 0$ tal que $|\Delta(h)| \leq M|h|$ para h suficientemente pequeño. En particular, de esto (o directamente de la definición de $\Delta(h)$) se sigue que $\lim_{h\to 0} \Delta(h) = 0$. Por otra parte, por ser $\eta(k) = o(|k|)$ podemos poner $\eta(k) = |k|\,\psi(k)$, con $\lim_{k\to 0} \psi(k) = 0$. En consecuencia,

$$\frac{\eta(\Delta(h))}{|h|} = \frac{|\Delta(h)|}{|h|} \cdot \psi(\Delta(h)) \leq M\psi(\Delta(h))@ >> h \to 0 > 0,$$

lo cual concluye la demostración. *Q.E.D.*

Ejemplo 2.14. Si $f : \mathbb{R}^n \to \mathbb{R}$ es diferenciable en $a \in \mathbb{R}^n$, y $f(a) \neq 0$, entonces $1/f$ es diferenciable en a. En efecto, $1/f = \varphi \circ f$, siendo $\varphi : \mathbb{R} \to \mathbb{R}$ la función definida por $\varphi(t) = 1/t$, $\forall t \neq 0$. Como esta función es derivable en su dominio y $f(a) \neq 0$, aplicando la regla de la cadena se obtiene efectivamente que $1/f$ es diferenciable en a. Además, la derivada de $1/f$ está dada por

$$D(1/f)(a) \cdot h = D\varphi(f(a)) \cdot \big(Df(a) \cdot h\big) = -\frac{1}{f(a)^2} Df(a) \cdot h.$$

Por tanto,

$$D(1/f)(a) = -\frac{1}{f(a)^2} Df(a),$$

ó

$$\nabla(1/f)(a) = -\frac{1}{f(a)^2} \nabla f(a),$$

en términos del gradiente. En particular, si $g = P/Q$ es una función racional, de lo anterior se sigue que g es derivable en el conjunto de puntos en que no se anula el denominador Q (que evidentemente es un abierto de $\mathbb{R}^n$).

La regla de la cadena puede formularse también matricialmente. En efecto, con las mismas hipótesis que en el teorema anterior, recordando que la matriz de un producto de aplicaciones lineales es el producto de las matrices de dichas aplicaciones (en las bases adecuadas) se obtiene la fórmula

$$J(f \circ g)(a) = Jf(g(a)) \cdot Jg(a),$$

donde el punto denota el producto de matrices. Si $f = \big(f_1, \dots, f_p\big)$ y $g = \big(g_1, \dots, g_m\big)$, tomando el elemento de matriz ij de la igualdad anterior se obtiene

$$D_j(f_i \circ g)(a) = \sum_{k=1}^{m} D_k f_i\big(g(a)\big)\, D_j g_k(a), \qquad 1 \leq i \leq p,\ 1 \leq j \leq n.$$

Abusando ligeramente de la notación, es costumbre escribir $y = g(x)$, $z = f(y) = f(g(x))$, $D_j g_k(a) = \frac{\partial y_k}{\partial x_j}(a)$, $D_k f_i(g(a)) = \frac{\partial z_i}{\partial y_k}(y(a))$ y $D_j(f_i \circ g)(a) = \frac{\partial z_i}{\partial x_j}(a)$; la fórmula anterior se escribe entonces en la forma más familiar

$$\frac{\partial z_i}{\partial x_j}(a) = \sum_{k=1}^{m} \frac{\partial z_i}{\partial y_k}(y(a))\, \frac{\partial y_k}{\partial x_j}(a).$$

En particular, sea $f : \mathbb{R}^n \to \mathbb{R}$ una función escalar, y llamemos $F(x) = f(g_1(x), \ldots, g_m(x))$, $\forall x \in \mathbb{R}^n$. Si cada una de las funciones $g_i : \mathbb{R}^n \to \mathbb{R}$ son de clase 1 en un entorno de $x \in \mathbb{R}^n$ y f es de clase 1 en un entorno de $g(x) = \big(g_1(x), \ldots, g_m(x)\big)$, entonces $F = f \circ g$ es derivable en x (regla de la cadena), y se tiene

$$\frac{\partial F}{\partial x_i}(x) = \sum_{j=1}^{m} \frac{\partial f}{\partial y_j}(g(x))\, \frac{\partial g_j}{\partial x_i}(x), \qquad 1 \le i \le n,$$

donde hemos puesto, como es costumbre, $\frac{\partial f}{\partial y_j} = D_j f$.

Ejemplo 2.15. Sea $f : \mathbb{R}^n \to \mathbb{R}$ una función diferenciable, y sea $\gamma : \mathbb{R} \to \mathbb{R}^n$ una curva diferenciable. Evaluando f a lo largo de la curva γ se obtiene la función $F = f \circ \gamma : \mathbb{R} \to \mathbb{R}$, que es diferenciable en $\mathbb{R}$ por la regla de la cadena. Aplicando la regla de la cadena se obtiene

$$\begin{aligned} DF(t) \cdot h &= F'(t)\, h = Df(\gamma(t)) \cdot \big(D\gamma(t) \cdot h\big) = Df(\gamma(t)) \cdot \big(\gamma'(t)\, h\big) \\ &= h\, Df(\gamma(t)) \cdot \gamma'(t), \end{aligned}$$

de donde

$$F'(t) = Df(\gamma(t)) \cdot \gamma'(t).$$

En términos del gradiente, esta fórmula se escribe

$$F'(t) = \nabla f(\gamma(t)) \cdot \gamma'(t). \tag{2.12}$$

Nótese que las fórmulas anteriores son una generalización de (2.6) (tómese $\gamma(t) = a + t\, v$ y evalúese F' en $t = 0$).

Ejemplo 2.16. *Cálculo del laplaciano en coordenadas polares*
Sea $f : \mathbb{R}^2 \to \mathbb{R}$, y supongamos que existen las derivadas parciales $f_x = \partial f/\partial x, f_y = \partial f/\partial y : \mathbb{R}^2 \to \mathbb{R}$ y son de clase 1. Entonces existen y son continuas las derivadas parciales de segundo orden $f_{xx} = \partial f_x/\partial x, f_{xy} = \partial f_x/\partial y, f_{yx} = \partial f_y/\partial x, f_{yy} = \partial f_y/\partial y : \mathbb{R}^2 \to \mathbb{R}$. Además, veremos a continuación que con las hipótesis hechas sobre f se verifica la igualdad $f_{xy} = f_{yx}$. Se define entonces el *laplaciano de* f mediante

$$(\triangle f)(x, y) = f_{xx}(x, y) + f_{yy}(x, y).$$

Sea ahora $g : \mathbb{R}^2 \to \mathbb{R}^2$ la función definida por $g(r,\theta) = (r\cos\theta, r\operatorname{sen}\theta)$, $\forall r > 0$, $\forall\theta \in (0,2\pi)$. La función g aplica biyectivamente la franja abierta $(0,\infty) \times (0,2\pi)$ en el abierto $U = \mathbb{R}^2 - \big\{(x,0) : x \geq 0\big\}$; por definición, se dice que (r,θ) son las *coordenadas polares* del punto $(r\cos\theta, r\operatorname{sen}\theta) \in U$. Entonces la composición $F = f \circ g$ no es otra cosa que la expresión de f en coordenadas polares, y nuestro objetivo es escribir $(\triangle f) \circ g$—es decir, la expresión de $\triangle f$ en coordenadas polares—en términos de las derivadas parciales de F con respecto a r y θ.

Para ello, calculamos primero F_r y F_θ utilizando la regla de la cadena:

$$F_r = f_x\,\frac{\partial x}{\partial r} + f_y\frac{\partial y}{\partial r} = \cos\theta\, f_x + \operatorname{sen}\theta\, f_y, \tag{2.13}$$

$$F_\theta = f_x\,\frac{\partial x}{\partial \theta} + f_y\frac{\partial y}{\partial \theta} = r\,(-\operatorname{sen}\theta\, f_x + \cos\theta\, f_y). \tag{2.14}$$

Nótese que las derivadas parciales respecto de r y θ están calculadas en (r,θ), mientras que las parciales respecto de x e y se evalúan en $(x,y) = g(r,\theta) = (r\cos\theta, r\operatorname{sen}\theta)$. Aplicando de nuevo la regla de la cadena a F_r y F_θ se obtiene:

$$\begin{aligned}
F_{rr} &= \cos\theta\,(f_{xx}\cos\theta + f_{xy}\operatorname{sen}\theta) + \operatorname{sen}\theta\,(f_{yx}\cos\theta + f_{yy}\operatorname{sen}\theta),\\
\frac{1}{r}\,F_{\theta\theta} &= -\cos\theta\, f_x - \operatorname{sen}\theta\, f_y - r\operatorname{sen}\theta\,(-f_{xx}\operatorname{sen}\theta + f_{xy}\cos\theta)\\
&\quad + r\cos\theta(-f_{yx}\operatorname{sen}\theta + f_{yy}\cos\theta).
\end{aligned}$$

Utilizando la expresión anterior de F_r, y teniendo en cuenta la igualdad $f_{xy} = f_{yx}$ podemos escribir

$$\begin{aligned}
F_{rr} &= \cos^2\theta\, f_{xx} + \operatorname{sen}(2\theta)\, f_{xy} + \operatorname{sen}^2\theta\, f_{yy},\\
\frac{1}{r^2}F_{\theta\theta} + \frac{1}{r}F_r &= \operatorname{sen}^2\theta\, f_{xx} - \operatorname{sen}(2\theta)\, f_{xy} + \cos^2\theta\, f_{yy},
\end{aligned}$$

y sumando queda

$$F_{rr} + \frac{1}{r^2}F_{\theta\theta} + \frac{1}{r}F_r = f_{xx} + f_{yy} = (\triangle f)\circ g,$$

que es la expresión que buscábamos. En particular, supongamos que f es *armónica*, es decir satisface la *ecuación de Laplace* $\triangle f = 0$, y F sólo depende de r. Entonces se tiene

$$0 = (\triangle f)\circ g = F''(r) + \frac{1}{r}F'(r).$$

Integrando una vez queda $\log|F'(r)| = -\log r + c$ con c constante, o equivalentemente $F'(r) = \frac{a}{r}$, donde a es constante ($a = \pm e^c$). Integrando de nuevo queda

$$F(r) = a\log r + b$$

con a y b constantes arbitrarias, que es la expresión más general de una función armónica independiente del ángulo polar θ en $\mathbb{R}^2$.

Nota. El potencial eléctrico (o gravitatorio) $V : \mathbb{R}^3 \to \mathbb{R}$ satisface la ecuación de Laplace $\triangle V = V_{xx} + V_{yy} + V_{zz}$ en la región del espacio en que no existen cargas (masas). Si consideramos el potencial eléctrico creado por un hilo infinito de carga situado en el eje Oz, entonces $\triangle V = 0$ fuera del eje Oz. Por simetría, V no depende de z, es decir $V(x, y, z) = f(x, y)$, donde F tampoco depende de θ. Aplicando el resultado anterior obtenemos que la expresión de V en coordenadas cilíndricas es

$$V(r\,\cos\theta, r\,\operatorname{sen}\theta, z) = a\,\log r + b$$

donde $r = \sqrt{x^2 + y^2}$ es la distancia al eje Oz (es decir, al hilo de carga), y la constante b puede ponerse igual a cero (pues físicamente sólo interesa la diferencia de potencial). El campo eléctrico $\mathbf{E}(x, y, z)$ creado por la carga está dado por

$$\begin{aligned}\mathbf{E} &= -\nabla V = -(V_x, V_y, 0) = -(f_x, f_y, 0) = -(\cos\theta\, F'(r), \operatorname{sen}\theta\, F'(r), 0)\\ &= -\frac{a}{r}\,(\cos\theta, \operatorname{sen}\theta, 0).\end{aligned}$$

en el punto $(x, y, z) = (r\,\cos\theta, r\,\operatorname{sen}\theta, z)$. La constante a se determina aplicando el teorema de Gauss; se obtiene $a = -\frac{q}{2\pi\epsilon_0}$, siendo q la densidad lineal de carga del hilo y ϵ_0 la permitividad del vacío. (Para el campo gravitatorio se obtiene el mismo resultado, donde ahora $a = 2\,G\,\mu$, siendo μ la densidad lineal de masa y G la constante de la gravitación de Newton.)

2.4. Interpretación geométrica del gradiente y la derivada

Definición 2.17. Si $f : \mathbb{R}^n \to \mathbb{R}$ es una función escalar y $c \in \mathbb{R}$, el *conjunto de nivel* asociado a c es el subconjunto $L_c(f)$ de $\mathbb{R}^n$ definido por

$$L_c(f) = f^{-1}(c) = \big\{x \in \mathbb{R}^n : f(x) = c\big\} \subset \operatorname{dom}(f).$$

Escribiremos sencillamente L_c, a menos que no esté claro por el contexto quién es la función f. Claramente, $L_c = \emptyset$ si $c \notin \operatorname{Im}(f)$. Veremos más adelante que si $a \in L_c$ ($\Leftrightarrow f(a) = c$), f es diferenciable en a y $\nabla f(a) \neq 0$, entonces L_c es una hipersuperficie "suave" (variedad diferenciable) en un entorno de a. En particular, si $n = 2$ ó $n = 3$ las hipersuperficies de nivel son respectivamente curvas y superficies de nivel. (A veces interesa considerar hipersuperficies en espacios de dimensión mayor que 3: por ejemplo, el *hiperboloide de masas* en relatividad es la hipersuperficie de nivel $L_{m^2} \subset \mathbb{R}^4$ asociada a la función $f(t, x, y, z) = t^2 - x^2 - y^2 - z^2$.)

Consideremos una curva diferenciable cualquiera $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$ cuya imagen esté contenida en la hipersuperficie L_c, es decir tal que $\gamma(t) \in L_c$, $\forall t \in I$, o equivalentemente $f(\gamma(t)) = c$, $\forall t \in I$. Si la curva pasa por el punto

$a \in L_c$, es decir si $\gamma(t_0) = a$ para algún $t_0 \in I$, diremos que el vector $\gamma'(t_0)$ es *tangente a* L_c *en* a. El conjunto de todos los vectores tangentes a L_c en a es el *espacio tangente a* L_c *en* a, denotado por $T_a(L_c)$. Con las hipótesis que estamos haciendo sobre f, puede probarse que $T_a(L_c)$ es un subespacio vectorial de $\mathbb{R}^n$ de dimensión $n-1$. Llamaremos *hiperplano tangente a* L_c *en* a al hiperplano $H_a(L_c)$ paralelo a $T_a(L_c)$ que pasa por a. En otras palabras, un punto $x \in H_a(L_c)$ si y sólo si $x - a \in T_a(L_c)$.

Sea $\gamma : I \to \mathbb{R}^n$ una curva en L_c que pase por $a \in L_c$ para $t = t_0$, y sea $v = \gamma'(t_0) \in T_a(L_c)$; como $F = f \circ \gamma$ es constante para todo $t \in I$ (es igual a c), aplicando (2.12) queda

$$0 = F'(t_0) = \nabla f(\gamma(t_0)) \cdot \gamma'(t_0) = \nabla f(a) \cdot v.$$

Por tanto, hemos probado que $\nabla f(a)$ es ortogonal a cualquier vector tangente a L_c en a. Equivalentemente, $\nabla f(a)$ *es ortogonal a la hipersuperficie* L_c *en* a. Esta es una de las propiedades geométricas fundamentales del gradiente.

Ejemplo 2.18. Sea $f : \mathbb{R}^n \to \mathbb{R}$ definida por $f(x) = |x|^2$, $\forall x \in \mathbb{R}^n$. Si $c > 0$, el conjunto de nivel L_c es una esfera de radio $\sqrt{c} > 0$. Si $a \in L_c$, es decir si $|a| = \sqrt{c}$, entonces f es diferenciable en a (por ser un polinomio) y se tiene $\nabla f(a) = 2a \neq 0$ (pues $|a| \neq 0$); por tanto, L_c es una hipersuperficie. Además, por lo visto anteriormente el radio vector a es perpendicular a la esfera L_c en a, lo cual es bien conocido en los casos $n = 2, 3$.

Sea de nuevo $a \in L_c(f)$, con f diferenciable en a y $\nabla f(a) \neq 0$. Vimos anteriormente que $T_a(L_c) \perp \nabla f(a)$, o lo que es lo mismo $T_a(L_c) \subset \operatorname{lin}\{\nabla f(a)\}^\perp$. Si admitimos que $T_a(L_c)$ es $(n-1)$-dimensional, entonces

$$T_a(L_c) = \operatorname{lin}\{\nabla f(a)\}^\perp.$$

Esto nos permite obtener fácilmente las ecuaciones implícitas del hiperplano tangente a L_c en a. En efecto,

$$x \in H_a(L_c) \iff x - a \in T_a(L_c) \iff (x - a) \cdot \nabla f(a) = 0;$$

por tanto,

$$H_a(L_c) = \big\{x \in \mathbb{R}^n : (x - a) \cdot \nabla f(a) = 0\big\}. \tag{2.15}$$

Equivalentemente, las ecuaciones implícitas de $H_a(L_c)$ son

$$\sum_{i=1}^{n} (x_i - a_i) \frac{\partial f}{\partial x_i}(a) = 0.$$

Las ecuaciones del espacio tangente $T_a(L_c)$ son simplemente

$$\sum_{i=1}^{n} x_i \frac{\partial f}{\partial x_i}(a) = 0.$$

Por ejemplo, las ecuaciones del hiperplano tangente a la esfera $S_r(0) = L_{r^2}(|x|^2)$ en un punto $a \in S_r(0)$ son $(x - a) \cdot a = 0$ ó, teniendo en cuenta que $a \in S_r(0) \Leftrightarrow |a|^2 = r^2$, $a \cdot x = r^2$. Por otra parte, el espacio tangente a $S_r(0)$ en a es sencillamente el complemento ortogonal de $\text{lin}\{a\}$, es decir $T_a\big(S_r(0)\big)$ es el subespacio de ecuación $a \cdot x = 0$.

Ejemplo 2.19. Hallar las ecuaciones del plano y del espacio tangente a la superficie de ecuación $3x^2y - xy^2z = 2$ en el punto $(1, 1, 1)$.

La superficie dada es la superficie de nivel $L_2(f)$ de la función $f : \mathbb{R}^3 \to \mathbb{R}$ dada por $f(x, y, z) = 3x^2y - xy^2z$ (o la $L_0(g)$ asociada a $g(x, y, z) = f(x, y, z) - 2$). La función f es diferenciable en todo $\mathbb{R}^3$, y el punto dado $a = (1, 1, 1)$ pertenece a dicha superficie, ya que efectivamente $f(1, 1, 1) = 3 - 1 = 2$. El gradiente de f en $(1, 1, 1)$ es

$$\nabla f(1, 1, 1) = (6xy - y^2z, 3x^2 - 2xyz, -xy^2)\big|_{(x,y,z)=(1,1,1)} = (5, 1, -1) \neq 0.$$

Por tanto, la ecuación del hiperplano tangente a la superficie en $(1, 1, 1)$ es

$$(5, 1, -1) \cdot (x - 1, y - 1, z - 1) = 0 \iff 5x + y - z = 5.$$

En cuanto al espacio tangente, sus ecuaciones son simplemente

$$\nabla f(1, 1, 1) \cdot (x, y, z) = 0 \iff (5, 1, -1) \cdot (x, y, z) = 0 \iff 5x + y - z = 0.$$

Un caso particular muy importante de todo lo anterior lo proporciona la gráfica de una función $g : \mathbb{R}^n \to \mathbb{R}$. Por definición, la *gráfica de g* es el subconjunto $\text{gr}(g)$ de $\mathbb{R}^{n+1}$ definido por

$$\text{gr}(g) = \{(x, g(x)) : x \in \mathbb{R}^n\} \subset \mathbb{R}^{n+1}.$$

Es evidente que $\text{gr}(f) = L_0(f)$, con $f : \mathbb{R}^{n+1} \to \mathbb{R}$ definida por

$$f(x_1, \ldots, x_n, x_{n+1}) \equiv f(x, x_{n+1}) = g(x) - x_{n+1}.$$

Sea $(a, g(a))$ un punto de $\text{gr}(g)$. Claramente, f es diferenciable en $(a, g(a))$ si y sólo si g es diferenciable en a, y en tal caso se tiene

$$\nabla f(a, g(a)) = (\nabla g(a), -1) \in \mathbb{R}^{n+1};$$

en particular, nótese que automáticamente se cumple que $\nabla f(a, g(a)) \neq 0$ $\forall a$ (¡aunque sea $\nabla g(a) = 0$!). Por tanto, si g es diferenciable en a automáticamente $\text{gr}(g)$ es una hipersuperficie. Además, por (2.15) un punto (x, x_{n+1}) pertenece al hiperplano tangente a $\text{gr}(f)$ en $(a, g(a))$ si y sólo si

$$(x - a, x_{n+1} - g(a)) \cdot (\nabla g(a), -1) = 0 \iff x_{n+1} - g(a) = \nabla g(a) \cdot (x - a).$$

Por lo tanto, la ecuación del hiperplano tangente a la gráfica de g en un punto $(a, g(a)) \in \text{gr}(g)$ es

$$x_{n+1} = g(a) + \nabla g(a) \cdot (x - a)$$

o también
$$x_{n+1} = g(a) + Dg(a) \cdot (x - a).$$
Esta es otra propiedad geométrica importante de $\nabla g(a)$, o mejor de $Dg(a)$; de hecho, la existencia de $Dg(a)$ (no meramente la de $\nabla g(a)$) es geométricamente equivalente a la existencia del hiperplano tangente a la gráfica de g en a.

Por último, otra propiedad importante de $\nabla f(a)$ está ligada a la velocidad de variación de $f : \mathbb{R}^n \to \mathbb{R}$ en a. En efecto, podemos representar cualquier dirección en $\mathbb{R}^n$ por un vector v con $|v| = 1$. La velocidad de variación de f en $a \in \mathbb{R}^n$ a lo largo de la recta que pasa por a en la dirección de v es $\frac{d}{dt}\big|_{t=0} f(a + t\,v) = D_v f(a)$. Si f es diferenciable en a, $D_v f(a) = Df(a) \cdot v = \nabla f(a) \cdot v = |\nabla f(a)| \cdot \cos\theta(v)$, siendo $\theta(v)$ el ángulo que forman los vectores $\nabla f(a)$ y v. Por consiguiente, si $\nabla f(a) \neq 0$ el máximo de $D_v f(a)$ es $|\nabla f(a)| > 0$, y se alcanza cuando $\theta(v) = 0$, es decir en la dirección $\nabla f(a)/|\nabla f(a)|$ marcada por el vector $\nabla f(a)$. En otras palabras, *si $\nabla f(a) \neq 0$ entonces $\nabla f(a)$ es un vector cuya dirección es aquella a lo largo de la cuál f crece más rápidamente en a, y cuyo módulo es igual a la máxima velocidad de crecimiento de f en a.* Análogamente, si $\nabla f(a) \neq 0$ el mínimo de $D_v f(a)$ es $-|\nabla f(a)| < 0$, y se alcanza en la dirección opuesta a $\nabla f(a)$.

2.5. Derivadas parciales de orden superior

Sea $f : \mathbb{R}^n \to \mathbb{R}^m$, y supongamos que existen las n derivadas parciales de primer orden $D_i f$, $1 \leq i \leq n$. (Por ejemplo, esto ocurrirá si f es diferenciable). Una *derivada parcial de segundo orden* de f es una derivada parcial (si existe) de una de las derivadas parciales de primer orden $D_i f$. En otras palabras, una derivada parcial de segundo orden de f es una función del tipo
$$D_j(D_i f) \equiv D_{ij} f \equiv \frac{\partial^2 f}{\partial x_j \partial x_i}, \qquad 1 \leq i, j \leq n$$
(nótese el orden de los subíndices en $D_{ij} f$). Obsérvese que, por las propiedades elementales de las derivadas parciales, $D_{ij} f$ existe si y sólo si existe $D_{ij} f_k$ para todo $k = 1, 2, \ldots, m$, y se tiene
$$D_{ij} f = (D_{ij} f_1, \ldots, D_{ij} f_m).$$
En principio, $D_{ij} f \neq D_{ji} f$ si $j \neq i$, y puede incluso existir una de estas derivadas parciales en algún punto sin que exista la otra. Sin embargo, veremos más adelante una condición suficiente que garantiza la igualdad de estas derivadas parciales cruzadas bajo hipótesis muy generales.

Definición 2.20. Si $U \subset \mathbb{R}^n$ y $f : \mathbb{R}^n \to \mathbb{R}^m$, diremos que f es *de clase 2 en U*, y escribiremos $f \in C^2(U)$, si existen y son continuas las n^2 derivadas parciales de segundo orden de f en U.

Evidentemente, $f \in C^2(U)$ si y sólo si las funciones componentes $f_i \in C^2(U)$, $\forall i = 1, 2, \ldots, m$. Si $f \in C^2(U)$, las n funciones $D_i f$ son por definición de clase 1 en U, y en particular son continuas (pues $C^1(U) \subset C^0(U)$). Por tanto, si f es de clase 2 en U automáticamente f es de clase 1 en U (y por tanto es diferenciable), es decir $C^2(U) \subset C^1(U) \subset C^0(U)$.

En general, una *derivada parcial de orden p de f* es una función del tipo

$$D_{i_1 i_2 \ldots i_p} f = D_{i_p}\Big(D_{i_{p-1}}\big(\ldots(D_{i_1} f)\ldots\big)\Big) \equiv \frac{\partial^p f}{\partial x_{i_p} \ldots \partial x_{i_1}},$$

con $1 \leq i_1, \ldots, i_p \leq n$. Como en el caso $p = 2$, es evidente que la derivada parcial de orden p $D_{i_1 i_2 \ldots i_p} f$ existe si y sólo si existen las derivadas parciales $D_{i_1 \ldots i_p} f_k$, $1 \leq k \leq m$, de las componentes de f, y

$$D_{i_1 i_2 \ldots i_p} f = \Big(D_{i_1 \ldots i_p} f_1, \ldots, D_{i_1 \ldots i_p} f_m\Big)$$

Definición 2.21. Se dice que $f : \mathbb{R}^n \to \mathbb{R}^m$ es *de clase p* en un abierto U de $\mathbb{R}^n$, y se escribe $f \in C^p(U)$, si existen y son continuas en U todas las derivadas parciales de orden p de f. La función f es *de clase* ∞ *en* U (denotado $f \in C^\infty(U)$) si f es de clase p en U para todo $p \in \mathbb{N}$.

En otras palabras, $C^\infty(U) = \cap_{p=0}^\infty C^p(U)$. Por ejemplo, las funciones racionales son de clase ∞ en el abierto en que no se anula el denominador. En particular, los polinomios son funciones C^∞ en todo $\mathbb{R}^n$.

Es inmediato probar que $C^{p+1}(U) \subset C^p(U)$ para todo $p \in \mathbb{N}$. En efecto, si f es de clase $p+1$ en U, como las derivadas parciales de orden $p+1$ de f son las derivadas parciales de primer orden de las derivadas parciales de orden p, todas las derivadas parciales de orden p de f son de clase 1, y por tanto continuas, en U. En consecuencia $f \in C^p(U)$, como queríamos demostrar.

Probaremos a continuación que si f es de clase 2 en un entorno de $x \in \mathbb{R}^n$ entonces las derivadas parciales cruzadas de segundo orden de f en x son iguales: $D_{ij} f = D_{ji} f$, $\forall i \neq j$. Esto será consecuencia del resultado ligeramente más general siguiente:

Teorema (lema de Schwarz)**.** *Sea $f : \mathbb{R}^2 \to \mathbb{R}$ una función definida en un entorno V de $(x_0, y_0) \in \mathbb{R}^2$. Si las derivadas parciales cruzadas $D_{12} f$ y $D_{21} f$ existen en V y son continuas en (x_0, y_0), entonces $D_{12} f(x_0, y_0) = D_{21} f(x_0, y_0)$.*

Demostración. Antes de empezar, obsérvese que la existencia de $D_{12} f = D_2(D_1 f)$ y $D_{21} f = D_1(D_2 f)$ en V garantiza la existencia de las derivadas parciales de primer orden de f en V (¡aunque no su continuidad!) La demostración se basa en estudiar el cociente incremental de segundo orden

$$A(u) = \frac{1}{u^2}\big(f(x_0 + u, y_0 + u) - f(x_0 + u, y_0) - f(x_0, y_0 + u) + f(x_0, y_0)\big).$$

Si $V = B_\delta(x_0, y_0)$, entonces $A(u)$ está definido si $2u^2 < \delta^2$; como en lo que sigue sólo estamos interesados en valores positivos de u, tomaremos u en el intervalo $(0, \delta/\sqrt{2})$. Fijado un u en dicho intervalo, sea

$$g(t) = f(t, y_0 + u) - f(t, y_0), \qquad x_0 - \epsilon < t < x_0 + u + \epsilon,$$

con $\epsilon > 0$ suficientemente pequeño para que $(t, y_0), (t, y_0 + u) \in V$ (basta tomar $\epsilon^2 < \delta^2 - 2u^2$). Entonces g es derivable en su dominio, pues

$$g'(t) = D_1 f(t, y_0 + u) - D_1 f(t, y_0)$$

existe si $t \in (x_0 - \epsilon, x_0 + u + \epsilon)$, al existir $D_1 f$ en V por hipótesis. En particular, g es continua en el intervalo $[x_0, x_0 + u]$, y diferenciable en el interior de dicho intervalo. Aplicando el teorema del valor medio a g en $[x_0, x_0 + u]$ se obtiene

$$A(u) = \frac{g(x_0 + u) - g(x_0)}{u^2} = \frac{D_1 f(\xi_u, y_0 + u) - D_1 f(\xi_u, y_0)}{u},$$

para algún ξ_u tal que $x_0 < \xi_u < x_0 + u$. A su vez, la función $h : \mathbb{R} \to \mathbb{R}$ definida mediante $h(t) = D_1 f(\xi_u, t)$ tiene derivada $h'(t) = D_{12} f(\xi_u, t)$ para todo $t \in (y_0 - \epsilon, y_0 + u + \epsilon)$ con $\epsilon > 0$ suficientemente pequeño (basta tomar de nuevo $\epsilon^2 < \delta^2 - 2u^2$), debido a que $D_{12} f$ por hipótesis existe en V. Aplicando de nuevo el teorema del valor medio a h en el intervalo $[y_0, y_0 + u]$ queda

$$A(u) = D_{12} f(\xi_u, \eta_u)$$

para algún $\eta_u \in (y_0, y_0 + u)$. Por otra parte, también podemos escribir

$$A(u) = \frac{G(y_0 + u) - G(y_0)}{u^2},$$

con $G(t) = f(x_0 + u, t) - f(x_0, t)$. Aplicando el teorema del valor medio a G en el intervalo $[y_0, y_0 + u]$ obtenemos

$$A(u) = \frac{D_2 f(x_0 + u, \tilde{\eta}_u) - D_2 f(x_0, \tilde{\eta}_u)}{u}$$

para algún $\tilde{\eta}_u \in (y_0, y_0 + u)$. Finalmente, aplicando otra vez el teorema del valor medio a $D_2 f(\cdot, \tilde{\eta}_u)$ en el intervalo $[x_0, x_0 + u]$, lo cual puede hacerse por la existencia de $D_{21} f$ en V, queda

$$A(u) = D_{21} f(\tilde{\xi}_u, \tilde{\eta}_u)$$

para algún $\tilde{\xi}_u \in (x_0, x_0 + u)$. En definitiva, hemos probado que para todo $u \in (0, \delta/\sqrt{2})$ existen dos puntos $(\xi_u, \eta_u), (\tilde{\xi}_u, \tilde{\eta}_u) \in (x_0, x_0 + u) \times (y_0, y_0 + u) \subset V$ tales que

$$A(u) = D_{12} f(\xi_u, \eta_u) = D_{21} f(\tilde{\xi}_u, \tilde{\eta}_u).$$

Si ahora hacemos u tender a $0+$ se tiene $(\xi_u, \eta_u) \to (x_0, y_0)$, $(\tilde{\xi}_u, \tilde{\eta}_u) \to (x_0, y_0)$, y por la continuidad de $D_{12}f$ y $D_{21}f$ en $(x_0, y_0) \in V$ queda

$$\lim_{u \to 0+} A(u) = D_{12}f(x_0, y_0) = D_{21}f(x_0, y_0).$$

Q.E.D.

Corolario 2.22. *Sea $f : \mathbb{R}^n \to \mathbb{R}^m$, y sea $a \in \mathbb{R}^n$. Si existen las derivadas parciales $D_{ij}f$ y $D_{ji}f$ (con $i \neq j$) en un entorno de a, y son continuas en dicho punto, entonces $D_{ij}f(a) = D_{ji}f(a)$.*

Demostración. Sin pérdida de generalidad, podemos suponer que $i < j$. Entonces basta aplicar el lema de Schwarz a cada una de las n funciones $g_k : \mathbb{R}^n \to \mathbb{R}$ definidas por

$$g_k(x, y) = f_k(a_1, \ldots, a_{i-1}, x, a_{i+1}, \ldots, a_{j-1}, y, a_{j+1}, \ldots, a_n).$$

Q.E.D.

En particular, del corolario anterior se sigue que si f es de clase 2 en un abierto $U \subset \mathbb{R}^n$ entonces todas las derivadas parciales cruzadas de f son iguales en U:

$$D_{ij}f(x) = D_{ji}f(x), \qquad \forall x \in U, \quad 1 \leq i, j \leq n.$$

Más generalmente, si f es de clase $p > 1$ se tiene el resultado siguiente:

Proposición 2.23. *Si $f : \mathbb{R}^n \to \mathbb{R}^m$ es de clase $p > 1$ en un abierto $U \subset \mathbb{R}^n$, entonces todas las derivadas parciales cruzadas de orden $q \leq p$ de f en U son iguales:*

$$D_{i_1 \ldots i_q} f(x) = D_{i_{\sigma(1)} \ldots i_{\sigma(q)}} f(x), \qquad \forall x \in V,$$

para toda permutación σ de $\{1, 2, \ldots, q\}$.

Para finalizar, mencionemos sin demostración dos condiciones suficientes que garantizan la igualdad de las derivadas parciales cruzadas $D_{ij}f(a)$ y $D_{ji}f(a)$ de una función $f : \mathbb{R}^n \to \mathbb{R}^m$ en un punto $a \in \mathbb{R}^n$:

I) Si $D_i f$ y $D_j f$ son diferenciables en a, entonces $D_{ij}f(a) = D_{ji}f(a)$

II) Si $D_i f$, $D_j f$ y (por ejemplo) $D_{ij}f$ son continuas en un entorno de a, entonces existe $D_{ji}f(a)$, y $D_{ij}f(a) = D_{ji}f(a)$

Capítulo 3

Funciones inversas e implícitas

3.1. Teorema de la función inversa

Sea $f : \mathbb{R}^n \to \mathbb{R}^n$ una función cuyo dominio contiene un abierto $U \subset \mathbb{R}^n$. Si f es *inyectiva* (uno a uno) en U, es decir si

$$x, y \in U,\ f(x) = f(y) \Longrightarrow x = y$$

entonces la restricción de f a U es biyectiva considerada como una función $f|_U : U \to f(U)$, y por tanto es invertible: $\exists (f|_U)^{-1} : f(U) \to U$ tal que

$$f|_U \circ (f|_U)^{-1} = I_{f(U)}, \quad (f|_U)^{-1} \circ f|_U = I_U. \tag{3.1}$$

En otras palabras,

$$x \in U,\ y = f(x) \iff x = (f|_U)^{-1}(y).$$

Se dice entonces que la función f es *globalmente invertible* en U, siendo la inversa de f en U la función $(f|_U)^{-1} : f(U) \to U$. Se dice que f es *localmente invertible* en un punto $a \in \mathbb{R}^n$ si existe un entorno $B_r(a)$ de a tal que f es globalmente invertible en $B_r(a)$. Análogamente, f es localmente invertible en un abierto U si es localmente invertible en cualquier punto de dicho abierto.

Ejemplo 3.1. Sea $f : \mathbb{R} \to \mathbb{R}$ la función dada por $f(x) = x^2$, $\forall x \in \mathbb{R}$. Entonces f no es globalmente invertible en $\mathbb{R}^2$, es globalmente invertible en $\mathbb{R}_+ = \{x : x > 0\}$ y en $\mathbb{R}_- = \{x : x < 0\}$, y es localmente invertible en todo punto salvo en $x = 0$.

Evidentemente, si f es globalmente invertible en U entonces f es localmente invertible en U, pero el recíproco no tiene por qué ser cierto, como demuestra el siguiente ejemplo:

Ejemplo 3.2. Sea $f : \mathbb{R}^2 \to \mathbb{R}^2$ la función definida por $f(x, y) = (e^x \cos y, e^x \operatorname{sen} y)$, $\forall (x, y) \in \mathbb{R}^2$. Como f es periódica en y, es decir

$$f(x, y) = f(x, y + 2k\pi), \quad \forall k \in \mathbb{Z},$$

está claro que f no es globalmente invertible en $U = \mathbb{R}^2$. Sin embargo, f es localmente invertible en cualquier punto $(a, b) \in \mathbb{R}^2$, pues es globalmente invertible en la banda $V_b = \mathbb{R} \times (b - \pi, b + \pi)$. En efecto, $f(V_b) = \mathbb{R}^2 - S_b$, siendo S_b la semirrecta cerrada que forma un ángulo $b + \pi$ con el eje Ox positivo. Resolviendo las ecuaciones

$$u = e^x \cos y, \quad v = e^x \operatorname{sen} y$$

con $(u, v) \in f(V_b)$ queda

$$x = \frac{1}{2} \log(u^2 + v^2), \quad y = b + \arg_\pi(u + i\,v).$$

Como f es globalmente invertible en el abierto V_b que contiene a (a, b), f es globalmente invertible en una bola centrada en (a, b), como queríamos demostrar.

Si la función $f : \mathbb{R}^n \to \mathbb{R}^n$ es suficientemente sencilla, se puede determinar algunas veces si f es invertible en un abierto U y construir su inversa directamente, como en el ejemplo anterior. Hay que resolver la ecuación (n ecuaciones escalares)

$$y = f(x)$$

para cada $y \in f(U)$; si dicha ecuación tiene una única solución $x = g(y)$, $\forall y \in f(U)$, entonces f es invertible en U y $(f|_U)^{-1} = g$. Sin embargo, en general es imposible (o muy engorroso) en la práctica proceder directamente. Interesa por tanto tener algún criterio para poder averiguar si f es, por lo menos, localmente invertible en un abierto, y hallar las propiedades de f^{-1} sin necesidad de calcularla; esto se consigue con el teorema de la función inversa, que veremos más adelante.

Empecemos considerando el caso $n = 1$, que es particularmente sencillo (y en parte conocido). En primer lugar, es fácil probar que una aplicación $f : \mathbb{R} \to \mathbb{R}$ es inyectiva (y por tanto invertible globalmente) en un *intervalo* abierto $U \subset \mathbb{R}$ si y sólo si f es *estrictamente monótona* en dicho intervalo. Si además f es diferenciable en U, entonces es bien conocido que la función inversa $g = (f|_U)^{-1}$ es derivable en un punto $y = f(x)$ si y sólo si $f'(x) \neq 0$, y se tiene $g'(y) = 1/f'(x)$. En particular, si f es derivable y f' tiene signo constante en todos los puntos de un intervalo abierto $U \subset \mathbb{R}$ entonces f es globalmente invertible en U y $(f|_U)^{-1}$ es derivable en $f(U)$.

En el caso $n > 2$, la situación es bastante más complicada. Antes de pasar a este caso, sin embargo, es importante notar que f puede ser globalmente invertible en un abierto U y diferenciable (o incluso de clase ∞) en dicho abierto sin que $(f|_U)^{-1}$ sea ni siquiera diferenciable en $f(U)$:

Ejemplo 3.3. Sea $f : \mathbb{R} \to \mathbb{R}$ definida por $f(x) = x^3$, $\forall x \in \mathbb{R}$. Entonces f es claramente invertible globalmente y de clase C^∞ en $\mathbb{R}$. La función inversa de f está dada por $f^{-1}(y) = y^{1/3}$, $\forall y \in f(\mathbb{R}) = \mathbb{R}$, que no es derivable en $y = 0$.

Definición 3.4. Sean U y V abiertos de $\mathbb{R}^n$. Una aplicación $f : U \subset \mathbb{R}^n \to V \subset \mathbb{R}^n$ es un *difeomorfismo* de U en V si $V = f(U)$, f es globalmente invertible en U, y tanto f como $f^{-1} : V \to U$ son diferenciables (respectivamente en U y V). Bajo las mismas hipótesis, f es un *difeomorfismo de clase C^p* si f es un difeomorfismo, y tanto f como f^{-1} son de clase C^p (en U y V, respectivamente). Finalmente, f es un *difeomorfismo local* (difeomorfismo local de clase C^p) en un punto $a \in \mathbb{R}^n$ si f es un difeomorfismo (difeomorfismo de clase C^p) en un entorno $B_r(a)$ de a.

Por la Proposición 2.7, un difeomorfismo es automáticamente un homeomorfismo, pero es claro que el recíproco no es cierto en general. Por ejemplo, la función $x \mapsto x^3$ es un homeomorfismo de $\mathbb{R}$ en $\mathbb{R}$, pero no es un difeomorfismo.

Sea $f : U \subset \mathbb{R}^n \to V = f(U) \subset \mathbb{R}^n$ (con U, V abiertos) derivable y globalmente invertible en U; si $x \in U$, y f^{-1} es derivable en $y = f(x) \in V$, derivando las relaciones (3.1) se obtiene (regla de la cadena)

$$Df(x) \cdot D(f^{-1})\big(f(x)\big) = D(f^{-1})\big(f(x)\big) \cdot Df(x) = I,$$

siendo $I : \mathbb{R}^n \to \mathbb{R}^n$ la identidad. Por tanto, $Df(x)$ es una aplicación lineal invertible, con inversa

$$Df(x)^{-1} = D(f^{-1})\big(f(x)\big),$$

o equivalentemente

$$Jf(x)^{-1} = J(f^{-1})\big(f(x)\big)$$

Por tanto, si f es globalmente invertible y diferenciable en U, una condición *necesaria* para que f^{-1} sea derivable en $f(U)$ es que $Df(x)$ sea invertible $\forall x \in U$, es decir que

$$\det Df(x) = \det Jf(x) \neq 0, \qquad \forall x \in U.$$

En particular, se tiene:

Proposición 3.5. *Si $f : U \subset \mathbb{R}^n \to V \subset \mathbb{R}^n$ (con U, V abiertos) es un difeomorfismo, entonces $\det Df(x) \neq 0$, $\forall x \in U$.*

Si $f : \mathbb{R}^n \to \mathbb{R}^n$ es diferenciable en U, la no anulación de $\det Df$ en U no garantiza que f sea globalmente invertible. Por ejemplo, si $f : \mathbb{R}^2 \to \mathbb{R}^2$ es la aplicación del ejemplo 3.2, entonces

$$\det Df(x,y) = \det Jf(x,y) = \begin{vmatrix} e^x \cos y & -e^x \operatorname{sen} y \\ e^x \operatorname{sen} y & e^x \cos y \end{vmatrix} = e^{2x} \neq 0,\ \forall (x,y) \in \mathbb{R}^2,$$

aunque f no es globalmente invertible en $\mathbb{R}^2$. Sin embargo, nótese que f si es localmente invertible en $\mathbb{R}^2$. El teorema que vamos a enunciar a continuación (teorema de la función inversa) garantiza que esto es cierto en general, es decir que una función de clase 1 con determinante jacobiano $\det Df$ no nulo en un punto es localmente invertible en dicho punto. Por último, nótese que si f es de clase C^1 en un entorno de a y $\det Df(a) \neq 0$, entonces existe un entorno de a en el cual $\det Df$ no se anula. En efecto, $\det Df : \mathbb{R}^n \to \mathbb{R}$ es una función polinómica de las n^2 derivadas parciales de f, que a su vez son funciones continuas en un entorno de a por hipótesis. Por tanto $\det Df$ es continua en dicho entorno (composición de funciones continuas), de donde se sigue que el conjunto de puntos donde $\det Df$ no se anula es abierto. Esto prueba nuestra afirmación.

Teorema (de la función inversa)**.** *Sea $f : \mathbb{R}^n \to \mathbb{R}^n$ una función de clase C^p (con $1 \leq p \leq \infty$) en un entorno de un punto $a \in \mathbb{R}^n$. Si $\det Df(a) \neq 0$, existe un entorno U de a tal que $f|_U : U \to f(U)$ es un difeomorfismo de clase C^p.*

En otras palabras, $f(U)$ es abierto, f es globalmente invertible en U y $(f|_U)^{-1} : f(U) \to U$ es de clase C^p. Además, por la observación anterior la derivada de $(f|_U)^{-1}$ en un punto $y \in f(U)$ es simplemente $Df\big(f^{-1}(y)\big)^{-1}$.

El teorema de la función inversa es uno de los pilares del Análisis. Nosotros omitiremos su demostración, que es relativamente complicada, y nos limitaremos a deducir algunas de sus consecuencias más importantes. Nótese que el que la condición $\det Df(a) \neq 0$ garantice la invertibilidad local de f en a se puede justificar heurísticamente observando que si f es diferenciable en a entonces

$$f(x) = f(a) + Df(a) \cdot (x - a) + o(|x - a|),$$

siendo $\det Df(a) \neq 0$ la condición necesaria y suficiente para que la aproximación lineal a f

$$x \mapsto f(a) + Df(a) \cdot (x - a)$$

sea invertible.

Proposición 3.6. *Si $f : \mathbb{R}^n \to \mathbb{R}^n$ es de clase C^1 en un abierto $U \subset \mathbb{R}^n$ y $\det Df(x) \neq 0$ $\forall x \in U$, entonces para todo $V \subset U$ abierto en $\mathbb{R}^n$ $f(V)$ es abierto.*

Demostración. En efecto, sea $a \in V \subset U$. Por el teorema de la función inversa, $f|_W$ es un difeomorfismo de un entorno W de a en $f(W)$. Podemos suponer (achicando W si es preciso) que $W \subset V$. Entonces $f(W) = f|_W(W) \subset f(V)$ es un abierto (pues los difeomorfismos son homeomorfismos) que contiene a $f(a)$. Al ser $f(W)$ abierto y $f(a) \in W$, hay un entorno X de $f(a)$ contenido en $f(W)$, y por tanto en $f(V)$. Luego $f(V)$ contiene un entorno X de uno cualquiera de sus puntos $f(a)$, es decir $f(V)$ es abierto. *Q.E.D.*

Ejemplo 3.7. Sea $f : \mathbb{R}^2 \to \mathbb{R}^2$ definida por $f(s,t) = (s^2+t^2, 2st)$, $\forall (s,t) \in \mathbb{R}^2$. Si ponemos $(x,y) = f(s,t)$, como

$$x + y = (s+t)^2, \quad x - y = (s-t)^2,$$

entonces $f(\mathbb{R}^2)$ está contenido en el cuadrante cerrado $C = \big\{(x,y) : x+y \geq 0, x-y \geq 0\big\}$. Por otra parte, si $(x,y) \in C$ entonces

$$s + t = \epsilon_1 \sqrt{x+y}, \quad s - t = \epsilon_2 \sqrt{x-y}$$

con $\epsilon_{1,2} = \pm 1$ independientes. Por tanto, $f(\mathbb{R}^2) = C$, y además todo punto $(x,y) \in C$ tiene en general cuatro preimágenes dadas por

$$s = \frac{1}{2}\Big(\epsilon_1\sqrt{x+y} + \epsilon_2\sqrt{x-y}\Big), \quad t = \frac{1}{2}\Big(\epsilon_1\sqrt{x+y} - \epsilon_2\sqrt{x-y}\Big).$$

Luego $f : \mathbb{R}^2 \to C$ no es globalmente invertible en $\mathbb{R}^2$. Sin embargo, f es globalmente invertible en cada uno de los cuatro cuadrantes abiertos de ecuaciones $\epsilon_1(s+t) > 0$, $\epsilon_2(s-t) > 0$ correspondientes a las cuatro elecciones posibles de $\epsilon_{1,2}$. Por ejemplo, tomando $\epsilon_1 = -\epsilon_2 = +1$ obtenemos el cuadrante

$$U = \big\{(s,t) : -t < s < t\big\} = \big\{(s,t) : |s| < t\big\}$$

y $f : U \to \mathring{C}$ es globalmente invertible, con inversa $g = (f|_U)^{-1}$ dada por

$$g(x,y) = \frac{1}{2}\Big(\sqrt{x+y} - \sqrt{x-y}, \sqrt{x+y} + \sqrt{x-y}\Big).$$

Claramente, f es localmente invertible en todo punto (s,t) fuera de las rectas $t = \pm s$, ya que dicho punto está en uno de los los cuatro cuadrantes abiertos mencionados anteriormente en que f es invertible. Del mismo modo, es fácil ver que f no es localmente invertible en ningún punto de las dos rectas $t = \pm s$. La primera de estas dos últimas conclusiones se puede verificar fácilmente aplicando el teorema de la función inversa. En efecto, es claro que f es derivable en todo $\mathbb{R}^2$, con matrix jacobiana

$$Jf(s,t) = \begin{pmatrix} 2s & 2t \\ 2t & 2s \end{pmatrix},$$

y por tanto

$$\det Df(t) = 4(s^2 - t^2).$$

Luego, por el teorema de la función inversa, f es un difeomorfismo local (y por tanto es localmente invertible) si $s^2 \neq t^2$, es decir para $t \neq \pm s$. Nótese, sin embargo, que el teorema de la función inversa no dice nada acerca de la invertibilidad de f en un entorno de los puntos en que $\det Df$ se anula, a excepción de que f^{-1} (si existe) no puede ser diferenciable en las imágenes de dichos puntos. Por ejemplo, en este caso f no es invertible ni siquiera localmente en los puntos en que $\det Df$ se anula, mientras que si f es la función $x \mapsto x^3$ f es globalmente invertible.

3.2. Teorema de la función implícita

Sea $\Phi = (\Phi_1, \dots, \Phi_m) : \mathbb{R}^{n+m} \to \mathbb{R}^m$, y representemos por (x, y), con $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, un punto cualquiera de $\mathbb{R}^{n+m}$. Si Φ está definida y es de clase C^p en un abierto $U \subset \mathbb{R}^{n+m}$, nuestro objetivo es estudiar el conjunto definido por la relación $\Phi = 0$ en U, es decir

$$\big\{(x, y) \in U : \Phi(x, y) = 0\big\}.$$

Nótese que la ecuación "vectorial"$\Phi(x, y) = 0$ es en realidad equivalente a las m ecuaciones escalares

$$\Phi_i(x, y) = 0, \qquad 1 \le i \le m. \tag{3.2}$$

Intutitivamente, las m ecuaciones anteriores deberían permitir, al menos bajo ciertas condiciones sobre Φ, despejar las m variables $y = (y_1, \dots .y_m)$ en función de las restantes variables $x = (x_1, \dots, x_n)$. En otras palabras, debería exisitir una función $f : \mathbb{R}^n \to \mathbb{R}^m$ definida en un abierto V_n de $\mathbb{R}^n$ de forma que

$$(x, y) \in U,\ \Phi(x, y) = 0 \iff x \in V_n,\ y = f(x). \tag{3.3}$$

Definición 3.8. Si $\Phi : \mathbb{R}^{n+m} \to \mathbb{R}^m$ está definida en un abierto $U \subset \mathbb{R}^{n+m}$, diremos que la relación $\Phi(x, y) = 0$ define implícitamente a y como función de x en U si existe un abierto $V_n \subset \mathbb{R}^n$ y una función $f : V_n \to \mathbb{R}^m$ tal que se verifica (3.3).

Nótese que (3.3) es equivalente a

$$\big\{(x,y) \in U : \Phi(x,y) = 0\big\} = \big\{(x, f(x)) : x \in V_n\big\};$$

además, esta igualdad o directamente (3.3) implican que

$$\Phi(x, f(x)) = 0, \qquad \forall x \in V_n.$$

En otras palabras, $\Phi(x,y) = 0$ define implícitamente a y como función de x en U si el subconjunto $\Phi = 0$ de $U \subset \mathbb{R}^{n+m}$ es la *gráfica* de una función $f : V_n \subset \mathbb{R}^n \to \mathbb{R}^m$ (cf. la definición de gráfica para funciones escalares del capítulo anterior). En particular, como la gráfica de una función no puede contener puntos de la forma (x, y_1), (x, y_2) con $y_1 \neq y_2$, si el conjunto $\Phi = 0$ no tiene esta propiedad la relación $\Phi(x,y) = 0$ no define a y como función de x en U.

Ejemplo 3.9. Consideremos el caso $n = m = 1$, y sea $\Phi : \mathbb{R}^2 \to \mathbb{R}$ definida por $\Phi(x,y) = x^2 + y^2 - 1$. Claramente, $\Phi(x,y) = 0$ no define a y como función de x en $\mathbb{R}^2$, pues dado un punto (x,y) que satisface $\Phi(x,y) = 0$ el punto $(x,-y)$ también cumple dicha relación. Si U_+ es el semiplano $y > 0$, entonces $\Phi(x,y) = 0$ define a y como función de x en U_+; en efecto,

$$y > 0,\ x^2 + y^2 - 1 = 0 \iff y = \sqrt{1 - x^2},\ -1 < x < 1.$$

En otras palabras, el abierto $V_n \equiv V_1$ es el intervalo $(-1,1)$, y $f : (-1,1) \to \mathbb{R}$ está dada por $f(x) = \sqrt{1-x^2}$. Análogamente, $\Phi(x,y) = 0$ define implícitamente a y como función de x en el semiplano $U_- = -U_+$, siendo $V_1 = (-1,1)$ y $f(x) = -\sqrt{1-x^2}$. Sin embargo, y no está definida implícitamente como función de x en ningún entorno de los puntos $(-1,0)$ ó $(1,0)$, pues en cualquier entorno de dichos puntos hay puntos de la forma (x,y), $(x,-y)$ con $y \neq 0$ que satisfacen $\Phi(x,y) = 0$. También se puede estudiar cuando $\Phi(x,y) = 0$ define implícitamente a x como función de y. Como antes, se demuestra fácilmente que esto ocurre en los semiplanos abiertos $x > 0$ ó $x < 0$, pero no en todo $\mathbb{R}^2$ ni en ningún entorno de los puntos $(0, \pm 1)$.

En el ejemplo anterior, los puntos en un entorno de los cuales y no está definida implícitamente como función de x son precisamente aquellos puntos del conjunto $\Phi = 0$ en que $\partial\Phi/\partial y$ se anula. En el caso general, si $\Phi = (\Phi_1, \dots, \Phi_m) : \mathbb{R}^{n+m} \to \mathbb{R}^m$ la matriz jacobiana de Φ es una matriz $m \times (n+m)$ que escribiremos por cajas como sigue:

$$Jf(x,y) = \begin{pmatrix} \dfrac{\partial\Phi}{\partial x}(x,y) & \dfrac{\partial\Phi}{\partial y}(x,y) \end{pmatrix}$$

siendo $\frac{\partial \Phi}{\partial y}(x,y)$ la matriz cuadrada de orden m dada por

$$\frac{\partial \Phi}{\partial y}(x,y) = \begin{pmatrix} \dfrac{\partial \Phi_1}{\partial y_1} & \cdots & \dfrac{\partial \Phi_1}{\partial y_m} \\ \vdots & & \vdots \\ \dfrac{\partial \Phi_m}{\partial y_1} & \cdots & \dfrac{\partial \Phi_m}{\partial y_m} \end{pmatrix},$$

y análogamente $\frac{\partial \Phi}{\partial x}(x,y)$. El análogo de la condición $\frac{\partial \Phi}{\partial y}(x,y) \neq 0$ en el caso $n = m = 1$ es, evidentemente, la desigualdad $\det \frac{\partial \Phi}{\partial y}(x,y) \neq 0$. El teorema siguiente garantiza que si Φ es suficientemente regular, y $\det \frac{\partial \Phi}{\partial y}(x_0,y_0) \neq 0$ en un punto (x_0,y_0) del conjunto $\Phi = 0$, entonces $\Phi(x,y) = 0$ define implícitamente a y como función de x en un entorno de (x_0,y_0):

Teorema (de la función implícita)**.** *Sea $\Phi : \mathbb{R}^{n+m} \to \mathbb{R}^m$ una función de clase C^p en un entorno de un punto $(x_0,y_0) \in \mathbb{R}^n \times \mathbb{R}^m = \mathbb{R}^{n+m}$. Si $\frac{\partial \Phi}{\partial y}(x_0,y_0) \neq 0$, entonces la relación $\Phi(x,y) = 0$ define implícitamente a y como una función f de x en un entorno de (x_0,y_0), siendo además f de clase C^p en su dominio.*

Demostración. Hay que probar que existe un entorno $U \subset \mathbb{R}^{n+m}$ de (x_0,y_0), un abierto $V_n \subset \mathbb{R}^n$ y una función $f : V_n \to \mathbb{R}^m$ de clase C^p en V_n tal que se verifica (3.3). Para ello recurriremos al teorema de la función inversa, aplicado a la función auxiliar $F : U \subset \mathbb{R}^{n+m} \to \mathbb{R}^{n+m}$ definida por

$$F(x,y) = (x, \Phi(x,y)), \qquad \forall (x,y) \in U.$$

Claramente F es de clase C^p en el entorno W de (x_0,y_0) en que Φ es de clase C^p, al serlo sus componentes; las primeras n componentes son aplicaciones lineales, mientras que las m últimas son las componentes de Φ. Además, la matriz jacobiana de F tiene la estructura

$$JF(x,y) = \begin{pmatrix} \mathbb{1}_n & 0 \\ \frac{\partial \Phi}{\partial x}(x,y) & \frac{\partial \Phi}{\partial y}(x,y) \end{pmatrix},$$

por lo que

$$\det DF(x,y) = \det JF(x,y) = \det \frac{\partial \Phi}{\partial y}(x,y)$$

no se anula en (x_0,y_0), por hipótesis. Por tanto, podemos aplicar el teorema de la función inversa a F en (x_0,y_0): existe un entorno U de (x_0,y_0) en $\mathbb{R}^{n+m}$ tal que F es globalmente invertible en U, con inversa $G : F(U) \to U$ de clase C^p en el abierto $F(U)$. Por la forma de F, es claro que G tiene la estructura

$$G(u,v) = (u, g(u,v)), \qquad \forall (u,v) \in F(U) \subset \mathbb{R}^{n+m},$$

con $g : F(U) \subset \mathbb{R}^{n+m} \to \mathbb{R}^m$ de clase C^p. Entonces se tiene

$$\begin{aligned}(x, y) \in U,\ \Phi(x, y) = 0 &\iff (x, y) \in U,\ F(x, y) = (x, 0)\\ &\iff (x, 0) \in F(U),\ G(x, 0) = (x, y)\\ &\iff (x, 0) \in F(U),\ g(x, 0) = y.\end{aligned}$$

Si $i : \mathbb{R}^n \to \mathbb{R}^{n+m}$ es la función lineal definida por $i(x) = (x, 0)$ (que identifica $\mathbb{R}^n$ con el subespacio $\big\{(x, 0) : x \in \mathbb{R}^n\big\}$ de $\mathbb{R}^{n+m}$) entonces

$$(x, 0) \in F(U) \iff i(x) \in F(U) \iff x \in V_n \equiv i^{-1}\big(F(U)\big),$$

con V_n abierto por serlo $F(U)$ y ser i continua (lineal). Por tanto, hemos probado que

$$(x, y) \in U,\ \Phi(x, y) = 0 \iff x \in V_n,\ y = f(x)$$

siendo U un entorno de (x_0, y_0) en $\mathbb{R}^{n+m}$, V_n un abierto de $\mathbb{R}^n$ y $f : V_n \to \mathbb{R}^m$ dada por $f(x) = g(x, 0)$. Por último, la función f es de clase C^p en V_n, pues

$$\frac{\partial^q f}{\partial x_{i_1} \dots \partial x_{i_q}} = \frac{\partial^q g}{\partial x_{i_1} \dots \partial x_{i_q}} \circ i, \qquad 1 \le q \le p,$$

e $i(V_n) \subset F(U)$. *Q.E.D.*

Cuando el teorema de la función implícita es aplicable, puede calcularse el valor de las derivadas parciales de orden $\le p$ de la función implícita f en el punto x_0 sin necesidad de tener que determinar f explícitamente. Para ello, basta derivar la identidad

$$\Phi(x, f(x)) = 0, \qquad \forall x \in V_n,$$

respecto de una cualquiera de las coordenadas x_i de x, obteniéndose

$$0 = \frac{\partial \Phi_k}{\partial x_i}(x, f(x)) + \sum_{j=1}^{m} \frac{\partial \Phi_k}{\partial y_j}(x, f(x)) \frac{\partial f_j}{\partial x_i}(x), \qquad 1 \le i \le n,\ 1 \le k \le m. \tag{3.4}$$

Las nm igualdades escalares anteriores son equivalente a la igualdad matricial

$$\frac{\partial \Phi}{\partial x}(x, f(x)) + \frac{\partial \Phi}{\partial y}(x, f(x)) \cdot Jf(x) = 0;$$

como, por hipótesis, el determinante de $\frac{\partial \Phi}{\partial y}(x_0, y_0)$ no se anula, sustituyendo $x = x_0$ en la igualdad anterior y despejando $Jf(x_0)$ se obtiene finalmente

$$Jf(x_0) = -\Big(\frac{\partial \Phi}{\partial y}(x_0, y_0)\Big)^{-1} \cdot \frac{\partial \Phi}{\partial x}(x_0, y_0).$$

Si $p \geq 2$, para calcular las derivadas segundas de f en x_0 derivamos (3.4) una vez más y evaluamos el resultado para $x = x_0$, obteniendo una expresión del tipo

$$\sum_{j=1}^{m} \frac{\partial \Phi_k}{\partial y_j}(x_0, y_0) \frac{\partial^2 f_j}{\partial x_l \partial x_i}(x_0) + \cdots = 0,$$

donde los términos omitidos dependen de $Jf(x_0)$ y de las derivadas parciales de orden ≤ 2 de Φ en (x_0, y_0), que son conocidas. De nuevo por la invertibilidad de $\frac{\partial \Phi}{\partial y}(x_0, y_0)$, se pueden despejar en la ecuación anterior las derivadas parciales de segundo orden de f en x_0, y así sucesivamente. En la práctica, no es necesario recordar las fórmulas anteriores, sino que es preferible derivar directamente las igualdades $\Phi_k(x, f(x)) = 0$.

Nota. En los puntos (x_0, y_0) en que $\det \frac{\partial \Phi}{\partial y}(x_0, y_0) = 0$, no puede afirmarse nada acerca de si la relación $\Phi(x, y) = 0$ define o no a y como función de x. Por ejemplo, si $\Phi(x, y) = y^2 - x$ entonces $\Phi(x, y) = 0$ no define a y como función de x en un entorno de $(0, 0)$, mientras que si $\Phi(x, y) = y^3 - x$ la relación $\Phi(x, y) = 0$ define globalmente a y como función de x.

Ejemplo 3.10. Consideremos las relaciones

$$\begin{cases} \Phi_1(x, y, z, u) = xy^2 + xzu + yu^2 - 3 = 0 \\ \Phi_2(x, y, z, u) = u^3yz + 2xy - 2u^2v^2 - 1 = 0 \end{cases}$$

En primer lugar, veamos en qué puntos puede asegurarse que las relaciones anteriores definen a u y v implícitamente como funciones de (x, y, z). Como Φ es evidentemente de clase C^∞ en $\mathbb{R}^5$, basta imponer la condición

$$\det \frac{\partial(\Phi_1, \Phi_2)}{\partial(u, v)} \neq 0,$$

es decir

$$\begin{vmatrix} xz + 2yu & 0 \\ 3u^2yz - 4uv^2 & -4u^2v \end{vmatrix} = -4u^2v(xz + 2yu) \neq 0.$$

Por el teorema de la función implícita, $\Phi = 0$ define implícitamente a (u, v) como función de clase C^∞ de (x, y, z) en un entorno de cualquier punto (x, y, z, u, v) que satisfaga las condiciones

$$\Phi(x, y, z, u, v) = 0, \qquad uv(xz + 2yu) \neq 0;$$

por ejemplo, esto ocurre en un entorno del punto $(1, 1, 1, 1, 1)$. Calculemos a continuación $\partial v / \partial y$ en el punto $(1, 1, 1)$. Para ello, basta derivar respecto de y las relaciones $\Phi_i = 0$, $1 \leq i \leq 2$, obteniéndose

$$\begin{aligned} 2xy + u^2 + (xz + 2yu)u_y &= 0 \\ u^3z + 2x + (3u^2yz - 4uv^2)u_y - 4u^2vv_y &= 0 \end{aligned}$$

de donde

$$\begin{aligned} 1 + u_y(1,1,1) &= 0 \\ 3 - u_y(1,1,1) - 4v_y(1,1,1) &= 0 \end{aligned}$$

Despejando u_y y v_y de estas relaciones se obtiene

$$u_y(1,1,1) = -1, \qquad v_y(1,1,1) = 1.$$

Para calcular, por ejemplo, v_{yy}, basta derivar las relaciones anteriores respecto de y:

$$\begin{aligned} 2x + 2uu_y + (2u + 2yu_y)u_y + (xz + 2yu)u_{yy} &= 0 \\ 3u^2zu_y + \left[3u^2z + (6uyz - 4v^2)u_y - 8uvv_y\right] u_y & \\ + (3u^2yz - 4uv^2)u_{yy} - 4(2uvu_y + u^2v_y)v_y - 4u^2vv_{yy} &= 0. \end{aligned}$$

Teniendo en cuenta los valores de u_y y v_y en el punto $(1,1,1)$ calculados anteriormente, se obtiene fácilmente

$$u_{yy}(1,1,1) = 0, \qquad v_{yy}(1,1,1) = 2.$$

3.3. Independencia funcional

El objetivo de esta sección es estudiar cuando, dadas m funciones escalares $f_i : \mathbb{R}^n \to \mathbb{R}$, $1 \le i \le m$, hay alguna relación no trivial de dependencia funcional del tipo $\Phi(f_1, \dots, f_m) = 0$ entre ellas. Para facilitar el análisis—fundamentalmente, para poder utilizar los teoremas de la función inversa e implícita—suponderemos a partir de ahora que las funciones en cuestión son de clase C^1:

Definición 3.11. Dadas m funciones escalares $f_i : \mathbb{R}^n \to \mathbb{R}$, $1 \le i \le m$, de clase C^1 en un abierto $U \subset \mathbb{R}^n$, diremos que $f_1, \dots, f_m$ son *funcionalmente dependientes* en U si existe una función $\Phi : \mathbb{R}^m \to \mathbb{R}$ de clase C^1 en un abierto $V \supset f(U)$, que no se anula en ningún subconjunto abierto de V y tal que $\Phi \circ f = 0$ en U. En caso contrario, diremos que $f_1, \dots, f_m$ son *funcionalmente independientes* en U.

Por ejemplo, las funciones $f_i : \mathbb{R}^3 \to \mathbb{R}$ definidas por $f_1(x,y,z) = x+y+z$, $f_2(x,y,z) = xy + xz + yz$, $f_3(x,y,z) = x^2 + y^2 + z^2$ son funcionalmente dependientes. En efecto, se satisface idénticamente la relación $f_1^2 - 2f_2 - f_3 = 0$, y la función $\Phi : \mathbb{R}^3 \to \mathbb{R}$ definida por $\Phi(u,v,w) = u^2 - 2v - w$ es de clase C^∞ en todo $\mathbb{R}^3$ y no se anula en ninguna bola abierta (y por tanto en ningún abierto) de $\mathbb{R}^3$ (¿por qué?).

¿Como averiguar si m funciones dadas son funcionalmente dependientes sin necesidad de hallar explícitamente la relación de dependencia funcional entre ambas (lo que puede ser complicado o engorroso en la práctica)?

En primer lugar, nótese que si $f(U)$ es abierto entonces es inmediato que $f_1, \ldots, f_m$ son funcionalmente independientes; en efecto, la función Φ ha de anularse idénticamente en $f(U)$, que por hipótesis es abierto. Por ejemplo, $m \leq n$ funciones coordenadas cualesquiera $x_{i_1}, \ldots, x_{i_m}$ en $\mathbb{R}^n$ son funcionalmente independientes, ya que en este caso $f(\mathbb{R}^n) = \mathbb{R}^m$. En particular, si $m = n$ y $f = (f_1, \ldots, f_m)$ es de clase C^1 en U y cumple la condición $\det Df(x) \neq 0$, $\forall x \in U$, entonces las funciones $f_1, \ldots, f_m$ son funcionalmente independientes, por la Proposición 3.6. De hecho, este resultado se generaliza sin dificultad al caso $m < n$, reemplazando la condición $\det Df \neq 0$ por su análoga $\operatorname{rank} Df = m$ en U:

Proposición 3.12. *Si las $m \leq n$ funciones $f_i : \mathbb{R}^n \to \mathbb{R}$, $1 \leq i \leq m$, son de clase C^1 en una abierto U de $\mathbb{R}^n$, y cumplen la condición*

$$\operatorname{rank} Df(x) = m, \qquad \forall x \in U,$$

entonces $f(U)$ es abierto, y $f_1, \ldots, f_m$ son funcionalmente independientes en U.

Demostración. Por los comentarios anteriores, podemos suponer que $m < n$, y basta probar que $f(U)$ es abierto. Sea entonces $a = (a_1, \ldots, a_n)$ un punto cualquiera de U; hay que probar que hay un entorno V de $f(a)$ contenido en $f(U)$. Por la condición sobre el rango de Df, hay un menor de orden m de Df que no se anula en a; por ejemplo, podemos suponer que

$$\det \left(\frac{\partial f_i}{\partial x_j}(a) \right)_{1 \leq i,j \leq m} \neq 0.$$

Si utilizamos la notación $x = (\hat{x}, \tilde{x}) \in \mathbb{R}^m \times \mathbb{R}^{n-m} = \mathbb{R}^n$ para denotar los puntos de $\mathbb{R}^n$, y definimos $i : \mathbb{R}^m \to \mathbb{R}^n$ mediante $i(\hat{x}) = (\hat{x}, \tilde{a})$, entonces la aplicación $F = f \circ i : \mathbb{R}^m \to \mathbb{R}^m$ es de clase C^1 en el abierto $i^{-1}(U)$, y

$$\det DF(\hat{x}) = \det \left(\frac{\partial f_i}{\partial x_j}(\hat{x}, \tilde{a}) \right)_{1 \leq i,j \leq m}$$

no se anula en $\hat{a}$. Por el teorema de la función inversa, existe un entorno $W_m \subset i^{-1}(U)$ de $\hat{a}$ en $\mathbb{R}^m$ en el cual F es un difeomorfismo; en particular, $F(W_m)$ es abierto en $\mathbb{R}^m$. Por tanto, $f(a) = F(\hat{a}) \in F(W_m) = f\big(i(W_m)\big) \subset f(U)$, y $f(U)$ contiene un entorno de uno cualquiera de sus puntos $f(a)$. *Q.E.D.*

De hecho, el recíproco de la proposición anterior también se cumple (localmente); su demostración, que es bastante más laboriosa, será omitida aquí por razones de espacio:

Proposición 3.13. *Sean $f_1, \ldots, f_m : \mathbb{R}^n \to \mathbb{R}$ de clase C^1 en un entorno U de $a \in \mathbb{R}^n$, y sea $\operatorname{rank} Df(x) = r < m$, $\forall x \in U$. Entonces las funciones*

$f_1, \dots, f_m$ son funcionalmente dependientes en un entorno $W \subset U$ de a. Además, existen exactamente $m - r$ relaciones funcionalmente independientes de la forma $\Phi(f_1, \dots, f_m)$ en W.

Nota. Nótese que para poder aplicar la proposición anterior el rango de Df ha de ser *constante* en U.

Por ejemplo, si $f_1(x,y,z) = x + y + z$, $f_2(x,y,z) = xy + xz + yz$, $f_3(x,y,z) = x^2 + y^2 + z^2$, entonces

$$\det Df(x,y,z) = \begin{vmatrix} 1 & 1 & 1 \\ y+z & x+z & x+y \\ 2x & 2y & 2z \end{vmatrix} = 2 \begin{vmatrix} 1 & 1 & 1 \\ 0 & x-y & x-z \\ 0 & y-x & z-x \end{vmatrix} = 0$$

(las dos últimas filas son proporcionales). Por tanto, $\operatorname{rank} Df < 3$ en $\mathbb{R}^3$. El rango de Df es 2 en $\mathbb{R}^3 - \big\{(x,y,z) : x = y = z\big\}$, y es igual a 1 en la recta $x = y = z$. Por tanto, la proposición anterior nos asegura que f_1, f_2 y f_3 son funcionalmente dependientes en un entorno de cualquier punto fuera de la recta $x = y = z$, y que sólo hay una relación funcionalmente independiente entre estas tres funciones en un entorno de dichos puntos. Nótese que la proposición anterior no es aplicable en los puntos de la recta $x = y = z$, ya que el rango de Df no es constante en un entorno de dichos puntos. Sin embargo, procediendo directamente vimos antes que las funciones f_1, f_2 y f_3 son funcionalmente dependientes en todo $\mathbb{R}^3$.

Capítulo 4

Fórmula de Taylor. Extremos

4.1. Teorema del valor medio

Dados dos puntos $x, y \in \mathbb{R}^n$, designaremos por $[x, y]$ al segmento cerrado que une dichos puntos, es decir:

$$[x, y] = \{x + t(y - x) : 0 \leq t \leq 1\}.$$

Nótese que $[x, y]$ es cerrado (compacto), al ser la imagen del compacto $[0, 1] \subset \mathbb{R}$ bajo una función continua. Análogamente,

$$(x, y) = \{x + t(y - x) : 0 < t < 1\}$$

denotará al segmento abierto (sin los extremos) entre x e y. (Nótese que (x, y) *no* es un subconjunto abierto de $\mathbb{R}^n$.) El teorema del valor medio para funciones de $\mathbb{R}$ en $\mathbb{R}$ admite entonces la siguiente generalización para funciones escalares:

Teorema (del valor medio)**.** *Sea $f : \mathbb{R}^n \to \mathbb{R}$ una función continua en el segmento cerrado $[x_0, x_0 + h]$ y diferenciable en $(x_0, x_0 + h)$. Entonces $\exists s \in (0, 1)$ tal que*

$$f(x_0 + h) = f(x_0) + Df(x_0 + sh) \cdot h.$$

Nota. El teorema del valor medio recibe a veces el nombre de *teorema de los incrementos finitos*. Obsérvese que el punto $x_0 + sh$ es un punto del segmento $[x_0, x_0 + h]$ intermedio entre los extremos x_0 y $x_0 + h$.

Demostración. Sea $\phi : \mathbb{R} \to \mathbb{R}$ la restricción de f al segmento $[x_0, x_0 + h]$, definida por $\phi(t) = f(x_0 + th)$, $\forall t \in [0, 1]$. Entonces $\phi = f \circ g$, siendo $g(t) = x_0 + th$. Por el teorema sobre composición de funciones continuas y la regla de la cadena, ϕ es continua en $[0, 1]$ y diferenciable en $(0, 1)$. Aplicando a ϕ el teorema del valor medio para funciones de $\mathbb{R}$ en $\mathbb{R}$, y teniendo en cuenta que

$$\phi'(t) = Df(x_0 + th) \cdot h$$

queda

$$f(x_0 + h) - f(x_0) = \phi(1) - \phi(0) = \phi'(s) = Df(x_0 + sh) \cdot h, \qquad 0 < s < 1.$$

Q.E.D.

El teorema del valor medio no tiene una generalización inmediata a funciones con valores vectoriales, es decir a funciones $\mathbb{R}^n \to \mathbb{R}^m$. Para este tipo de funciones, lo más que puede demostrarse es la siguiente acotación de la norma del incremento $f(x_0 + h) - f(x_0)$:

Proposición 4.1. *Sea $f : \mathbb{R}^n \to \mathbb{R}^m$ una función continua en el segmento cerrado $[x_0, x_0 + h]$ y con derivada acotada en $(x_0, x_0 + h)$:*

$$|\nabla f_i(x_0 + th)| < M, \qquad 1 \leq i \leq n.$$

Entonces se tiene

$$|f(x_0 + h) - f(x_0)| < mM\,|h|\,.$$

Demostración. Por las hipótesis hechas, podemos aplicar el teorema del valor medio a cada una de las funciones componentes f_i, obteniendo

$$f_i(x_0 + h) - f_i(x_0) = Df_i(x_0 + s_i h) \cdot h = \nabla f_i(x_0 + s_i h) \cdot h, \qquad 0 < s_i < 1.$$

Por tanto

$$\begin{aligned} |f(x_0 + h) - f(x_0)| &\leq \sum_{i=1}^{m} |f_i(x_0 + h) - f_i(x_0)| = \sum_{i=1}^{m} |\nabla f_i(x_0 + s_i h) \cdot h| \\ &\leq |h| \sum_{i=1}^{m} |\nabla f_i(x_0 + s_i h)| < |h| \cdot mM. \end{aligned}$$

Q.E.D.

Un subconjunto $A \subset \mathbb{R}^n$ es *convexo* si dados dos puntos cualesquiera $x, y \in A$, el segmento cerrado $[x, y]$ está contenido en A. Por ejemplo, cualquier bola (abierta o cerrada) es convexa, ya que si por ejemplo $|x - a| < r$ y $|y - a| < r$ entonces

$$\begin{aligned} |x + t(y - x) - a| &= |t(y - a) + (1 - t)(x - a)| \leq t\,|y - a| + (1 - t)\,|x - a| \\ &< tr + (1 - t)r = r, \qquad \forall t \in [0, 1]. \end{aligned}$$

Si $f : \mathbb{R}^n \to \mathbb{R}$ es diferenciable en un abierto convexo $A \subset \mathbb{R}^n$, entonces $\forall x, y \in A$ podemos aplicar el teorema del valor medio al segmento $[x, y] \subset A$, obteniéndose

$$f(y) - f(x) = Df(x + s(y - x)) \cdot (y - x)$$

para algún $s \in (0, 1)$ (que por supuesto dependerá de x e y). Análogamente, si $f : \mathbb{R}^n \to \mathbb{R}^m$ es diferenciable en un abierto convexo A y $|\nabla f_i| < M$ en A, $\forall i = 1, 2, \ldots, m$, entonces

$$|f(y) - f(x)| < mM\,|y - x|\,, \qquad \forall x, y \in A.$$

El teorema del valor medio permite probar fácilmente el siguiente resultado importante:

Proposición 4.2. *Sea $A \subset \mathbb{R}^n$ un abierto conexo, y sea $f : \mathbb{R}^n \to \mathbb{R}^m$ diferenciable en A. Entonces f es constante en A si y sólo si $Df(x) = 0$, $\forall x \in A$.*

Demostración. En primer lugar, es claro que f constante en A implica $Df = 0$ en A. Para probar la recíproca, obsérvese que basta hacerlo para el caso de funciones escalares, ya que el caso general se obtiene aplicando el caso escalar a cada componente de f. Sea entonces $f : \mathbb{R}^n \to \mathbb{R}$, sea a un punto cualquiera de A, y sea

$$Z = \big\{x \in A : f(x) = f(a)\big\}.$$

Por ser f continua en A (al ser diferenciable), Z es cerrado en A. Por otra parte, por ser A abierto, si $x \in Z \subset A$ existe $r > 0$ tal que $B_r(x) \subset A$. Como $B_r(x)$ es convexa, aplicando el teorema del valor medio obtenemos

$$f(y) - f(x) = f(y) - f(a) = Df(x + t(y - x)) \cdot (y - x) = 0, \qquad \forall y \in B_r(x).$$

Por tanto $B_r(x) \subset Z$, y Z es abierto en $\mathbb{R}^n$. Como A es abierto, Z es abierto en A. Al ser A conexo y Z simultáneamente abierto y cerrado en A, $Z = \emptyset$ ó $Z = A$. Como $a \in Z$ por construcción, ha de ser $Z = A$, de donde se sigue que

$$f(x) = f(a), \qquad \forall x \in A.$$

Q.E.D.

4.2. Fórmula de Taylor

Una función $f : \mathbb{R} \to \mathbb{R}$ es de clase C^p en un intervalo cerrado $[a, b]$ si f es de clase C^p en (a, b), y si todas las derivadas de orden $\leq p$ de f son continuas en los extremos a y b, es decir si existen los límites

$$\lim_{h \to 0+} f^{(i)}(a + h) \equiv f^{(i)}(a), \quad \lim_{h \to 0+} f^{(i)}(b - h) \equiv f^{(i)}(b), \qquad 0 \leq i \leq p.$$

Si $f : \mathbb{R} \to \mathbb{R}$ es una función de clase C^{p-1} en $[a, b]$ y $f^{(p)}$ existe en (a, b)—en particular, si f es de clase C^p en $[a, b]$—es bién sabido que es posible desarrollar $f(b)$ en potencias de $b - a$ como sigue:

$$f(b) = \sum_{i=0}^{p-1} \frac{f^{(i)}(a)}{i!}(b - a)^i + R_p,$$

con

$$R_p = \frac{f^{(p)}(c)}{p!}(b-a)^p$$

para algún $c \in (a,b)$. El resultado anterior se conoce como *fórmula de Taylor de orden $p-1$ con resto.* La demostración es muy sencilla: si $G(x)$ es la función definida por

$$G(x) = f(b) - \sum_{i=0}^{p-1} \frac{f^{(i)}(x)}{i!}(b-x)^i - \frac{K}{p!}(b-x)^p,$$

con K escogido de modo que $G(a) = 0$, basta probar que $K = f^{(p)}(c)$, para algún $c \in (a,b)$. Para ello, aplicamos el teorema del valor medio a G. Utilizando la regla de Leibniz, se obtiene inmediatamente la igualdad

$$G'(x) = \frac{(b-x)^{p-1}}{(p-1)!}\left(K - f^{(p)}(x)\right),$$

y el teorema del valor medio implica entonces que $\exists c \in (a,b)$ tal que

$$G(b) - G(a) = 0 = (b-a)G'(c) \Longrightarrow 0 = G'(c) = \frac{(b-c)^{p-1}}{(p-1)!}\left(K - f^{(p)}(c)\right).$$

Al ser $c \in (a,b)$, de esta última igualdad se sigue efectivamente que $K = f^{(p)}(c)$ para algún $c \in (a,b)$.

Para generalizar el teorema de Taylor a una función escalar $f : \mathbb{R}^n \to \mathbb{R}$, debemos primero introducir un poco más de notación en relación con las derivadas parciales de f. Si f es de clase C^p en un entorno de un punto x y $1 \le q \le p$, definimos un operador q-lineal $D^q f(x) : \mathbb{R}^n \times \underset{q \text{ veces}}{\cdots} \times \mathbb{R}^n \to \mathbb{R}$ por la fórmula

$$D^q f(x)(v^1, \dots, v^q) = \sum_{i_1,\dots,i_q=1}^{n} D_{i_1\dots i_q} f(x)\, v^1_{i_1} \dots v^q_{i_q}, \qquad \forall v^1, \dots, v^q \in \mathbb{R}^n. \tag{4.1}$$

Por ejemplo, si $n = 2$, $p = 2$, $v = (v_1, v_2)$ y $w = (w_1, w_2)$ entonces

$$\begin{aligned} D^2 f(x,y)(v,w) &= f_{xx}(x,y)\, v_1 w_1 + f_{xy}(x,y)\, v_1 w_2 + f_{yx}(x,y)\, v_2 w_1 + f_{yy}(x,y)\, v_2 w_2 \\ &= f_{xx}(x,y)\, v_1 w_1 + f_{xy}(x,y)\, (v_1 w_2 + v_2 w_1) + f_{yy}(x,y)\, v_2 w_2, \end{aligned}$$

donde la última igualdad es consecuencia del lema de Schwarz. En lo que sigue, utilizaremos también la notación

$$D^q f(x) \cdot h^q = D^q f(x)(h, \underset{q \text{ veces}}{\dots}, h), \qquad h \in \mathbb{R}^n;$$

por ejemplo, en el caso particular $n = p = 2$ se tiene

$$D^2 f(x,y) \cdot h^2 = f_{xx}(x,y)(h_1)^2 + 2 f_{xy}(x,y)\, h_1 h_2 + f_{yy}(x,y)(h_2)^2.$$

Si $p = 1$, es evidente que $D^1 f(x) = Df(x)$. Si $p = 2$, para cada x fijo $D^2 f(x)$ es una forma bilineal, y $D^2 f(x) \cdot h^2$ es una forma cuadrática en h.

Sea ahora $f : \mathbb{R}^n \to \mathbb{R}$ una función de clase C^p en un abierto $U \subset \mathbb{R}^n$, y sean $x_0, x_0 + h \subset U$ tales que $[x_0, x_0 + h] \in U$; por ejemplo, esta última condición se cumplirá automáticamente si U es convexo. La función $\phi : [0, 1] \to \mathbb{R}$ dada por $\phi(t) = f(x_0 + th)$ es entonces de clase C^p en $[0, 1]$, pues aplicando repetidamente la regla de la cadena se tiene:

$$\phi'(t) = Df(x_0 + th) \cdot h = \sum_{i=1}^{n} D_i f(x_0 + th) h_i,$$

$$\phi''(t) = \sum_{i=1}^{n} h_i \frac{d}{dt} D_i f(x_0 + th) = \sum_{i,j=1}^{n} D_{ij} f(x_0 + th) h_i h_j = D^2 f(x_0 + th) \cdot h^2$$

y, en general, se prueba por inducción que

$$\phi^{(k)}(t) = D^k f(x_0 + th) \cdot h^k, \qquad 1 \leq k \leq p.$$

Aplicando el teorema de Taylor a la función ϕ en $[0, 1]$ se obtiene

$$\phi(1) - \phi(0) = f(x_0 + h) - f(x_0) = \sum_{i=1}^{p-1} \frac{\phi^{(i)}(0)}{i!} + \frac{\phi^{(p)}(s)}{p!}$$

con $s \in (0, 1)$, de donde se sigue que

$$f(x_0 + h) = f(x_0) + \sum_{i=1}^{p-1} \frac{1}{i!} D^i f(x_0) \cdot h^i + R_p(h), \tag{4.2}$$

con

$$R_p(h) = \frac{1}{p!} D^i f(x_0 + sh) \cdot h^p, \qquad s \in (0, 1).$$

Este es el teorema de Taylor para funciones escalares:

Teorema de Taylor. *Si $f : \mathbb{R}^n \to \mathbb{R}$ es de clase C^p en un abierto U, y el segmento $[x_0, x_0 + h]$ está contenido en U, entonces se verifica la* fórmula de Taylor de orden $p - 1$ *(4.2).*

En particular, si U es un abierto convexo y $f : \mathbb{R}^n \to \mathbb{R}$ es de clase $C^p(U)$, la fórmula de Taylor de orden $p - 1$ puede aplicarse a dos puntos cualesquiera $x, y \in U$.

Si f satisface las hipótesis del teorema de Taylor, es inmediato probar que el resto $R_p(h)$ es $o(|h|^{p-1})$. De hecho, vamos a probar el resultado más fuerte siguiente:

$$R_p(h) = \frac{1}{p!} D^p f(x_0) \cdot h^i + o(|h|^p). \tag{4.3}$$

En efecto, multiplicando por $p!$ se tiene

$$\begin{aligned}
\frac{|p!R_p(h) - D^p f(x_0)\cdot h^p|}{|h|^p} &= \frac{|D^p f(x_0+sh)\cdot h^p - D^p f(x_0)\cdot h^p|}{|h|^p} \\
&\leq \sum_{i_1,\ldots,i_p=1}^{n} \left|D_{i_1\ldots i_p} f(x_0+sh) - D_{i_1\ldots i_p} f(x_0)\right| \frac{|h_{i_1}|}{|h|}\cdots\frac{\left|h_{i_p}\right|}{|h|} \\
&\leq \sum_{i_1,\ldots,i_p=1}^{n} \left|D_{i_1\ldots i_p} f(x_0+sh) - D_{i_1\ldots i_p} f(x_0)\right| @>> h \to 0 > 0,
\end{aligned}$$

al ser f de clase C^p en un entorno de x_0. Por tanto, hemos probado la siguiente versión del teorema de Taylor:

Teorema 4.3. *Si f es de clase C^p en un abierto U, y el segmento $[x_0, x_0+h]$ está contenido en U, entonces se tiene*

$$f(x_0+h) = f(x_0) + \sum_{i=1}^{p} \frac{1}{i!} D^i f(x_0)\cdot h^p + o(|h|^p).$$

La ecuación anterior recibe el nombre de *desarrollo de Taylor de orden p de $f(x_0+h)$ en x_0*. Nótese que, al ser f de clase C^p en un entorno de x_0, la igualdad de las derivadas parciales cruzadas de f de orden $k \leq p$ permite expresar $D^k f(x_0)\cdot h^k$ para $k \leq p$ en la forma más concisa siguiente:

$$D^k f(x_0)\cdot h^k = \sum_{j_1+\cdots+j_n=k} \frac{k!}{j_1!\ldots j_n!} \frac{\partial^k f}{\partial x_1^{j_1}\ldots \partial x_n^{j_n}}(x_0)\cdot h_1^{j_1}\ldots h_n^{j_n}.$$

Ejemplo 4.4. Hallemos el desarrollo de Taylor de orden 2 de la función $f:\mathbb{R}^2 \to \mathbb{R}$ definida por

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

en el punto $(0,0)$. Es claro, en primer lugar, que f es de clase C^∞ en todo $\mathbb{R}^2$, luego se puede calcular su desarrollo de Taylor de cualquier orden en $(0,0)$. Calculando las derivadas parciales de primer orden de f obtenemos

$$\begin{aligned}
f_x &= e^{x+y}\left[\cos(x-y^2) - \operatorname{sen}(x-y^2)\right] &&\Longrightarrow\quad f_x(0,0) = 1 \\
f_y &= e^{x+y}\left[\cos(x-y^2) + 2y\operatorname{sen}(x-y^2)\right] &&\Longrightarrow\quad f_y(0,0) = 1.
\end{aligned}$$

Las derivadas segundas también se calculan fácilmente:

$$\begin{aligned}
f_{xx} &= -2e^{x+y}\operatorname{sen}(x-y^2) \\
f_{xy} &= e^{x+y}\left[\cos(x-y^2) - \operatorname{sen}(x-y^2) + 2y\left(\operatorname{sen}(x-y^2) + \cos(x-y^2)\right)\right] \\
f_{yy} &= e^{x+y}\left[\cos(x-y^2) + 2(1+2y)\operatorname{sen}(x-y^2) - 4y^2\cos(x-y^2)\right],
\end{aligned}$$

de donde

$$f_{xx}(0,0) = 0, \quad f_{xy}(0,0) = f_{yy}(0,0) = 1.$$

Por tanto, el desarrollo de Taylor de orden 2 de f en el origen es

$$f(x,y) = 1 + x + y + \frac{1}{2}(2\,x\,y + y^2) + o(x^2 + y^2).$$

4.3. Serie de Taylor y funciones analíticas reales

Sea $f : \mathbb{R}^n \to \mathbb{R}$ una función de clase C^∞ en un abierto convexo U (por ejemplo, en una bola abierta), y sea x_0 un punto fijo de U. Dado un punto cualquiera $x \in U$ y un entero $p > 0$, por las hipótesis hechas es válida la fórmula de Taylor de orden p con resto

$$f(x) = \sum_{i=0}^{p} D^i f(x_0) \cdot (x - x_0)^i + R_{p+1}(f, x_0; x),$$

siendo $D^0 f(x_0) \equiv f(x_0)$ y

$$R_{p+1}(f, x_0; x) = \frac{1}{(p+1)!} D^{p+1} f(c) \cdot (x - x_0)^{p+1}, \qquad c \in (x_0, x).$$

El polinomio en x

$$P_p(f, x_0; x) = \sum_{i=0}^{p} \frac{1}{i!} D^i f(x_0) \cdot (x - x_0)^i$$

recibe el nombre de *polinomio de Taylor de orden* p *de* f *en* x_0. La pregunta natural para una función C^∞ f es entonces para qué puntos x se tiene

$$\lim_{p \to \infty} P_p(f, x_0; x) = f(x),$$

ó en otras palabras para qué valores de x se puede escribir

$$f(x) = \sum_{i=0}^{\infty} \frac{1}{i!} D^i f(x_0) \cdot (x - x_0)^i.$$

El miembro derecho de la ecuación anterior recibe el nombre de *serie de Taylor de* $f(x)$ *en* x_0.

Definición 4.5. Una función $f : \mathbb{R}^n \to \mathbb{R}$ de clase C^∞ en un abierto $U \subset \mathbb{R}^n$ es *analítica en* $x_0 \in U$ si hay un entorno $B_r(x_0) \subset U$ de x_0 tal que la serie de Taylor de $f(x)$ en x_0 converge a $f(x)$, $\forall x \in B_r(x_0)$. Se dirá que f es analítica en U si f es analítica en x_0, para todo $x_0 \in U$.

Ejemplo 4.6. Los polinomios son funciones analíticas en todo el espacio. En efecto, para un polinomio P de grado q todas las derivadas parciales de orden $> q$ se anulan idénticamente. Por tanto, la serie de Taylor de $P(x)$ en cualquier punto x_0 es una serie finita. Además, es inmediato probar que la serie de Taylor de $P(x)$ en x_0 converge a $P(x)$ para todo $x \in \mathbb{R}^n$, es decir que

$$P(x) = \sum_{i=0}^{q} \frac{1}{i!} D^i P(x_0) \cdot (x - x_0)^i, \qquad \forall x \in \mathbb{R}^n,$$

para todo polinomio P de grado q, sin más que aplicar la fórmula de Taylor de orden q.

Sin embargo, hay funciones de clase C^∞ (de hecho "la mayoría") que no son analíticas, como veremos en el próximo ejemplo. Nótese que la analiticidad de $f : \mathbb{R}^n \to \mathbb{R}$ en un punto x_0 significa que se han de cumplir las dos condiciones siguientes:

I) La serie de Taylor de $f(x)$ en x_0 es convergente para todo x en un entorno V de x_0

II) La serie de Taylor de $f(x)$ en x_0 converge a $f(x)$, $\forall x \in V$

La segunda condición es independiente de la primera, como muestra el siguiente ejemplo:

Ejemplo 4.7. Sea $f : \mathbb{R} \to \mathbb{R}$ la función definida por

$$f(x) = \begin{cases} 0, & x = 0 \\ e^{-\frac{1}{x^2}}, & x \neq 0. \end{cases}$$

Es inmediato comprobar que f es de clase C^∞ en un entorno de cualquier punto $x \neq 0$, puesto que

$$f^{(k)}(x) = P_k(1/x)\, e^{-\frac{1}{x^2}}, \qquad x \neq 0,$$

siendo P_k un polinomio. (Esta última igualdad se demuestra fácilmente por inducción sobre k.) Para ver que f es de clase C^∞ en $\mathbb{R}$, necesitamos el siguiente lema:

Lema 4.8. *Sea $f : \mathbb{R}^n \to \mathbb{R}$ continua en un abierto U y de clase C^1 en $U - \{x_0\}$, siendo $x_0 \in U$. Si existe $\lim_{x \to x_0} D_i f(x) = l_i$, entonces $D_i f(x_0) = l_i$, y f es de clase C^1 en U.*

Demostración. Por ser U abierto, hay una bola abierta $B_r(x_0)$ centrada en x_0 contenida en U. Para $|t| < r$, por el teorema del valor medio existe $s(t) \in (0, 1)$ tal que

$$D_i f(x_0) = \lim_{t \to 0} \frac{f(x_0 + te_i) - f(x_0)}{t} = \lim_{t \to 0} D_i f\big(x_0 + s(t)te_i\big) = l_i,$$

Q.E.D.

Utilizando el lema anterior, se prueba por inducción que si f es de clase C^∞ en $U-\{x_0\}$, y todas las derivadas parciales de f (incluyendo a f) tienen límite en x_0, entonces f es de clase C^∞ en x_0. Por tanto, en nuestro caso basta probar que f y todas sus derivadas parciales tienen límite en el origen. Pero esto es sencillo, ya que

$$\lim_{x\to 0} f^{(k)}(x) = \lim_{t\to\infty} P_k(t)e^{-t^2} = 0, \qquad \forall k \in \mathbb{N}.$$

Por la observación anterior, de esto se se sigue que f es C^∞ en todo $\mathbb{R}$, y

$$f^{(k)}(0) = 0, \qquad \forall k \geq 0. \tag{4.4}$$

En virtud de (4.4), la serie de Taylor de $f(x)$ en 0 es trivialmente convergente a 0 para todo x (pues todos sus términos son nulos). Esto implica que f no es analítica en el origen, pues $f(x) > 0$ si $x \neq 0$.

Daremos a continuación una condición suficiente que asegura la analiticidad de f en un punto x_0. Para ello, supongamos que f es de clase C^∞ en un abierto U, y sea $B_r(x_0)$ una bola centrada en x_0 cuyo cierre esté contenido en U. Hay que probar que se verifica

$$\lim_{p\to\infty} R_p(f, x_0; x) = 0,$$

para todo x en un cierto entorno de x_0. Por el teorema de Taylor, sabemos que $R_p(f, x_0; x)$ puede expresarse como

$$R_p(f, x_0; x) = \frac{1}{p!} D^p f(x_0 + s(x - x_0)) \cdot (x - x_0)^p, \qquad s \in (0, 1).$$

Como cada derivada parcial de f es continua en U, y $\overline{B}_r(x_0)$ es un compacto contenido en U, todas las derivadas parciales de f están acotadas en $\overline{B}_r(x_0)$, y por tanto en $B_r(x_0)$. En consecuencia, existe una constante C_p tal que

$$\left|D_{i_1\ldots i_p} f(y)\right| \leq C_p, \qquad \forall y \in B_r(x_0), \quad \forall i_1, \ldots, i_p.$$

Al ser $x_0 + s(x - x_0) \in B_r(x_0)$ si $x \in B_r(x_0)$, llamando $h = x - x_0$ se tiene entonces

$$\begin{aligned} |D^p f(x_0 + sh) \cdot h^p| &\leq \sum_{i_1,\ldots,i_p=1}^{n} \left|D_{i_1\ldots i_p} f(x_0 + sh)\right| |h_{i_1}| \ldots \left|h_{i_p}\right| \\ &\leq C_p \sum_{i_1,\ldots,i_p=1}^{n} |h_{i_1}| \ldots \left|h_{i_p}\right| \\ &= C_p \left(\sum_{i=1}^{p} |h_i|\right)^p \leq C_p\, n^{p/2}\, |h|^p , \end{aligned} \tag{4.5}$$

donde la última desigualdad se obtiene aplicando la desigualdad de Cauchy–Schwarz al producto escalar de los vectores $\sum_{i=1}^{p} |h_i|\, e_i$ y $\sum_{i=1}^{p} e_i$.

La desigualdad (4.5) es totalmente general. Si ahora hacemos la hipótesis adicional de que las constantes C_p que acotan a las derivadas parciales de orden p de f no crecen más rápido que la p-ésima potencia de una constante *independiente de* p, es decir si se cumple que $C_p \leq M^p$ $\forall p \in N$, para alguna constante positiva M, entonces (4.5) implica que

$$|R_p(f, x_0; x)| \leq \frac{(\sqrt{n}\, M\, |x - x_0|)^p}{p!} @ >> p \to \infty > 0.$$

Se cumple por tanto la siguiente proposición:

Proposición 4.9. *Si $f : \mathbb{R}^n \to \mathbb{R}$ es de clase C^∞ en un entorno V de x_0, y se cumplen las desigualdades*

$$\left| D_{i_1 \dots i_p} f(x) \right| \leq M^p, \qquad \forall x \in V, \quad 1 \leq i_1, \dots, i_p \leq n, \quad \forall p \in \mathbb{N},$$

entonces la serie de Taylor de $f(x)$ converge a $f(x)$ para todo $x \in V$. En particular, f es analítica en x_0.

Nota. La definición de función analítica *real* que acabamos de ver no debe confundirse con la de función analítica de variable compleja (*función holomorfa*), que es mucho más restrictiva. En efecto, dar una función $f : \mathbb{C} \to \mathbb{C}$ es equivalente a dar dos funciones $u = \operatorname{Re} f, v = \operatorname{Im} f : \mathbb{R}^2 \to \mathbb{R}$. Si f es una función analítica compleja, se puede probar que u y v son funciones analíticas reales, pero el recíproco no es cierto. Por ejemplo, si $f(z) = \bar{z}$ entonces $u(x, y) + iv(x, y) = x - iy$, y por tanto $u(x, y) = x$, $v(x, y) = -y$ son funciones analíticas reales (polinomios). Sin embargo, f *no* es una función analítica compleja, ya que u y v no cumplen las ecuaciones de Cauchy–Riemann.

4.4. Extremos

Una de las aplicaciones fundamentales del Análisis es el cálculo de extremos (máximos y mínimos) de funciones escalares de varias variables. Comenzaremos esta sección con las siguientes definiciones:

Definición 4.10. Sea $f : \mathbb{R}^n \to \mathbb{R}$ una función definida en un conjunto $D \subset \mathbb{R}^n$, y sea $x_0 \in D$.

i) x_0 es un *máximo absoluto* (ó *global*) de f en D si

$$f(x) \leq f(x_0), \qquad \forall x \in D$$

ii) x_0 es un *máximo absoluto estricto* de f en D si

$$f(x) < f(x_0), \qquad \forall x \neq x_0, \quad x \in D$$

III) x_0 es un *máximo relativo* (ó *local*) de f en D si existe un entorno U de x_0 tal que x_0 es un máximo absoluto de f en $U \cap D$, es decir

$$f(x) \leq f(x_0), \qquad \forall x \in U \cap D$$

IV) x_0 es un *máximo relativo estricto* de f en D si existe un entorno U de x_0 tal que x_0 es un máximo absoluto estricto de f en $U \cap D$, es decir

$$f(x) < f(x_0), \qquad \forall x \neq x_0, \quad x \in U \cap D$$

V) x_0 es un *mínimo* de cualquiera de los tipos anteriores de f en D si es un máximo del mismo tipo de $-f$ en D

VI) x_0 es un *extremo* de cualquiera de los tipos anteriores de f en D si es un máximo ó un mínimo de ese tipo de f en D

VII) x_0 es un *punto de silla* de f en D si x_0 no es un extremo relativo de f en D, es decir si para todo $r > 0$ hay dos puntos $x_1, x_2 \in B_r(x_0) \cap D$ tales que

$$f(x_1) < f(x_0) < f(x_2)$$

Ejemplo 4.11. Veamos algunos ejemplos de funciones de $\mathbb{R}^2$ en $\mathbb{R}$ que ilustran las definiciones anteriores:

I) $f(x,y) = -\sqrt{x^2+y^2}$ tiene un máximo absoluto estricto en $(0,0)$ (que *no* se puede encontrar utilizando el cálculo diferencial, ya que f no tiene derivadas parciales en el origen)

II) $f(x,y) = -x^2 - 2x\,y - y^2$ tiene un máximo absoluto, pero no estricto, en $(0,0)$

III) $f(x,y) = -x^2 - y^2 + x^2\,y$ tiene un máximo relativo estricto, pero no un máximo absoluto, en $(0,0)$. Para ver esto, basta darse cuenta de que para $(x,y) \neq (0,0)$ se tiene

$$\frac{f(0,0) - f(x,y)}{x^2+y^2} = 1 + o(1).$$

Por tanto, existe $\epsilon > 0$ tal que el segundo término es menor que, por ejemplo, $1/2$ en $B_\epsilon(0,0)$, por lo que se tiene

$$\frac{f(0,0) - f(x,y)}{x^2+y^2} > 1 - \frac{1}{2} = \frac{1}{2} > 0, \qquad \forall (x,y) \in B_\epsilon(0,0) - \{(0,0)\}$$

IV) $f(x,y) = x^2 - y^3$ tiene un punto de silla en $(0,0)$ (aunque el término de orden más bajo en $(0,0)$, $g(x,y) = x^2$, tiene un mínimo absoluto no estricto en el origen).

Naturalmente, cambiando f por $-f$ en los tres primeros ejemplos se obtienen ejemplos análogos reemplazando la palabra "máximo"por "mínimo".

El resultado fundamental que permite aplicar el cálculo diferencial para hallar los extremos de una función es el siguiente:

Teorema 4.12. *Sea $f : \mathbb{R}^n \to \mathbb{R}$, sea $D \subset \mathbb{R}^n$, y sea $x_0 \in \overset{\circ}{D}$ un extremo relativo de f en D. Si f es diferenciable en x_0, entonces $Df(x_0) = 0$.*

Demostración. Como f es diferenciable en x_0, f está definida en un entorno de x_0. Por ser, además, x_0 un extremo relativo de f en el abierto $\overset{\circ}{D}$, podemos encontrar otro entorno $B_r(x_0)$ de x_0 contenido en el anterior tal que f restringida a $B_r(x_0)$ tiene un extremo absoluto en x_0. Si $v \neq 0$ es un vector de norma uno, la función $g : (-r, r) \to \mathbb{R}$ definida por $g(t) = f(x_0 + tv)$ tiene un extremo relativo en $t = 0$, ya que $x_0 + tv \in B_r(x_0)$ si $|t| < r$. En consecuencia, se cumple

$$0 = g'(0) = Df(x_0) \cdot v$$

para todo vector v de norma igual a uno. Al ser $Df(x_0)$ lineal, la igualdad anterior se ha de verificar para todo vector v, lo que concluye la demostración. *Q.E.D.*

El teorema anterior motiva la siguiente definición:

Definición 4.13. Un *punto crítico* de una función $f : \mathbb{R}^n \to \mathbb{R}$ es un punto $x_0 \in \mathbb{R}^n$ tal que $Df(x_0) = 0$.

Notas

i) Evidentemente, la condición $Df(x_0) = 0$ es equivalente a $\nabla f(x_0) = 0$, ó a las n ecuaciones escalares

$$\frac{\partial f}{\partial x_i}(x_0) = 0, \qquad 1 \leq i \leq n.$$

ii) Si $x_0 \notin \overset{\circ}{D}$ (es decir, $x_0 \in \operatorname{fr} D$) es un extremo relativo de f en D, aunque f sea diferenciable en x_0 dicho punto *no* tiene por qué ser un punto crítico de f. (Para que tenga sentido $Df(x_0)$, es necesario que f esté definida en un entorno de x_0.) Por ejemplo, sea $f : \mathbb{R}^2 \to \mathbb{R}$ la función definida por $f(x, y) = x^2 + y^2$, y sea D el cuadrado cerrado $[-1, 1]^2$. Geométricamente, es evidente que f tiene un mínimo absoluto estricto en el origen y cuatro máximos absolutos no estrictos (máximos relativos estrictos) en los vértices del cuadrado, es decir en los puntos de la forma (ϵ_1, ϵ_2) con $\epsilon_1, \epsilon_2 \in \{-1, 1\}$. El origen es obviamente un punto de $\overset{\circ}{D}$, por lo que $Df(0,0)$ ha de anularse de acuerdo con el teorema anterior. Sin embargo, en los cuatro máximos Df (ó equivalentemente ∇f) no se anula, pues

$$\nabla f(\epsilon_1, \epsilon_2) = 2(\epsilon_1, \epsilon_2) \neq 0.$$

Esto no contradice al teorema anterior, ya que los cuatro vértices del cuadrado son puntos frontera de D.

III) También es importante observar que la anulación de $Df(x_0)$ en un punto interior al dominio de f es condición necesaria, pero no suficiente, para que f tenga un extremo relativo en x_0. En otras palabras, los puntos críticos de una función no son necesariamente extremos relativos de dicha función. Por ejemplo, la función $f : \mathbb{R}^2 \to \mathbb{R}$ dada por $f(x,y) = x^2 - y^3$ no tiene un extremo en el origen, aunque $\nabla f(x,y) = (2x, -3y^2)$ se anula en el origen.

En el caso de funciones de una variable, es posible utilizar las derivadas de orden superior de una función para averiguar si un punto crítico de una función es un máximo ó un mínimo. La generalización de estos resultados a funciones de varias variables se puede resumir en la siguiente proposición:

Proposición 4.14. *Sea x_0 un punto crítico de $f : \mathbb{R}^n \to \mathbb{R}$, sea f de clase m en un entorno de x_0, y supongamos que*

$$D^k f(x_0) = 0, \qquad \forall k = 1, 2, \ldots, m-1,$$

y $D^m f(x_0) \neq 0$. Entonces se verifica:

I) *Si x_0 es un mínimo relativo de f, entonces m es par y $D^m f(x_0) \cdot v^m$ es semidefinida positiva, es decir*

$$D^m f(x_0) \cdot v^m \geq 0, \qquad \forall v \in \mathbb{R}^n.$$

II) *Si m es par y $D^m f(x_0) \cdot v^m$ es definida positiva, es decir*

$$D^m f(x_0) \cdot v^m > 0, \qquad \forall v \neq 0, \quad v \in \mathbb{R}^n,$$

entonces x_0 es un mínimo relativo estricto.

Nota. Cambiando el sentido de las desigualdades en la proposición anterior se obtiene una proposición análoga para máximos relativos de f.

Demostración. En primer lugar, nótese que la anulación de la aplicación k-lineal $D^k f(x_0)$ es equivalente a la anulación de todas las derivadas parciales de orden k de f en x_0. Análogamente, la no anulación de $D^m f(x_0)$ es equivalente a la no anulación de alguna de las derivadas parciales de orden

m de f en x_0, lo que implica la existencia de algún vector $v \in \mathbb{R}^n$ tal que $D^m f(x_0) \cdot v^m \neq 0$.
i) Basta considerar, para cada vector $v \neq 0$, la función $g : \mathbb{R} \to \mathbb{R}$ definida en un entorno de 0 por $g(t) = f(x_0 + tv)$. Por las hipótesis hechas, g es de clase m en un entorno de 0, tiene un mínimo relativo en 0 y cumple

$$\begin{aligned} g^{(k)}(0) &= D^k f(x_0) \cdot v^k = 0, \qquad k = 1, 2, \dots, m-1; \\ g^m(0) &= D^m f(x_0) \cdot v^m. \end{aligned}$$

Escogiendo v tal que $D^m f(x_0) \cdot v^m \neq 0$, se concluye que m tiene que ser par. Si $v \neq 0$ es arbitrario, las ecuaciones anteriores implican que $g^m(0) = D^m f(x_0) \cdot v^m \geq 0$ para que g tenga un mínimo relativo en $t = 0$.
ii) Aplicando el teorema 4.3—lo cual es posible, ya que por hipótesis f es de clase C^m en un entorno U de x_0—y teniendo en cuenta la anulación de $D^k f(x_0)$ para $1 \leq k \leq m-1$ se obtiene

$$f(x) = f(x_0) + \frac{1}{m!} D^m f(x_0) \cdot (x - x_0)^m + o(|x - x_0|^m), \qquad \forall x \in U,$$

o equivalentemente

$$\frac{f(x) - f(x_0)}{|x - x_0|^m} = \frac{1}{m!} D^m f(x_0) \cdot \left(\frac{x - x_0}{|x - x_0|} \right)^m + o(1).$$

La función $\phi : \mathbb{R}^n \to \mathbb{R}$ definida por $\phi(v) = \frac{1}{m!} D^m f(x_0) \cdot v^m$ es de clase C^∞ (¡es un polinomio en v!), y por tanto ha de alcanzar un valor mínimo M sobre la esfera unidad, que es un conjunto compacto. Como, por la hipótesis sobre $D^m f(x_0)$, ϕ es estrictamente positiva sobre la esfera unidad, el mínimo M ha de ser estrictamente positivo, es decir

$$\phi(v) = \frac{1}{m!} D^m f(x_0) \cdot v^m \geq M > 0, \qquad \forall v \in \mathbb{R}^n, \quad |v| = 1.$$

Por otra parte, tomando $\epsilon > 0$ suficientemente pequeño se puede conseguir que $B_\epsilon(x_0) \subset U$ y que $|o(1)| < M/2$. Por tanto, para todo $x \in B_\epsilon(x_0) - \{x_0\}$ se cumple

$$\frac{f(x) - f(x_0)}{|x - x_0|^m} \geq M - \frac{M}{2} = \frac{M}{2} > 0,$$

y f tiene un mínimo relativo estricto en x_0. *Q.E.D.*

Nota. Obsérvese que la proposición anterior no dice nada acerca de la naturaleza del punto crítico x_0 cuando $D^m f(x_0) \cdot v^m$ es semidefinida (positiva o negativa). Este caso es muy complicado, pues la naturaleza del punto crítico depende esencialmente de los términos de orden $> m$ en el desarrollo de Taylor de f en x_0. Por ejemplo, si $f(x, y) = x^2 - y^4$ entonces se cumplen las hipótesis de la proposición anterior con $m = 2$ en el origen. Aquí $D^2 f(0, 0) \cdot v = 2v_1^2$ es semidefinida positiva, pero el origen es un punto silla de f. Si tomamos $f(x,y) = x^2+y^4$ entonces de nuevo $D^2f(0,0)\cdot v = 2v_1^2$ es semidefinida positiva, pero ahora el origen es obviamente un mínimo absoluto estricto.

El caso más frecuente en que se aplica la proposición anterior, y en el que nos centraremos a partir de ahora, es el caso $m = 2$. En este caso f es de clase C^2 en un entorno del punto crítico x_0, y hay que estudiar la forma cuadrática en $v = (v_1, \dots, v_n)$ no idénticamente nula dada por

$$D^2f(x_0)\cdot v^2 = \sum_{i,j=1}^{n} D_{ij}f(x_0)v_iv_j = V^\top Hf(x_0)V,$$

siendo $V = (v_1 \dots v_n)^\top$ y $Hf(x_0)$ la matriz cuadrada de orden n

$$Hf(x_0) = (D_{ij}f(x_0))_{1\le i,j\le n} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x_0) & \frac{\partial^2 f}{\partial x_2\partial x_1}(x_0) & \cdots & \frac{\partial^2 f}{\partial x_n\partial x_1}(x_0) \\ \dots\dots & \dots\dots & \dots\dots & \dots\dots \\ \frac{\partial^2 f}{\partial x_1 x_n}(x_0) & \frac{\partial^2 f}{\partial x_2\partial x_n}(x_0) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x_0) \end{pmatrix},$$

llamada el *hessiano* de f en el punto x_0. Obsérvese que, por el lema de Schwarz, $Hf(x_0)$ es una matriz simétrica: $Hf(x_0)^\top = Hf(x_0)$. Como se recordará del curso de Álgebra Lineal, la matriz $Hf(x_0)$ no es otra cosa que la matriz de la forma cuadrática $D^2f(x_0)\cdot v^2$ en la base canónica de $\mathbb{R}^n$.

Antes de proseguir, es conveniente recordar algunas definiciones y hechos elementales acerca de las formas cuadráticas y sus matrices. Una forma cuadrática $Q : \mathbb{R}^n \to \mathbb{R}$ es *semidefinida positiva* si

$$Q(v) \ge 0, \qquad \forall v \in \mathbb{R}^n,$$

y *definida positiva* si

$$Q(v) > 0, \qquad \forall v \ne 0, \quad v \in \mathbb{R}^n.$$

Análogamente, Q es semidefinida negativa o definida negativa si $-Q$ es semidefinida positiva o definida positiva, respectivamente. Finalmente, se dice que Q es *indefinida* si no es semidefinida positiva ni negativa, es decir si existen dos vectores $v_1, v_2 \in \mathbb{R}^n$ tales que

$$Q(v_1) > 0, \quad Q(v_2) < 0.$$

Si Q_0 es la matriz de Q en la base canónica de $\mathbb{R}^n$, los autovalores de Q son por definición los autovalores de Q_0. La matriz Q_0 es una matriz cuadrada simétrica real, y por tanto es diagonalizable en una base ortonormal y todos sus autovalores son reales. Al hacer un cambio lineal de coordenadas de matriz A, la matriz de Q se transforma en $A^\top \cdot Q_0 \cdot A$; por tanto, los autovalores de Q se pueden calcular de hecho a partir de su matriz en cualquier base *ortonormal* de $\mathbb{R}^n$. El número de autovalores (contando la multiplicidad) positivos, negativos y cero de Q es invariante bajo cambios lineales de coordenadas arbitrarios, es decir se puede calcular a partir de la matriz de Q en cualquier base. Una forma cuadrática Q es *degenerada* si su matriz (en cualquier base) no es invertible, es decir si

$$\det Q_0 = 0.$$

Equivalentemente, Q es degenerada si 0 es autovalor de Q.

En términos de los autovalores, Q es semidefinida positiva (negativa) si todos sus autovalores son no negativos (no positivos), y definida positiva (negativa) si son estrictamente positivos (negativos). En la práctica, para averiguar si una forma cuadrática es definida positiva (sobre todo en dimensión baja) se suele utilizar el siguiente criterio, debido a Sylvester: Q es definida positiva si y sólo si todos los *menores principales*

$$d_k = \det(q_{ij})_{1\le i,j\le k}, \qquad 1 \le k \le n,$$

de su matriz $Q_0 = (q_{ij})_{1\le i,j\le n}$ son estrictamente positivos. Evidentemente, cambiando Q por $-Q$ se obtiene un criterio análogo para formas cuadráticas definidas negativas: Q es definida negativa si y sólo si

$$\operatorname{sign} d_k = (-1)^k, \qquad 1 \le k \le n.$$

Este criterio es particularmente fácil de formular en el caso $n = 2$: la condición necesaria y suficiente para que Q sea definida positiva en este caso se reduce a

$$q_{11} > 0, \quad \det Q_0 > 0,$$

mientras que la condición necesaria y suficiente para que Q sea definida negativa es

$$q_{11} < 0, \quad \det Q_0 > 0.$$

Nótese que si $\det Q_0 > 0$, entonces $q_{11} > 0 \iff q_{22} > 0 \iff \operatorname{tr} Q_0 > 0$. En particular, de lo anterior se deduce que una forma cuadrática en dimensión dos con determinante negativo es necesariamente indefinida.

Aplicando la Proposición 4.14 al caso $m = 2$, obtenemos el siguiente resultado:

Proposición 4.15. *Sea x_0 un punto crítico de $f : \mathbb{R}^n \to \mathbb{R}$, sea f de clase C^2 en un entorno de x_0, y supongamos que $D^2 f(x_0) \neq 0$. Entonces se cumple:*

I) *Si la forma cuadrática $D^2 f(x_0) \cdot h^2$ es definida positiva, entonces x_0 es un mínimo relativo estricto de f*

II) *Si $D^2 f(x_0) \cdot h^2$ es definida negativa, entonces x_0 es un máximo relativo estricto de f*

iii) *Si* $D^2 f(x_0) \cdot h^2$ *es indefinida, entonces* x_0 *es un punto de silla de* f

No se puede afirmar nada *a priori* acerca de la naturaleza del punto crítico x_0 cuando $D^2 f(x_0) \cdot h^2$ es semidefinida, es decir cuando $Hf(x_0)$ tiene algún autovalor nulo, y todos los autovalores distintos de cero tienen el mismo signo. En este caso, el determinante $\det Hf(x_0)$ se anula, y se dice entonces que x_0 es un *punto crítico degenerado*. Si x_0 es un punto crítico no degenerado, la proposición anterior permite decidir sin ambigüedad su naturaleza. Nótese, sin embargo, que cuando x_0 es un punto crítico degenerado, pero hay al menos dos autovalores de signos opuestos, se puede aplicar la proposición anterior para concluir que x_0 es un punto de silla de f.

En el caso $n = 2$ la proposición anterior admite la siguiente formulación sencilla:

Proposición 4.16. *Sea* (x_0, y_0) *un punto crítico de una función* $f : \mathbb{R}^2 \to \mathbb{R}$, *sea* f *de clase* C^2 *en un entorno de* (x_0, y_0), *y supongamos que* $Hf(x_0, y_0) \neq 0$. *Entonces se verifica:*

i) *Si*

$$\det Hf(x_0, y_0) = f_{xx}(x_0, y_0) f_{yy}(x_0, y_0) - f_{xy}(x_0, y_0)^2 < 0,$$

entonces (x_0, y_0) *es un punto de silla de* f

ii) *Si*

$$f_{xx}(x_0, y_0) > 0, \quad \det Hf(x_0, y_0) > 0,$$

entonces (x_0, y_0) *es un mínimo relativo estricto de* f

iii) *Si*

$$f_{xx}(x_0, y_0) < 0, \quad \det Hf(x_0, y_0) > 0,$$

entonces (x_0, y_0) *es un máximo relativo estricto de* f

En esto caso, cuando (x_0, y_0) es un punto crítico degenerado $D^2 f(x_0, y_0)$ es necesariamente semidefinida, por lo que la Proposición 4.15 no dice nada sobre la naturaleza de dicho punto crítico.

Veamos, para finalizar este capítulo, un ejemplo sencillo que ilustra todo lo anterior:

Ejemplo 4.17. Sea $f : \mathbb{R}^2 \to \mathbb{R}$ definida por $f(x, y) = x^3 - 3xy^2 - 2y^2 + x^4$, $\forall (x, y) \in \mathbb{R}^3$. Calculemos, en primer lugar, los puntos críticos de f. Para ello, hay que resolver el sistema

$$f_x(x, y) = 3x^2 - 3y^2 + 4x^3 = 0$$
$$f_y(x, y) = -2y(3x + 2) = 0.$$

De la segunda ecuación obtenemos que $y = 0$ ó $x = -2/3$. Si $y = 0$, sustituyendo en la primera ecuación se obtiene

$$x = 0 \quad \text{ó} \quad x = -\frac{3}{4}.$$

Por otra parte, si $x = -2/3$ la primera ecuación proporciona

$$y = \pm\frac{2}{9}.$$

Por tanto, f tiene exactamente cuatro puntos críticos:

$$(0,0), \quad \left(-\frac{3}{4}, 0\right), \quad \left(-\frac{2}{3}, -\frac{2}{9}\right), \quad \left(-\frac{2}{3}, \frac{2}{9}\right),$$

cuya naturaleza vamos a determinar a continuación.

En general, el hessiano de f es igual a

$$Hf(x,y) = \begin{pmatrix} 6x + 12x^2 & -6y \\ -6y & -6x - 4 \end{pmatrix}.$$

En el origen

$$Hf(0,0) = \begin{pmatrix} 0 & 0 \\ 0 & -4 \end{pmatrix},$$

de donde se sigue que $(0,0)$ es un punto crítico degenerado, al que no se puede aplicar la proposición anterior. Sin embargo, en este caso es fácil ver que el origen es un punto de silla de f, ya que por ejemplo

$$f(x,0) = x^3 + x^4,$$

que tiene un punto de inflexión en $x = 0$.

En el segundo punto crítico

$$Hf\left(-\frac{3}{4}, 0\right) = \begin{pmatrix} \frac{9}{4} & 0 \\ 0 & \frac{1}{2} \end{pmatrix},$$

de donde se sigue que $(-3/4, 0)$ es un mínimo relativo estricto de f. Finalmente, en los dos puntos críticos restantes

$$Hf\left(-\frac{2}{3}, \pm\frac{2}{9}\right) = \frac{4}{3}\begin{pmatrix} 1 & \mp 1 \\ \mp 1 & 0 \end{pmatrix}$$

tiene determinante negativo, por lo que ambos puntos críticos son puntos de silla.

Aunque f tiene un mínimo relativo estricto en $(-3/4, 0)$, es fácil ver que f no tiene extremos absolutos en $\mathbb{R}^2$. En efecto, es claro que f no puede tener ningún máximo absoluto, pues no tiene máximos relativos. Por otra parte,

$$\lim_{y\to\infty} f(1,y) = \lim_{y\to\infty} 2 - 5y^2 = -\infty,$$

por lo que f no tiene tampoco nigún mínimo absoluto. Nótese que, para funciones diferenciables de una variable, la existencia de un mínimo local y tres puntos de silla implica que el mínimo local es de hecho un mínimo absoluto. (La diferencia con en el caso multidimensional es que en este último en un punto de silla hay direcciones a lo largo de las cuales f crece y otras en las que f decrece.)

Capítulo 5

Variedades diferenciables. Extremos condicionados

5.1. Subvariedades diferenciables de $\mathbb{R}^n$

En términos generales y poco precisos, una variedad diferenciable de dimensión r es un espacio topológico abstracto M que es "localmente equivalente" a (un abierto de) $\mathbb{R}^r$ en cada punto (en particular, M ha de ser localmente homeomorfo a $\mathbb{R}^r$). En esta sección sólo estamos interesados en un tipo especial de variedades diferenciables, las subvariedades diferenciables de $\mathbb{R}^n$, que como veremos a continuación generalizan el concepto de hipersuperficie de nivel visto en el Capítulo 2:

Definición 5.1. Sea $1 \leq r < n$, $q \geq 1$. Una *subvariedad diferenciable de* $\mathbb{R}^n$ *de dimensión* r *y clase* q es un subconjunto $M \subset \mathbb{R}^n$ con la siguiente propiedad: para todo punto $a \in M$ hay un entorno U de a y una función $\Phi : \mathbb{R}^n \to \mathbb{R}^{n-r}$ de clase $C^q(U)$ que cumple

$$\operatorname{rank} D\Phi(x) = n - r, \qquad \forall x \in U,$$

y

$$U \cap M = \big\{x \in U : \Phi(x) = 0\big\}.$$

Nótese que tanto el entorno U como la función Φ en esta definición dependen en general del punto $a \in M$. La definición anterior no tiene sentido si $r = n$. En este caso, diremos que $M \subset \mathbb{R}^n$ es una subvariedad de $\mathbb{R}^n$ de dimensión n si M es un abierto de $\mathbb{R}^n$.

Si $M \subset \mathbb{R}^n$ es una subvariedad diferenciable de dimensión $r < n$, por definición M se puede describir localmente en un entorno de cualquier punto $a \in M$ por $n - r$ ecuaciones implícitas funcionalmente independientes

$$\Phi_i(x) = 0, \qquad 1 \leq i \leq n - r.$$

El teorema de la función implícita permite probar que, en un entorno de a, M se puede parametrizar en función de r parámetros independientes. En efecto, sea $a \in M$ y U, Φ como en la definición de subvariedad diferenciable, y supongamos que por ejemplo

$$\det \left(\frac{\partial \Phi_i}{\partial x_j}(a) \right)_{\substack{1 \le i \le n-r \\ r+1 \le j \le n}} \neq 0.$$

Escribamos $a = (\hat{a}, \tilde{a}) \in \mathbb{R}^r \times \mathbb{R}^{n-r}$. Por el teorema de la función implícita aplicado a Φ en $a \in U$, hay un entorno $W \subset U$ de a en $\mathbb{R}^n$, un abierto $V \subset \mathbb{R}^r$ que contiene a $\hat{a}$ y una función $\psi : V \to \mathbb{R}^{n-r}$ de clase C^q en V tal que

$$M \cap W = \big\{ \big(u, \psi(u)\big) : u \in V \big\}.$$

En otras palabras, existe un entorno W de a en el cual la subvariedad M está descrita por las ecuaciones paramétricas

$$\begin{aligned} x_1 &= u_1 \\ &\dots \\ x_r &= u_r \\ x_{r+1} &= \psi_1(u_1, \dots, u_r) \\ &\dots \\ x_n &= \psi_{n-r}(u_1, \dots, u_r). \end{aligned}$$

La demostración del teorema de la función implícita también permite entender en qué sentido una subvariedad diferenciable r-dimensional de $\mathbb{R}^n$ es localmente equivalente a un abierto de $\mathbb{R}^r$ en cada punto. En efecto, la función de clase C^q $\Psi : V \subset \mathbb{R}^r \to \mathbb{R}^n$ definida por $\Psi(u) = (u, \psi(u))$ establece una biyección entre $M \cap W$ y el abierto V de $\mathbb{R}^r$.

Ejemplo 5.2. El conjunto

$$M = \big\{ (x, y) : x^2 - y^2 = 1 \big\} \cup \big\{ (x, y) : x = y \big\} \equiv M_1 \cup M_2$$

es una subvariedad diferenciable de dimensión 1 y clase ∞ de $\mathbb{R}^2$. En efecto, si $(x_0, y_0) \in M_1$ sea U un entorno de (x_0, y_0) que no corte a M_2, y tomemos $\Phi(x, y) = x^2 - y^2 - 1$. Entonces

$$U \cap M = U \cap M_1 = \big\{ (x, y) \in U : \Phi(x, y) = 0 \big\},$$

Φ es de clase C^∞ en U, y el rango de $D\Phi$ es igual a 1 en U, ya que $\nabla \Phi$ sólo se anula en $(0, 0) \in M_2$. Análogamete, si $(x_0, y_0) \in M_2$ basta tomar como U un entorno de (x_0, y_0) que no corte a M_1, y definir $\Phi(x, y) = x - y$. Nótese que en este ejemplo tanto U como Φ dependen del punto $(x_0, y_0) \in M$ considerado.

Ejemplo 5.3. El conjunto

$$\{(x,y) : x^2 - y^2 = 0\}$$

no es una subvariedad diferenciable de $\mathbb{R}^2$. En efecto, no es una subvariedad diferenciable de dimensión 2, ya que no es abierto (es cerrado, y $\mathbb{R}^2$ es conexo). Tampoco es una subvariedad de dimensión 1, ya que, por lo visto anteriormente, si lo fuera existirían un entorno W de $(0,0)$, un abierto $V \subset \mathbb{R}$ que contiene a 0 y una función $\psi : \mathbb{R} \to \mathbb{R}$ tales que

$$W \cap M = \{(x, \psi(x)) : x \in V\}$$

ó

$$W \cap M = \{(\psi(y), y) : x \in V\},$$

lo cual evidentemente no ocurre.

Una forma muy habitual de construir subvariedades diferenciables de $\mathbb{R}^n$ consiste en tomar una función $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ (con $1 \leq m < n$) de clase C^q (con $q \geq 1$) en un abierto $D \subset \mathbb{R}^n$, y definir

$$M = \{x \in D : \Phi(x) = 0 \text{ y } \operatorname{rank} D\Phi(x) = m\}.$$

Si M no es vacío, entonces es fácil ver que M es una subvariedad diferenciable de $\mathbb{R}^n$ de dimensión $n - m$ y clase q. En efecto, si $a \in M$ entonces $\operatorname{rank} D\Phi(a) = m$, por lo que existirá un menor μ de orden m de $J\Phi$ no nulo en a. Como Φ es de clase $q \geq 1$ en D, las derivadas parciales de primer orden de Φ son de clase $q - 1 \geq 0$ en D, y por tanto μ es una función continua en D. El conjunto de puntos en que no se anula una función continua es un abierto, por lo que debe existir un entorno $U \subset D$ de a en que μ no se anula. En particular, $\operatorname{rank} D\Phi(x) = m$ para todo $x \in U$. Con esta elección de U y Φ se satisfacen las condiciones de la definición de subvariedad diferenciable, con $r = n - m$. Llamaremos a M la *subvariedad diferenciable definida por la función* Φ. Nótese que en este tipo de subvariedades el abierto U depende del punto, mientras que Φ es la misma para todos los puntos de M. En particular, dada $f : \mathbb{R}^n \to \mathbb{R}$ de clase q la subvariedad diferenciable de dimensión $n - 1$ y clase q definida por la función $f - c$ es la hipersuperficie de nivel $L_c(f)$ menos el conjunto de puntos críticos de f.

Por ejemplo, la esfera de centro a y radio r en $\mathbb{R}^n$

$$S_r(a; n) = \{x \in \mathbb{R}^n : |x - a| = r\}$$

es la subvariedad de dimensión $n - 1$ y clase ∞ definida por la función $\Phi : \mathbb{R}^n \to \mathbb{R}$ dada por

$$\Phi(x) = \sum_{i=1}^{n} (x_i - a_i)^2 - r^2.$$

En efecto, Φ es de clase C^∞, $S_r(a; n)$ es el conjunto de ceros de Φ y el rango de Φ es uno en $S_r(a; n)$, ya que $\nabla\Phi$ sólo se anula en el origen.

5.2. Vectores tangentes y normales a una subvariedad

Sea M una subvariedad diferenciable de $\mathbb{R}^n$, y sea $a \in M$. Un vector $v \in \mathbb{R}^n$ es *tangente a M en a* si existe una curva diferenciable $\gamma : (-\delta, \delta) \to M$ contenida en M tal que

$$\gamma(0) = a, \qquad \gamma'(0) = v. \tag{5.1}$$

Esta definición es totalmente análoga a la vista en el Capítulo 2 para una hipersuperficie de $\mathbb{R}^n$. Llamaremos *espacio tangente a M en a* al conjunto

$$T_a(M) = \big\{ v \in \mathbb{R}^n : v \text{ es un vector tangente a } M \text{ en } a \big\}$$

formado por todos los vectores tangentes a M en a.

Proposición 5.4. *Sea M una subvariedad diferenciable de dimensión r de $\mathbb{R}^n$, sea $a \in M$, y sean U y Φ como en la definición de subvariedad diferenciable. Entonces*

$$T_a(M) = \ker D\Phi(a). \tag{5.2}$$

En particular, $T_a(M)$ es un subespacio de dimensión r de $\mathbb{R}^n$.

Demostración. Obsérvese que si probamos (5.2) entonces $T_a(M)$ es un subespacio vectorial de $\mathbb{R}^n$, y de la fórmula

$$\dim \ker D\Phi(a) + \operatorname{rank} D\Phi(a) = \dim \mathbb{R}^n = n$$

y la condición $\operatorname{rank} D\Phi(a) = n - r$ se sigue automáticamente que $\dim T_a(M) = r$. Por tanto, basta probar (5.2).

En primer lugar, veamos que $T_a(M) \subset \ker D\Phi(a)$. En efecto, si $v \in T_a(M)$ y la curva $\gamma : (-\delta, \delta) \to M$ satisface las condiciones (5.1) entonces $\Phi \circ \gamma = 0$ en un entorno de $t = 0$, por lo que

$$0 = (\Phi \circ \gamma)'(0) = D\Phi(a) \cdot \gamma'(0) = D\Phi(a) \cdot v.$$

Veamos ahora que también $\ker D\Phi(a) \subset T_a(M)$. En otras palabras, dado un vector $v \in \ker D\Phi(a)$ hay que construir una curva $\gamma : (-\delta, \delta) \to M$ que cumpla las condiciones (5.1). Para ello, si $x \in \mathbb{R}^n$ escribamos $x = (\hat{x}, \tilde{x})$, con $\hat{x} \in \mathbb{R}^r$ y $\tilde{x} \in \mathbb{R}^{n-r}$, y sea

$$\frac{\partial \Phi}{\partial \tilde{x}} = \left(\frac{\partial \Phi_i}{\partial x_j} \right)_{\substack{1 \le i \le n-r \\ r+1 \le j \le n}}.$$

Por lo visto en la sección anterior, podemos suponer que $\frac{\partial \Phi}{\partial \tilde{x}}(a)$ es invertible. Entonces hay un entorno W de a en $\mathbb{R}^n$, un abierto $V \subset \mathbb{R}^r$ que contiene a $\hat{a}$ y una función $\psi : \mathbb{R}^r \to \mathbb{R}^{n-r}$ de clase q en V (si $\Phi \in C^q(U)$) tales que

$$M \cap W = \big\{ \big(u, \psi(u)\big) : u \in V \big\}.$$

Como la imagen de γ tiene que estar contenida en M, escribiremos

$$\gamma(t) = (g(t), \psi(g(t))),$$

donde $g : (-\delta, \delta) \to \mathbb{R}^r$ es una función diferenciable que hay que calcular de forma que γ satisfaga las condiciones (5.1). La primera condición implica que

$$g(0) = \hat{a}.$$

Al ser

$$\gamma'(t) = (g'(t), D\psi(g(t))) \cdot g'(t)),$$

la condición sobre $\gamma'(0)$ es equivalente al par de ecuaciones

$$\begin{aligned} g'(0) &= \hat{v} \\ D\psi(\hat{a}) \cdot \hat{v} &= \tilde{v}. \end{aligned}$$

En particular, si tomamos $g(t) = \hat{a} + t\,\hat{v}$, todo se reduce a probar que se cumple la última de las igualdades anteriores. Para ello, utilizaremos el hecho de que, por construcción de γ,

$$\Phi(\gamma(t)) = 0, \qquad \forall t \in (-\delta, \delta).$$

Derivando esta igualdad en $t = 0$ obtenemos

$$0 = D\Phi(a) \cdot \gamma'(0) = D\Phi(a) \cdot (\hat{v}, D\psi(\hat{a}) \cdot \hat{v}).$$

Por otra parte, como $D\Phi(a) \cdot v = D\Phi(a) \cdot (\hat{v}, \tilde{v}) = 0$ restando obtenemos

$$D\Phi(a) \cdot (0, D\psi(\hat{a}) \cdot \hat{v} - \tilde{v}) = 0,$$

o lo que es lo mismo (con un ligero abuso de notación)

$$\frac{\partial \Phi}{\partial \tilde{x}}(a) \cdot (D\psi(\hat{a}) \cdot \hat{v} - \tilde{v}).$$

Como la matriz $\frac{\partial \Phi}{\partial \tilde{x}}(a)$ es invertible, esto implica que

$$D\psi(\hat{a}) \cdot \hat{v} - \tilde{v} = 0,$$

como habíamos afirmado. *Q.E.D.*

Si $M \subset \mathbb{R}^n$ es una subvariedad diferenciable de dimensión r y $a \in M$, diremos que $v \in \mathbb{R}^n$ es un *vector normal a M en a* si v es ortogonal a todo vector tangente a M en a, es decir si

$$v \cdot h = 0, \qquad \forall h \in T_a(M).$$

El conjunto $N_a(M)$ de todos los vectores normales a M en a es pues el complemento ortogonal $T_a(M)^\perp$ de $T_a(M)$; en particular,

$$\dim N_a(M) = n - r. \tag{5.3}$$

Si Φ es como en la definición de subvariedad diferenciable, entonces

$$\begin{aligned} h \in T_a(M) &\iff D\Phi(a) \cdot h = 0 \\ &\iff J\Phi(a) \cdot h = 0, \qquad 1 \le i \le n-r \\ &\iff \nabla\Phi_i(a) \cdot h = 0, \qquad 1 \le i \le n-r. \end{aligned}$$

Por tanto, los $n - r$ vectores

$$N_i = \nabla\Phi_i(a), \qquad 1 \le i \le n-r, \tag{5.4}$$

son normales a M en a. Dichos vectores son linealmente independientes, pues no son otra cosa que (los vectores correspondientes a) las $n-r$ filas de la matriz $D\Phi(a)$, que por construcción tiene rango $n-r$. De (5.3) se deduce entonces que los vectores (5.4) forman una base de $N_a(M)$. En particular, si M es una subvariedad $(n-1)$-dimensional de $\mathbb{R}^n$ el vector $\nabla\Phi(a)$ genera el subespacio normal a M en a, que es de dimensión uno en este caso.

Como en el caso de las hipersuperficies de nivel, dada una subvariedad diferenciable $M \subset \mathbb{R}^n$ de dimensión r y un punto $a \in M$, definimos el *r-plano tangente a M en a* como el r-plano $P_a(M)$ que pasa por a y es paralelo a $T_a(M)$, es decir

$$\begin{aligned} P_a(M) = a + T_a(M) &= \big\{x \in \mathbb{R}^n : x - a \in T_a(M)\big\} \\ &= \big\{x \in \mathbb{R}^n : D\Phi(a) \cdot (x-a) = 0\big\}. \end{aligned}$$

Equivalentemente,

$$P_a(M) = \big\{x \in \mathbb{R}^n : \nabla\Phi_i(a) \cdot (x-a) = 0,\ 1 \le i \le n-r\big\}.$$

5.3. Multiplicadores de Lagrange

Sea M una subvariedad diferenciable de $\mathbb{R}^n$, y sea $f : \mathbb{R}^n \to \mathbb{R}$ una función diferenciable en algún abierto que contenga a M. En esta sección nos ocuparemos de desarrollar un método para calcular los extremos de f en M (lo que clásicamente se denomina un problema de *extremos condicionados*). Por lo visto en el capítulo anterior, los extremos relativos de f en M no tienen por qué ser puntos críticos de f, lo cual no permite aplicar los métodos desarrollados en dicho capítulo en este caso. Cuando M es fácilmente parametrizable, es decir si podemos expresar M como

$$M = \big\{\Psi(u) : u \in V\big\}$$

(siendo $V \subset \mathbb{R}^r$ abierto y $\Psi : V \to M$ biyectiva y de clase $C^1(V)$) entonces los extremos relativos de f en M son simplemente los extremos relativos de la función $f \circ \Psi$ en el abierto V, que se calculan a partir de los puntos críticos de $f \circ \Psi$. El problema es que muchas veces, en la práctica, no es posible encontrar una parametrización sencilla de M. Interesa hallar por tanto una condición necesaria de extremo, análoga a la anulación de Df para problemas de extremos no condicionados, que no exija encontrar explícitamente una parametrización de M.

Para resolver este problema, sea r la dimensión de M, y sea $a \in M$ un extremo relativo de f en M. Sea $\gamma : (-\delta, \delta) \to M$ una curva diferenciable contenida en M que pase por a con vector tangente h en $t = 0$, y sea Φ como en la definición de subvariedad diferenciable. Como f tiene un extremo relativo en a, la función $f \circ \gamma$, definida en un entorno de $t = 0$, tiene un extremo relativo en $t = 0$. Por la regla de la cadena se tiene entonces

$$0 = (f \circ \gamma)'(0) = Df(a) \cdot \gamma'(0) = Df(a) \cdot h.$$

Como cualquier vector $h \in T_a(M)$ puede obtenerse por definición como el vector tangente en a a una curva diferenciable γ contenida en M, hemos probado que si f tiene un extemo relativo en $a \in M$ entonces

$$Df(a) \cdot h = 0, \qquad \forall h \in T_a(M).$$

Esta es pues una condición necesaria para que f tenga un extremo relativo en a. Sin embargo, esta condición es un poco engorrosa todavía, por lo que procederemos a simplificarla utilizando los resultados de la sección anterior. En efecto, dicha ecuación se puede reescribir como

$$\nabla f(a) \cdot h = 0, \qquad \forall h \in T_a(M),$$

lo que es equivalente a la condición $\nabla f(a) \in N_a(M)$. A su vez, esto es lo mismo que decir que el vector $\nabla f(a)$ es una combinación lineal de los vectores $N_i = \nabla \Phi_i(a)$, $1 \leq i \leq n - r$. Por tanto, existirán $n - r$ escalares $\mu_1, \ldots, \mu_{n-r}$ tales que

$$\nabla f(a) = \sum_{i=1}^{n-r} \mu_i \nabla \Phi_i(a),$$

o bien

$$\nabla \left(f + \sum_{i=1}^{n-r} \lambda_i \Phi_i \right)(a) = 0,$$

donde hemos puesto $\lambda_i = -\mu_i$. Por tanto, hemos probado el siguiente resultado:

Teorema 5.5. *Sea M una subvariedad diferenciable de dimensión r de $\mathbb{R}^n$, sea Φ como en la definición de subvariedad diferenciable, y sea $f : \mathbb{R}^n \to \mathbb{R}$ una función diferenciable en un abierto que contenga a M. Si $a \in M$ es un extremo relativo de f en M, existen $n - r$ escalares $\lambda_1, \dots, \lambda_{n-r}$ tales que la función $F = f + \sum_{i=1}^{n-r} \lambda_i \Phi_i$ tiene un punto crítico en a.*

Este teorema es especialmente útil si, como ocurre en la mayor parte de las aplicaciones, M es la subvariedad definida por una función $\Phi : \mathbb{R}^n \to \mathbb{R}^m$, pues entonces Φ es conocida y es la misma para todos los puntos de M. En tal caso, para encontrar los extremos relativos de f en M hay que encontrar primero todos los puntos $a \in M$ que son puntos críticos de la función auxiliar $F = f + \sum_{i=1}^{m} \lambda_i \Phi_i$ para algún valor de los parámetros λ_i. Con un cierto abuso de lenguaje, llamaremos a estos puntos *puntos críticos de f restringida a M*. Para hallar dichos puntos, hay que resolver el siguiente sistema de $n+m$ ecuaciones

$$\frac{\partial F}{\partial x_i}(a) = \frac{\partial f}{\partial x_i}(a) + \sum_{j=1}^{m} \lambda_j \frac{\partial \Phi_j}{\partial x_i}(a) = 0, \qquad 1 \le i \le n; \tag{5.5}$$

$$\Phi_k(a) = 0, \qquad 1 \le k \le m \tag{5.6}$$

en las $n + m$ incógnitas $a_1, \dots, a_n, \lambda_1, \dots, \lambda_m$ (con la condición adicional $\operatorname{rank} D\Phi(a) = m$). A los m parámetros auxiliares λ_i se les llama *multiplicadores de Lagrange*, y el método que acabamos de exponer para encontrar los extremos de f en M recibe el nombre de *método de los multiplicadores de Lagrange.*

Ejemplo 5.6. Hallemos los extremos relativos de la función $f : \mathbb{R}^2 \to \mathbb{R}$ definida por $f(x, y) = x^2 + y^2$ en la curva

$$C = \big\{(x, y) : x^3 + y^3 - 6xy = 0\big\}.$$

Obsérvese que en este caso no hay una forma sencilla de parametrizar M, por lo que es prácticamente obligado aplicar el método de los multiplicadores de Lagrange que acabamos de ver.

En primer lugar, veamos si C es la subvariedad diferenciable M de dimensión 1 definida por la función $\Phi(x, y) = x^3 + y^3 - 6xy$. Para ello, basta recordar que

$$M = C - \big\{(x, y) : \nabla\Phi(x, y) = (0, 0)\big\}.$$

de donde se sigue fácilmente que

$$M = C - \{(0, 0), (2, 2)\} = C - \{(0, 0)\},$$

pues $(2, 2) \notin C$. Nótese que el punto $(0, 0)$ es obviamente un mínimo absoluto estricto de f en C. Por lo tanto, los restantes extremos relativos de f en C

hay que buscarlos en el conjunto de los puntos críticos de f restringida a la subvariedad diferenciable M.

Aplicando el método de los multiplicadores de Lagrange, buscamos los extremos auxiliares de la función $F(x,y) = f(x,y) + \lambda\Phi(x,y)$ que satisfacen la condición $\Phi(x,y) = 0$ (y $(x,y) \neq (0,0)$). Debemos por tanto resolver el siguiente sistema de 3 ecuaciones en las tres incógnitas (x,y,λ):

$$2x + \lambda(3x^2 - 6y) = 0 \tag{5.7}$$

$$2y + \lambda(3y^2 - 6x) = 0 \tag{5.8}$$

$$x^3 + y^3 - 6xy = 0. \tag{5.9}$$

De las dos primeras ecuaciones deducimos que

$$x(y^2 - 2x) = y(x^2 - 2y) \iff 2(x-y)(x+y) = xy^2 - yx^2 = xy(y-x),$$

o equivalentemente

$$(x-y)(2x + 2y + xy) = 0 \quad \iff \quad x = y \quad \text{ó} \quad 2x + 2y + xy = 0.$$

Si $x = y$, sustituyendo en la última ecuación del sistema obtenemos la ecuación

$$x^3 - 3x^2 = 0,$$

que proporciona los puntos críticos $(0,0)$ (que no pertenece a M, pero ya hemos tenido en cuenta) y $(3,3)$. Si $x \neq y$, entonces

$$2x + 2y + xy = 0 \iff y = -\frac{2x}{x+2}.$$

Sustituyendo esta relación en la condición $\Phi(x,y) = 0$ y quitando denominadores, se demuestra fácilmente que la ecuación resultante sólo tiene la raiz real $x = 0$, que conduce de nuevo al punto $(0,0)$. Por tanto, los únicos puntos de C en los cuáles f puede tener un extremo relativo son $(0,0)$ (que es un mínimo absoluto estricto) y $(3,3)$. Por consideraciones geométricas (representando esquemáticamente la curva C) se puede ver que $(3,3)$ es un máximo relativo estricto, pero no absoluto, de f en C.

Si a es un punto crítico de la función f restringida a la subvariedad diferenciable M, existe también en este caso un criterio, análogo al de la derivada segunda visto en el capítulo anterior, que permite determinar en ciertos casos si a es un extremo relativo de f en M. Concretamente, se tiene:

Proposición 5.7. *Sea M una subvariedad diferenciable de $\mathbb{R}^n$ de dimensión r y clase $q \geq 2$, sea $f : \mathbb{R}^n \to \mathbb{R}$ una función de clase C^2 en un abierto de $\mathbb{R}^n$ que contenga a M, y sea Φ como en la definición de subvariedad diferenciable. Si $a \in M$ es un punto crítico de f restringida a M, sean $\lambda_1, \dots, \lambda_{n-r}$ los multiplicadores de Lagrange asociados a este punto, y sea $F = f + \sum_{i=1}^{n-r} \lambda_i \Phi_i$. Entonces se cumple:*

i) *Si*

$$D^2F(a) \cdot h^2 > 0, \qquad \forall h \in T_a(M), \quad h \neq 0,$$

entonces a es un mínimo relativo estricto de f en M

ii) *Si*

$$D^2F(a) \cdot h^2 < 0, \qquad \forall h \in T_a(M), \quad h \neq 0,$$

entonces a es un máximo relativo estricto de f en M

iii) *Si la forma cuadrática $D^2F(a) \cdot h^2$ es indefinida en $T_a(M)$, es decir*

$$\exists h_1, h_2 \in T_a(M) \quad \textit{tales que} \quad D^2F(a) \cdot (h_1)^2 < 0 < D^2F(a) \cdot (h_2)^2,$$

entonces a es un punto de silla de f en M.

Demostración. La demostración se basa en el siguiente lema sencillo: si $\gamma : (-\delta, \delta) \to \mathbb{R}^n$ es cualquier curva de clase C^2 tal que $\gamma(0) = a$ y $\gamma'(0) = h$, entonces

$$D^2F(a) \cdot h^2 = (F \circ \gamma)''(0). \tag{5.10}$$

En efecto, F es de clase C^2 por serlo f y Φ. Aplicando la regla de la cadena dos veces a $f \circ \gamma$ se obtiene

$$\begin{aligned}
(F \circ \gamma)'(t) &= DF(\gamma(t)) \cdot \gamma'(t) = \sum_{i=1}^{n} D_iF(\gamma(t))\gamma_i'(t) \\
(F \circ \gamma)''(t) &= \sum_{i=1}^{n} D_iF(\gamma(t))\gamma_i''(t) + \sum_{i=1}^{n} D_{ij}F(\gamma(t))\gamma_i'(t)\gamma_j'(t) \\
&= DF(\gamma(t))\gamma''(t) + D^2F(\gamma(t)) \cdot (\gamma'(t))^2.
\end{aligned}$$

Evaluando esta última igualdad en $t = 0$, y teniendo en cuenta que $DF(a) = 0$, se obtiene (5.10). Nótese que la anulación de DF en a es fundamental; en particular, (5.10) no es válida en general para f, pues $Df(a)$ no tiene por qué anularse.

Para probar la última parte, probaremos que si a es, por ejemplo, un mínimo relativo de f en M entonces

$$D^2f(a) \cdot h^2 \geq 0, \qquad \forall h \in T_a(M).$$

Si $h \in T_a(M)$, sea $\gamma : (-\delta, \delta) \to M$ una curva de clase C^2 tal que $\gamma(0) = a$ y $\gamma'(0) = h$. (Siempre se puede encontrar una curva de clase C^2 que cumpla esto; por ejemplo, utilícese la representación paramétrica de M en un entorno de a.) Entonces la función $f \circ \gamma = F \circ \gamma$ tiene un mínimo relativo en $t = 0$. Por tanto, se ha de cumplir la condición

$$0 \leq (F \circ \gamma)''(0) = D^2F(a) \cdot h^2,$$

donde hemos aplicado (5.10).

Veamos ahora que si, por ejemplo, se cumple la condición i), entonces a es un mínimo relativo estricto de f en M. En efecto, si $\psi : V \subset \mathbb{R}^n$ es la función que da la representación paramétrica de M en un entorno W de a mediante

$$M \cap W = \{(u, \psi(u)) : u \in V\},$$

y $\Psi(u) = (u, \psi(u))$, basta probar que la función $\varphi = f \circ \Psi = F \circ \Psi$ tiene un mínimo relativo estricto en $\hat{a}$. En primer lugar, $\hat{a}$ es un punto crítico de φ, ya que

$$D\varphi(\hat{a}) = DF(a) \cdot D\Psi(\hat{a}) = 0.$$

A continuación aplicaremos el criterio de la derivada segunda a φ, que es de clase C^2 por serlo F y ψ. Para ello, vamos a probar que

$$D^2\varphi(\hat{a}) \cdot v^2 > 0, \qquad \forall v \in \mathbb{R}^r, \quad v \neq 0.$$

En efecto, sea $v \in \mathbb{R}^r$, y sea $g : (-\delta, \delta) \to \mathbb{R}^r$ una curva cualquiera de clase C^2 tal que $g(0) = \hat{a}$ y $g'(0) = v$. Si $\gamma = \Psi \circ g$, entonces γ es una curva contenida en M que cumple $\gamma(0) = a$, y por tanto $h = \gamma'(0) \in T_a(M)$. Aplicando (5.10) a φ (que también cumple $D\varphi(\hat{a}) = 0$) y F obtenemos

$$D^2\varphi(\hat{a}) \cdot v^2 = (\varphi \circ g)''(0) = (F \circ \gamma)''(0) = D^2F(a) \cdot h^2 \geq 0.$$

Por último, para probar que la desigualdad anterior es estricta si $v \neq 0$, basta observar que en tal caso se cumple

$$h = \gamma'(0) = D\Psi(\hat{a}) \cdot v = (v, D\psi(\hat{a}) \cdot v) \neq 0.$$

Por tanto, el criterio de la derivada segunda implica que $f \circ \Psi = F \circ \Psi$ tiene un mínimo relativo estricto en $\hat{a}$, como habíamos afirmado. *Q.E.D.*

En relación con la proposición anterior, es importante darse cuenta de dos detalles. En primer lugar, la forma cuadrática que hay que estudiar es la derivada segunda de la función auxiliar F, en lugar de la de f. Esto es debido a que F, pero no necesariamente f, tiene un punto crítico en a. En segundo lugar, lo que hay que estudiar no es simplemente la forma cuadrática $D^2F(a) \cdot h^2$, sino dicha forma cuadrática *restringida a* $T_a(M)$. En particular, esta última forma cuadrática puede ser definida aunque $D^2F(a) \cdot h^2$ sea indefinida en $\mathbb{R}^n$.