Manual de CÁLCULO DIFERENCIAL

Facultad de Ciencias Matemáticas. UCM

José María Martínez Ansemil
Socorro Ponte Miramontes

Introducción

Este Manual se ha preparado para su utilización por los alumnos de la asignatura CÁLCULO DIFERENCIAL de los distintos Grados de la Facultad de Matemáticas. En él se desarrollan los temas del programa de esa asignatura siguiendo fundamentalmente el libro de J. E. Marsden y M. J. Hoffmann "Análisis Clásico Elemental".

Se trata de una versión provisional del Manual que seguramente tendrá bastantes errores y erratas. Se agradece cualquier comentario que ayude a mejorarlo.

Madrid, Julio de 2019

Los autores:

José María Martínez Ansemil, ansemil@mat.ucm.es
Socorro Ponte Miramontes, ponte@mat.ucm.es

Índice general

Capítulo 1

Topología en el espacio euclídeo

1.1. El espacio euclídeo

Como es habitual, y así se ha hecho en la asignatura de Análisis de Variable Real, denotaremos por $\mathbb{R}$ al **cuerpo** de los números reales.

Fijado un número natural $n = 2, 3, 4...$ denotaremos por $\mathbb{R}^n$ al producto cartesiano de $\mathbb{R}$ por sí mismo n veces. Esto es,

$$\mathbb{R}^n = \{(x_1, ..., x_n) : x_1, ..., x_n \in \mathbb{R}\}.$$

Para $n = 1$ se define $\mathbb{R}^n = \mathbb{R}$.

El conjunto $\mathbb{R}^n$ se considerará siempre dotado de la estructura de espacio vectorial sobre el cuerpo $\mathbb{R}$ definida por las siguientes operaciones de **suma** y **producto por escalares**:

$$(x_1, ..., x_n) \boxplus (y_1, ..., y_n) = (x_1 + y_1, ..., x_n + y_n)$$

$$\alpha \boxdot (x_1, ..., x_n) = (\alpha \cdot x_1, ..., \alpha \cdot x_n)$$

para todo $(x_1, ..., x_n)$, $(y_1, ..., y_n) \in \mathbb{R}^n$ y para todo $\alpha \in \mathbb{R}$. Nótese que las operaciones de suma y producto que aparecen en los términos de la derecha son las operaciones habituales de suma y producto de números reales.

En lo que sigue se denotarán indistintamente con $+$ las operaciones de suma, tanto entre elementos de $\mathbb{R}^n$ como de números reales, y con $\cdot$ tanto el producto de un número real y un elemento de $\mathbb{R}^n$ como de números reales. Es habitual el suprimir el $\cdot$ en el producto y "pegar" los términos que se multiplican. Así, se escribirá $\alpha(x_1, ..., x_n)$ en lugar de $\alpha \cdot (x_1, ..., x_n)$

Recordamos que un espacio vectorial sobre el cuerpo $\mathbb{R}$, es un conjunto V (aquí $V = \mathbb{R}^n$) en el que se consideran dos operaciones, una interna $+$ entre elementos de V y otra externa $\cdot$, entre elementos de V y del cuerpo $\mathbb{R}$, que satisfacen los siguientes axiomas:

i) $u + v = v + u$; para todo $u, v \in V$ (propiedad conmutativa)

ii) $(u + v) + w = u + (v + w)$ para todo $u, v, w \in V$ (propiedad asociativa)

iii) Existe un elemento en V, normalmente denotado por 0 y llamando **elemento neutro de la suma**, tal que

$$v + 0 = v \text{ para todo } v \in V$$

iv) Para todo $v \in V$ existe un elemento en V, usualmente denotado por $-v$ y llamado **elemento opuesto** de v tal que

$$v + (-v) = 0$$

v) Para todo $\alpha \in \mathbb{R}$ y para todo $u, v \in V$

$$\alpha(u + v) = \alpha u + \alpha v$$

vi) Para todo $\alpha, \beta \in \mathbb{R}$ y para todo $v \in V$, $\alpha(\beta v) = (\alpha\beta)v$

vii) Para todo $\alpha, \beta \in \mathbb{R}$ y para todo $v \in V$, $(\alpha + \beta)v = \alpha v + \beta v$

viii) $1v = v$ para todo $v \in V$ (1 es el elemento unidad para el producto de números reales).

Es inmediato el comprobar que $\mathbb{R}^n$ con las operaciones que hemos definido en él es un espacio vectorial sobre $\mathbb{R}$. El elemento neutro para la suma es el $(0, ..., 0)$ y el opuesto de un elemento $(x_1, ..., x_n)$ es $(-x_1, ..., -x_n)$.

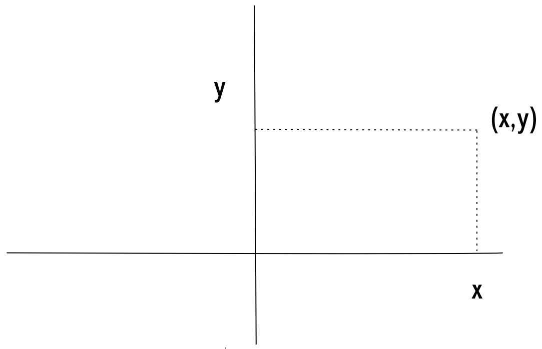
El conjunto formado por los vectores $e_1, ..., e_n$, definidos por

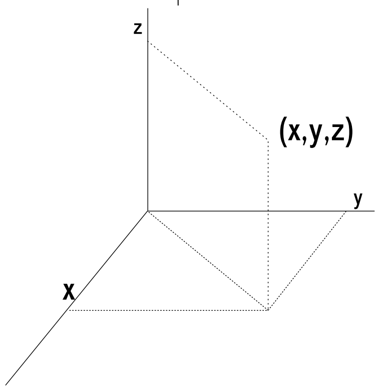
$$e_j = (0, ..., 0, 1, 0, ..., 0)$$

donde el 1 se encuentra en la j-ésima posición constituye una base del espacio vectorial $\mathbb{R}^n$. Esto es, $\{e_1, ..., e_n\}$ es un conjunto formado por n vectores linealmente independientes y cada elemento (vector) de $\mathbb{R}^n$ se puede escribir como una combinación lineal (con escalares reales) de elementos de $\{e_1, ..., e_n\}$. La independencia de los e_i es clara: si $\alpha_1 e_1 + \cdots + \alpha_n e_n = (0, ..., 0)$, entonces $(\alpha_1, ..., \alpha_n) = (0, ..., 0)$, esto es, cada α_i es 0. Por otra parte dado $(x_1, ..., x_n) \in \mathbb{R}^n$ se tiene que $(x_1, ..., x_n) = x_1 e_1 + \cdots + x_n e_n$.

Evidentemente $\mathbb{R}^n$ tiene muchas bases, la que nosotros estamos considerando aquí, se llama la **base canónica**. Como acabamos de ver, los coeficientes de un vector de $\mathbb{R}^n$ respecto a la base canónica son precisamente sus componentes, llamadas **coordenadas**.

Los elementos de $\mathbb{R}^2$ se representan como puntos del plano y los de $\mathbb{R}^3$ como puntos del espacio. Es habitual, y así lo haremos nosotros, denotar por (x, y) los puntos de $\mathbb{R}^2$ y por (x, y, z) los puntos de $\mathbb{R}^3$. Es más cómodo hacerlo así que usar los subíndices.



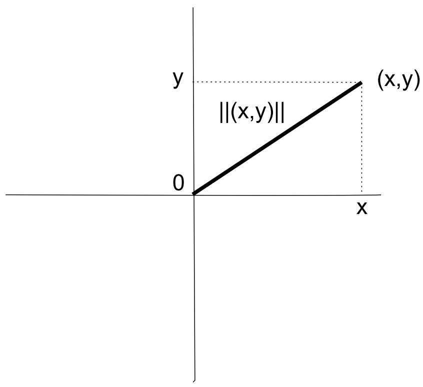


1.1.1. Métricas, normas y productos euclídeos

Definición 1.1.1 *Se llama* ***longitud o norma euclídea*** *del vector* $x = (x_1, ..., x_n) \in \mathbb{R}^n$ *al número real, mayor o igual que 0, definido por*

$$\|x\| := \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}}$$

(se considera siempre la raíz cuadrada con signo positivo). Si $n = 2$*, por el conocido teorema de Pitágoras,* $\|(x, y)\|$ *no es otra cosa que la longitud del segmento de extremos* $(0, 0)$ *y* (x, y)*.*

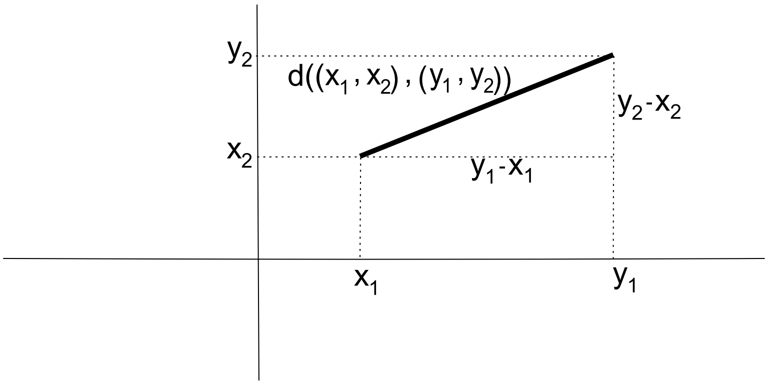


Ejemplo 1.1.2 *La norma del vector* $(1,1,2)$ *de* $\mathbb{R}^3$ *es* $\sqrt{6}$.

Definición 1.1.3 *Se llama* ***distancia euclídea*** *entre dos vectores* $x = (x_1, ..., x_n)$ *e* $y = (y_1, ..., y_n)$ *de* $\mathbb{R}^n$ *al número real mayor o igual que 0 definido por*

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

De nuevo, si $n = 2$, *el teorema de Pitágoras nos dice que la distancia euclídea entre dos puntos* (x_1, x_2) *e* (y_1, y_2) *de* $\mathbb{R}^2$ *es precisamente la longitud del segmento con extremos en esos puntos.*



Ejemplo 1.1.4 *La distancia entre los vectores* $(1,2,1)$ *y* $(3,2,0)$ *de* $\mathbb{R}^3$ *es* $\sqrt{5}$.

Definición 1.1.5 *Se llama* ***producto escalar (o interior) euclídeo*** *de dos vectores* $x = (x_1, ..., x_n)$ *e* $y = (y_1, ..., y_n)$ *de* $\mathbb{R}^n$ *al número real (escalar), no necesariamente mayor o igual que 0, definido por*

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

Si estamos en $\mathbb{R}$ el producto escalar de dos números reales es exactamente su producto. En $\mathbb{R}^2$ resulta que

$$\langle x, y\rangle = \|x\|\,\|y\|\cos\theta$$

siendo θ el ángulo que forma los vectores x e y. Veremos esto enseguida, antes obtenemos algunas propiedades del producto escalar euclídeo que necesitaremos.

Propiedades del producto escalar euclídeo:

i) $\langle x, x\rangle$ es mayor o igual que 0 para todo $x \in \mathbb{R}^n$.

Esto es inmediato pues $\langle x, x\rangle = \sum_{i=1}^n x_i^2$. Observamos que $\langle x, x\rangle$ tiene un estrecha relación con la norma de x, de hecho $\|x\| = \langle x, x\rangle^{\frac{1}{2}}$.

ii) $\langle x, x\rangle = 0$ si y sólo si $x = 0$.

Nótese que, según acabamos de ver, $\langle x, x\rangle = \sum_{i=1}^n x_i^2$ y que $\sum_{i=1}^n x_i^2 = 0 \Leftrightarrow x_i = 0$ para todo $i = 1, ..., n$.

iii) Para todo $x, y, z \in \mathbb{R}^n$ se verifica que

$$\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle.$$

En efecto,

$$\begin{aligned}\langle x + y, z\rangle &= \sum_{i=1}^n (x_i + y_i)z_i = \sum_{i=1}^n (x_i z_i + y_i z_i) \\ &= \sum_{i=1}^n x_i z_i + \sum_{i=1}^n y_i z_i = \langle x, z\rangle + \langle y, z\rangle.\end{aligned}$$

iv) $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle$ para todo $x, y \in \mathbb{R}^n$ y para todo $\alpha \in \mathbb{R}$.

En efecto $\langle \alpha x, y\rangle = \sum_{i=1}^n (\alpha x_i)y_i = \alpha \sum_{i=1}^n x_i y_i = \alpha\langle x, y\rangle$.

v) $\langle y, x\rangle = \langle x, y\rangle$ para todo $x, y \in \mathbb{R}^n$.

Esto es obvio teniendo en cuenta que el producto de números reales es conmutativo.

De las propiedades anteriores se deduce fácilmente que:

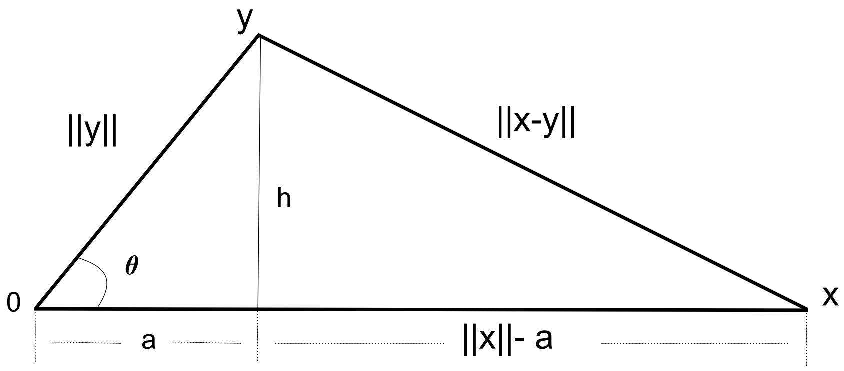
vi) $\langle x, 0\rangle = \langle 0, y\rangle = 0$ para cualesquiera x e $y \in \mathbb{R}^n$.

vii) $\langle \alpha x + \beta y, \alpha x + \beta y\rangle = \alpha^2\langle x, x\rangle + 2\alpha\beta\,\langle x, y\rangle + \beta^2\langle y, y\rangle$ para cualesquiera $\alpha, \beta \in \mathbb{R}$ y $x, y \in \mathbb{R}^n$.

Veamos ahora la fórmula anunciada que relaciona el producto interior de un par de vectores de $\mathbb{R}^2$ con sus normas y el coseno del ángulo que forman.

Teorema 1.1.6 *Para cualesquiera x e $y \in \mathbb{R}^2$ se verifica que $\langle x, y\rangle = \|x\|\,\|y\|\cos\theta$ siendo θ el ángulo que forman los vectores x e y.*

Demostración. Dados dos vectores x e y de $\mathbb{R}^2$, que supondremos distintos de 0 (pues si uno de ellos es 0 el resultado es inmediato), consideremos el triángulo de vértices $0, x, y$:



Utilizando trigonometría elemental tenemos que

$$\cos\theta = \frac{a}{\|y\|}$$

y que

$$\|y\|^2 - a^2 = h^2 = \|x-y\|^2 - (\|x\| - a)^2.$$

Con lo que

$$\|x-y\|^2 = \|y\|^2 - a^2 + \|x\|^2 - 2a\,\|x\| + a^2 = \|y\|^2 + \|x\|^2 - 2\cos\theta\,\|x\|\,\|y\|\,.$$

Si nos fijamos un poco nos damos cuenta de que se trata de una igualdad que relaciona las longitudes de los tres lados de un triángulo no necesariamente rectángulo, suele darse en los cursos de enseñanza secundaria.

Si ahora usamos las propiedades del producto interior obtenemos que

$$\|x-y\|^2 = \langle x-y, x-y\rangle = \|x\|^2 - 2\langle x,y\rangle + \|y\|^2$$

de donde se deduce, teniendo en cuenta el valor previamente obtenido de $\|x-y\|^2$, que

$$\langle x,y\rangle = \|x\|\,\|y\|\cos\theta.$$

□

Nota 1.1.7 *Hemos supuesto que $0 < \theta \le \frac{\pi}{2}$ (véase el dibujo que nos ha servido para despejar los valores de $\cos\theta$ y de h^2), pero el resultado es válido en general para $\frac{\pi}{2} < \theta < \pi$. Tanto si $\theta = 0$ como si $\theta = \pi$ el resultado es igualmente cierto y la demostración es más directa. En efecto, si $\theta = 0$ entonces $x = \alpha y$ para un cierto $\alpha > 0$. Luego $\langle x,y\rangle = \langle \alpha y, y\rangle = \alpha\,\|y\|^2$ y $\|x\|\,\|y\| = \|\alpha y\|\,\|y\| = \alpha\,\|y\|^2$ de donde*

$$\langle x,y\rangle = \|x\|\,\|y\|\cos 0.$$

De forma análoga se hace el caso $\theta = \pi$.

Ejemplo 1.1.8 *El producto escalar de* $(1,2)$ *y* $(3,2)$ *es* 7. *El ángulo que forman estos vectores es* $\theta =$*arcos*$(\frac{7}{\sqrt{5}\sqrt{13}})$.

Asociado al concepto de producto escalar se encuentra el de ortogonalidad:

Definición 1.1.9 *Se dice que dos vectores* x *e* y *de* $\mathbb{R}^n$ *son* ***ortogonales*** *si* $\langle x, y\rangle = 0$.

Obsérvese que para $n = 2$ la fórmula $\langle x, y\rangle = \|x\| \|y\| \cos\theta$ que hemos obtenido anteriormente nos dice que el que x e y sean ortogonales significa que forman un ángulo de $\pm\frac{\pi}{2}$.

Ejemplo 1.1.10 *Los vectores* $(1,0)$ *y* $(0,1)$ *de* $\mathbb{R}^2$ *son ortogonales.*

Propiedades de la norma euclídea:

i) $\|x\| \geq 0$ para todo $x \in \mathbb{R}^n$.

Esto es obvio por la propia definición de $\|x\|$.

ii) $\|x\| = 0$ si y sólo si $x = 0$.

También esto es inmediato al ser $\|x\| = (\sum_{i=1}^n x_i^2)^{\frac{1}{2}}$.

iii) $\|\alpha x\| = |\alpha| \|x\|$ para todo $\alpha \in \mathbb{R}$ y para todo $x \in \mathbb{R}^n$.

En efecto,

$$\|\alpha x\| = \left(\sum_{i=1}^n (\alpha x_i)^2\right)^{\frac{1}{2}} = \left(\alpha^2 \sum_{i=1}^n x_i^2\right)^{\frac{1}{2}} = |\alpha| \left(\sum_{i=1}^n x_i^2\right)^{\frac{1}{2}} = |\alpha| \|x\|.$$

iv) $\|x+y\| \leq \|x\| + \|y\|$ para todo $x, y \in \mathbb{R}^n$.

La prueba de esta propiedad no es inmediata, la obtendremos a partir de la llamada "Desigualdad de Cauchy-Schwarz" que damos a continuación.

Teorema 1.1.11 ***(Desigualdad de Cauchy-Schwarz)*** *Para todo* $x, y \in \mathbb{R}^n$ *se cumple que:*

$$|\langle x, y\rangle| \leq \|x\| \|y\|$$

o, equivalentemente,

$$\left|\sum_{i=1}^n x_i y_i\right| \leq \left(\sum_{i=1}^n x_i^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^n y_i^2\right)^{\frac{1}{2}}.$$

Demostración. Fijemos $x, y \in \mathbb{R}^n$. Para cada $\alpha \in \mathbb{R}$ se tiene que

$$\langle \alpha x + y, \alpha x + y \rangle \geq 0.$$

De donde, usando la propiedad (vii) del producto escalar, se obtiene que

$$\langle \alpha x + y, \alpha x + y \rangle = \alpha^2 \langle x, x \rangle + 2\alpha \langle x, y \rangle + \langle y, y \rangle \geq 0.$$

Si hacemos $A = \langle x, x \rangle,\ B = 2\langle x, y \rangle$ y $C = \langle y, y \rangle$, resulta que

$$A\alpha^2 + B\alpha + C \geq 0 \text{ para todo } \alpha \in \mathbb{R}.$$

Si $A = 0$ entonces $x = 0$ y la desigualdad de Cauchy-Schwarz es inmediata. Supongamos en consecuencia que $A \neq 0$.

La ecuación de segundo grado $A\alpha^2 + B\alpha + C = 0$ debe tener su discriminante $B^2 - 4AC$ menor o igual que 0, pues de lo contrario tendría dos raíces reales distintas y entonces $A\alpha^2 + B\alpha + C$ tomaría algún valor menor que 0 (si α_1 y α_2 son tales raíces, $A\alpha^2 + B\alpha + C = A(\alpha - \alpha_1)(\alpha - \alpha_2)$). Se sigue entonces que $B^2 \leq 4AC$ lo que nos da que:

$$4\langle x, y \rangle^2 \leq 4\langle x, x \rangle \langle x, y \rangle = 4 \|x\|^2 \|y\|^2 .$$

de donde se sigue la desigualdad requerida.

Probaremos ahora que $\|x + y\| \leq \|x\| + \|y\|$ comprobando que $\|x + y\|^2 \leq (\|x\| + \|y\|)^2$, lo que es equivalente. Usando propiedades del producto escalar obtenemos que

$$\|x + y\|^2 = \langle x + y, x + y \rangle = \|x\|^2 + 2\langle x, y \rangle + \|y\|^2 .$$

Por la desigualdad de Cauchy-Schwarz $\langle x, y \rangle \leq \|x\| \|y\|$ con lo que

$$\|x + y\|^2 \leq \|x\|^2 + 2 \|x\| \|y\| + \|y\|^2 = (\|x\| + \|y\|)^2.$$

□

Propiedades de la distancia euclídea:

i) $d(x, y) \geq 0$, lo que es obvio por la propia definición de la distancia.

ii) $d(x, y) = 0 \Longleftrightarrow x = y$.

En efecto,

$$d(x, y) = \|x - y\| = 0 \Longleftrightarrow x - y = 0 \Longleftrightarrow x = y.$$

iii) $d(x, y) \leq d(x, z) + d(z, y)$ para todo $x, y, z \in \mathbb{R}^n$.

En efecto,

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

Esta última desigualdad se llama **desigualdad triangular** y en $\mathbb{R}^2$ expresa el hecho de que la longitud de cualquier lado de un triángulo es menor que la suma de las longitudes de los otros dos.

iv) $d(x,y) = d(y,x)$ para todo $x, y \in \mathbb{R}^n$. Nótese que

$$d(x,y) = \|x-y\| = \|(-1)(x-y)\| = \|y-x\| = d(y,x).$$

1.1.2. Métricas, normas y productos escalares

En $\mathbb{R}^n$ hemos considerado la distancia euclídea d, la norma euclídea $\|\ \|$ y el producto escalar euclídeo $\langle\ ,\ \rangle$. Recordamos que para $x,\ y \in \mathbb{R}^n$,

$$d(x,y) := \left(\sum_{i=1}^{n}(x_i - y_i)^2\right)^{\frac{1}{2}},\ \|x\| := \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}} \text{ y } \langle x,y\rangle := \sum_{i=1}^{n} x_i y_i.$$

Estos conceptos pueden considerarse en muchos otros contextos. Veremos ahora cómo se puede definir el concepto de métrica en un conjunto arbitrario y los conceptos de norma y producto interior en espacios vectoriales arbitrarios.

Definición 1.1.12 *Se llama* ***métrica***[1] *sobre un conjunto arbitrario M a cualquier aplicación $d : M \times M \to \mathbb{R}$ que verifique las propiedades que tenía la métrica euclídea que hemos considerado anteriormente en $\mathbb{R}^n$, a saber:*

i) $d(x,y) \geq 0$, para todo $x, y \in M$

ii) $d(x,y) = 0 \Longleftrightarrow x = y$

iii) $d(x,y) \leq d(x,z) + d(z,y)$ para todo $x, y, z \in M$

iv) $d(x,y) = d(y,x)$ para todo $x, y \in M$.

Llamaremos **espacio métrico** a cualquier conjunto en el que se haya considerado una métrica d. Normalmente no se hace referencia a la métrica pero debe sobreentenderse en el contexto de qué métrica se trata, pues sobre un mismo conjunto se pueden definir varias métricas. Así, si decimos, sea M un espacio métrico; se estará sobreentendiendo que sobre el conjunto M se está considerando una métrica, que normalmente se denotará por d.

[1]La palabra "métrica" viene del griego "metreo" que significa mido.

Nota 1.1.13 *Las condiciones (iii) y (iv) equivalen (ambas juntas) a la condición*

(*iii* ′) $d(x, y) \leq d(x, z) + d(y, z)$ para todo $x, y, z \in M$.

Desde luego es claro que (*iii*) + (*iv*) implican (*iii* ′). Por otra parte, si suponemos cierta (*iii* ′) y la aplicamos a x, y, x y después a y, x, y obtenemos:

$$d(x, y) \leq d(x, x) + d(y, x) \text{ y } d(y, x) \leq d(y, y) + d(x, y),$$

de donde $d(x, y) = d(y, x)$, pues $d(x, x) = d(y, y) = 0$. Es claro finalmente que $(iii)' +$ (*iv*) implican (*iii*).

Nota 1.1.14 *Las condiciones* $(i), (ii)$ *y* (iii) *no bastan por sí solas para definir una métrica. En efecto, si consideramos en el conjunto* M *formado por los números naturales* 1 *y* 2*, la aplicación* $d(1, 1) = 0$*,* $d(2, 2) = 0$*,* $d(1, 2) = 1$ *y* $d(2, 1) = 2$*, verifica las propiedades* $(i), (ii)$ *y* (iii)*, pero* $d(1, 2)$ *no es igual a* $d(2, 1)$*.*

Nota 1.1.15 *Las condiciones* $(ii), (iii)$ *y* (iv) *implican la* (i) *y por lo tanto esta condición puede suprimirse de la definición. Veámoslo: para todo* $x,\ y \in M$*,*

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

con lo que $d(x, y) \geq 0$*.*

Algunos ejemplos de métricas:

1. La métrica euclídea en $\mathbb{R}^n$.
2. $d_1(x, y) = \sum_{i=1}^{n} |x_i - y_i|$ en $\mathbb{R}^n$ (llamada *métrica del taxi*. Normalmente un taxi no puede ir en linea recta de un sitio a otro, va por calles, haciendo zig-zag y cobra, claro está, por la suma de los tramos rectos que recorre).
3. $d_\infty(x, y) =$máx$\{|x_i - y_i| : i = 1, ..., n\}$ en $\mathbb{R}^n$.
4. $d(f, g) = \int_a^b |f(x) - g(x)|dx$, en el espacio $\mathcal{C}[a, b]$ de las funciones continuas en un intervalo $[a, b]$ de la recta real. Observamos que si para una función continua no negativa su integral es 0, ésta debe ser idénticamente nula (de tomar un valor mayor que 0 en un punto también lo tomaría en un entorno de ese punto y entonces la integral sería mayor que 0).
5. $d_\infty(f, g) = \sup\{|f(x) - g(x)| : x \in [a, b]\}$, en el espacio $\mathcal{B}[a, b]$ de las funciones acotadas en $[a, b]$. Una función $f : [a, b] \to \mathbb{R}$ es **acotada** si existe $M \geq 0$ tal que $|f(x)| \leq M$ para todo $x \in [a, b]$.

6. Sea M el conjunto de todas las expresiones del tipo $\omega = (\omega_1, ..., \omega_8)$ donde cada ω_i es 0 ó 1. Para cualesquiera u y v de M se define

   $$d(u, v) = \text{número de coordenadas en las que } u \text{ y } v \text{ son diferentes.}$$

   d se conoce con el nombre de **métrica lexicográfica** y es útil en informática. Por supuesto que el que los elementos de M tengan 8 coordenadas es irrelevante, podría ser cualquier otro número natural.

7. $d(x, y) = \begin{cases} 0 \text{ si } x = y \\ 1 \text{ si } x \neq y \end{cases}$ en cualquier conjunto M.

Esta métrica se llama la métrica **discreta**. Se usará varias veces para proporcionar ejemplos y contraejemplos, en alguna ocasión sorprendentes.

Definición 1.1.16 *Se llama* ***diámetro*** *de un subconjunto S de un espacio métrico (M, d) a*

$$\text{diám}(S) := \sup\{d(x, y) : x, y \in S\}$$

si el conjunto de números reales $\{d(x, y) : x, y \in S\}$ es acotado (superiormente), y se define $\text{diám}(S) = \infty$ si no lo es. Cuando el diámetro es finito se dice que el conjunto es **acotado**.

Ejemplos:

1. Cualquier intervalo delimitado por dos números reales es acotado para la métrica euclídea. Su diámetro es precisamente la diferencia entre el extremo superior y el inferior del intervalo.

2. El conjunto de los números naturales no es un subconjunto acotado de $\mathbb{R}$ con la métrica euclídea, pues $d(1, n + 1) = n$ para todo $n \in \mathbb{N}$.

3. El conjunto S de los pares de números reales (x, y) tales que $0 < x \leq 2$, $y \geq 4$ no es acotado para la métrica euclídea pues para cada $n \in \mathbb{N}$, $n \geq 4$, el punto $(1, n)$ está en S y $d((1, 4), (1, n)) = n - 4$.

4. Cualquier conjunto es acotado en relación con la métrica discreta.

Definición 1.1.17 *Sea V un espacio vectorial sobre el cuerpo $\mathbb{R}$. Se llama* ***norma*** *en V a toda aplicación $\| \ \| : V \to \mathbb{R}$ que verifique las siguientes condiciones:*

i) $\|v\| \geq 0$ para todo $v \in V$.

ii) $\|v\| = 0$ si y sólo si $v = 0$.

iii) $\|\alpha v\| = |\alpha| \, \|v\|$ para todo $\alpha \in \mathbb{R}$ y para todo $v \in V$.

iv) $\|u + v\| \leq \|u\| + \|v\|$ para todo $u, v \in V$.

Ejemplos de normas:

1. $\|x\| = |x|$ es una norma en $\mathbb{R}$.
2. $\|x\|_2 = \left(\sum_{j=1}^{n} x_j^2\right)^{\frac{1}{2}}$ es una norma en $\mathbb{R}^n$. Se trata de la norma euclídea que hemos considerado anteriormente.
3. $\|x\|_1 = \sum_{j=1}^{n} |x_j|$ en una norma en $\mathbb{R}^n$.
4. $\|x\|_\infty =$máx$\{|x_j| : j = 1, ..., n\}$ es una norma en $\mathbb{R}^n$.
5. $\|f\|_\infty = \sup\{|f(x)| : x \in [a, b]\}$ en $\mathcal{B}[a, b]$.

Hemos observado, al hablar de la métrica y la norma euclídea en $\mathbb{R}^n$, que entre ellas existe la relación: $d(x, y) = \|x - y\|$. Pues bien, si tenemos un espacio normado arbitrario $(V, \| \ \|)$ entonces la aplicación $d : V \times V \to \mathbb{R}$, dada por $d(x, y) = \|x - y\|$ define una métrica en V.

Conviene notar que aunque, como acabamos de ver, toda norma da lugar a una métrica, no se verifica, en general, que dada una métrica exista una norma de la cual provenga. El problema no está sólo en que la métrica pueda estar definida en un conjunto que no tenga estructura de espacio vectorial y entonces no podamos hablar de norma. Consideremos, por ejemplo, la métrica discreta d en un espacio vectorial arbitrario V. Es claro que no existe ninguna norma $\| \ \|$ en V tal que $d(x, y) = \|x - y\|$. Basta considerar que si eso fuera posible, al ser $d(x, 0) = 1$ para todo $x \neq 0$, de las propiedades de la norma y de la relación $d(x, y) = \|x - y\|$, se llegaría al absurdo de que para todo $x \neq 0$,

$$1 = d(2x, 0) = \|2x\| = 2 \|x\| = 2d(x, 0) = 2.$$

Nótese que lo mismo ocurre si d es cualquier métrica acotada (existe $C \geq 0$ tal que $d(x, y) \leq C$ para todo x, y), pues para cualquier $x \neq 0$ se tendría que

$$d(nx, 0) = \|nx\| = n \|x\|$$

y este valor tiende a ∞ cuando n tiende a ∞.

Definición 1.1.18 *Sea V un espacio vectorial sobre $\mathbb{R}$. Llamaremos* ***producto escalar*** *o* ***interior*** *en V a toda aplicación $\langle \ , \ \rangle : V \times V \to \mathbb{R}$ que verifique las siguientes condiciones:*

i) $\langle v, v\rangle \geq 0$ para todo $v \in V$

ii) $\langle v, v\rangle = 0$ si y sólo si $v = 0$

iii) $\langle \alpha u, v\rangle = \alpha\langle u, v\rangle$ para todo $u, v \in V$ y para todo $\alpha \in \mathbb{R}$

iv) $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$ para todo $u, v, w \in V$.

v) $\langle v, u\rangle = \langle u, v\rangle$ para todo $u, v \in V$

Ejemplos de productos escalares:

1. $\langle x, y\rangle = xy$ en $\mathbb{R}$
2. $\langle x, y\rangle = \sum_{j=1}^{n} x_j y_j$ en $\mathbb{R}^n$
3. $\langle f, g\rangle = \int_a^b f(x)g(x)dx$ en el espacio $\mathcal{C}[a, b]$ de las funciones reales continuas en un intervalo $[a, b]$ de $\mathbb{R}$.

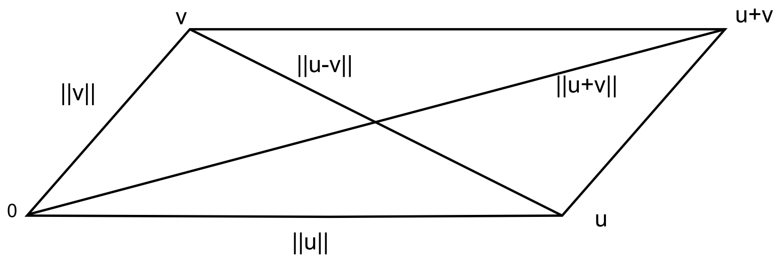
Todo producto escalar da lugar a una norma definiendo $\|v\| = \langle v, v\rangle^{\frac{1}{2}}$ (recuérdese que esta relación ligaba la norma y el producto escalar euclídeos). La verificación de esto se hace igual que en el caso de $\mathbb{R}^n$.

Así como no todas las métricas provienen de una norma, no todas las normas provienen de un producto escalar. Para ver esto obtendremos primero una interesante relación que verifican todas las normas que provienen de un producto escalar. Después daremos un ejemplo de norma que no verifica esa igualdad, luego no puede provenir de un producto escalar.

La igualdad a la que me refiero es la llamada **igualdad del paralelogramo** y nos dice que si $\|\ \|$ es una norma que procede de un producto interior entonces

$$\|u + v\|^2 + \|u - v\|^2 = 2(\|u\|^2 + \|v\|^2).$$

Si pensamos por un momento en lo que esto significa en $\mathbb{R}^2$ nos damos cuenta de que esta igualdad establece el hecho de que en un paralelogramo la suma de los cuadrados de las diagonales es la suma de los cuadrados de los (cuatro) lados:



Demostración.

$$\begin{aligned}\|u+v\|^2+\|u-v\|^2 &= \langle u+v,u+v\rangle+\langle u-v,u-v\rangle\\ &= \|u\|^2+\langle u,v\rangle+\langle v,u\rangle+\|v\|^2+\|u\|^2\\ &\quad -(\langle u,v\rangle+\langle v,u\rangle)+\|v\|^2\\ &= 2\|u\|^2+2\|v\|^2.\end{aligned}$$

Comprobemos ahora, ayudándonos de la propiedad anterior, que la norma $\| \ \|_\infty$ en $\mathbb{R}^2$ no proviene de ningún producto escalar. En efecto, si tomamos $u=(1,0)$ y $v=(0,1)$, entonces

$$\|u+v\|_\infty=\|(1,1)\|_\infty=1,$$

$$\|u-v\|_\infty=\|(1,-1)\|_\infty=1,$$

$$\|u\|_\infty=1 \text{ y } \|v\|_\infty=1.$$

Luego es claro que $\| \ \|_\infty$ no verifica la igualdad del paralelogramo pues

$$1^2+1^2\neq 2(1^2+1^2).$$

□

1.1.3. Nociones de topología

Vamos a introducir ahora unos conceptos que serán necesarios para más adelante estudiar la convergencia, la compacidad, la continuidad y la diferenciabilidad.

Definición 1.1.19 *Dados $x_o\in\mathbb{R}^n$ y un número real $r>0$, llamaremos **bola abierta** de centro x_o y radio $r>0$ al conjunto*

$$B(x_o,r):=\{x\in\mathbb{R}^n : d(x,x_o)<r\},$$

donde d es la métrica que se está considerando en $\mathbb{R}^n$ (que, si no se especifica otra, será la euclídea)

Ejemplos:

1. Si $n=1$, entonces $B(x_o,r)=(x_o-r,x_o+r)$.
2. En $\mathbb{R}^2$,
$$B((x_o,y_o),r)=\{(x,y)\in\mathbb{R}^2 : \sqrt{(x-x_o)^2+(y-y_o)^2}<r\}.$$
Esto es, $B((x_o,y_o),r)$ es el círculo de centro (x_o,y_o) y radio r sin la circunferencia de centro (x_o,y_o) y radio r.

3. En $\mathbb{R}^3$ las bolas abiertas son esferas sin la superficie esférica exterior.

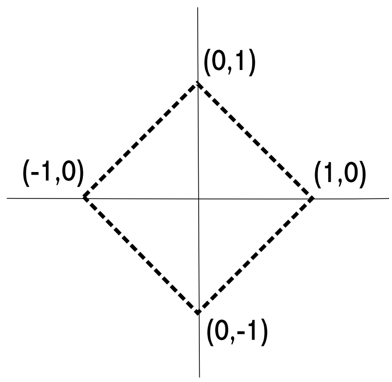
En los tres ejemplos hemos trabajado con la métrica euclídea. Si trabajamos con otras métricas las bolas son diferentes. Consideremos en $\mathbb{R}^2$, por ejemplo, la métrica que hemos denotado en el tema anterior por d_1 :

$$d_1((x,y),(x',y')) = |x-x'| + |y-y'|.$$

Entonces

$$B_1((0,0),1) = \{(x,y) \in \mathbb{R}^2; |x| + |y| < 1\}.$$

Gráficamente esa bola es el rombo de vértices $(-1,0), (0,-1), (1,0), (0,1)$.



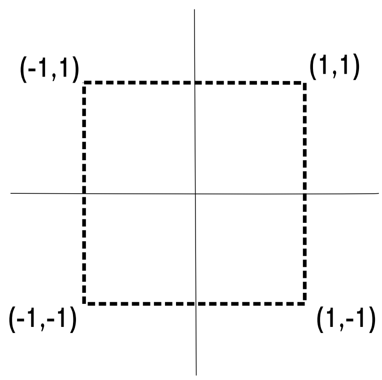
Si consideramos en $\mathbb{R}^2$ la métrica d_∞ :

$$d_\infty((x,y),(x',y')) = máx\{|x-x'|, |y-y'|\},$$

entonces

$$B_\infty((0,0),1) = \{(x,y) \in \mathbb{R}^2;\ máx\{|x|,|y|\} < 1\}.$$

Gráficamente esa bola es el cuadrado de vértices $(-1,1), (-1,-1), (1,-1)$ y $(1,1)$.



Definición 1.1.20 *Se dice que un subconjunto A de $\mathbb{R}^n$ es **abierto** si para todo $x \in A$ existe $r_x > 0$ tal que $B(x, r_x) \subset A$.*

Nota 1.1.21 *A veces escribiremos r en vez de r_x, pero debe quedar claro que, en general, el radio de la bola depende de x.*

Nota 1.1.22 *El concepto de conjunto abierto se puede definir, de manera análoga, en cualquier espacio métrico, en particular en un subconjunto de $\mathbb{R}^n$, que es un espacio métrico con la métrica inducida por una métrica d en $\mathbb{R}^n$. Así si S es un subconjunto de $\mathbb{R}^n$, un subconjunto B de S es abierto (en el espacio métrico S) si para todo $x \in B$ existe una bola $B_S(x, r) := \{y \in S : d(y, x) < r\}$ contenida en B. Sucede que $B \subset S \subset \mathbb{R}^n$ es abierto (en el espacio métrico S) si y sólo si existe un abierto A en $\mathbb{R}^n$ tal que $B = A \cap S$ (Ejercicio).*

Ejemplos:

1. Toda bola abierta $B(x_0, r)$ es un conjunto abierto. Observad que esto no es una trivialidad, hay que demostrar que para todo punto x de $B(x_0, r)$ existe $r_x > 0$ tal que $B(x, r_x) \subset B(x_0, r)$ (para $x = x_0$ basta tomar $r_x = r$). Sea entonces x un punto arbitrario de $B(x_0, r)$. Dado que $d(x, x_0) < r, r_x := r - d(x, x_0)$ es mayor que 0, y para todo $y \in B(x, r_x)$ se tiene que

$$d(y, x_o) \leq d(y, x) + d(x, x_0) < r - d(x, x_0) + d(x, x_0) = r.$$

   Entonces $y \in B(x_0, r)$. La arbitrariedad de y en $B(x, r_x)$ nos da que $B(x, r_x) \subset B(x_0, r)$.

2. En $\mathbb{R}^2$ el conjunto $A = \{(x, y) : x > 0\}$ es abierto. En efecto, si $(x_0, y_o) \in A$ entonces $x_0 > 0$. Si hacemos $r = x_0$ resulta que $B((x_0, y_o), r) \subset A$, pues si un punto (x, y) de $\mathbb{R}^2$ verifica que $(x - x_0)^2 + (y - y_o)^2 < x_0^2$, entonces $|x - x_0| < x_0$ por lo que $-x + x_0 < x_0$ y, en consecuencia, $x > 0$, es decir, $(x, y) \in A$.

3. En $\mathbb{R}$ el intervalo $(0, 1]$ no es un conjunto abierto pues es imposible encontrar una bola (intervalo) centrada en 1 que está contenida en $(0, 1]$.

4. Cualquier subconjunto de $\mathbb{R}^n$ es abierto para la métrica discreta, pues con esa métrica, $B(x, 1) = \{x\}$ para todo $x \in \mathbb{R}^n$.

Propiedades de los conjuntos abiertos:

i) El conjunto vacío y el total son abiertos.

ii) La unión de cualquier colección de conjuntos abiertos es un abierto.

iii) La intersección de una colección **finita** de abiertos es un abierto.

En efecto:

i) Como el conjunto vacío no tiene puntos, la condición de ser abierto se verifica automáticamente. Si $A = \mathbb{R}^n$ entonces cualquier bola centrada en un punto de A está contenida en $A = \mathbb{R}^n$.

ii) Sea $\{A_\alpha : \alpha \in \Lambda\}$ una colección arbitraria de abiertos de $\mathbb{R}^n$. Dado cualquier $x \in \cup_{\alpha\in\Lambda} A_\alpha$ existe un $\alpha \in \Lambda$ tal que $x \in A_\alpha$. Al ser A_α abierto existe un $r > 0$ tal que $B(x,r) \subset A_\alpha$. Entonces es claro que $B(x,r) \subset \cup_{\alpha\in\Lambda} A_\alpha$.

iii) Sea $A_1, ..., A_k$ una colección finita de abiertos, y sea $x \in \cap_{j=1}^k A_j$. Entonces $x \in A_j$ para todo $j = 1, ..., k$ y por lo tanto existe, para cada j, un $r_j > 0$ tal que $B(x, r_j) \subset A_j$. Si hacemos $r = \text{mín}\{r_j, j = 1, ..., k\}$, es claro que $B(x,r) \subset \cap_{j=1}^k A_j$.

Nótese que podemos asegurar que $r = \text{mín}\{r_j, j = 1, ..., k\}$ es mayor que 0 pues todos los r_j lo son. Otra cosa sería si se tratase de una cantidad infinita de r_j. Queremos decir con esto que puede no ser cierto, en general, que la intersección de una colección infinita de abiertos sea un abierto. En efecto, si por cada $j = 1, 2...$ hacemos $A_j = (-\frac{1}{j}, \frac{1}{j})$, resulta que $\cap_{j=1}^\infty A_j = \{0\}$ que no es un conjunto abierto en $\mathbb{R}$ (con la métrica euclídea).

Nota 1.1.23 *Así como hicimos en su momento abstracción de las propiedades de la métrica, norma o producto escalar euclídeos en* $\mathbb{R}^n$*, para introducir esos conceptos en contextos muy generales, podemos hacer ahora abstracción de las propiedades de los abiertos de un espacio métrico para definir el concepto de* ***espacio topológico*** *que se estudia con detalle en otras asignaturas del Grado. Se llama espacio topológico a cualquier conjunto* X *en el que se ha prefijado una colección de subconjuntos, que se llaman los abiertos de* X *tal que esos subconjuntos verifiquen las propiedades* $i), ii)$ *y* $iii)$ *de los conjuntos abiertos de* $\mathbb{R}^n$*.*

Definición 1.1.24 *Se dice que un punto* x *de un subconjunto* S *de* $\mathbb{R}^n$ *es un* ***punto interior*** *de* S *si existe una bola abierta centrada en* x *contenida en* S*. Denotaremos por* $\overset{o}{S}$ *o por* S^o *al conjunto de los puntos interiores de* S*.*

$\overset{o}{S}$ puede ser vacío, por ejemplo si S es un subconjunto de $\mathbb{R}$ con un único punto o si S es una sucesión de puntos de $\mathbb{R}$. Observamos que si $S \subset T$ entonces $\overset{o}{S} \subset \overset{o}{T}$.

Si $S = (1, 2] \subset \mathbb{R}$ entonces $\overset{o}{S} = (1, 2)$. Por otra parte, si $S = B(x, r)$ en un espacio métrico, entonces $\overset{o}{S} = B(x, r)$.

Para cualquier $S \subset \mathbb{R}^n$ se verifica que

1. $\overset{o}{S}$ es el mayor abierto (de $\mathbb{R}^n$) contenido en S.

2. $\overset{o}{S}$ es precisamente la unión de todos los abiertos (de $\mathbb{R}^n$) contenidos en S.

3. S es abierto si y sólo si $S = \overset{o}{S}$.

Veamos esto:

1. $\overset{o}{S}$ es abierto, pues dado $x \in \overset{o}{S}$ si tomamos $r > 0$ tal que $B(x,r) \subset S$, entonces sucede que $B(x,r) \subset \overset{o}{S}$, pues al ser $B(x,r)$ abierto, $B(x,r) = [B(x,r)]^o \subset \overset{o}{S}$. Por otra parte, si A es un abierto de $\mathbb{R}^n$ contenido en S, entonces para todo punto de A hay una bola centrada en él contenida en A y por lo tanto en S, luego todos los puntos de A están en $\overset{o}{S}$.

2. Es claro que el mayor abierto contenido en S es la unión de todos los abiertos contenidos en S, nótese que la unión de abiertos es un abierto.

3. Si S es abierto, entonces él es el mayor abierto contenido en S, luego $S = \overset{o}{S}$. Por otra parte, si $S = \overset{o}{S}$ entonces S es abierto por serlo $\overset{o}{S}$.

Definición 1.1.25 *Dados $x_0 \in \mathbb{R}^n$ y un número real $r > 0$, llamaremos **bola cerrada** de centro x_0 y radio r al conjunto*

$$\overline{B}(x_0, r) = \{x \in \mathbb{R}^n : d(x, x_0) \leq r\}.$$

Si la métrica que se considera en $\mathbb{R}^n$ no es necesariamente la euclídea, pero proviene de una norma (como le ocurre a la euclídea), entonces,

$$[\overline{B}(x_0, r)]^o = B(x_0, r).$$

En efecto, "$\subset$"

Si $x \in [\overline{B}(x_0, r)]^o$ entonces existe $r_x > 0$ tal que $B(x, r_x) \subset \overline{B}(x_0, r)$. Si ocurriese que $\|x - x_0\| = r$ tomemos $y = x + \frac{x - x_0}{\|x - x_0\|}\frac{1}{2}r_x$; entonces resulta que $\|y - x\| = \frac{1}{2}r_x < r_x$ por lo que $y \in B(x, r_x)$ y sin embargo

$$\|y - x_0\| = \|x - x_0\| \left(1 + \frac{1}{\|x - x_0\|}\frac{1}{2}r_x\right) > \|x - x_0\| = r$$

lo que es absurdo.

"$\supset$"
Esta inclusión es inmediata pues $B(x_0, r) \subset \overline{B}(x_0, r)$ y al ser $B(x_0, r) = [B(x_0, r)]^o$ resulta que $B(x_0, r) \subset [\overline{B}(x_0, r)]^o$.

Para la métrica discreta en $\mathbb{R}^n$ se verifica que $[\overline{B}(0,1)]^o = \mathbb{R}^n$, mientras que $B(0,1) = \{0\}$.

Definición 1.1.26 *Se dice que un subconjunto C de $\mathbb{R}^n$ es **cerrado** si su complementario (respecto de $\mathbb{R}^n$) es abierto.*

Ejemplos:

1. $[a,b]$ es un cerrado de $\mathbb{R}$ pues su complementario es $(-\infty,a)\cup(b,+\infty)$ que es abierto al ser la unión de dos intervalos abiertos de $\mathbb{R}$.
2. Las bolas cerradas son conjuntos cerrados. En efecto, el complementario de una bola cerrada $\overline{B}(x_0,r)$ es $\{x \in \mathbb{R}^n : d(x,x_0) > r\}$ y este conjunto es un abierto de $\mathbb{R}^n$, pues si $x \in \mathbb{R}^n$ y $d(x,x_0) > r$, entonces $d(x,x_0) - r > 0$ y
$$B(x, d(x,x_0) - r) \subset \mathbb{R}^n \backslash \overline{B}(x_0,r)$$
ya que si $y \in B(x, d(x,x_0) - r)$ se tiene que
$$d(y,x_0) \geq d(x,x_0) - d(y,x) > d(x,x_0) - (d(x,x_0) - r) = r$$
(la primera desigualdad es consecuencia de la desigualdad triangular).
3. Cualquier subconjunto de $\mathbb{R}^n$ es cerrado para la métrica discreta.

Nota 1.1.27 *Puede haber conjuntos que no sean ni abiertos ni cerrados. Por ejemplo $(0,1]$ no es ni abierto ni cerrado en $\mathbb{R}$. Cualquier subconjunto de $\mathbb{R}^n$ es a la vez abierto y cerrado para la métrica discreta en $\mathbb{R}^n$.*

Propiedades de los conjuntos cerrados:

i) El conjunto vacío y $\mathbb{R}^n$ son cerrados.

ii) La unión de una cantidad **finita** de cerrados es un conjunto cerrado.

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

En efecto:

i) El vacío es cerrado por ser el complementario de $\mathbb{R}^n$ y $\mathbb{R}^n$ lo es por ser complementario del vacío.

ii) Si tenemos una cantidad finita $C_1, ..., C_k$ de conjuntos cerrados. El complementario de $\cup_{j=1}^k C_j$ es abierto, pues $\mathbb{R}^n \backslash \left(\cup_{j=1}^k C_j\right) = \cap_{j=1}^k (\mathbb{R}^n \backslash C_j)$ que es una intersección finita de abiertos y por lo tanto un conjunto abierto.

iii) Si tenemos una colección $\{C_\alpha : \alpha \in \Lambda\}$ de conjuntos cerrados, el complementario de $\cap_{\alpha\in\Lambda} C_\alpha$ es abierto, pues $\mathbb{R}^n \backslash (\cap_{\alpha\in\Lambda} C_\alpha) = \cup_{\alpha\in\Lambda}(\mathbb{R}^n \backslash C_\alpha)$ que es una unión de abiertos y por lo tanto un conjunto abierto, luego $\cap_{\alpha\in\Lambda} C_\alpha$ es cerrado.

Obsérvese que no es cierto, en general, que una unión arbitraria de conjuntos cerrados sea un conjunto cerrado. Considérese por ejemplo

$$\bigcup_{j=1}^{\infty} \left[-1 + \frac{1}{j}, 1 - \frac{1}{j}\right] = (-1, 1).$$

Definición 1.1.28 *Se dice que un punto x de $\mathbb{R}^n$ es un* ***punto de acumulación*** *de un subconjunto S de $\mathbb{R}^n$ si toda bola abierta centrada en x contiene algún punto de S distinto del x (si es que x pertenece a S). Esto es, para todo $r > 0$ se verifica que $B(x, r) \cap S \neq \emptyset, \{x\}$. Se suele denotar por S' al conjunto de los puntos de acumulación de S. Si $S \subset T$, entonces $S' \subset T'$.*

Ejemplos:

1. Todo punto del intervalo $[0, 1]$ es un punto de acumulación de $S = (0, 1)$.
2. El punto 0 es un punto de acumulación de $S = \left\{1, \frac{1}{2}, \frac{1}{3}, ...\right\}$.
3. Si $S = \mathbb{Q} \cap [0, 1]$, entonces $S' = [0, 1]$.

Teorema 1.1.29 *S es cerrado si y sólo si S contiene a todos sus puntos de acumulación.*

Demostración. "$\Longrightarrow$" Sea $x \in S'$. Si $x \notin S$ entonces $\mathbb{R}^n \backslash S$ no es abierto pues $x \in \mathbb{R}^n \backslash S$ y cualquier bola de centro en x contiene puntos de S, luego no hay ninguna bola centrada en x totalmente contenida en $\mathbb{R}^n \backslash S$.

"$\Longleftarrow$" Veamos como $\mathbb{R}^n \backslash S$ es abierto. Dado $x \in \mathbb{R}^n \backslash S$ existe una bola abierta centrada en x contenida en $\mathbb{R}^n \backslash S$, de lo contrario toda bola abierta centrada en x tendría puntos de S y entonces x sería de acumulación de S por lo que x tendría que estar en $S' \subset S$. pero $x \in \mathbb{R}^n \backslash S$. $\square$

Definición 1.1.30 *Se dice que un punto x de $\mathbb{R}^n$ es* ***adherente*** *a un subconjunto S de $\mathbb{R}^n$ si toda bola con centro en x tiene algún punto de S. Esto es, para todo $r > 0$ se verifica que $B(x, r) \cap S \neq \emptyset$.*

Obsérvese la diferencia que existe entre punto adherente y punto de acumulación. Para que un punto sea de acumulación, en toda bola centrada en él debe haber puntos de S **distintos** de él. Para que el punto sea adherente sólo se pide que haya puntos S. Así, si tomamos un punto $x \in S$ ese punto ya es automáticamente adherente, pero puede no ser de acumulación.

Definición 1.1.31 *Se llama* ***adherencia*** *o* ***clausura*** *de un subconjunto S de $\mathbb{R}^n$ al conjunto de sus puntos adherentes. Se suele denotar por $\overline{S}$. Es claro que $\overline{S} = S \cup S'$ y que si $S \subset T$ entonces $\overline{S} \subset \overline{T}$.*

Para cualquier $S \subset \mathbb{R}^n$ se verifica que:

1. $\overline{S}$ es cerrado y es el menor cerrado de $\mathbb{R}^n$ que contiene a S.
2. $\overline{S}$ es precisamente la intersección de todos los cerrados que lo contienen.
3. S es cerrado si y sólo si $S = \overline{S}$.

Veamos esto:

1. Veamos como su complementario $\mathbb{R}^n \setminus \overline{S}$ es abierto. Sea $x \in \mathbb{R}^n \setminus \overline{S}$, entonces existe $r > 0$ tal que $B(x,r) \cap S = \emptyset$. Para todo $y \in B(x,r)$ se verifica que $B(y, r - d(y,x)) \cap S = \emptyset$. Esto implica que $y \notin \overline{S}$ y por lo tanto $B(x,r) \subset \mathbb{R}^n \setminus \overline{S}$. Por otra parte, si C es un cerrado que contiene a S, entonces C debe contener a $\overline{S}$ pues C contiene a S y a S' debido a que $S' \subset C' \subset C$ por ser C cerrado.
2. Es claro que el menor cerrado que contiene a un conjunto es la intersección de todos los cerrados que lo contienen, nótese que la intersección de cerrados es cerrada.
3. Si S es un cerrado es claro que él es el menor cerrado que lo contiene, luego $S = \overline{S}$. Recíprocamente, por lo visto en el punto 1, $\overline{S}$ es cerrado, luego si $S = \overline{S}$, S es cerrado.

Definición 1.1.32 *Dados un conjunto S y un punto $x \in \mathbb{R}^n$ se llama* ***distancia*** *de x a S al número real no negativo*

$$d(x,S) = \text{\'inf}\{d(x,y) : y \in S\}.$$

Teorema 1.1.33 *Un punto $x \in \mathbb{R}^n$ está en la adherencia de un subconjunto S de $\mathbb{R}^n$ si y sólo si $d(x,S) = 0$.*

Demostración. "$\Longrightarrow$" Si $x \in \overline{S}$ entonces para todo $r > 0$ existe $y \in S$ tal que $d(x,y) < r$. Es claro entonces que $\text{ínf}\{d(x,y) : y \in S\} = 0$. Obsérvese que puede que este ínfimo sea un mínimo, eso ciertamente ocurre si y sólo si $x \in S$.

"$\Longleftarrow$" El que $\text{ínf}\{d(x,y) : y \in S\} = 0$ quiere decir que para todo $\varepsilon > 0$ existe un $y \in S$ tal que $d(x,y) < \varepsilon$. Entonces $B(x,\varepsilon) \cap S \neq \emptyset$, esto es, $x \in \overline{S}$. $\square$

Definición 1.1.34 *Se llama* ***punto frontera*** *de un conjunto S de $\mathbb{R}^n$ a todo punto x de $\mathbb{R}^n$ con la propiedad de que para todo $r > 0$ se verifique que*

$$B(x,r) \cap S \neq \emptyset \text{ y } B(x,r) \cap (\mathbb{R}^n \backslash S) \neq \emptyset.$$

Ejemplo 1.1.35 *Los puntos x de $\mathbb{R}^n$ tales que $d(x, x_0) = r$ son puntos frontera de $B(x_0, r)$ y de $\overline{B}(x_0, r)$.*

Definición 1.1.36 *Se dice que un punto $x \in S$ es un* ***punto aislado*** *de S si existe $r > 0$ tal que $B(x,r) \cap S = \{x\}$.*

Ejemplo 1.1.37 *Cada punto de la forma $\frac{1}{k}$ es aislado de $S = \{1, \frac{1}{2}, \frac{1}{3}, ...\}$.*

1.2. Convergencia

Definición 1.2.1 *Se llama* ***sucesión*** *en $\mathbb{R}^n$ a toda aplicación $x : \mathbb{N} \to \mathbb{R}^n$. Como se hace en el caso de sucesiones de números reales, se identificará x con la "tira" de sus valores $(x(1), x(2), \ldots)$. Normalmente se escribirá x_k en vez de $x(k)$ y las sucesiones se representarán por $\{x_k\}_{k=1}^{\infty}$, $\{x_k\}$, $(x_k)_{k=1}^{\infty}$ o (x_k). Obsérvese que cada x_k es un punto de $\mathbb{R}^n$ y por lo tanto es de la forma $x_k = (x_{k,1}, ..., x_{k,n})$.*

Definición 1.2.2 *Se dice que una sucesión $\{x_k\}$* ***converge*** *a un punto $x_0 \in \mathbb{R}^n$ si la sucesión de números reales $\{d(x_k, x_0)\}$ converge a 0. Esto es,*

$$\forall \varepsilon > 0 \; \exists \nu \in \mathbb{N} \mid k \geq \nu \Longrightarrow d(x_k, x_0) < \varepsilon$$

Si $n = 1$, la anterior condición se puede expresar así:

$$\forall \varepsilon > 0 \; \exists \nu \in \mathbb{N} \mid k \geq \nu \Longrightarrow |x_k - x_0| < \varepsilon.$$

Observamos que si $\{x_k\}$ converge a x_0 entonces no puede converger también a otro punto y_0, siendo $y_0 \neq x_0$. Si lo hiciese, tomando como ε el número real $d(x_0, y_0)/2$ que es mayor que 0, obtendríamos números naturales ν_1 y ν_2 tales que $d(x_k, x_0) < \varepsilon$ si $k \geq \nu_1$ y $d(x_k, y_0) < \varepsilon$ si $k \geq \nu_2$. La desigualdad triangular nos daría entonces que para $k \geq$máx(ν_1, ν_2),

$$d(x_0, y_0) \leq d(x_k, x_0) + d(x_k, y_0) < 2\varepsilon = d(x_0, y_0).$$

Para indicar que la sucesión $\{x_k\}$ converge a x_0 escribiremos $x_k \to x_0$, lím $x_k = x_0$, o $\lim\limits_{k \to \infty} x_k = x_0$.

Nota 1.2.3 *Sucede que* $x_k \to x_0$ *si y sólo si para todo* $i = 1, ..., n$ *se verifica que* $x_{k,i} \to x_{0,i}$. *En efecto, por una parte, si* $x_k \to x_0$ *entonces dado* $\varepsilon > 0$ $\exists \nu \in \mathbb{N} \mid k \geq \nu \Longrightarrow$

$$d(x_k, x_0) = \sqrt{(x_{k,1} - x_{0,1})^2 + \cdots + (x_{k,n} - x_{0,n})^2} < \varepsilon,$$

con lo que para todo $i = 1, ..., n$ *se verifica que*

$$|x_{k,i} - x_{0,i}| < \varepsilon \ \forall k \geq \nu.$$

Por otra parte, si dado $\varepsilon > 0$ *consideramos* $\frac{\varepsilon}{\sqrt{n}}$, *el hecho de que cada* $x_{k,i} \to x_{0,i}$ *nos da la existencia de un* $\nu \in \mathbb{N}$ *tal que* $|x_{k,i} - x_{0,i}| < \frac{\varepsilon}{\sqrt{n}}$ $\forall k \geq \nu$ *y* $\forall i = 1, ..., n$. *En consecuencia*

$$\begin{aligned} \sqrt{(x_{k,1} - x_{0,1})^2 + \cdots + (x_{k,n} - x_{0,n})^2} &< \sqrt{\left(\frac{\varepsilon}{\sqrt{n}}\right)^2 + \cdots + \left(\frac{\varepsilon}{\sqrt{n}}\right)^2} \\ &= \sqrt{\frac{n\varepsilon^2}{n}} = \varepsilon \end{aligned}$$

para todo $k \geq \nu$.

Nota 1.2.4 *Si* $x_k \to x_0$ *entonces* x_0 *es un punto adherente al conjunto formado por los puntos de la sucesión.*

Ejercicio 1.2.5 *Demostrar que* x_0 *es un punto adherente a un conjunto* S *si y sólo si existe una sucesión de puntos de* S *convergente a él.*

Ejercicio 1.2.6 *Demostrar que* x_0 *es un punto de acumulación de un conjunto* S *si y sólo si existe una sucesión de elementos de* S *con todos sus términos distintos convergente a él.*

Definición 1.2.7 *Se llama* ***subsucesión*** *de una sucesión dada* $\{x_k\}$ *a toda aplicación* $m \in \mathbb{N} \mapsto x_{k_m}$, *donde* x_{k_m} *es uno de los términos de la sucesión* $\{x_k\}$, *que verifique la condición de que la aplicación* $m \in \mathbb{N} \mapsto k_m$ *sea estrictamente creciente. Las sucesiones suelen representarse por* $\{x_{k_m}\}_{m=1}^{\infty}$, $\{x_{k_m}\}$, $(x_{k_m})_{m=1}^{\infty}$ *o* (x_{k_m}).

Ejercicio 1.2.8 *Demostrar que si* x_0 *es un punto de acumulación de una sucesión* $\{x_k\}$ *entonces existe una subsucesión de* $\{x_k\}$ *con todos los puntos distintos convergente a* x_0.

Teorema 1.2.9 *Si una sucesión* $\{x_k\}$ *es convergente a un punto* x_0 *entonces toda subsucesión de ella converge a* x_0.

Demostración. El hecho de que $\{x_k\}$ es convergente a x_0 nos dice que

$$\forall\varepsilon > 0\ \exists\nu \in \mathbb{N} \mid k \geq \nu \Longrightarrow d(x_k, x_0) < \varepsilon$$

Sea $\{x_{k_m}\}$ una subsucesión de $\{x_k\}$. Dado $\varepsilon > 0$ consideremos un $\nu \in \mathbb{N} \mid k \geq \nu \Longrightarrow d(x_k, x_0) < \varepsilon$. Entonces para todo $m \geq \nu$ se tiene que $k_m \geq m \geq \nu$ y por lo tanto $d(x_{k_m}, x_0) < \varepsilon$. □

De entre todas las posibles sucesiones de $\mathbb{R}^n$ son particularmente importantes las llamadas **sucesiones de Cauchy**[2].

Definición 1.2.10 *Se dice que una sucesión $\{x_k\}$ es de Cauchy si*

$$\forall\varepsilon > 0\ \exists\nu \in \mathbb{N} \mid p, q \geq \nu \Longrightarrow d(x_p, x_q) < \varepsilon$$

Si $n = 1$ la condición es:

$$\forall\varepsilon > 0\ \exists\nu \in \mathbb{N} \mid p, q \geq \nu \Longrightarrow |x_p - x_q| < \varepsilon.$$

Teorema 1.2.11 *Una sucesión de $\mathbb{R}^n$ es convergente si y sólo si es de Cauchy.*

Demostración. Por una parte, sea $\{x_k\}$ una sucesión convergente y sea x_o su punto límite (como ya sabemos de haber un límite, éste es único). Como $x_k \to x_0$, dado $\varepsilon > 0$ existe $\nu \in \mathbb{N}$ tal que para todo $k \in \mathbb{N}$, $k \geq \nu$, se verifica que $d(x_k, x_0) < \frac{\varepsilon}{2}$ (nótese que realmente hemos tomado $\frac{\varepsilon}{2}$ en vez de ε al expresar el hecho de que $x_k \to x_0$). Tomando p y q mayores o iguales que ν, la propiedad triangular de la métrica d nos da que:

$$d(x_p, x_q) \leq d(x_p, x_0) + d(x_q, x_0) < \varepsilon$$

lo que nos proporciona la condición requerida para que la sucesión $\{x_k\}$ sea de Cauchy.

Por otra parte si $\{x_k\}$ es una sucesión de Cauchy, entonces

$$\forall\varepsilon > 0\ \exists\nu \in \mathbb{N}\ |p, q \geq \nu \Longrightarrow d(x_p, x_q) < \varepsilon.$$

De aquí se deduce que para cada $i = 1, 2, ..., n$, $\{x_{k,i}\}_{k=1}^{\infty}$ es una sucesión de Cauchy de números reales y por lo tanto, como se sabe del curso de Análisis de Variable Real, existe $x_{0,i} \in \mathbb{R}$ tal que $x_{k,\,i} \to x_{0,i}$. Consideremos el elemento de $\mathbb{R}^n$ dado por $x_0 = (x_{0,1}, ..., x_{0,n})$. Por lo que hemos visto en la Nota 1.2.3 resulta que $x_k \to x_0$. □

Ejercicio 1.2.12 *Si $x_k \to x_0$ e $y_k \to y_0$, demostrar que $x_k + y_k \to x_0 + y_0$. Demostrar también que si $x_k \to x_0$ y $\alpha \in \mathbb{R}$, entonces $\alpha x_k \to \alpha x_0$.*

[2]Agustin L. Cauchy fue un matemático francés que vivió entre los años 1789 y 1857, su aportación a las matemáticas fue realmente significativa, algunos de sus resultados los encontraréis en distintas asignaturas de la licenciatura.

1.3. Compacidad y conexión

1.3.1. Compacidad

Para introducir el importante concepto de compacidad damos previamente la siguiente definición:

Definición 1.3.1 *a) Llamaremos* ***recubrimiento*** *de un subconjunto S de $\mathbb{R}^n$ a cualquier colección de subconjuntos de $\mathbb{R}^n$ cuya unión contenga a S.*

b) Se dice que un recubrimiento es ***abierto*** *si los conjuntos que lo forman son abiertos.*

c) Llamaremos ***subrecubrimiento*** *de un recubrimiento de un conjunto S a toda colección de elementos del recubrimiento que sea un recubrimiento de S.*

d) Finalmente hablaremos de ***recubrimiento finito*** *de un conjunto S de $\mathbb{R}^n$ cuando se trate de un recubrimiento de S formado por una cantidad finita de subconjuntos de $\mathbb{R}^n$.*

Podemos ya definir ahora el concepto al que nos referíamos anteriormente:

Definición 1.3.2 *Se dice que un subconjunto K de $\mathbb{R}^n$ es* ***compacto*** *si de todo recubrimiento abierto de él se puede extraer un subrecubrimiento finito.*

Ejemplos:

1. $\left\{0, 1, \frac{1}{2}, \frac{1}{3}, ...\right\}$ es conjunto compacto de $\mathbb{R}$. En efecto, si tomamos un recubrimiento abierto $\{A_\alpha : \alpha \in \Lambda\}$ (Λ es un conjunto arbitrario) de nuestro conjunto, entonces el 0 debe estar en un cierto A_{α_0}. Al ser A_{α_0} abierto existe un intervalo centrado en 0 contenido en A_{α_0}. Como la sucesión $\{1/k\}$ converge a 0 existe ν tal que los x_k, con $k \geq \nu$, pertenecen a A_{α_0}. Luego si consideramos una cantidad finita de abiertos A_α que contengan a los puntos $1, \frac{1}{2}, ..., \frac{1}{\nu-1}$ y el A_{α_0} tendremos un subrecubrimiento finito del conjunto $\{0, 1, \frac{1}{2}, \frac{1}{3}, ...\}$. El mismo argumento prueba que el conjunto formado por los puntos de una sucesión convergente y su límite es un conjunto compacto.
2. El conjunto $\{1, 1/2, 1/3, ...\}$ no es compacto pues del recubrimiento de él formado por los abiertos $A_i : i \in \mathbb{N}$, donde $A_i = (\frac{1}{i} - \frac{1}{i+1}, \frac{1}{i} + \frac{1}{i+1})$, no se puede extraer ningún subrecubrimiento finito.
3. El conjunto $(0, 2]$ no es compacto, pues del recubrimiento abierto
$$\left\{(1,3), \left(\frac{1}{2}, 3\right), \left(\frac{1}{3}, 3\right), ...\right\}$$
de él no se puede extraer ningún subrecubrimiento finito.

4. El conjunto $B(0,1)$ no es compacto, pues del recubrimiento abierto formado por la unión de las bolas $B(0,1-\frac{1}{j})$ no se puede extraer ninguno finito.

La compacidad de un conjunto (cuando la hay) es una propiedad que da lugar a interesantes resultados, por ejemplo, veremos más adelante que toda función continua definida en conjunto compacto alcanza en él el supremo, resultado que ya es conocido para las funciones continuas en un intervalo cerrado de $\mathbb{R}$.

Teorema 1.3.3 *Los subconjuntos compactos de $\mathbb{R}^n$ son cerrados y todo subconjunto cerrado contenido en un compacto es compacto.*

Demostración. Sea K un compacto, vamos a probar que $\mathbb{R}^n \backslash K$ es abierto. Para ello sea $x \in \mathbb{R}^n \backslash K$, como $x \notin K$ entonces para todo $y \in K$ existen $r_y > 0$ tal que $B(x, r_y) \cap B(y, r_y) = \emptyset$ (basta tomar $r_y = \frac{d(x,y)}{2}$). Es claro que K está contenido en $\cup_{y \in K} B(y, r_y)$ y por lo tanto existen puntos $y_1, ..., y_k$ en K tales que $K \subset \cup_{i=1}^k B(y_i, r_{yi})$. Entonces la bola de centro x y radio igual al mínimo de los $r_{yi},\ i = 1, ..., k$ está contenida en $\mathbb{R}^n \backslash K$.

Sea ahora C un subconjunto cerrado de $\mathbb{R}^n$ contenido en el compacto K. Consideremos un recubrimiento $\{A_\alpha : \alpha \in \Lambda\}$ de C formado por abiertos. Llamemos A al complementario de C en $\mathbb{R}^n$. Entonces $\{A_\alpha \cup A : \alpha \in \Lambda\}$ es un recubrimiento abierto de K por lo que existirán $\alpha_1, ..., \alpha_m \in \Lambda$ tales que $K \subset A_{\alpha_1} \cup \cdots \cup A_{\alpha_m} \cup A$. Evidentemente $C \subset A_{\alpha_1} \cup \cdots \cup A_{\alpha_m}$. □

Teorema 1.3.4 *Todo conjunto compacto es acotado.*

Demostración. Sea K un conjunto compacto. Del recubrimiento por bolas abiertas $\{B(0,k), k = 1, 2, ...\}$ se tiene que poder extraer uno finito. Esto es, existe $k_0 \in \mathbb{N}$ tal que $K \subset B(0, k_0)$, con lo que

$$\sup\{d(x,y) :\ x, y \in K\} \leq 2k_0$$

y entonces K es acotado. □

Un poco más adelante veremos que un subconjunto de $\mathbb{R}^n$ es compacto si y sólo si es cerrado y acotado. El siguiente teorema da una caracterización de los subconjuntos compactos:

Teorema 1.3.5 (de Bolzano-Weierstrass de compacidad por sucesiones) *Un subconjunto K de $\mathbb{R}^n$ es compacto si y sólo si toda sucesión de elementos de K tiene una subsucesión convergente a un punto de él.*

Demostración. Sea K un subconjunto compacto de $\mathbb{R}^n$ y sea $\{x_k\}$ una sucesión de elementos de K. Supongamos que tal sucesión no tiene ninguna subsucesión convergente a un punto de K. Ciertamente $\{x_k\}$ tiene que tener infinitos elementos distintos, pues en caso contrario tendría subsucesiones convergentes a un punto de K. Denotemos por $\{x_{k_m}\}$ a una subsucesión de $\{x_k\}$ cuyos (infinitos) términos sean todos distintos entre sí. Sea S el conjunto formado por los puntos de esa sucesión. Entonces $S' = \emptyset$, pues si existiese un punto x_0 en S' (que está contenido en K', que a su vez está contenido en K, pues éste es cerrado), entonces habría una subsucesión de puntos de S (en particular de $\{x_k\}$) convergente a $x_0 \in K$ (Ejercicio 1.2.8) y estamos suponiendo que esto no es cierto. Esto nos permite afirmar, en particular, que para cada $m \in \mathbb{N}$ existe $r_m > 0$ tal que $B(x_{k_m}, r_m) \cap S = \{x_{k_m}\}$ (como conjunto) y que S es cerrado (pues contiene a $S' = \emptyset$) y como S está contenido en el compacto K, es también compacto. Sin embargo esto no es posible, pues del recubrimiento de S formado por las bolas $B(x_{k_m}, r_m)$, $m \in \mathbb{N}$, es imposible extraer ninguno finito.

Recíprocamente, supongamos ahora que K tiene la propiedad de que toda sucesión de elementos de él tiene una subsucesión convergente a un punto de él. Entonces:

I. Veamos en primer lugar como para todo $r > 0$ del recubrimiento por abiertos definido por las bolas de centro en los puntos de K y radio r existe un subrecubrimiento finito, esto es, existe una cantidad finita $x_1, ..., x_m$ de puntos de K tal que

$$K \subset \cup_{j=1}^{m} B(x_j, r).$$

Si esto no fuese cierto existiría $r_0 > 0$ tal que K no se podría recubrir por ninguna cantidad finita de bolas de radio r_0 centradas en puntos de K. Fijemos un punto $x_1 \in K$. Como $B(x_1, r_0)$ no recubre a K, existe $x_2 \in K \backslash B(x_1, r_0)$. Como tampoco $B(x_1, r_0) \cup B(x_2, r_0)$ recubre a K, existe $x_3 \in K \backslash B(x_1, r_0) \cup B(x_2, r_0)$. Reiterando este proceso obtenemos una sucesión $\{x_k\}$ de puntos de K que evidentemente no tiene ninguna subsucesión convergente, pues para cada p y $q \in \mathbb{N}$, $d(x_p, x_q) \geq r_0$.

II. Sea $\{A_\alpha : \alpha \in \Lambda\}$ un recubrimiento abierto de K. Con el fin de ver que de él podemos extraer un subrecubrimiento finito vamos a ver primero como existe $r^* > 0$ tal que para todo $x \in K$ existe $\alpha_x \in \Lambda$ verificándose que $B(x, r^*) \subset A_{\alpha_x}$. En efecto, en caso contrario existiría para todo $k \in \mathbb{N}$ un punto $x_k \in K$ tal que $B(x_k, \frac{1}{k})$ no estaría contenida en A_β cualquiera que sea $\beta \in \Lambda$. La sucesión $\{x_k\}$ así formada deberá tener una subsucesión $\{x_{k_m}\}_{m=1}^{\infty}$ convergente a un punto $x_0 \in K$ (por la hipótesis). Ese punto x_0 deberá estar en algún A_{α_o}, y al ser éste un conjunto abierto existe un $r_0 > 0$ tal que $B(x_0, r_0) \subset A_{\alpha_o}$. Tomemos m bastante grande para que $d(x_{k_m}, x_0) < \frac{r_0}{2}$ y $\frac{1}{k_m} < \frac{r_0}{2}$. Entonces $B(x_{k_m}, 1/k_m) \subset B(x_0, r_0) \subset A_{\alpha_o}$ lo que contradice el que $B(x_{k_m}, 1/k_m)$ no está contenida en A_β para ningún $\beta \in \Lambda$.

III. Fijemos finalmente un $r^* > 0$ tal que para todo $x \in K$ existe $\alpha_x \in \Lambda$ de forma que $B(x, r^*) \subset A_{\alpha_x}$. Según hemos visto anteriormente existen puntos $x_1, ..., x_k \in K$ tales que $K \subset \cup_{i=1}^{k} B(x_i, r^*)$. Dado que cada $B(x_i, r^*)$ está contenida en algún $A_{\alpha_{x_i}}$, vemos que una cantidad finita de A_α ya recubren a K. $\square$

Corolario 1.3.6 (Teorema de Bolzano-Weierstrass para conjuntos) *Un conjunto K es compacto si y sólo si todo subconjunto de él con infinitos elementos tiene un punto de acumulación en K.*

Demostración. Si K es compacto y consideramos un subconjunto S de él con infinitos elementos entonces podemos formar una sucesión de elementos de S con todos sus términos distintos. Como por el teorema anterior, toda sucesión de elementos de S tiene una subsucesión convergente a un punto de K, si consideramos una tal subsucesión y denotamos por x_0 su límite, resulta que x_0 es un punto de acumulación de S y por lo tanto de K.

Recíprocamente, dada un sucesión de elementos de K pueden suceder dos casos: que tenga infinitos términos distintos, en cuyo caso tiene, por hipótesis, un punto de acumulación en K y en consecuencia una subsucesión convergente a un punto de K, o que no tenga infinitos términos distintos, en cuyo caso necesariamente hay algún término que se repite infinitas veces, con lo que ya tiene una subsucesión convergente a un elemento de K. $\square$

Nota 1.3.7 *Observamos que el teorema anterior y su corolario se verifican en cualquier espacio métrico, mientras que el siguiente se verifica en $\mathbb{R}^n$, pero no, en general, en un espacio métrico arbitrario. Piénsese por ejemplo que el conjunto $\mathbb{N}$ es cerrado y acotado pero no es compacto con respecto a la métrica discreta.*

Teorema 1.3.8 (de Heine-Borel) *Un subconjunto de $\mathbb{R}^n$ es compacto si y sólo si es cerrado y acotado.*

Demostración. Ya hemos visto anteriormente que los conjuntos compactos son cerrados y acotados. Veamos el recíproco. Sea K un subconjunto de $\mathbb{R}^n$ cerrado y acotado, probaremos que K es compacto demostrando que toda sucesión de elementos de él tiene una subsucesión convergente a un punto de K (teorema de Bolzano-Weierstrass de compacidad por sucesiones). Consideremos una sucesión $\{x_k\}$ de puntos de K y por cada $i = 1, ..., n$ consideremos la sucesión de números reales formada por las coordenadas de lugar i de los términos de $\{x_k\}$, nos estamos refiriendo a la sucesión $\{x_{k,i}\}_{k=1}^{\infty}$. Tomemos $i = 1$. Se sabe del curso de Análisis de Variable Real que al ser la sucesión $\{x_{k,1}\}$ acotada de $\mathbb{R}$, tiene una subsucesión $\{x_{k_l,1}\}_{l=1}^{\infty}$ convergente a un punto que denotaremos por $x_{0,1}$ (teorema de Bolzano-Weierstrass para sucesiones de números reales; puede verse una demostración en el libro de Marsden y Hoffman, "Análisis Clásico Elemental", Teorema 1.4.3). La sucesión $\{x_{k_l,2}\}_{l=1}^{\infty}$ tiene, por la misma razón, una subsucesión convergente $\{x_{k_{l_r},2}\}_{r=1}^{\infty}$ a un cierto $x_{0,2}$. Reiterando este proceso obtendremos finalmente una cierta sucesión $\left\{x_{k_{l_{r_{\cdot_{\cdot_{\cdot_t}}}}},n}\right\}$ convergente a un cierto $x_{0,n}$. Entonces $\{x_k\}$ tiene una subsucesión, la $\left\{x_{k_{l_{r_{\cdot_{\cdot_{\cdot_t}}}}}}\right\}$, convergente al punto $x_0 = (x_{0,1}, ..., x_{0,n})$. Este punto es adherente a $\{x_k\}$ y por lo tanto a K; como K es cerrado, necesariamente $x_0 \in K$. $\square$

1.3.2. Conexión

Definición 1.3.9 *Sea S un subconjunto de $\mathbb{R}^n$. Se dice que un par de abiertos (U, V) de $\mathbb{R}^n$* ***separa*** *a S si verifican:*

(a) $U \cap V \cap S = \emptyset$, (b) $U \cap S \neq \emptyset$, (c) $V \cap S \neq \emptyset$ y (d) $S \subset U \cup V$.

Cuando S admite una separación diremos que S es **disconexo**, cuando no admite ninguna separación diremos que S es **conexo**.

Ejemplo 1.3.10 *El conjunto $S = [1, 2] \cup [3, 4]$ admite la separación formada por los abiertos $U = (0, \frac{5}{2}), V = (\frac{5}{2}, 5)$. Por tanto S es disconexo. En general, si S es un subconjunto de $\mathbb{R}$, a y b son dos puntos de S tales que $a < b$ y existe un punto $c \in (a, b)$ tal que $c \notin S$, entonces S es disconexo. En efecto, basta considerar $U = (-\infty, c)$, $V = (c, +\infty)$.*

Podemos afirmar entonces que los subconjuntos de $\mathbb{R}$ conexos tiene que ser intervalos, o bien un sólo punto. Más aún, veremos a continuación que todos los intervalos de $\mathbb{R}$ son conjuntos conexos, luego un subconjunto de $\mathbb{R}$ es conexo si y sólo si es un intervalo o un punto.

Teorema 1.3.11 *Sea I un intervalo de $\mathbb{R}$, entonces I es conexo.*

Demostración. Supongamos que I admite una separación (U, V) y tomemos puntos $u \in U \cap I$ y $v \in V \cap I$. Desde luego $u \neq v$ (pues en caso contrario $U \cap V \cap I$ no sería vacío. Supondremos que $u < v$ (si $v < u$ se haría lo mismo). Sea y_0 el supremo del conjunto $U \cap [u, v]$ (nótese que $U \cap [u, v]$ es no vacío y está acotado superiormente). Como $[u, v] \subset I \subset U \cup V$, y_0 debe pertenecer a $U \cup V$ (téngase en cuenta que $y_0 \in [u, v]$ por ser éste un conjunto cerrado).

Ahora bien, si $y_0 \in U$, entonces $y_0 < v$ (pues si $y_0 = v$, $U \cap V \cap I$ no sería vacío). Al ser U abierto existe $\delta > 0$ tal que $(y_0 - \delta, y_0 + \delta) \subset U$ y δ puede tomarse suficientemente pequeño para que $[y_0, y_0+\delta) \subset U \cap [u, v]$. Entonces $y_0+\frac{\delta}{2}$, que es mayor que y_0, pertenecería a $U \cap [u, v]$ en contra del hecho de que $y_0 = \sup(U \cap [u, v])$.

Si $y_0 \in V$, necesariamente $u < y_0$ (pues si $y_0 = u$, $U \cap V \cap I$ no sería vacío) y al ser V abierto existe $\delta' > 0$ tal que $(y_0 - \delta', y_0] \subset V \cap [u, v]$. Como $y_0 - \delta'$ no es cota superior de $U \cap [u, v]$ (por ser $y_0 = \sup(U \cap [u, v])$) existe $a \in U \cap [u, v]$ tal que $y_0 - \delta' < a \leq y_0$. Entonces $a \in U \cap V \cap I$ en contra de lo que hemos visto anteriormente.

Resulta, en consecuencia, que y_0 no pertenece a $U \cup V$ lo que es absurdo. □

Teorema 1.3.12 *Sea $\{S_\alpha : \alpha \in \Lambda\}$ una colección de conjuntos conexos con intersección no vacía, entonces su unión es también un conjunto conexo.*

Demostración. Supongamos que $\cup_{\alpha\in\Lambda}S_\alpha$ admite una separación (U,V) y sea $x_0 \in \cap_{\alpha\in\Lambda}S_\alpha$. Como $\cup_{\alpha\in\Lambda}S_\alpha \subset U\cup V$, x_0 debe pertenecer a $U\cup V$, supongamos que $x_0\in U$. También existe $y_0 \in (\cup_{\alpha\in\Lambda}S_\alpha)\cap V$ luego existe $\alpha_0\in\Lambda$ tal que $y_0\in S_{\alpha_0}\cap V$ y por lo tanto (U,V) es una separación de S_{α_0} lo que es absurdo. □

Hay un tipo de conexión, la llamada conexión por poligonales que está relacionada con la anterior y que se define de la siguiente manera:

Definición 1.3.13 *Se dice que un subconjunto S de $\mathbb{R}^n$ es* ***conexo por poligonales*** *si para cualesquiera puntos x e y de S existe una poligonal contenida en S que los une. Una* ***poligonal*** *no es más que una concatenación de* ***segmentos*** *y éstos son subconjuntos de la forma $L[a,b]=\{a+t(b-a) : t\in[0,1]\}$, siendo a y b dos elementos de $\mathbb{R}^n$. Así, una poligonal $\Gamma[x,y]$ en S es la concatenación de una colección finita de segmentos $L[x,a_1]$, $L[a_1,a_2],...,L[a_{m-1},y]$, todos ellos contenidos en S.*

Cualquier bola $B(x_0,r)$ es conexa por poligonales: Dados dos puntos x e y de $B(x_0,r)$, el segmento que los une ya está contenido en la bola, pues para todo $t\in[0,1]$, se tiene que

$$\begin{aligned}
||x+t(y-x)-x_0|| &= ||(1-t)x+ty-(tx_0+(1-t)x_0|| \\
&= ||(1-t)(x-x_0)+t(y-x_0)|| \\
&\leq (1-t)||x-x_0||+t||y-x_0|| \\
&< (1-t)r+tr=r.
\end{aligned}$$

Los conjuntos S que tienen la propiedad de que dados dos puntos de S el segmento que los une está contenido en S se llaman conjuntos **convexos** y se considerarán más adelante.

Nota 1.3.14 *Sucede que todo subconjunto S de $\mathbb{R}^n$ que sea conexo por poligonales es conexo. Podría hacerse aquí la prueba de esto, pero es mejor dejarla para un poco más adelante, donde se obtendrá de forma muy sencilla.*

Teorema 1.3.15 *Todo subconjunto abierto A de $\mathbb{R}^n$ que sea conexo es conexo por poligonales.*

Demostración. Supongamos que A es conexo y fijemos un punto x_0 en A. Consideremos los conjuntos

$$\begin{aligned}
U &= \{x\in A : x_0 \text{ se puede unir con } x \text{ por una poligonal en } A\} \\
V &= \{x\in A : x_0 \textbf{ no} \text{ se puede unir con } x \text{ por una poligonal en } A\}.
\end{aligned}$$

Entonces U y V son abiertos de $\mathbb{R}^n$. En efecto, sea $x\in U$ y sea $r>0$ tal que $B(x,r)\subset A$ (téngase en cuenta que A es abierto). Todo punto y de $B(x,r)$ se puede unir con x mediante el segmento $L[y, x]$ y así, si $\Gamma[x_0, x]$ es una poligonal contenida en A que une x_0 con x, entonces $\Gamma[x_0, x] \cup L[x, y]$ es una poligonal en A que une x_0 con y. Esto implica que $B(x, r) \subset U$ y prueba que U es abierto.

Sea ahora $x \in V$ y sea $r > 0$ tal que $B(x, r) \subset A$. Entonces x_0 no puede unirse con ningún punto y de $B(x, r)$ mediante una poligonal en A, pues si x_0 se pudiese unir con algún $y \in B(x, r)$ mediante una poligonal $\Gamma[x_0, y]$ entonces x_0 también se podría unir con x mediante la poligonal $\Gamma[x_0, y] \cup \Gamma[y, x]$. Así, $B(x, r) \subset V$ y V es abierto.

Por la propia definición de U se tiene que $U \cap V = \emptyset$, $U \cap A \neq \emptyset$ y $A \subset U \cup V$. Como A es conexo necesariamente debe ocurrir que $V \cap A = \emptyset$. Esto es, no hay ningún punto de A con el que no se pueda unir x_0 mediante una poligonal en A, y si x_0 puede unirse con todo punto de A mediante una poligonal $\Gamma[x_0, x]$ en A entonces todo par de puntos x e y de A se pueden unir por la poligonal $\Gamma[x, x_0] \cup \Gamma[x_0, y]$, donde $\Gamma[x, x_0]$ es la poligonal definida a partir de una poligonal $\Gamma[x_0, x] = L[x_0, a_1] \cup L[a_1, a_2] \cup \cdots \cup L[a_{n-1}, x]$ por

$$\Gamma[x, x_0] = L[x, a_{n-1}] \cup L[a_{n-1}, a_{n-2}] \cup \cdots \cup L[a_2, a_1] \cup L[a_1, x_0].$$

□

Nota 1.3.16 *Los conceptos de conexión y de conexión por poligonales pueden darse en cualquier espacio normado y también en ese contexto son ciertos los dos teoremas anteriores.*

Nota 1.3.17 *Hay ejemplos de conjuntos no abiertos de $\mathbb{R}^n$ que son conexos pero no son conexos por poligonales. Por ejemplo, el conjunto*

$$S = \bigcup_{n=1}^{\infty} \left\{ \left((x, \frac{1}{n} x \right) \in \mathbb{R}^2 : x \in [0, 1] \right\} \cup \{(1, 0)\}$$

tiene esa característica.

1.4. Continuidad. Imágenes de conjuntos compactos y conexos

Comenzamos definiendo el concepto de *límite* de una aplicación de un subconjunto de $\mathbb{R}^n$ con valores en $\mathbb{R}^m$, siendo m un número natural, que puede coincidir con n o no, este concepto será la base para la introducción de la continuidad.

Consideremos un subconjunto S de $\mathbb{R}^n$, un punto $x_0 \in S'$, un punto $l \in \mathbb{R}^m$ y una aplicación $f : S \to \mathbb{R}^m$.

Definición 1.4.1 *Se dice que f tiene por **límite** l en el punto x_0 si*

$$\forall \varepsilon > 0 \ \exists \delta > 0 \mid \forall x \in S, \ x \neq x_0, \ d(x, x_0) < \delta \Longrightarrow d(f(x), l) < \varepsilon.$$

Esto es,

$$\forall \varepsilon > 0 \ \exists \delta > 0 \mid \forall x \in B(x_0, \delta) \cap S, \ x \neq x_0 \Rightarrow \ f(x) \in B(l, \varepsilon).$$

Obsérvese que d representa, en principio, tanto a la métrica euclídea en $\mathbb{R}^n$ como en $\mathbb{R}^m$.

En el caso particular de que $m = n = 1$, la condición es

$$\forall \varepsilon > 0 \ \exists \delta > 0 \mid \forall x \in S, \ x \neq x_0, \ |x - x_0| < \delta \Longrightarrow |f(x) - l| < \varepsilon.$$

Nota 1.4.2 *La definición anterior puede darse de forma análoga en cualquier espacio métrico.*

Nota 1.4.3 *El punto x_0 debe ser un punto de acumulación de S, pues de lo contrario no tiene interés lo anterior, puede no haber puntos de S "cerca" del x_0 distintos de él. Por otra parte nótese que no se exige que x_0 pertenezca a S y que aunque pertenezca a S no tiene por qué ocurrir que $l = f(x_0)$. Por ejemplo la función $f : \mathbb{R} \to \mathbb{R}$ definida por*

$$f(x) = \begin{cases} 1 & si\ x \neq 0 \\ 2 & si\ x = 0 \end{cases}$$

tiene límite 1 en el punto 0, mientras que $f(0) = 2$.

Para expresar que f tiene por límite l en el punto x_0 se escribe $\lim_{x \to x_0} f(x) = l$, o bien, $f(x) \to l$ cuando $x \to x_0$.

Ejercicio 1.4.4 *a) Demostrar que se verifica que si $\lim_{x \to x_0} f(x) = l$ y $\lim_{x \to x_0} g(x) = l'$, entonces $\lim_{x \to x_0} (f + g)(x) = l + l'$*

b) Demostrar que si $\lim_{x \to x_0} f(x) = l$ y $\alpha \in \mathbb{R}$, entonces $\lim_{x \to x_0} \alpha f(x) = \alpha l$

c) Demostrar que para $f : S \subset \mathbb{R}^n \to \mathbb{R}^m$ se verifica que f tiene límite $l = (l_1, ..., l_m)$ en el punto x_0 si y sólo si tienen límite l_j sus coordenadas $f_j, \ j = 1, ..., m$ definidas por $f_j(x) = (f(x))_j$ en el punto x_0.

Ejercicio 1.4.5 *Demuéstrese que si existe el límite de una aplicación en un punto, éste es único.*

Teorema 1.4.6 *Si las funciones[3] f y g toman valores reales, $\lim_{x \to x_0} f(x) = l$ y $\lim_{x \to x_0} g(x) = l'$, entonces $\lim_{x \to x_0} (fg)(x) = l{\cdot}l'$.*

Demostración. Observamos que

$$\begin{aligned} |f(x)g(x) - l{\cdot}l'| &\leq |f(x)g(x) - f(x)l'| + |f(x)l' - l{\cdot}l'| \\ &= |f(x)||g(x) - l'| + |l'||f(x) - l|. \end{aligned}$$

Como $\lim_{x \to x_0} f(x) = l$ existe $\delta_1 > 0$ tal que $|f(x)| < |l| + 1$ si $0 < d(x, x_0) < \delta_1$ $(x \in S)$.

[3]Suele usarse el término función para las aplicaciones que toman valores reales.

Como $\lim_{x \to x_0} g(x) = l'$ existe $\delta_2 > 0$ tal que $|g(x) - l'| < \frac{\varepsilon}{2(|l|+1)}$ si $0 < d(x, x_0) < \delta_2$ $(x \in S)$.

El que $\lim_{x \to x_0} f(x) = l$ nos da además que existe $\delta_3 > 0$ tal que $|f(x) - l| < \frac{\varepsilon}{2(|l'|+1)}$ si $0 < d(x, x_0) < \delta_3$ $(x \in S)$.

Entonces, si hacemos $\delta = \min\{\delta_1, \delta_2, \delta_3\}$ y tomamos $x \in S$ tal que $0 < d(x, x_0) < \delta$ se tiene que

$$|f(x)g(x) - l{\cdot}l'| < \varepsilon$$

□

Nota 1.4.7 *Se pone $|l'|+1$ cuando se usa por segunda vez el que $\lim_{x \to x_0} f(x) = l$ para evitar un posible 0 en el denominador.*

Teorema 1.4.8 *Si la función f toma valores reales, $\lim_{x \to x_0} f(x) = l$ y $l \neq 0$, entonces $\lim_{x \to x_0} \frac{1}{f}(x) = \frac{1}{l}$. Observamos que si $l \neq 0$ entonces necesariamente, como se verá enseguida, $f(x) \neq 0$ para todos los x, $x \neq x_0$ de una cierta bola centrada en x_0.*

Demostración. Observamos primero que al ser $l \neq 0$, existe $\delta_1 > 0$ tal que $f(x) \neq 0$ para todo $x \in B(x_0, \delta_1) \cap S$, $x \neq x_0$, nótese que para un cierto $\delta_1 > 0$ se verifica que $|f(x) - l| < \frac{1}{2}|l|$ para todo $x \in B(x_0, \delta_1) \cap S$, $x \neq x_0$, con lo que $|f(x)| > \frac{1}{2}|l| > 0$. Para los $x \in B(x_0, \delta_1) \cap S$ se verifica que

$$\left| \frac{1}{f(x)} - \frac{1}{l} \right| = \left| \frac{f(x) - l}{f(x)l} \right| \leq \frac{|f(x) - l|}{\frac{1}{2}|l|^2}$$

Dado un $\varepsilon > 0$ arbitrario, para un δ_2 adecuado se puede conseguir que $|f(x) - l| < \varepsilon\frac{1}{2}|l|^2$ para todo $x \in B(x_0, \delta_2) \cap S$, $x \neq x_0$. Luego para $\delta = \min\{\delta_1, \delta_2\}$ y $x \in B(x_0, \delta) \cap S$, $x \neq x_0$, se tiene que

$$\left| \frac{1}{f(x)} - \frac{1}{l} \right| < \varepsilon.$$

□

Teorema 1.4.9 (criterio del límite por sucesiones) *Una aplicación $f : S \subset \mathbb{R}^n \to \mathbb{R}^m$ tiene límite l en x_0 si y sólo si para toda sucesión $\{x_k\}$ de puntos de S convergente a x_0, con $x_k \neq x_0$ para todo k, se verifica que $f(x_k) \to l$.*

Demostración. Sean $x_0 \in S$ y $\{x_k\} \subset S$ una sucesión que converja a x_0, con $x_k \neq x_0$ para todo k. Dado $\varepsilon > 0$ tomemos un $\delta > 0$ tal que si $x \in S$ y $0 < d(x, x_0) < \delta$ entonces $d(f(x), l) < \varepsilon$. Como $x_k \to x_0$ existe $\nu \in \mathbb{N}$ tal que si $k \geq \nu$ entonces $d(x_k, x_0) < \delta$, por lo que $d(f(x_k), l) < \varepsilon$. Esto prueba que $f(x_k) \to l$.

Recíprocamente, si suponemos que f no tiene por límite l en x_0, entonces existe $\varepsilon_0 > 0$ tal que para todo $k \in \mathbb{N}$ existe $x_k \in B(x_0, \frac{1}{k}) \cap S$, $x_k \neq x_0$, verificando que $d(f(x_k), l) \geq \varepsilon_0$. Es claro entonces que $x_k \to x_0$ y sin embargo $\{f(x_k)\}$ no converge a l. □

Ejercicio 1.4.10 *Utilizando el teorema anterior comprobar que la función* $f(x) = \operatorname{sen}\left(\frac{1}{x}\right)$ *no tiene límite en 0.*

Definición 1.4.11 *Sean* S *un subconjunto de* $\mathbb{R}^n$, x_0 *un punto de* $S \cap S'$ *y* f *una aplicación definida en* S *con valores en* $\mathbb{R}^m$. *Se dice que* f *es* ***continua*** *en* x_0 *si existe* $\lim_{x \to x_0} f(x)$ *y su valor es precisamente* $f(x_0)$.

Esto quiere decir que

$$\forall \varepsilon > 0 \; \exists \delta > 0 \mid x \in S, \; d(x, x_0) < \delta \Longrightarrow d(f(x), f(x_0)) < \varepsilon.$$

Esto es,

$$\forall \varepsilon > 0 \; \exists \delta > 0 \mid f(x) \in B(f(x_0), \varepsilon) \text{ si } x \in B(x_0, \delta) \cap S$$

En el caso particular de que $n = m = 1$ la condición es

$$\forall \varepsilon > 0 \; \exists \delta > 0 \mid \forall x \in S, \; |x - x_0| < \delta \Longrightarrow |f(x) - f(x_0)| < \varepsilon.$$

Teorema 1.4.12 (criterio de continuidad por sucesiones) *Sean* S *un subconjunto de* $\mathbb{R}^n$, x_0 *un punto de* $S \cap S'$ *y* $f : S \to \mathbb{R}^m$. *Entonces* f *es* ***continua*** *en* x_0 *si y sólo si para toda sucesión* $\{x_k\}$ *de puntos de* S *convergente a* x_0 *se verifica que* $f(x_k) \to f(x_0)$.

Demostración. Es consecuencia directa del teorema anterior. □

Una aplicación $f : S \to \mathbb{R}^m$ es continua en x_0 si y sólo si son continuas en ese punto todas las componentes $f_1, ..., f_m$ de f.

Definición 1.4.13 *Diremos que una aplicación* $f : S \subset \mathbb{R}^n \to \mathbb{R}^m$ *es continua en* S *si lo es en cada punto de* S *(esto lleva implícito que* $S \subset S'$*). Se considera que todas las aplicaciones son continuas en los puntos aislados de su conjunto de definición.*

Si tenemos dos aplicaciones $f, g : S \subset \mathbb{R}^n \to \mathbb{R}^m$, tiene sentido considerar su suma; pues bien, si f y g son continuas en $x_0 \in S$, entonces $f + g$ es también continua en x_0 (Ejercicio). Si tenemos una aplicación $f : S \subset \mathbb{R}^n \to \mathbb{R}^m$ continua en $x_0 \in S$ y $\alpha \in \mathbb{R}$, entonces αf es también continua en x_0 (Ejercicio). Por otra parte si tenemos dos aplicaciones $f, g : S \subset \mathbb{R}^n \to \mathbb{R}$ podemos multiplicarlas; pues bien, si f y g son continuas en $x_0 \in S$, entonces fg es continua en x_0. También sucede que si $f : S \subset \mathbb{R}^n \to \mathbb{R}$ es continua en x_0 y $f(x) \neq 0$ para todo x, entonces $\frac{1}{f}$ es continua en x_0 (Ejercicio).

Cuando tenemos aplicaciones $f : S \subset \mathbb{R}^n \to \mathbb{R}^m$, $g : T \subset \mathbb{R}^m \to \mathbb{R}^p$ tales que $f(S) \subset T$. Tiene sentido componer esas aplicaciones; pues bien, si f es continua en un punto $x_0 \in S$ y g lo es en el punto $f(x_0) \in T$, entonces $g \circ f$ es continua en x_0 (Ejercicio).

Damos a continuación un teorema que caracteriza la continuidad en términos de conjuntos abiertos.

Teorema 1.4.14 *Para una aplicación $f : S \subset \mathbb{R}^n \to \mathbb{R}^m$ equivalen:*

(a) f es continua en S

(b) Para todo abierto A de $\mathbb{R}^m$ existe un abierto B de $\mathbb{R}^n$ tal que $f^{-1}(A) = B \cap S$. Esto es, $f^{-1}(A)$ es un abierto de S (veáse la Nota 1.1.22)

Recordamos que $f^{-1}(A) = \{x \in S : f(x) \in A\}$.

Demostración. $(a) \Longrightarrow (b)$ Sea A un abierto de $\mathbb{R}^m$. Si $f^{-1}(A) = \emptyset$ basta tomar $B = \emptyset$. Si $f^{-1}(A) \neq \emptyset$, por cada $x \in S$ tal que $f(x) \in A$ consideramos un $\varepsilon_x > 0$ tal que $B(f(x), \varepsilon_x) \subset A$. La continuidad de f en x nos da que existe un $\delta_x > 0$ tal que $f(B(x, \delta_x) \cap S) \subset B(f(x), \varepsilon_x)$. El conjunto

$$B = \bigcup_{x \in f^{-1}(A)} B(x, \delta_x)$$

es un abierto de $\mathbb{R}^n$ y se verifica que $f^{-1}(A) = B \cap S$. En efecto, si $x_0 \in f^{-1}(A)$ es obvio, por la propia definición de B, que $x_0 \in B \cap S$. Por otra parte, si $x_0 \in B \cap S$ entonces existe $x \in f^{-1}(A)$ tal que $x_0 \in B(x, \delta_x)$ y, como $f(B(x, \delta_x) \cap S) \subset B(f(x), \varepsilon_x) \subset A$, resulta que $f(x_0) \in A$, es decir, $x_0 \in f^{-1}(A)$.

$(b) \Longrightarrow (a)$ Dados $x_0 \in S$ y $\varepsilon > 0$, consideremos en $\mathbb{R}^m$ la bola abierta $A = B(f(x_0), \varepsilon)$. Sea B un abierto de $\mathbb{R}^n$ tal que $f^{-1}(A) = B \cap S$. Como $x_0 \in f^{-1}(A)$ entonces $x_0 \in B \cap S$; dado que B es abierto existe $\delta > 0$ tal que $B(x_0, \delta) \subset B$, con lo que $f(B(x_0, \delta) \cap S) \subset B(f(x_0), \varepsilon)$. Esto es, si $x \in S$ y $d(x, x_0) < \delta$ entonces $d(f(x), f(x_0)) < \varepsilon$. $\square$

Nota 1.4.15 *En el caso particular de que S sea abierto, f es continua si y sólo si $f^{-1}(A)$ es un abierto de $\mathbb{R}^n$ para todo abierto A de $\mathbb{R}^m$.*

El que f sea continua en un conjunto **no** implica que $f(A)$ sea abierto para los abiertos A de $\mathbb{R}^n$ contenidos en S. En efecto, basta considerar la aplicación $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \operatorname{sen}(x)$ que lleva el abierto $\mathbb{R}$ en $[-1, 1]$ que no es abierto de $\mathbb{R}$.

Cuando una aplicación biyectiva entre dos subconjuntos, uno de $\mathbb{R}^n$ y otro de $\mathbb{R}^m$ es continua y también lo es su inversa, se dice que tal aplicación es un **homeomorfismo**. A la vista de lo anterior una aplicación biyectiva y continua entre dos abiertos, uno de $\mathbb{R}^n$ y otro de $\mathbb{R}^m$ es un homeomorfismo si y sólo si lleva abiertos de $\mathbb{R}^n$ en abiertos de $\mathbb{R}^m$.

Vamos ahora a relacionar la continuidad con la compacidad para obtener unos interesantes resultados que usaremos más adelante.

Teorema 1.4.16 *Si $f : K \subset \mathbb{R}^n \to \mathbb{R}^m$ es una aplicación continua en K, siendo K un compacto de $\mathbb{R}^n$, entonces $f(K)$ es un compacto de $\mathbb{R}^m$.*

Demostración. Sea $\{A_\alpha : \alpha \in \Lambda\}$ un recubrimiento abierto de $f(K)$, entonces $K \subset \cup_{\alpha\in\Lambda} f^{-1}(A_\alpha)$. Para cada $\alpha \in \Lambda$ sea G_α un abierto de $\mathbb{R}^n$ tal que $f^{-1}(A_\alpha) = G_\alpha \cap K$. Entonces $K \subset \cup_{\alpha\in\Lambda} G_\alpha$. La compacidad de K nos da la existencia de un conjunto finito $\mathcal{F} \subset\Lambda$ tal que $K \subset \cup_{\alpha\in\mathcal{F}} G_\alpha$. En consecuencia $f(K) \subset \cup_{\alpha\in\mathcal{F}} A_\alpha$. □

Nota 1.4.17 *No tiene por qué verificarse que la imagen inversa de un compacto por una aplicación continua sea un compacto (lo es la imagen directa como acabamos de ver). Por ejemplo si $f(x) = \operatorname{sen}(x)$ entonces $f^{-1}([-1, 1]) = \mathbb{R}$ que no es compacto.*

Teorema 1.4.18 *(del máximo y el mínimo) Sea $f : K \subset \mathbb{R}^n \to \mathbb{R}$ continua, siendo K un conjunto compacto. Entonces existen a y $b \in K$ tales que*

$$f(a) = \acute{\imath}nf\{f(x) : x \in K\} \quad y \quad f(b) = sup\{f(x) : x \in K\}.$$

Demostración. Observamos en primer lugar que al ser $f(K)$ compacto (teorema anterior) este conjunto es acotado en $\mathbb{R}$, por lo que tiene sentido hablar de ínfimo y supremo de $f(K)$. El ínfimo de un conjunto es siempre un punto adherente al conjunto; como $f(K)$ es cerrado (por ser compacto) ese punto pertenece al conjunto, luego es de la forma $f(a)$ para algún $a \in K$. Lo mismo se hace con el supremo. □

Vemos ahora como la conexión se mantiene por aplicaciones continuas.

Teorema 1.4.19 *Si $f : S \subset \mathbb{R}^n \to \mathbb{R}^m$ es una aplicación continua y S es un subconjunto conexo de $\mathbb{R}^n$, entonces $f(S)$ es un subconjunto conexo de $\mathbb{R}^m$.*

Demostración. Supongamos que $f(S)$ es disconexo y sea (U_2, V_2) una separación de $f(S)$. Entonces existen abiertos U_1 y V_1 de $\mathbb{R}^n$ tales que $f^{-1}(U_2) = U_1 \cap S$ y $f^{-1}(V_2) = V_1 \cap S$. Es claro que (U_1, V_1) es una separación del conexo S de $\mathbb{R}^n$ lo que es un absurdo. □

Nota 1.4.20 *Como consecuencia del teorema anterior se obtiene que todos los segmentos son conexos y que, en general, lo son las imágenes de los intervalos por aplicaciones continuas.*

Podemos demostrar ahora fácilmente el resultado que anticipamos en la Nota 1.3.14 referente a la conexión de los conjuntos conexos por polígonales.

Teorema 1.4.21 *Todo conjunto conexo por poligonales es conexo.*

Demostración. Sea S un conjunto conexo por poligonales y fijemos un punto x_0 de S. Dado cualquier punto x de S consideremos una poligonal Γ en S que los una, esa poligonal es conexa por ser los segmentos conexos y el teorema 1.3.12. Resulta así que S es unión de poligonales con x_0 como punto común, luego es conexo. □

1.5. Continuidad sobre compactos. Continuidad uniforme

Estudiamos aquí un tipo particular de continuidad, la continuidad uniforme. Es un tipo de continuidad más restrictivo que, en los conjuntos compactos, equivale al de continuidad.

Definición 1.5.1 *Sea $f : S \subset \mathbb{R}^n \to \mathbb{R}^m$. Se dice que f es **uniformemente continua** en S si*

$$\forall \varepsilon > 0 \; \exists \delta > 0 \mid x, y \in S, d(x,y) < \delta \Longrightarrow d(f(x), f(y)) < \varepsilon.$$

Obsérvese que hay una diferencia esencial con respecto a la continuidad en cada punto de S, aquí dado un ε debe encontrarse un δ que valga para todo par de puntos de S que estén a distancia menor que δ. Nótese también que estamos denotando con d tanto la métrica de $\mathbb{R}^n$ como la de $\mathbb{R}^m$, por cierto esta métrica será normalmente la euclídea pero puede ser cualquier otra y la definición puede darse para espacios métricos arbitrarios, no necesariamente los euclídeos.

La identidad es uniformemente continua, mientras que en $\mathbb{R}$ la aplicación $f(x) = x^2$ no es uniformemente continua. En efecto, si lo fuese, dado $\varepsilon = 1$ existiría un $\delta > 0$ tal que si $|x - y| < \delta$ entonces $|x^2 - y^2| < 1$. Ahora bien, si tomamos $x = \frac{1}{\delta} + \frac{\delta}{3}$ e $y = \frac{1}{\delta} - \frac{\delta}{3}$, resulta que $|x - y| = \frac{2\delta}{3} < \delta$ y sin embargo

$$|x^2 - y^2| = |x - y||x + y| = \frac{2\delta}{3}\frac{2}{\delta} = \frac{4}{3} > 1.$$

Otro típico ejemplo de función no uniformemente continua es la definida por $f(x) = \frac{1}{x}$ en $S = (0, +\infty)$. Para verlo tomemos, como antes, $\varepsilon = 1$ y supongamos que existe un $\delta > 0$ tal que si $x > 0,\ y > 0$ y $|x-y| < \delta$ entonces $|\frac{1}{x} - \frac{1}{y}| < 1$. Sean $x = \frac{1}{N}$ e $y = \frac{1}{N+n}$ con N y n bastante grandes para que $|x - y| < \delta$ (nótese que $\{\frac{1}{k}\}$ es de Cauchy). Entonces, para esos x e y se verifica que

$$\left|\frac{1}{x} - \frac{1}{y}\right| = |N - (N + n)| = n \geq 1.$$

Veremos a continuación que las aplicaciones uniformemente continuas tienen la propiedad de transformar sucesiones de Cauchy en sucesiones de Cauchy. Esto no le pasa, en general, a las que solamente son continuas, por ejemplo a la $f(x) = \frac{1}{x}$ que acabamos de considerar, pues $\{\frac{1}{k}\}$ es de Cauchy y

$$\left|\frac{1}{\frac{1}{p}} - \frac{1}{\frac{1}{q}}\right| = |p - q| \geq 1 \text{ si } p \neq q.$$

Teorema 1.5.2 *Si $f : S \subset \mathbb{R}^n \to \mathbb{R}^m$ es uniformemente continua en S y $\{x_k\}$ es una sucesión de Cauchy en S, entonces $\{f(x_k)\}$ es una sucesión de Cauchy en $\mathbb{R}^m$.*

Demostración. Dado $\varepsilon > 0$ sea $\delta > 0$ tal que

$$x, y \in S, d(x, y) < \delta \Longrightarrow d(f(x), f(y)) < \varepsilon.$$

Consideremos ahora un $\nu \in \mathbb{N}$ tal que $p, q \geq \nu \Longrightarrow d(x_p, x_q) < \delta$. Entonces para $p, q \geq \nu$ se tiene que $d(f(x_p), f(x_q)) < \varepsilon$. □

Aunque, como ya hemos visto, no todas las aplicaciones continuas son uniformemente continuas se tiene el siguiente resultado.

Teorema 1.5.3 *Si $f : K \subset \mathbb{R}^n \to \mathbb{R}^m$ es continua en el compacto K, entonces f es uniformemente continua en K.*

Demostración. Sea $\varepsilon > 0$. Por cada $x \in K$ existe un $\delta_x > 0$ tal que si $y \in K$ y $d(y, x) < \delta_x$, entonces $d(f(y), f(x)) < \frac{\varepsilon}{2}$. La colección de bolas $B(x, \frac{\delta_x}{2})$, cuando x varía en K, recubre a K, luego existen $x_1, ..., x_k \in K$ tales que

$$K \subset \bigcup_{j=1}^{k} B\left(x_j, \frac{\delta_{x_j}}{2}\right).$$

Sea δ el mínimo de los $\frac{\delta_{x_j}}{2}$, $j = 1, ..., k$. Si dados x e $y \in K$, tales que $d(x, y) < \delta$ consideramos un $j = 1, ..., k$ tal que $x \in B\left(x_j, \frac{\delta_{x_j}}{2}\right)$, resulta que $y \in B(x_j, \delta_{x_j})$ pues

$$d(y, x_j) \leq d(y, x) + d(x, x_j) < \delta_{x_j}.$$

En consecuencia

$$d(f(y), f(x)) \leq d(f(y), f(x_j)) + d(f(x_j), f(x)) < \varepsilon.$$

□

Las aplicaciones uniformemente continuas tienen una propiedad de prolongación a la adherencia de su conjunto de definición, cosa que no tienen, en general, todas las aplicaciones continuas; piénsese por ejemplo en la función $f(x) = \frac{1}{x}$ definida en $(0, 1]$.

Teorema 1.5.4 *Sea $f : S \subset \mathbb{R}^n \to \mathbb{R}^m$ una aplicación uniformemente continua en S, entonces existe una aplicación $\overline{f} : \overline{S} \to \mathbb{R}^m$, uniformemente continua en $\overline{S}$ tal que $\overline{f} = f$ en S.*

Demostración. Sea $x_0 \in \overline{S}$. Sabemos que existe una sucesión $\{x_k\}$ de puntos de S tal que $x_k \to x_0$. La sucesión $\{x_k\}$ es de Cauchy (por ser convergente), entonces, por la continuidad uniforme de f se tiene que $\{f(x_k)\}$ de Cauchy en $\mathbb{R}^m$, y por lo tanto es convergente. Pues bien, definamos $\overline{f}(x_0)$ =lím$\{f(x_k)\}$. Veamos como ese límite no depende de la sucesión convergente a x_0 elegida: Si $\{x'_k\}$ es otra sucesión de puntos de S convergente a x_0, consideramos la sucesión $\{x''_k\}$ donde $x''_k = x_j$ si $k = 2j$ e $x''_k = x'_j$

si $k = 2j - 1$. Resulta que $x_k'' \to x_0$ y $\{f(x_k'')\}$ debe ser convergente, por ser de Cauchy; como una subsucesión de ella converge a $\overline{f}(x_0)$ ella debe converger también a $\overline{f}(x_0)$ y por lo tanto también debe converger a ese valor la subsucesión $\{f(x_k')\}$.

Claramente $\overline{f} = f$ en S.

Veamos finalmente como $\overline{f}$ así definida es uniformemente continua en $\overline{S}$:

Dado $\varepsilon > 0$ sea $\delta > 0$ tal que $d(f(x), f(y)) < \frac{\varepsilon}{3}$ si $x,\ y \in S$ y $d(x, y) < \delta$ (téngase en cuenta que f es uniformemente continua en S). Dados $x,\ y \in \overline{S}$ consideremos sucesiones $x_k \to x$ e $y_k \to y$. Para cualquier $k \in \mathbb{N}$ se tiene que

$$d(\overline{f}(x), \overline{f}(y)) \leq d(\overline{f}(x), f(x_k)) + d(f(x_k), f(y_k)) + d(f(y_k), \overline{f}(y)) \qquad (*)$$

Veamos como para k adecuado podemos conseguir que cada uno de esos sumandos sea menor que $\frac{\varepsilon}{3}$ cuando $d(x, y) < \frac{\delta}{3}$. Por una parte existe $k_1 \in \mathbb{N}$ tal que $d(\overline{f}(x), f(x_k)) < \frac{\varepsilon}{3}$ y $d(f(y_k), \overline{f}(y)) < \frac{\varepsilon}{3}$ para todo $k \geq k_1$. Por otra parte existe $k_2 \in \mathbb{N}$ tal que $d(x_k, x) < \frac{\delta}{3}$ y $d(y_k, y) < \frac{\delta}{3}$ para todo $k \geq k_2$.

Si tomamos $x,\ y \in \overline{S}$ tales que $d(x, y) < \frac{\delta}{3}$ se verifica que

$$d(x_k, y_k) \leq d(x_k, x) + d(x, y) + d(y, y_k) < \delta \quad \forall k \geq k_2$$

Si tomamos $k \geq$máx$\{k_1, k_2\}$ resulta que cada uno de los sumandos de (*) es menor que $\frac{\varepsilon}{3}$ y por lo tanto

$$d(\overline{f}(x), \overline{f}(y)) < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon \quad \forall x,\ y \in \overline{S} \mid d(x, y) < \frac{\delta}{3}.$$

□

1.6. El teorema del punto fijo para aplicaciones contractivas

Obtendremos aquí un resultado importante relativo a la existencia de puntos fijos para aplicaciones contractivas.

Definición 1.6.1 *Se dice que una aplicación $f : S \subset \mathbb{R}^n \to \mathbb{R}^m$ es* ***contractiva*** *si existe $\lambda \in [0, 1)$ tal que $d(f(x), f(y)) \leq \lambda d(x, y)$ para todo x e y de S.*

Obsérvese que toda aplicación contractiva es uniformemente continua.

Ejemplo 1.6.2 *La aplicación $x \mapsto \frac{1}{2}x$ es contractiva de $\mathbb{R}^n$ en $\mathbb{R}^n$.*

El siguiente resultado demuestra la existencia de puntos fijos para aplicaciones contractivas de un conjunto cerrado en si mismo.

Teorema 1.6.3 ***(del punto fijo)*** *Sea S un subconjunto cerrado de $\mathbb{R}^n$ y sea $f : S \to S$ una aplicación contractiva. Entonces existe un punto x^* en S tal que $f(x^*) = x^*$. Además sólo hay un punto en S con esa propiedad.*

Demostración. Fijemos un punto x_0 en S y consideremos la sucesión de elementos de S definida por:

$$x_1 = f(x_0),\ x_2 = f(x_1)...,\ x_k = f(x_{k-1})...$$

Vamos a demostrar que esta sucesión es de Cauchy en S. Dados cualesquiera p y $q \in \mathbb{N}$ distintos, uno será mayor que el otro, supongamos que $q > p$, entonces se tiene que

$$d(x_p, x_q) \leq d(x_p, x_{p+1}) + d(x_{p+1}, x_{p+2}) + \cdots + d(x_{q-1}, x_q).$$

Ahora bien, por la contractividad de f tenemos que para todo $k \in \mathbb{N}$,

$$\begin{aligned} d(x_k, x_{k+1}) &= d(f(x_{k-1}), f(x_k)) \\ &\leq \lambda d(x_{k-1}, x_k) \leq \lambda^2 d(x_{k-2}, x_{k-1}) \\ &\leq \cdots \\ &\leq \lambda^k d(x_0, x_1). \end{aligned}$$

En consecuencia,

$$\begin{aligned} d(x_p, x_q) &\leq d(x_0, x_1)(\lambda^p + \lambda^{p+1} + \cdots + \lambda^{q-1}) \\ &\leq d(x_0, x_1) \sum_{j=p}^{\infty} \lambda^j = d(x_0, x_1) \frac{\lambda^p}{1-\lambda}. \end{aligned}$$

Así, si dado $\varepsilon > 0$ elegimos $\nu \in \mathbb{N}$ tal que $d(x_0, x_1)\frac{\lambda^\nu}{1-\lambda} < \varepsilon$ tenemos que para $q > p \geq \nu$, $d(x_p, x_q) < \varepsilon$.

Esta sucesión de Cauchy en $\mathbb{R}^n$ debe ser convergente a un punto x^* de $\mathbb{R}^n$ que necesariamente está en S, pues S es cerrado. Veamos como para ese punto se verifica que $f(x^*) = x^*$. En efecto, por una parte $f(x_k) = x_{k+1} \to x^*$, por otra parte, al ser f continua, $f(x_k) \to f(x^*)$, luego necesariamente $f(x^*) = x^*$.

Hemos probado entonces que f tiene un punto fijo, el x^*. Veamos como no tiene ningún otro. Supongamos que x^{**} fuese también un punto fijo de f. Entonces

$$d(x^{**}, x^*) = d(f(x^{**}), f(x^*)) \leq \lambda d(x^{**}, x^*)$$

Si $\lambda \neq 0$ necesariamente $d(x^{**}, x^*) = 0$ con lo que $x^{**} = x^*$, y si $\lambda = 0$ entonces la aplicación es constante, es decir, existe $c \in S$ tal que $f(x) = c$ para todo $x \in$, luego el único punto fijo es c. □

Observamos que el que S sea cerrado es importante, pues la aplicación contractiva $f(x) = \frac{1}{2}x$ en $(0, 1)$ no tiene ningún punto fijo. Notamos también que la existencia del $\lambda < 1$ que garantiza la contractividad de una aplicación es necesaria, no basta pedir que $d(f(x), f(y)) < d(x, y)$ para todo $x, y \in S, x \neq y$. La aplicación

$$f(x) = \frac{x + \sqrt{x^2 + 1}}{2}, x \in \mathbb{R}$$

verifica la condición anterior y no tiene ningún punto fijo. Es fácil ver esto último; en efecto, si para un cierto $x^* \in \mathbb{R}$ se verificase que $f(x^*) = x^*$ entonces

$$x^* + \sqrt{x^{*2} + 1} = 2x^*,$$

con lo que $x^{*2} + 1 = x^{*2}$ lo que es absurdo. La comprobación de que se verifica que $d(f(x), f(y)) < d(x, y)$ para todo $x, y \in \mathbb{R}$ puede obtenerse a partir del teorema del valor medio para funciones de una variable real teniendo en cuenta que $|f'(t)| < 1$ para todo $t \in \mathbb{R}$.

Capítulo 2

Aplicaciones diferenciables

Introducción


Leibniz

Newton


La fundación del Cálculo Diferencial (CD) se atribuye a Newton (1643-1727) y a Leibniz (1646-1716). Había ya algunos precedentes, de ahí que se diga que fueron los fundadores no los descubridores. En el s. XVI ya se habían tratado tres problemas que tienen mucho que ver con el CD: los de máximos y mínimos, los de tangentes y los del cálculo de áreas.

Entre los precursores del CD en una variable podemos citar a Arquímedes (285-212 a.C.), Cavalieri (1598-1647), Wallis (1616-1667), Fermat (1601-1651), Pascal (1623-1662) y Barrow (1630-1667), véase [1, p.p. 97-104].

En relación con el CD en varias variables, que es el que estudiamos aquí, éste se desarrolló a principios del s. XVIII. La idea de derivada parcial ya fue utilizada por Newton para obtener resultados sobre ecuaciones polinómicas del tipo $P(x, y) = 0$. También fue utilizada esta idea posteriormente por los hermanos J. (1667-1748) y N. Bernoulli (1687-1759). Sin embargo los verdaderos creadores de la teoría de derivación parcial fueron

Alexis des Bertins (1705-1771), Euler (1701-1783) y d'Alembert (1717-1784), véase [2, p. 425].

La fundamentación rigurosa del CD en varias variables fue llevada a cabo en el s. XIX siendo Cauchy (1789-1857) y Bolzano (1781-1848) los iniciadores de esta tarea. Otros destacados matemáticos que contribuyeron a ello fueron Abel (1802-1829), Dirichlet (1805-1859), Weierstrass (1815-1897), Riemann (1826-1866), Darboux (1842-1917), Poincaré (1854-1912) y Volterra (1860-1940), véase [1, p. 173].

Referencias

[1] Durán, A.J., "Historia, con personajes, de los conceptos del Cálculo". Alianza Universidad, Madrid, 1996.

[2] Kline, M., "Mathematica Thought from Ancient to Modern Times". Oxford University Press, New York, 1972.

2.1. Derivadas direccionales. Gradiente

Recordamos que se define la derivada de una función f definida en un subconjunto abierto A de $\mathbb{R}$ en un punto x_0 de A como el valor del límite que sigue, si es que ese límite existe y es finito:

$$\lim_{t\to 0}\frac{f(x_0+t)-f(x_0)}{t}$$

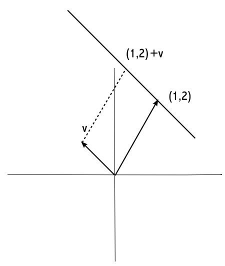
El valor de este límite (cuando existe) se denota por $f'(x_0)$.

El concepto de derivada para funciones de variable real no puede extenderse directamente a funciones $f : A \subset \mathbb{R}^n \to \mathbb{R}$ (A denotará en lo que sigue un subconjunto abierto de $\mathbb{R}^n$). Surge el problema de dividir números reales por elementos de $\mathbb{R}^n$, cosa que no tiene ningún sentido. En este tema veremos la forma de dar una definición "razonable" de este concepto.

Definición 2.1.1 *Llamaremos **dirección** en $\mathbb{R}^n$ a cualquier vector $v \in \mathbb{R}^n$, $v \neq 0$. Normalmente se supondrá que v tiene norma 1. Si lo hacemos así, para $n = 1$ tenemos sólo dos direcciones, la 1 y la -1. Para $n > 1$ tenemos infinitas direcciones. En el caso de $\mathbb{R}^2$ las direcciones de norma 1 pueden escribirse en la forma $(\cos\theta, \operatorname{sen}\theta)$, con $0 \le \theta < 2\pi$.*

Definición 2.1.2 *Dados un punto $x_0 \in \mathbb{R}^n$ y una dirección v, llamaremos **recta pasando por** x_0 **y de dirección** v a la recta $x_0 + tv,\ t \in \mathbb{R}$. Obsérvese que para $t = 0$ y vale x_0 y que para $t = 1$ y vale $x_0 + v$.*

Ejemplo 2.1.3 *Supongamos que x_0 es el punto $(1,2)$ de $\mathbb{R}^2$ y que $v = (-1/\sqrt{2}, 1/\sqrt{2})$. La ecuación de la recta pasando por $(1,2)$ de dirección $v = (-1/\sqrt{2}, 1/\sqrt{2})$ es entonces $t \to (1 + t(-1/\sqrt{2}), 2 + t(1/\sqrt{2}))$.*



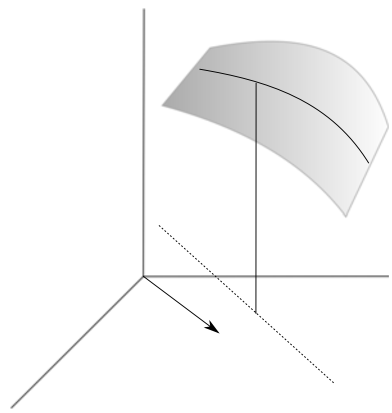
Definición 2.1.4 *Si f es una función definida en un subconjunto abierto A de $\mathbb{R}^n$, x_0 es un punto de A y v es una dirección en $\mathbb{R}^n$, se define la* ***derivada de f en x_0 en la dirección*** *v como el*

$$\lim_{t\to 0}\frac{f(x_0+tv)-f(x_0)}{t}$$

supuesto, claro está, que este límite exista (sea un número real). Cuando exista, su valor se denotará por $D_v f(x_0)$.

Observamos que si definimos $\phi(t) = f(x_0 + tv)$ para t próximos a 0 (para que $x_0 + tv$ sea próximo a x_0 y por lo tanto pertenezca a A), entonces f es derivable en x_0 en la dirección v si y sólo si ϕ es derivable en 0.

Idea geométrica



Ejemplo 2.1.5 *Sea f la función de $\mathbb{R}^2$ en $\mathbb{R}$ definida por*

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

Vamos a ver en qué direcciones es derivable en el punto $(0,0)$.

Sea v una dirección en $\mathbb{R}^2$, que suponemos de norma 1. Sabemos que v es de la forma $(\cos\theta, \operatorname{sen}\theta)$ con $\theta \in [0, 2\pi)$ y que f es derivable en $(0,0)$ en la dirección $v = (\cos\theta, \operatorname{sen}\theta)$ si y sólo si la función

$$\phi(t) = f((0,0) + t(\cos\theta, \operatorname{sen}\theta))$$

lo es en 0. Ahora bien,

$$\phi(t) = |t^2(\cos^2\theta - \operatorname{sen}^2\theta)|^{\frac{1}{2}} = |t||\cos^2\theta - \operatorname{sen}^2\theta|^{\frac{1}{2}}.$$

Luego, para aquellos $\theta \in [0, 2\pi)$ tales que $\cos^2\theta = \operatorname{sen}^2\theta$ resulta que ϕ es idénticamente nula y por lo tanto derivable en 0 y para los $\theta \in [0, 2\pi)$ tales que $\cos^2\theta \neq \operatorname{sen}^2\theta$ resulta que ϕ no es derivable en 0 pues "esencialmente" ϕ es la función valor absoluto. Resumiendo f es derivable en $(0, 0)$ solamente en las direcciones

$$\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right), \left(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right), \left(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right) \text{ y } \left(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right).$$

Ejercicio 2.1.6 *Dada la función $f(x, y, z) = |x + y + z|$ estúdiese en qué direcciones es derivable en un punto (x_0, y_0, z_0) tal que $x_0 + y_0 + z_0 = 0$.*

Nota 2.1.7 *Obsérvese que para cualquier dirección v, tanto ella como su opuesta $-v$ definen la misma recta pasando por x_0 (el vector de dirección determina una* ***orientación*** *en esa recta), sin embargo las derivadas en las direcciones v y $-v$ son de signo opuesto. En efecto,*

$$\begin{aligned}
D_{-v}f(x_0) &= \lim_{t\to 0}\frac{f(x_0 + t(-v)) - f(x_0)}{t} \\
&= \lim_{t\to 0}\frac{f(x_0 + (-t)v) - f(x_0)}{t} \\
&= \lim_{s\to 0}\frac{f(x_0 + sv) - f(x_0)}{-s} \\
&= -\lim_{s\to 0}\frac{f(x_0 + sv) - f(x_0)}{s} \\
&= -D_v f(x_0).
\end{aligned}$$

Consideremos en $\mathbb{R}^n$ las direcciones dadas por los vectores $e_i = (0, ..., 1, ..., 0)$ (con el 1 en el lugar i). Las derivadas en un punto x_0 de una función f en estas direcciones (si es que existen) se llaman **derivadas parciales** y se denotan normalmente por $D_i f(x_0)$ en vez de por $D_{e_i} f(x_0)$ que sería lo natural. También es muy frecuente usar la notación $\frac{\partial f}{\partial x_i}(x_0)$.

Ejemplo 2.1.8 *Calculemos las derivadas parciales de la función*

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

en un punto arbitrario (x_0, y_0, z_0) de $\mathbb{R}^3$.

Solución: Como hemos dicho,

$$D_i f((x_0, y_0, z_0) = \phi_i^{'}(0)$$

siendo $\phi_i(t) = f((x_0, y_0, z_0) + te_i)$, con lo que

$$\begin{aligned}
\phi_1(t) &= f((x_0 + t, y_0, z_0)) = (x_0 + t)^2 + y_0 + \cos(y_0^2 z_0) \\
\phi_2(t) &= f((x_0, y_0 + t, z_0)) = {x_0}^2 + (y_0 + t) + \cos((y_0 + t)^2 z_0) \\
\phi_3(t) &= f((x_0, y_0, z_0 + t)) = {x_0}^2 + y_0 + \cos(y_0^2(z_0 + t)).
\end{aligned}$$

Entonces,

$$\begin{aligned}
D_1 f((x_0, y_0, z_0)) &= \phi_1^{'}(0) = 2x_0 \\
D_2 f((x_0, y_0, z_0)) &= \phi_2^{'}(0) = 1 - 2y_0 z_0 \operatorname{sen}(y_0^2 z_0) \\
D_3 f((x_0, y_0, z_0)) &= \phi_3^{'}(0) = -y_0^2 \operatorname{sen}(y_0^2 z_0).
\end{aligned}$$

Definición 2.1.9 *Dada una función $f : A \subset \mathbb{R}^n \to \mathbb{R}$ que tenga todas las derivadas parciales en un punto $x_0 \in A$ se llama **gradiente** de f en x_0 al vector $(D_1 f(x_0), ..., D_n f(x_0))$ de $\mathbb{R}^n$. Para denotar el gradiente de f en x_0 suele usarse la notación $\nabla f(x_0)$.*

En el ejemplo anterior, $\nabla f(0, 0, 0) = (0, 1, 0)$.

Ejercicio 2.1.10 *Calcular el gradiente de $f(x, y) = x^2 + 2xy$ en un punto arbitrario de $\mathbb{R}^2$.*

2.2. Aplicaciones diferenciables. Representación matricial

Cuando una función real de una variable real f es derivable en un punto x_0, la curva que define en $\mathbb{R}^2$ (su gráfica) tiene una recta tangente en el punto $(x_0, f(x_0))$. Se trata precisamente de la recta que pasa por $(x_0, f(x_0))$ y tiene por pendiente $f'(x_0)$, su ecuación es:

$$y = f(x_0) + f'(x_0)(x - x_0).$$

En el caso de funciones de dos variables, la idea anterior se transforma en la idea de que las funciones f "derivables" en un punto (x_0, y_0) deberán ser tales que la superficie que ellas definen en $\mathbb{R}^3$ tenga un "plano tangente" en el punto $(x_0, y_0, f(x_0, y_0))$. Como veremos un poco más adelante esto no puede garantizarse aunque la función tenga derivadas en todas las direcciones en el punto (x_0, y_0). Precisaremos un poco más adelante lo que significa el concepto de "plano tangente" en un punto, la idea es la misma que la de recta tangente.

El concepto de "derivabilidad" para funciones de dos variables (o más) no es una generalización directa de ese mismo concepto para funciones de una variable, piénsese qué sentido tendría escribir

$$\lim_{t \to 0} \frac{f(x_0 + t) - f(x_0)}{t}$$

siendo t un vector de $\mathbb{R}^n$ (en $\mathbb{R}^n$ no hay división).

Supongamos que $f : A \subset \mathbb{R} \to \mathbb{R}$ es derivable en un punto $x_0 \in A$ y consideremos la aplicación lineal $L_{x_0} : t \to f'(x_0)t$ (se pone a L el subíndice x_0 para indicar que esa aplicación depende de x_0). El que f sea derivable en x_0 equivale a que

$$\lim_{t\to 0}\frac{f(x_0+t)-f(x_0)-L_{x_0}(t)}{t}=0$$

y esto equivale a que

$$\lim_{t\to 0}\frac{f(x_0+t)-f(x_0)-L_{x_0}(t)}{|t|}=0,$$

lo que se comprueba así,

$$\begin{aligned}
\lim_{t\to 0}\frac{f(x_0+t)-f(x_0)-L_{x_0}(t)}{t} &= 0 \Longleftrightarrow\\
\lim_{t\to 0}\left|\frac{f(x_0+t)-f(x_0)-L_{x_0}(t)}{t}\right| &= 0 \Longleftrightarrow\\
\lim_{t\to 0}\frac{|f(x_0+t)-f(x_0)-L_{x_0}(t)|}{|t|} &= 0 \Longleftrightarrow\\
\lim_{t\to 0}\left|\frac{f(x_0+t)-f(x_0)-L_{x_0}(t)}{|t|}\right| &= 0 \Longleftrightarrow\\
\lim_{t\to 0}\frac{f(x_0+t)-f(x_0)-L_{x_0}(t)}{|t|} &= 0.
\end{aligned}$$

Hemos visto entonces que una aplicación $f : A \subset \mathbb{R} \to \mathbb{R}$ es derivable en un punto $x_0 \in A$, si y sólo si existe una aplicación lineal $L_{x_0} : \mathbb{R} \to \mathbb{R}$ tal que

$$\lim_{t\to 0}\frac{f(x_0+t)-f(x_0)-L_{x_0}(t)}{|t|}=0$$

Esta idea sí que puede generalizarse para aplicaciones de más de una variable de la siguiente manera:

Definición 2.2.1 *Se dice que $f : A \subset \mathbb{R}^n \to \mathbb{R}$ es* ***diferenciable*** *en $x_0 \in A$ si existe una aplicación lineal $L_{x_0} : \mathbb{R}^n \to \mathbb{R}$ tal que*

$$\lim_{h\to 0}\frac{f(x_0+h)-f(x_0)-L_{x_0}(h)}{\|h\|}=0.$$

Preferimos cambiar la letra t por la h porque la t suele representar un número real y pudiera dar lugar a confusión, aquí h es un vector de $\mathbb{R}^n$. Observamos que la norma de h es un número real mayor que 0 (pues h nunca es 0) y por lo tanto no hay ningún problema con esa división. La anterior condición es equivalente a que

$$\lim_{x\to x_0}\frac{f(x)-f(x_0)-L_{x_0}(x-x_0)}{\|x-x_0\|}=0.$$

Nota 2.2.2 *No se usa habitualmente el término "derivable" para funciones de más de una variable, se usa en su lugar el término "diferenciable".*

Vamos a recordar en un momento lo que es una aplicación lineal de $\mathbb{R}^n$ en $\mathbb{R}$. Es una aplicación $L : \mathbb{R}^n \to \mathbb{R}$ tal que $L(x+y) = L(x) + L(y)$ para todo x e y de $\mathbb{R}^n$ y $L(\alpha x) = \alpha L(x)$ para todo $\alpha \in \mathbb{R}$ y para todo $x \in \mathbb{R}^n$. En particular,

$$\begin{aligned} L(x) &= L(x_1, ..., x_n) \\ &= L(x_1 e_1 + \cdots + x_n e_n) \\ &= x_1 L(e_1) + \cdots + x_n L(e_n) \\ &= \langle (x_1, ..., x_n), (L(e_1), ..., L(e_n)) \rangle . \end{aligned}$$

Observamos que todas las aplicaciones lineales de $\mathbb{R}^n$ en $\mathbb{R}$ son continuas, incluso lo son uniformemente, pues

$$\begin{aligned} |L(x) - L(y)| &= |L(x-y)| \\ &= |\langle (x_1 - y_1, ..., x_n - y_n), (L(e_1), ..., L(e_n)) \rangle| \\ &\leq \|(x_1 - y_1, ..., x_n - y_n)\| \, \|(L(e_1), ..., L(e_n))\| = C \|x - y\| \end{aligned}$$

siendo $C = \|(L(e_1), ..., L(e_n))\|$.

Si denotamos por $\mathcal{L}(\mathbb{R}^n, \mathbb{R})$ al espacio vectorial de las aplicaciones lineales de $\mathbb{R}^n$ en $\mathbb{R}$, resulta que la aplicación

$$L \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}) \mapsto \|L\| := \sup_{h \in \mathbb{R}^n, \|h\|=1} |L(h)| = \sup_{k \in \mathbb{R}^n, k \neq 0} \frac{|L(k)|}{\|k\|}.$$

es una norma en ese espacio. Veamos en primer lugar la igualdad anterior:

Por una parte, para todo $k \neq 0$ se verifica que

$$\frac{|L(k)|}{\|k\|} = \left| L\left(\frac{k}{\|k\|} \right) \right| \leq \sup_{h \in \mathbb{R}^n, \|h\|=1} |L(h)| .$$

Por otra parte, para todo h, tal que $\|h\| = 1$, se verifica que

$$|L(h)| = \frac{|L(h)|}{\|h\|} \leq \sup_{k \in \mathbb{R}^n, k \neq 0} \frac{|L(k)|}{\|k\|}.$$

Veamos como la aplicación

$$L \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}) \mapsto \|L\| = \sup_{h \in \mathbb{R}^n, \|h\|=1} |L(h)|$$

es una norma:

a) Desde luego que $\|L\| \geq 0$

b) Si $L = 0$ entonces $\|L\| = 0$ y si $\sup_{h\in\mathbb{R}^n, \|h\|=1} |L(h)| = 0$ entonces, si $h \neq 0$, $L(h) = ||h||L\left(\frac{h}{\|h\|}\right) = 0$, si $h = 0$ es claro que $L(h) = 0$

c) $||\alpha L|| = \sup_{h\in\mathbb{R}^n, \|h\|=1} |\alpha L(h)| = |\alpha| \sup_{h\in\mathbb{R}^n, \|h\|=1} |L(h)|$ para todo $\alpha \in \mathbb{R}$

d) $||L_1 + L_2|| = \sup_{h\in\mathbb{R}^n, \|h\|=1} |(L_1 + L_2)(h)| \leq \sup_{h\in\mathbb{R}^n, \|h\|=1} \{|L_1(h)| + |L_2(h)|\} \leq ||L_1|| + ||L_2||$.

Proposición 2.2.3 *Para toda aplicación lineal $L : \mathbb{R}^n \to \mathbb{R}$ se verifica que*

$$||L||_{\mathcal{L}(\mathbb{R}^n,\mathbb{R})} = ||(L(e_1), ..., L(e_n))||_2$$

(recordamos que $|| \ ||_2$ representa la norma euclídea en $\mathbb{R}^n$).

Demostración. Supondremos que L no es idénticamente nula, si lo fuese el resultado es trivial. Para todo $h \in \mathbb{R}^n$ con $||h||_2 = 1$ resulta que

$$\begin{aligned} |L(h)| &= |L(h_1e_1 + \cdots + h_ne_n)| \\ &= |h_1L(e_1) + \cdots + h_nL(e_n)| \\ &= |\langle (h_1, ..., h_n), (L(e_1), ..., L(e_n))\rangle| \\ &\leq ||(L(e_1), ..., L(e_n))||_2 \cdot ||h||_2 \end{aligned}$$

con lo que

$$|L(h)| \leq ||(L(e_1), ..., L(e_n))||_2.$$

□

Por otra parte para el vector

$$h = \frac{(L(e_1), ..., L(e_n))}{||(L(e_1), ..., L(e_n))||_2}$$

se verifica que

$$|L(h)| = \frac{L(e_1)^2 + \cdots + L(e_n)^2}{||(L(e_1), ..., L(e_n))||_2} = ||(L(e_1), ..., L(e_n))||_2$$

lo que nos da que

$$\sup_{h\in\mathbb{R}^n, ||h||_2=1} |L(h)| = ||(L(e_1), ..., L(e_n))||_2$$

y que el valor de ese supremo se alcanza en el vector de norma 1,

$$\frac{(L(e_1), ..., L(e_n))}{||(L(e_1), ..., L(e_n))||_2}.$$

Ejemplos de aplicaciones diferenciables: Cualquier aplicación constante, las proyecciones en una coordenada, esto es, $(x_1, ..., x_n) \mapsto x_j$ y, en general, cualquier aplicación lineal. Obsérvese que si L es una aplicación lineal de $\mathbb{R}^n$ en $\mathbb{R}$, entonces para todo $x_0 \in \mathbb{R}^n$, se tiene que $L_{x_0} = L$. Veremos más ejemplos en las clases prácticas.

Volviendo a la diferenciabilidad, observamos que una aplicación $f : A \subset \mathbb{R} \to \mathbb{R}$ es derivable en un punto $x_0 \in A$ si y sólo si es diferenciable en ese punto.

Como es lógico, la aplicación lineal L_{x_0} depende, en general, del punto x_0 pero sucede, también como parece natural, que de existir es única. Puede verse esto fácilmente suponiendo que hay más de una de tales aplicaciones pero nosotros lo veremos aquí ahora mismo de modo indirecto haciendo algo que también nos interesa.

Teorema 2.2.4 *Si $f : A \subset \mathbb{R}^n \to \mathbb{R}$ es diferenciable en un punto $x_0 \in A$, entonces de derivable en ese punto en todas las direcciones. Además si v es una dirección, que supondremos de norma 1, y una aplicación lineal $L : \mathbb{R}^n \to \mathbb{R}$ verifica que*

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

entonces $L(v) = D_v f(x_0)$.

Demostración. Tomando sólo vectores h de la forma tv con $t \in \mathbb{R}$, tenemos que

$$\lim_{t \to 0} \frac{f(x_0 + tv) - f(x_0) - tL(v)}{|t|} = 0$$

(nótese que $\|v\| = 1$) con lo que, con el mismo argumento que hemos utilizado al principio de esta sección, obtenemos que

$$\lim_{t \to 0} \frac{f(x_0 + tv) - f(x_0) - tL(v)}{t} = 0$$

y por lo tanto que

$$\lim_{t \to 0} \frac{f(x_0 + tv) - f(x_0)}{t} = L(v),$$

luego f es derivable en x_0 en la dirección v y $D_v f(x_0) = L(v)$. □

Nota 2.2.5 *El teorema anterior demuestra que si f es diferenciable en x_0 entonces es derivable en ese punto en todas las direcciones y además el valor de la derivada en la dirección v es precisamente el valor de L en v. También esto prueba que de haber una aplicación lineal verificando*

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

hay sólo una, pues para todo $j = 1, ..., n$, $L(e_i) = D_i f(x_0)$ y este valor no depende de L (nótese que L queda determinada por su valores en la base canónica $\{e_1, ..., e_n\}$).

Para representar la aplicación L (que es única, si existe) usaremos la notación $df(x_0)$. Obsérvese que tal L depende de x_0. A $df(x_0)$ se le llamará la **diferencial** de f en x_0. Dado que $df(x_0)(e_j) = D_j f(x_0)$, se verifica que

$$||df(x_0)||_{\mathcal{L}(\mathbb{R}^n,\mathbb{R})} = \|\nabla f(x_0)\|_{\mathbb{R}^n}.$$

A la vista de lo anterior resulta entonces que cuando f es diferenciable en x_0,

$$\frac{\partial f}{\partial x_i}(x_0) = D_i f(x_0) = df(x_0)(e_i).$$

Para un $h \in \mathbb{R}^n$ arbitrario podemos escribir

$$df(x_0)(h) = \sum_{i=1}^{n} h_i df(x_0)(e_i) = \sum_{i=1}^{n} h_i D_i f(x_0) \equiv \sum_{i=1}^{n} h_i \frac{\partial f}{\partial x_i}(x_0),$$

esto es,

$$df(x_0)(h) = \langle \nabla f(x_0), h \rangle$$

para todo $h \in \mathbb{R}^n$. Observamos que esto no quiere decir que si una aplicación tiene gradiente en un punto, esta aplicación tenga que ser diferenciable en ese punto, lo que dice es, que si la aplicación es diferenciable en un punto, entonces la diferencial en ese punto toma como valor, en cada $h \in \mathbb{R}^n$, el valor del producto escalar del gradiente de la función en el punto por h.

Para una dirección v de $\mathbb{R}^n$, $||v|| = 1$, tenemos en particular que

$$D_v f(x_0) = df(x_0)(v) = \langle \nabla f(x_0), v \rangle.$$

Como consecuencia de la proposición 2.2.3 se obtiene el siguiente corolario.

Corolario 2.2.6 *Sea $f : A \subset \mathbb{R}^n \to \mathbb{R}$ una aplicación diferenciable en un punto $x_0 \in A$ tal que $\nabla f(x_0) \neq 0$, siendo A un abierto de $\mathbb{R}^n$. Entonces el valor máximo de $|D_v f(x_0)|$ se alcanza en el vector unitario $v = \frac{\nabla f(x_0)}{||\nabla f(x_0)||_2}$ y ese valor máximo es $||\nabla f(x_0)||_2$.*

Demostración. Basta aplicar la proposición anterior a la aplicación lineal $df(x_o)$. $\square$

Cuando f es diferenciable en un punto x_0 llamaremos **espacio afín tangente a la superficie** $\{(x, f(x)), x \in A\} \subset \mathbb{R}^{n+1}$ en el punto $(x_0, f(x_0))$ a

$$T = \{(x, f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle) : x \in \mathbb{R}^n\}.$$

Nótese que T es un subconjunto de $\mathbb{R}^{n+1}$. En el caso particular en el que $n = 1$ el correspondiente espacio afín tangente en $(x_0, f(x_0))$ es

$$\begin{aligned} T &= \{(x, f(x_0) + f'(x_0)(x - x_0)) : x \in \mathbb{R}\} \\ &= \{(x, y) \in \mathbb{R}^2 : y = f(x_0) + f'(x_0)(x - x_0)\}, \end{aligned}$$

que es precisamente la gráfica de la recta tangente a la curva $\{(x, f(x)), x \in S\}$ en el punto $(x_0, f(x_0))$. Cuando $n = 2$ el correspondiente T es un plano, al que llamaremos **plano tangente.** Puede describirse así:

$$\begin{aligned} T &= \{(x,y,z) \in \mathbb{R}^3 : z = f(x_0,y_0) + D_1 f(x_0,y_0)(x - x_0) + D_2 f(x_0,y_0)(y - y_0)\} \\ &= \{(x,y,z) \in \mathbb{R}^3 : z = f(x_0,y_0) + ax + by - ax_0 - by_0\} \\ &= \{(x,y,z) \in \mathbb{R}^3 : ax + by - z = ax_0 + by_0 - f(x_0,y_0)\} \end{aligned}$$

donde $a = D_1 f(x_0, y_0)$ y $b = D_2 f(x_0, y_0)$.

En general T es lo que se llama un **hiperplano** en $\mathbb{R}^{n+1}$. En general, en $\mathbb{R}^m$, los hiperplanos son los conjuntos de la forma

$$H = \{x \in \mathbb{R}^m : \langle x, l_0 \rangle = d\}$$

para algún $l_0 \in \mathbb{R}^m$, $l_0 \neq 0$, y $d \in \mathbb{R}$. O bien

$$H = \{x \in \mathbb{R}^m : L(x) = d\}$$

siendo L una aplicación lineal y $d \in \mathbb{R}$. Nótese que como L es lineal en $\mathbb{R}^n$, entonces

$$L(x) = \langle (x_1, ..., x_n), (L(e_1), ..., L(e_n)) \rangle$$

y $(L(e_1, \ldots L(e_n)$ hace el papel de l_0. Cuando $m = 1$ los hiperplanos son puntos, pues dados l_0 y $d \in \mathbb{R}$, el conjunto de los puntos x de $\mathbb{R}$ que verifican $xl_0 = d$ está formado únicamente por el punto $x = \frac{d}{l_0}$. Cuando $m = 2$ se trata de rectas, pues dados $l_0 = (a, b) \in \mathbb{R}^2 \backslash \{(0, 0)\}$ y $d \in \mathbb{R}$, el conjunto de los puntos $(x, y) \in \mathbb{R}^2$ tales que $ax + by = d$ es una recta. Para $m = 3$, dados $l_0 = (a, b, c)$ y $d \in \mathbb{R}$, se trata del plano

$$\{(x, y, z) : ax + by + cz = d\}.$$

Observamos que los hiperplanos de $\mathbb{R}^m$ son espacios afines de dimensión $m - 1$.

Vamos a ver ahora como el espacio afín tangente a la superficie $\{(x, f(x)) : x \in A\} \subset \mathbb{R}^{n+1}$ en el punto $(x_0, f(x_0))$, $x_0 \in A$,

$$T = \{(x, y) \in \mathbb{R}^n \times \mathbb{R} : y = f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle\}$$

es un hiperplano. En efecto,

$$\begin{aligned} T &= \{(x,y) \in \mathbb{R}^{n+1} : \langle (x_1, ..., x_n, y), (D_1 f(x_0), ..., D_n f(x_0), -1) \rangle \\ &= \langle \nabla f(x_0), x_0 \rangle - f(x_0)\}, \end{aligned}$$

haciendo entonces $l_0 = (D_1 f(x_0), ..., D_n f(x_0), -1)$ y $d = \langle \nabla f(x_0), x_0 \rangle - f(x_0)$ tenemos comprobado lo que queríamos.

El concepto intuitivo de tangencia que tenemos para el caso de curvas se mantiene también para superficies en $\mathbb{R}^n$, pues si T es el espacio afín tangente a la superficie $\{(x, f(x)), x \in A\}$ en el punto $(x_0, f(x_0))$, entonces para cada punto $(x, f(x))$, $x \in A$ existe un punto $(x, y) \in T$ (y depende de x) tal que

$$\lim_{x \to x_0} \frac{f(x) - y}{\|x - x_0\|} = 0.$$

En efecto, basta tomar $y = f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle$.

Ejemplo 2.2.7 *Si* $f(x, y) = x^2 + y^2$, $(x, y) \in \mathbb{R}^2$. *El plano tangente a la superficie* $\{(x, y, f(x, y)); (x, y) \in \mathbb{R}^2\}$ *en el punto* $(0, 0, 0)$ *es el plano* $z = 0$. *Esta superficie se llama paraboloide.*

Ejemplo 2.2.8 *Si* $f(x, y) = \sqrt{x^2 + y^2}$, $(x, y) \in \mathbb{R}^2$ *no existe plano tangente a la superficie* $\{(x, y, f(x, y)); (x, y) \in \mathbb{R}^2\}$ *en el punto* $(0, 0, 0)$. *Esta superficie se llama cono.*

Sabemos del curso de Análisis de Variable Real que cuando una aplicación es derivable en un punto entonces es continua en ese punto. Lo mismo ocurre en el caso general de aplicaciones diferenciables en un punto de $\mathbb{R}^n$.

Teorema 2.2.9 *Si* $f : A \subset \mathbb{R}^n \to \mathbb{R}$ *es diferenciable en un punto* x_0 *del abierto* A *entonces* f *es continua en ese punto.*

Demostración. Escribamos

$$f(x) - f(x_0) = f(x) - f(x_0) - \langle \nabla f(x_0), x - x_0 \rangle + \langle \nabla f(x_0), x - x_0 \rangle.$$

Como

$$\lim_{x \to x_0} \frac{f(x) - f(x_0) - \langle \nabla f(x_0), (x - x_0) \rangle}{\|x - x_0\|} = 0$$

resulta que

$$\exists \delta_1 > 0 \mid \; \|x - x_0\| < \delta_1 \Longrightarrow |f(x) - f(x_0) - \langle \nabla f(x_0), x - x_0 \rangle| \leq \|x - x_0\|.$$

Dado que por otra parte, $|\langle \nabla f(x_0), x - x_0 \rangle| \leq \|\nabla f(x_0)\| \, \|x - x_0\|$, tenemos

$$|f(x) - f(x_0)| < (1 + \|\nabla f(x_0)\|) \, \|x - x_0\|$$

si $\|x - x_0\| < \delta_1$. Luego, si dado $\varepsilon > 0$ tomamos $\delta = \min\{\frac{\varepsilon}{1 + \|\nabla f(x_0)\|}, \delta_1\}$ resulta que si $\|x - x_0\| < \delta$ entonces $|f(x) - f(x_0)| < \varepsilon$. $\square$

Hemos comentado anteriormente que puede ocurrir que una aplicación $f : A \subset \mathbb{R}^n \to \mathbb{R}$ tenga derivadas en un punto $x_0 \in A$ en todas las direcciones y que sin embargo no sea diferenciable en ese punto. Vamos a dar ahora un ejemplo de una tal función, ella va a tener derivadas en $(0, 0)$ en todas las direcciones y no va a ser ni siquiera continua en ese punto (luego no puede ser diferenciable en él como acabamos de ver).

Ejemplo 2.2.10 *Consideremos la función*

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

Dada una dirección $(\cos\theta, \operatorname{sen}\theta)$, $\theta \in [0, 2\pi)$, consideremos la función

$$\phi(t) = f(t(\cos\theta, \operatorname{sen}\theta)) = \begin{cases} \frac{t\cos\theta \cdot t^2 \operatorname{sen}^2\theta}{t^2\cos^2\theta + t^4 \operatorname{sen}^4\theta} & \text{si } t(\cos\theta, \operatorname{sen}\theta) \neq (0,0) \\ 0 & \text{si } t(\cos\theta, \operatorname{sen}\theta) = (0,0) \end{cases}$$

Veamos como ϕ es derivable en 0.

$$\lim_{t\to 0} \frac{\phi(t) - \phi(0)}{t} = \lim_{t\to 0} \frac{t\cos\theta \cdot t^2 \operatorname{sen}^2\theta}{t(t^2\cos^2\theta + t^4\operatorname{sen}^4\theta)} = \lim_{t\to 0} \frac{\cos\theta \operatorname{sen}^2\theta}{\cos^2\theta + t^2 \operatorname{sen}^4\theta} = \frac{\operatorname{sen}^2\theta}{\cos\theta}.$$

Claro está que estamos suponiendo que $\cos\theta \neq 0$. Si $\cos\theta = 0$ entonces ϕ es idénticamente nula por lo que $\phi'(0) = 0$. Resumiendo

$$D_{(\cos\theta, \operatorname{sen}\theta)} f(0,0) = \begin{cases} \frac{\operatorname{sen}^2\theta}{\cos\theta} & \text{si } \cos\theta \neq 0 \\ 0 & \text{si } \cos\theta = 0. \end{cases}$$

Veamos como f no es continua en el $(0,0)$. Para ello calculamos el límite a lo largo de la curva $x = y^2$,

$$\lim_{y\to 0} f(y^2, y) = \lim_{y\to 0} \frac{y^4}{y^4 + y^4} = \frac{1}{2}.$$

Como el valor de este límite no coincide con el valor de la función en el punto $(0,0)$ (que es 0), la función f no es continua en $(0,0)$.

Obsérvese que aunque hubiésemos definido el valor de f en $(0,0)$ como $\frac{1}{2}$ la función seguiría sin ser continua, pues el límite de f en $(0,0)$ a lo largo de las rectas $x = 0$ ó $y = 0$ es 0.

2.3. Condición suficiente de diferenciabilidad

Hemos visto en el tema anterior dos condiciones necesarias para que una función $f : A \subset \mathbb{R}^n \to \mathbb{R}$, A abierto de $\mathbb{R}^n$, sea diferenciable en $x_0 \in A$, a saber, que f sea continua y que tenga todas las derivadas direccionales, y en particular las parciales, en x_0. El teorema siguiente nos da una condición que será suficiente para la diferenciabilidad.

Teorema 2.3.1 *Sean $A \subset \mathbb{R}^n$ un conjunto abierto, $f : A \to \mathbb{R}$ una aplicación que tiene todas las derivadas parciales en cada punto de A. Si para cada $i = 1, ..., n$ la función*

$$x \in A \mapsto \frac{\partial f}{\partial x_i}(x) \in \mathbb{R}$$

es continua en un punto $x_0 \in A$, entonces f es diferenciable en x_0.

Demostración. Fijemos el punto $x_0 \in A$ en el que todas las derivadas parciales de f sean continuas y consideremos un $r > 0$ tal que $B(x_0, r) \subset A$. Para cualquier punto x de esa bola se tiene que

$$\begin{aligned}
& f(x) - f(x_0) = f(x_1, x_2, ..., x_n) - f(x_{0_1}, ..., x_{0_n}) \\
= \;& f(x_1, x_2, ..., x_n) - f(x_{0_1}, x_2, ..., x_n) \\
& + f(x_{0_1}, x_2, ..., x_n) - f(x_{0_1}, x_{0_2}, x_3, ..., x_n) \\
& + f(x_{0_1}, x_{0_2}, x_3, ..., x_n) - f(x_{0_1}, x_{0_2}, x_{0_3}, ..., x_n) \\
& + \cdots \\
& + f(x_{0_1}, x_{0_2}, x_{0_3}, ..., x_{0_{n-2}}, x_{0_{n-1}}, x_n) - f(x_{0_1}, ..., x_{0_n}).
\end{aligned}$$

Aplicando el teorema del valor medio para funciones de una variable real a la función $\psi_1(s) = f(s, x_2, ..., x_n)$, que es continua en el intervalo determinado por x_{01} y x_1 y derivable en su interior (por ser f derivable en la dirección e_1 en todos los puntos de A), obtenemos la existencia de un punto u_1 entre $x_{0,1}$ y x_1 tal que

$$\psi_1(x_1) - \psi_1(x_{0_1}) = \psi_1'(u_1)(x_1 - x_{0_1})$$

(si $x_1 = x_{0_1}$ pasamos directamente a otra coordenada, pues en esta primera la diferencia $f(x_1, x_2, ..., x_n) - f(x_{0_1}, x_2, ..., x_n)$ sería nula). Ahora bien,

$$\psi_1'(u_1) = \frac{\partial f}{\partial x_1}(u_1, x_2, ..., x_n),$$

por lo que

$$f(x_1, x_2, ..., x_n) - f(x_{0_1}, x_2, ..., x_n) = \frac{\partial f}{\partial x_1}(u_1, x_2, ..., x_n)(x_1 - x_{0_1}).$$

Repitiendo el proceso con

$$\psi_2(s) = f(x_{0_1}, s, x_3, ..., x_n)$$

obtenemos un punto u_2 entre x_{0_2} y x_2 tal que

$$f(x_{0_1}, x_2, ..., x_n) - f(x_{0_1}, x_{0_2}, x_3, ..., x_n) = \frac{\partial f}{\partial x_2}(x_{0_1}, u_2, x_3, ..., x_n)(x_2 - x_{0_2}).$$

Continuando con el proceso anterior conseguimos la existencia de puntos $u_3, ..., u_n$ tales que

$$\begin{aligned}
& f(x_{0_1}, ..., x_{0_{i-1}}, x_i, x_{i+1}, ..., x_n) - f(x_{0_1}, ..., x_{0_{i-1}}, x_{0_i}, x_{i+1}, ..., x_n) \\
= \;& \frac{\partial f}{\partial x_i}(x_{0_1}, ..., x_{0_{i-1}}, u_i, x_{i+1}, ..., x_n)(x_i - x_{0_i})
\end{aligned}$$

para $i = 3, ..., n$. Tenemos entonces que,

$$\begin{aligned}
f(x) - f(x_0) \;=\; & \frac{\partial f}{\partial x_1}(u_1, x_2, ..., x_n)(x_1 - x_{0_1}) + \cdots \\
& + \frac{\partial f}{\partial x_n}(x_{0_1}, x_{0_2}, ..., x_{0_{n-1}}, u_n)(x_n - x_{0_n}).
\end{aligned}$$

En consecuencia,

$$\begin{array}{rl}
& \left| f(x) - f(x_0) - \left[\dfrac{\partial f}{\partial x_1}(x_0)(x_1 - x_{0_1}) + \cdots + \dfrac{\partial f}{\partial x_n}(x_0)(x_n - x_{0_n}) \right] \right| \\
\leq & \left| \dfrac{\partial f}{\partial x_1}(u_1, x_2, ..., x_n) - \dfrac{\partial f}{\partial x_1}(x_{0_1}, x_{0_2}, ..., x_{0_n}) \right| |x_1 - x_{0_1}| \\
& + \cdots \\
& + \left| \dfrac{\partial f}{\partial x_n}(x_{0_1}, x_{0_2}, ..., u_n) - \dfrac{\partial f}{\partial x_n}(x_{0_1}, x_{0_2}, ..., x_{0_n}) \right| |x_n - x_{0_n}| \,.
\end{array}$$

La continuidad de las derivadas parciales en x_0 junto con el hecho de que cada $|u_i - x_{0_i}|$ es menor o igual que $|x_i - x_{0_i}|$, nos da que para todo $\varepsilon > 0$ existe $\delta > 0$ tal que si $\|x - x_0\| < \delta$ cada uno de esos sumandos es menor que $\frac{\varepsilon}{n} \|x - x_0\|$ con lo que

$$\frac{|f(x) - f(x_0) - \langle \nabla f(x_0), x - x_0 \rangle|}{\|x - x_0\|} < \varepsilon.$$

Esto nos da que f es diferenciable en x_0. □

Nota 2.3.2 *Observamos que en las hipótesis del teorema, si todas las derivadas parciales son continuas en todos los puntos de A la aplicación $df : x \in A \longmapsto df(x) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ es continua (en $\mathcal{L}(\mathbb{R}^n, \mathbb{R})$ se considera la norma definida por $\|L\| = \sup\limits_{h \in \mathbb{R}^n, ||h||=1} |L(h)|$). En efecto,*

$$\begin{aligned}
|df(x)(h) - df(x_0)(h)| &= \left| \sum_{j=1}^{n} \left(\frac{\partial f}{\partial x_j}(x) - \frac{\partial f}{\partial x_j}(x_0) \right) h_j \right| \\
&\leq \sum_{j=1}^{n} \left| \left(\frac{\partial f}{\partial x_j}(x) - \frac{\partial f}{\partial x_j}(x_0) \right) \right| |h_j| \\
&\leq \sum_{j=1}^{n} \left| \left(\frac{\partial f}{\partial x_j}(x) - \frac{\partial f}{\partial x_j}(x_0) \right) \right| .
\end{aligned}$$

Como cada $\frac{\partial f}{\partial x_i}$ es continua en el punto x_0, puede conseguirse que esa suma sea menor que cualquier $\varepsilon > 0$ dado, tomando x suficiente próximo a x_0. Entonces

$$\begin{aligned}
\|df(x) - df(x_0)\| &= \sup_{h \in \mathbb{R}^n, ||h||=1} |[df(x) - df(x_0)](h)| \\
&\leq \sup_{h \in \mathbb{R}^n, ||h||=1} \sum_{j=1}^{n} \left| \left(\frac{\partial f}{\partial x_j}(x) - \frac{\partial f}{\partial x_j}(x_0) \right) \right| \|h\| \\
&= \sum_{j=1}^{n} \left| \left(\frac{\partial f}{\partial x_j}(x) - \frac{\partial f}{\partial x_j}(x_0) \right) \right| < \varepsilon.
\end{aligned}$$

Del teorema anterior se deduce la equivalencia entre la diferenciabilidad de f en A más la continuidad de la aplicación $df : A \to \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ y la existencia y continuidad de las derivadas parciales en A. Nótese que si df es continua entonces también lo son $\frac{\partial f}{\partial x_i} : x \in A \longmapsto \frac{\partial f}{\partial x_i}(x) = df(x)(e_i) \in \mathbb{R}$, pues $|df(x)(e_i) - df(x_0)(e_i)| \leq \|df(x) - df(x_0)\|_{\mathcal{L}(\mathbb{R}^n,\mathbb{R})}$.

La no continuidad de las parciales no impide la diferenciabilidad: La función

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

tiene derivadas parciales en todo punto de $\mathbb{R}^2$, no son continuas en $(0, 0)$ y sin embargo es diferenciable en ese punto.

Nota 2.3.3 *Puede probarse también la diferenciabilidad de una función en un abierto de $\mathbb{R}^n$ aunque no se sepa que una de sus n parciales es continua (véase F. Ballesteros, L. Martín, "Ejercicios de Análisis Matemático", p. 205).*

Cuando se consideran aplicaciones $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$, se define de manera análoga el concepto de aplicación diferenciable en un punto. En este caso la aplicación diferencial de f en un punto x_0 es la (única) aplicación lineal de $\mathbb{R}^n$ en $\mathbb{R}^m$ (si tal aplicación existe), que denotaremos también por $df(x_0)$, y que tiene la propiedad de que

$$\mathit{lím}_{x \to x_0} \frac{f(x) - f(x_0) - df(x_0)(x - x_0)}{\|x - x_0\|} = 0_{\mathbb{R}^m}.$$

Debe observarse que una tal aplicación es diferenciable en un punto x_0 si y sólo si lo son en ese punto las componentes $f_1, ..., f_m$ de f y se verifica que

$$df(x_0) = (df_1(x_0), df_2(x_0), ..., df_m(x_0)).$$

Así como las aplicaciones lineales L de $\mathbb{R}^n$ en $\mathbb{R}$ se representan por un elemento de $\mathbb{R}^n$, a saber, $(L(e_1), ..., L(e_n))$, las aplicaciones lineales $L = (L_1, ..., L_m) : \mathbb{R}^n \to \mathbb{R}^m$ verifican que:

$$\begin{pmatrix} L_1(x) \\ \cdots \\ L_m(x) \end{pmatrix} = \begin{pmatrix} L_1(e_1) & \cdots & L_1(e_n) \\ \cdots & \cdots & \cdots \\ L_m(e_1) & \cdots & L_m(e_n) \end{pmatrix} \begin{pmatrix} x_1 \\ \cdots \\ x_n \end{pmatrix}$$

Veamos como es esta matriz cuando L es la diferencial de una aplicación $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$. Empecemos con el caso $m = 1$:

$$L(e_i) = df(x_0)(e_i) = D_{e_i} f(x_0) = D_i f(x_0).$$

Esto es, la matriz asociada a $df(x_0)$ es precisamente el gradiente de f en x_0.

En el caso general se tiene lo siguiente:

$$(df(x_0))_j(e_i) = df_j(x_0)(e_i) = D_i f_j(x_0)$$

con lo que la matriz asociada a $df(x_0)$ es

$$J_f(x_0) = \begin{pmatrix} D_1 f_1(x_0) & \cdots & D_n f_1(x_0) \\ \vdots & \vdots & \vdots \\ D_1 f_m(x_0) & \cdots & D_n f_m(x_0) \end{pmatrix}.$$

Esta matriz se llama **matriz jacobiana**[1] de f en x_0 y será muy utilizada en lo que sigue. Como se ve es la matriz que tiene en la fila j el gradiente de la componente f_j de f.

Para aplicaciones lineales $L : \mathbb{R}^n \to \mathbb{R}^m$ se verifica que existe $C > 0$ tal que

$$||L(x)|| \leq C||x|| \text{ para todo } x \in \mathbb{R}^n.$$

En efecto, sabemos que para cada componente L_j de L sabemos que $|L_j(x)| \leq ||L_j||||x||$ para todo $x \in \mathbb{R}^n$. Entonces

$$|L_1(x)| + \cdots + |L_m(x)| \leq (||L_1|| + \cdots + ||L_m||)||x|| \text{ para todo } x \in \mathbb{R}^n.$$

Por lo tanto,

$$||L(x)|| \leq ||L(x)||_1 = |L_1(x)| + \cdots + |L_m(x)| \leq C||x|| \text{ para todo } x \in \mathbb{R}^n$$

siendo $C = ||L_1|| + \cdots + ||L_m||$.

Esto nos permite definir una norma en el espacio vectorial $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$:

$$||L|| = \sup\{||L(x)|| : ||x|| = 1\}.$$

Evidentemente, $||L(x)|| \leq ||L||||x||$ para todo $x \in \mathbb{R}^n$.

Se verifica que la suma de dos aplicaciones diferenciables en un punto es también diferenciable en ese punto y lo mismo ocurre si multiplicamos una función diferenciable por un número real. Además es también fácil ver que

$$d(f+g)(x_0) = df(x_0) + dg(x_0) \quad \text{y} \quad d(\alpha f)(x_0) = \alpha df(x_0).$$

También es cierto que, cuando $m = 1$, el producto de funciones diferenciables en un punto es diferenciable en ese punto verificándose que

$$d(fg)(x_0) = g(x_0)df(x_0) + f(x_0)dg(x_0).$$

Es sencillo el obtener esto directamente, pero nosotros lo obtendremos en el tema siguiente como consecuencia de un resultado de gran interés en el estudio de la diferenciabilidad, se trata de la llamada "Regla de la cadena", conocida del curso de Análisis de Variable Real para funciones de una variable.

[1] jacobiana viene de Jacobi, matemático alemán, 1804-1851.

2.4. Regla de la cadena

Comenzamos el tema obteniendo una propiedad de las aplicaciones diferenciables que se conoce con el nombre de *propiedad local de Lipschitz* y que usaremos para demostrar la Regla de la Cadena.

Teorema 2.4.1 *Sea A un abierto de $\mathbb{R}^n$ y $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ una aplicación diferenciable en un punto $x_0 \in A$. Entonces existen constantes M y $\delta_1 > 0$ tales que*

$$\|x - x_0\| < \delta_1 \Longrightarrow \|f(x) - f(x_0)\| \leq M \|x - x_0\| .$$

Nótese que, en particular, esto implica la continuidad de f en x_0, cosa que ya habíamos probado para $m = 1$. También puede obtenerse esta continuidad observando que cada f_j es continua, por ser diferenciable.

Demostración. Sea $\varepsilon = 1$, entonces existe $\delta_1 > 0$ tal que

$$0 < \|x - x_0\| < \delta_1 \Longrightarrow \|f(x) - f(x_0) - df(x_0)(x - x_0)\| < \|x - x_0\|$$

con lo que

$$\|f(x) - f(x_0)\| \leq \|df(x_0)(x - x_0)\| + \|x - x_0\| .$$

Dado que también por ser $df(x_0)$ lineal existe $C > 0$

$$\|df(x_0)(x - x_0)\| \leq C||x - x_0||$$

para todo $x \in \mathbb{R}^n$, resulta que

$$\|f(x) - f(x_0)\| \leq M \|x - x_0\|$$

para los x tales que $\|x - x_0\| < \delta_1$, donde $M = C + 1$. □

Obtenemos ahora la regla de la cadena que garantiza la diferenciabilidad de una composición y nos dice como calcular la correspondiente diferencial.

Teorema 2.4.2 ***(Regla de la cadena)***. *Sean A un abierto de $\mathbb{R}^n$, B un abierto de $\mathbb{R}^m$, $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$, $g : B \subset \mathbb{R}^m \to \mathbb{R}^p$ tales que $f(A) \subset B$, f es diferenciable en $x_0 \in A$ y g es diferenciable en $f(x_0)$. Entonces $g \circ f$ es diferenciable en x_0 y $d(g \circ f)(x_0) = dg(f(x_0)) \circ df(x_0)$.*

Demostración. Se tiene que

$$\begin{aligned}
& \frac{\|g\circ f(x)-g\circ f(x_0)-(dg(f(x_0))\circ df(x_0))(x-x_0)\|_{\mathbb{R}^p}}{\|x-x_0\|} \\
\leq \; & \frac{\|g(f(x))-g(f(x_0))-dg(f(x_0))(f(x)-f(x_0))\|}{\|x-x_0\|}+ \\
& \frac{\|dg(f(x_0))(f(x)-f(x_0)-df(x_0)(x-x_0))\|}{\|x-x_0\|}.
\end{aligned}$$

Al ser f diferenciable en x_0 existen, por el teorema anterior, $M>0$ y $\delta_1>0$ tales que

$$\|x-x_0\|<\delta_1 \Longrightarrow \|f(x)-f(x_0)\| \leq M\,\|x-x_0\|\,.$$

Por otra parte, al ser g diferenciable en $f(x_0)$, dado $\frac{\varepsilon}{2M}>0$ existe $\delta_2>0$ tal que

$$0<\|y-f(x_0)\|<\delta_2 \Longrightarrow \frac{\|g(y)-g(f(x_0))-dg(f(x_0))(y-f(x_0))\|}{\|y-f(x_0)\|} \leq \frac{\varepsilon}{2M}.$$

Tomando $\delta_3 = \text{mín}\{\delta_1, \frac{\delta_2}{M}\}$ tenemos que

$$\begin{aligned}
0 \; < \; & \|x-x_0\|<\delta_3 \Longrightarrow \frac{\|g(f(x))-g(f(x_0))-dg(f(x_0))(f(x)-f(x_0))\|}{\|x-x_0\|} \\
\leq \; & \frac{\varepsilon\,\|f(x)-f(x_0)\|}{2M||x-x_0||} \leq \frac{\varepsilon M\,\|x-x_0\|}{2M||x-x_0||} = \frac{\varepsilon}{2}
\end{aligned}$$

pues para esos x se tiene que $\|f(x)-f(x_0)\| \leq M\,\|x-x_0\| < M\frac{\delta_2}{M}=\delta_2$.

Además, por ser $dg(f(x_0))$ una aplicación lineal, existe $C^*>0$ tal que $\|dg(f(x_0))(y)\| \leq C^*\,\|y\|$ cualquiera que sea $y\in\mathbb{R}^m$. Por ser f diferenciable en x_0 existe $\delta_4>0$ tal que

$$0<\|x-x_0\|<\delta_4 \Longrightarrow \frac{\|f(x)-f(x_0)-df(x_0)(x-x_0)\|}{\|x-x_0\|} < \frac{\varepsilon}{2C^*}.$$

Así, para los x tales que $0<\|x-x_0\|<\delta_4$ se tiene que

$$\begin{aligned}
& \frac{\|dg(f(x_0))[f(x)-f(x_0)-df(x_0)(x-x_0)]\|}{\|x-x_0\|} \\
\leq \; & \frac{C^*||f(x)-f(x_0)-df(x_0)(x-x_0)||}{\|x-x_0\|} < \frac{C^*\varepsilon}{2C^*} = \frac{\varepsilon}{2}.
\end{aligned}$$

Si finalmente hacemos $\delta = \text{mín}\{\delta_3,\delta_4\}$ y tomamos x tales que $0<\|x-x_0\|<\delta$ resulta que

$$\begin{aligned}
& \frac{\|g(f(x))-g(f(x_0))-dg(f(x_0))\circ df(x_0)(x-x_0)\|}{\|x-x_0\|} \\
\leq \; & \frac{\|g(f(x))-g(f(x_0))-dg(f(x_0))(f(x)-f(x_0))\|}{\|x-x_0\|} \\
& +\frac{\|dg(f(x_0))(f(x)-f(x_0))-df(x_0)(x-x_0)\|}{\|x-x_0\|} \\
< \; & \varepsilon.
\end{aligned}$$

$\square$

La matriz jacobiana asociada a la composición es:

$$\begin{pmatrix} \frac{\partial(g\circ f)_1}{\partial x_1}(x_0) & \cdots & \frac{\partial(g\circ f)_1}{\partial x_n}(x_0) \\ \cdots & \cdots & \cdots \\ \frac{\partial(g\circ f)_p}{\partial x_1}(x_0) & \cdots & \frac{\partial(g\circ f)_p}{\partial x_n}(x_0) \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\partial g_1}{\partial y_1}(f(x_0)) & \cdots & \frac{\partial g_1}{\partial y_m}(f(x_0)) \\ \cdots & \cdots & \cdots \\ \frac{\partial g_p}{\partial y_1}(f(x_0)) & \cdots & \frac{\partial g_p}{\partial y_m}(f(x_0)) \end{pmatrix} \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x_0) & \cdots & \frac{\partial f_1}{\partial x_n}(x_0) \\ \cdots & \cdots & \cdots \\ \frac{\partial f_m}{\partial x_1}(x_0) & \cdots & \frac{\partial f_m}{\partial x_n}(x_0) \end{pmatrix}$$

El resultado es una matriz $p \times n$. Obsérvese que para cada $i = 1, ..., n$ y $j = 1, ..., p$, se tiene que

$$\frac{\partial(g\circ f)_j}{\partial x_i}(x_0) = \sum_{k=1}^{m} \frac{\partial g_j}{\partial y_k}(f(x_0))\frac{\partial f_k}{\partial x_i}(x_0).$$

Ejemplo 2.4.3 *Consideremos las funciones*

$$f(x,y,z) = (x^2+y^2, z); \; g(u,v) = (\cos u, 3v).$$

Entonces $g \circ f(x,y,z) = (\cos(x^2+y^2), 3z)$ *y la matriz jacobiana correspondiente a ella es*

$$\begin{pmatrix} -\operatorname{sen}(x^2+y^2) & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 2x & 2y & 0 \\ 0 & 0 & 1 \end{pmatrix}$$
$$= \begin{pmatrix} -2x\operatorname{sen}(x^2+y^2) & -2y\operatorname{sen}(x^2+y^2) & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

Nota 2.4.4 *De la regla de la cadena puede obtenerse fácilmente la regla de diferenciación para un producto de funciones definidas en un conjunto* A *abierto de* $\mathbb{R}^n$ *en* $\mathbb{R}$. *En efecto, consideremos primero el caso del producto de una* f *por si misma. Esta función es la composición de la función* f *con la función* $g : t \in \mathbb{R} \to t^2 \in \mathbb{R}$.

Por la regla de la cadena f^2 es diferenciable en todo punto donde lo sea f y en cada uno de sus puntos se verifica que

$$d(g\circ f)(x_0) = dg(f(x_0)) \circ df(x_0).$$

En efecto, para cada punto $h \in \mathbb{R}^n$

$$\begin{aligned} df^2(x_0)(h) &= d(g\circ f)(x_0)(h) \\ &= dg(f(x_0))(df(x_0)(h)) \\ &= g'(f(x_0)) \cdot (df(x_0)(h)) \\ &= 2f(x_0) \cdot (df(x_0)(h)) \\ &= (2f(x_0) \cdot df(x_0))(h). \end{aligned}$$

Si ahora f y g son dos funciones cualesquiera de definidas en un conjunto A abierto de $\mathbb{R}^n$ en $\mathbb{R}$, entonces

$$fg = \frac{1}{4}((f+g)^2 - (f-g)^2).$$

Entonces

$$\begin{aligned}
d(fg)(x_0) &= \frac{1}{4}(d((f+g)^2)(x_0) - d((f-g)^2)(x_0)) \\
&= \frac{1}{4}(2(f+g)(x_0)\cdot d(f+g)(x_0) - 2(f-g)(x_0)\cdot d(f-g)(x_0)) \\
&= \frac{1}{2}((f(x_0)+g(x_0))\cdot(df(x_0)+dg(x_0)) \\
&\quad -(f(x_0)-g(x_0))\cdot(df(x_0)-dg(x_0))) \\
&= \frac{1}{2}(2f(x_0)\cdot dg(x_0) + 2g(x_0)\cdot df(x_0)) \\
&= \frac{1}{2}(2f(x_0)\cdot dg(x_0) + 2g(x_0)\cdot df(x_0)) \\
&= f(x_0)\cdot dg(x_0) + g(x_0)\cdot df(x_0).
\end{aligned}$$

También de la regla de la cadena se obtiene que si una función $f : A \subset \mathbb{R}^n \to \mathbb{R}$ es diferenciable en x_0 y $f(x) \neq 0$ para todo $x \in A$, entonces $\frac{1}{f}$ es diferenciable en x_0 y

$$d\left(\frac{1}{f}\right)(x_0) = -\frac{df(x_0)}{(f(x_0))^2}.$$

Para demostrarlo basta considerar la composición de f con la aplicación de inversión $i : t \mapsto \frac{1}{t}$

Por la regla de la cadena

$$d(i\circ f)(x_0) = di[f(x_0)] \circ df(x_0) = i'(f(x_0))\cdot df(x_0) = -\frac{1}{[f(x_0)]^2}\cdot df(x_0).$$

Una última aplicación de la Regla de la Cadena que damos aquí es la siguiente: Sea $\varphi = (\varphi_1, ..., \varphi_m) : A \subset \mathbb{R}^1 \to \mathbb{R}^m$ diferenciable en un punto $t_0 \in A$ y sea $g : B \subset \mathbb{R}^m \to \mathbb{R}^1$, $\varphi(A) \subset B$ y g diferenciable en $\varphi(t_0)$. Entonces

$$(g\circ\varphi)'(t_0) = d(g\circ\varphi)(t_0))(1) = dg(\varphi(t_0))\circ d\varphi(t_0)(1) = dg(\varphi(t_0))(\varphi'(t_0)) = \langle\nabla g(\varphi(t_0)), \varphi'(t_0)\rangle$$

(donde $\varphi'(t_0) = (\varphi_1'(t_0), \ldots \varphi_m'(t_0))$.

2.5. Teorema del valor medio

Obtendremos aquí un teorema que generaliza al del valor medio para funciones de una variable. Con ese fin damos una definición y un lema.

Definición 2.5.1 *Se dice que un subconjunto S de $\mathbb{R}^n$ es* ***convexo*** *si para todo par de puntos x, $y \in S$, $x \neq y$, se verifica que el segmento $L[x,y] = \{x + t(y-x) : t \in [0,1]\}$ de extremos x e y está contenido en S.*

Las bolas son conjuntos convexos, pues si dada una bola $B(x_0, r)$ consideramos dos puntos x e y de ella, entonces para todo $t \in [0,1]$ se tiene que

$$\begin{aligned} ||x + t(y-x) - x_0|| &= ||(1-t)x + ty - tx_0 - (1-t)x_0|| = \\ &= (1-t)||x - x_0|| + t||y - x_0|| < (1-t)r + tr = r \end{aligned}$$

con lo que $L[x,y] \subset B(x_0, r)$.

Todos los conjuntos convexos son conexos, pues son conexos por poligonales (véase el Teorema 1.3.15), dado que los segmentos son un caso particular de poligonales.

Lema 2.5.2 *Sean A un subconjunto abierto de $\mathbb{R}^n$ y $f : A \to \mathbb{R}^m$ una aplicación diferenciable en cada punto de A. Entonces para todo $x, y \in A$ tales $L[x,y] \subset A$ y para todo $z \in \mathbb{R}^m$ se verifica que existe $c \in L[x,y]$, siendo c distinto de x e y, tal que*

$$\langle z, f(y) - f(x) \rangle = \langle z, df(c)(y-x) \rangle.$$

Obsérvese que, en principio, c depende de x, y, z.

Demostración. Dados $x, y \in A$ tales que $L[x,y] \subset A$, sea $\delta > 0$ tal que

$$x + t(y-x) \in A \text{ para todo } t \in (-\delta, 1+\delta)$$

(nótese que $L[x,y] \subset A$ y A es abierto). Dado $z \in \mathbb{R}^m$ consideremos la función real de variable real:

$$\psi(t) = \langle z, f(x + t(y-x)) \rangle; \quad t \in (-\delta, 1+\delta).$$

Esta función es derivable en $(-\delta, 1+\delta)$ y

$$\psi'(t) = \langle z, df(x + t(y-x))(y-x) \rangle,$$

pues

$$\psi(t) = z_1 f_1(x + t(y-x)) + \cdots + z_m f_m(x + t(y-x))$$

téngase en cuenta que se trata de la diferencial de la composición de f con la función $\varphi(t) = x + t(y-x)$).Luego, por el teorema del valor medio para funciones reales de una variable real existe $s \in (0,1)$ tal que

$$\psi(1) - \psi(0) = \psi'(s)(1-0)$$

de donde se sigue que

$$\begin{aligned}
\langle z, f(y) - f(x)\rangle &= \langle z, f(y)\rangle - \langle z, f(x)\rangle \\
&= \psi(1) - \psi(0) \\
&= \psi'(s) \\
&= \langle z, df(x + s(y - x))(y - x)\rangle.
\end{aligned}$$

Basta entonces tomar $c = x + s(y - x) \in L[x, y]$, $c \neq x, y$, para tener el resultado. Nótese que c depende de z, pues s depende de z. □

En los abiertos convexos el Lema anterior se verifica para todo $x, y \in A$.

Corolario 2.5.3 ***(Teorema del valor medio para funciones de*** $\mathbb{R}^n$ ***en*** $\mathbb{R}$***)*** *Sean A un subconjunto abierto de $\mathbb{R}^n$ y $f : A \to \mathbb{R}$ una función diferenciable en cada punto de A. Entonces para todo $x, y \in A$ tales $L[x, y] \subset A$ existe $c \in L[x, y]$, siendo c distinto de x e y, tal que*

$$f(y) - f(x) = df(c)(y - x).$$

Demostración. Como en este caso $m = 1$, si tomamos $z = 1$ el resultado se sigue directamente del Lema anterior, pues

$$f(y) - f(x) = \langle 1, f(y) - f(x)\rangle = \langle 1, df(c)(y - x)\rangle = df(c)(y - x).$$

□

Nota 2.5.4 *No hay un teorema del valor medio análogo al anterior para aplicaciones con valores en $\mathbb{R}^m$. En efecto, si consideramos la aplicación*

$$f(t) = (t^2, t^3);\ t \in \mathbb{R}.$$

Entonces $f(1) - f(0) = (1, 1) - (0, 0) = (1, 1)$, y como

$$df(u)(1 - 0) = (f_1'(u), f_2'(u)) = (2u, 3u^2)$$

debería ocurrir, que para un cierto $u \in (0, 1)$,

$$(1, 1) = (2u, 3u^2)$$

lo que ciertamente es imposible cualquiera que sea u.

No obstante para aplicaciones con valores en $\mathbb{R}^m$ se tiene el siguiente resultado.

Teorema 2.5.5 ***(desigualdad del valor medio para aplicaciones de*** $\mathbb{R}^n$ ***en*** $\mathbb{R}^m$, *también llamado* ***teorema de los incrementos finitos****). Sean A un abierto de $\mathbb{R}^n$ y $f : A \to \mathbb{R}^m$ una aplicación diferenciable en A. Entonces para todo $x, y \in A$ tales que $L[x,y] \subset A$ se verifica que*

$$\|f(y) - f(x)\| \leq \sup_{c \in L[x,y]} \|df(c)\| \cdot \|y - x\|$$

suponiendo que $\{\|df(c)\| : c \in L[x,y]\}$ es acotado. Si no lo es el resultado no tiene interés.

Demostración. Sean x e $y \in A$ tales que $L[x,y] \subset A$, podemos suponer que $f(y) \neq f(x)$, pues si son iguales la desigualdad es trivial. Sea

$$z = \frac{f(y) - f(x)}{||f(y) - f(x)||}.$$

Por el Lema 2.5 existe $c \in L[x,y]$, $c \neq x, y$, tal que

$$\langle z, f(y) - f(x) \rangle = \langle z, df(c)(y - x) \rangle \,,$$

de donde se sigue que

$$\begin{aligned} ||f(y) - f(x)|| &= \left\langle \frac{f(y) - f(x)}{||f(y) - f(x)||}, f(y) - f(x) \right\rangle \\ &= \left\langle \frac{f(y) - f(x)}{||f(y) - f(x)||}, df(c)(y - x) \right\rangle \\ &\leq \left\| \frac{f(y) - f(x)}{||f(y) - f(x)||} \right\| ||df(c)(y - x)|| \\ &= ||df(c)(y - x)|| \end{aligned}$$

(se ha usado la desigualdad de Cauchy-Schwarz y el hecho de que $||z|| = 1$). Como por la propia definición de la norma de la aplicación lineal $df(c)$ se tiene que

$$||df(c)(y - x)|| \leq ||df(c)|| \cdot ||y - x|| \leq \sup_{c \in L[x,y]} \|df(c)\| \cdot ||y - x||$$

se sigue el resultado. □

Para funciones con valores vectoriales se tiene también el siguiente resultado.

Teorema 2.5.6 *Sean A un abierto de $\mathbb{R}^n$ y $f : A \to \mathbb{R}^m$ una aplicación diferenciable en A. Entonces para todo $x, y \in A$ tal que $L[x,y] \subset A$ y para todo $j = 1, ..., m$ existe $c_j \in L[x,y]$ tal que*

$$f_j(y) - f_j(x) = df_j(c_j)(y - x)$$

Demostración. Basta aplicar el Teorema del Valor Medio para funciones con valores en $\mathbb{R}$ a cada componente de f. $\square$

Corolario 2.5.7 *Si A es un subconjunto abierto y conexo de $\mathbb{R}^n$, $f : A \to \mathbb{R}^m$ es diferenciable en A y $df(x) = 0$ para todo $x \in A$, entonces f es constante.*

Demostración. Fijemos un punto x_0 en A y sea x un punto arbitrario de A. Al ser A abierto y conexo existe una poligonal que une x_0 con x (Teorema 1.3.15). Esa poligonal está formada por la concatenación de segmentos $L[x_0, x_1]$, $L[x_1, x_2]$... $L[x_k, x]; x_1, \ldots, x_k \in \mathbb{R}^n$. Si aplicamos el teorema de los incrementos finitos a cada uno de esos segmentos obtenemos que

$$\begin{aligned}
||f(x_1) - f(x_0)|| &\leq \sup_{c_0 \in L[x_0,x_1]} \|df(c_0)\| \cdot \|x_1 - x_0\| = 0 \\
||f(x_2) - f(x_1)|| &\leq \sup_{c_1 \in L[x_1,x_2]} \|df(c_1)\| \cdot \|x_2 - x_1\| = 0 \\
&\cdots \\
||f(x) - f(x_k)|| &\leq \sup_{c_k \in L[x_k,x]} \|df(c_k)\| \cdot \|x - x_k\| = 0
\end{aligned}$$

con lo que $f(x) = f(x_k) = f(x_{k-1}) = \cdots = f(x_1) = f(x_0)$. Entonces f vale en cada $x \in A$ lo mismo que en x_0, esto es, f es constante en A. $\square$

2.6. Derivadas de orden superior. Teorema de Taylor. Aproximación

Supongamos que A es un conjunto abierto de $\mathbb{R}^n$ y que $f : A \to \mathbb{R}$ tiene derivada en una dirección v en todo punto de A. Tiene entonces sentido considerar, para un vector dado u de $\mathbb{R}^n$, $u \neq 0$, la derivada de la aplicación $D_v f : x \in A \mapsto D_v f(x) \in \mathbb{R}$ en un punto x_0 en la dirección u. Esto es

$$D_u(D_v f)(x_0) = \lim_{t \to 0} \frac{D_v f(x_0 + tu) - D_v f(x_0)}{t}.$$

Si en particular $v = e_i$ y $u = e_j$ $(i, j = 1, ..., n)$ escribiremos

$$D_j(D_i f)(x_0) \text{ ó } \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)(x_0).$$

Son también habituales las notaciones:

$$D_{ji} f(x_0) \text{ ó } \frac{\partial^2 f}{\partial x_j \partial x_i}(x_0).$$

Estos elementos de $\mathbb{R}$ (uno por cada elección de i y j) se llamarán las **derivadas parciales segundas** de f en x_0. Cuando $i=j$ suele escribirse

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

Ejemplo 2.6.1 *Calcular las derivadas parciales segundas en el punto* $(0,0)$ *de la función:*

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

$$\frac{\partial f}{\partial x}(x,y)=\begin{cases}\frac{[y(x^2-y^2)+2x^2y](x^2+y^2)-xy(x^2-y^2)2x}{(x^2+y^2)^2} & \text{si } (x,y)\neq(0,0)\\ \lim\limits_{t\to0}\frac{f(t(1,0))-0}{t}=0 & \text{si } (x,y)=(0,0),\end{cases}$$

$$\frac{\partial f}{\partial y}(x,y)=\begin{cases}\frac{[x(x^2-y^2)-2xy^2](x^2+y^2)-xy(x^2-y^2)2y}{(x^2+y^2)^2} & \text{si } (x,y)\neq(0,0)\\ \lim\limits_{t\to0}\frac{f(t(0,1))-0}{t}=0 & \text{si } (x,y)=(0,0),\end{cases}$$

$$\frac{\partial^2 f}{\partial x^2}(0,0)=\lim_{t\to0}\frac{\frac{\partial f}{\partial x}(t(1,0))-\frac{\partial f}{\partial x}(0,0)}{t}=0,$$

$$\frac{\partial^2 f}{\partial x\partial y}(0,0)=\lim_{t\to0}\frac{\frac{\partial f}{\partial y}(t(1,0))-\frac{\partial f}{\partial y}(0,0)}{t}=\lim_{t\to0}\frac{t^5}{t^5}=1,$$

$$\frac{\partial^2 f}{\partial y\partial x}(0,0)=\lim_{t\to0}\frac{\frac{\partial f}{\partial x}(t(0,1))-\frac{\partial f}{\partial x}(0,0)}{t}=\lim_{t\to0}-\frac{t^5}{t^5}=-1,$$

$$\frac{\partial f}{\partial y^2}(0,0)=\lim_{t\to0}\frac{\frac{\partial f}{\partial y}(t(0,1))-\frac{\partial f}{\partial y}(0,0)}{t}=0.$$

Se observa que $D_{12}f(0,0)\neq D_{21}f(0,0)$. Como veremos enseguida esto ocurre "pocas" veces.

Definición 2.6.2 *Sea* $f:A\subset\mathbb{R}^n\to\mathbb{R}$, *siendo* A *un abierto de* $\mathbb{R}^n$. *Se dice que* f *es de* ***clase*** C^1 *en* A *si existen todas las derivadas parciales primeras en todo punto de* A *y éstas son continuas en* A.

Nota 2.6.3 *Como hemos visto cuando hemos dado una condición suficiente para la diferenciabilidad de una función en un abierto de $\mathbb{R}^n$, se tiene que si f es de C^1 en A entonces es diferenciable en todo punto de A. Notamos además que cuando f es de C^1 en A entonces $df : x \in A \to df(x) \in L(\mathbb{R}^n, \mathbb{R})$ es continua (esto lo vimos al estudiar la condición mencionada). Además, sabemos también que la existencia de $df(x)$ en todo punto x de A y la continuidad de df en A equivale que f sea de clase C^1 en A.*

Usaremos la notación $C^1(A)$ para denotar al conjunto de las funciones $f : A \to \mathbb{R}$ que sean de clase C^1 en A. Claramente $C^1(A)$ es un espacio vectorial.

Definición 2.6.4 *Se dice que $f : A \subset \mathbb{R}^n \to \mathbb{R}$ es de clase C^2 en el abierto A si f tiene todas las derivadas parciales de primer y segundo orden y estas últimas son continuas.*

Nótese que si f va a tener derivadas parciales segundas las primeras tienen que existir y ser continuas (pues la existencia y continuidad de las funciones $\frac{\partial^2 f}{\partial x_j \partial x_i}$ ya implica la diferenciabilidad de las funciones $\frac{\partial f}{\partial x_i}$ y por lo tanto de todas las $D_v f$, pues $D_v f = v_1 \frac{\partial f}{\partial x_1} + \cdots + v_n \frac{\partial f}{\partial x_n}$). Luego las aplicaciones de clase C^2 también son de clase C^1. Se usará la notación $C^2(A)$ para denotar al conjunto (que es espacio vectorial) de las funciones de clase C^2 en A. Obsérvese que $f \in C^2(A)$ si y sólo si $f \in C^1(A)$ y $\frac{\partial f}{\partial x_i} \in C^1(A)$ para todo $i = 1, ..., n$.

En general, para $p = 1, 2, 3...$ se dice que $f : A \subset \mathbb{R}^n \to \mathbb{R}$ se dice que f es de **clase** C^p en A si f tiene derivadas parciales hasta el orden p y éstas son continuas. Se representará por $C^p(A)$ al espacio vectorial de las funciones de clase C^p en A. Cuando f tiene derivadas parciales de todos los órdenes se dice que es una función de clase C^∞ en A.

En el ejemplo de cálculo de las derivadas parciales segundas de una función $f : \mathbb{R}^2 \to \mathbb{R}$ que acabamos de hacer se ha obtenido que $D_{12}f(0,0) \neq D_{21}f(0,0)$. Vamos a ver ahora un teorema, conocido como el "**Teorema de Schwarz**" que afirma, en particular, que cuando una función es de clase C^2 en un abierto de $\mathbb{R}^2$ entonces $D_{12}f(x) = D_{21}f(x)$ para todo $x \in A$. Esto justifica el que "pocas" veces no coincidan $D_{12}f(x)$ y $D_{21}f(x)$. Probaremos primero el siguiente lema:

Lema 2.6.5 *Sean A un abierto de $\mathbb{R}^n$ y $f : A \to \mathbb{R}$, una función de clase C^2 en A. Fijado un punto x_0 de A sea $\delta > 0$ tal que $B(x_0, \delta) \subset A$. Entonces para cada u y $v \in B(0, \frac{\delta}{2}) \backslash \{0\}$ existen α y $\beta \in (0,1)$ tales que*

$$f(x_0 + u + v) - f(x_0 + u) - (f(x_0 + v) - f(x_0)) = D_v(D_u f)(x_0 + \alpha u + \beta v).$$

Demostración. Consideremos la función

$$g : B\left(x_0, \frac{\delta}{2}\right) \to \mathbb{R}$$

definida por $g(x) = f(x+v) - f(x)$. Esta función g es diferenciable en $B\left(x_0, \frac{\delta}{2}\right)$ y $dg(x) = df(x+v) - df(x)$. Aplicamos el teorema del valor medio para funciones con valores en $\mathbb{R}$ a g en relación con el segmento $L[x_0, x_0+u]$ y obtenemos la existencia de un punto c entre x_0 y x_0+u (distinto de ellos) tal que

$$g(x_0+u) - g(x_0) = dg(c)(u).$$

Esto es,

$$f(x_0+u+v) - f(x_0+u) - (f(x_0+v) - f(x_0)) = (df(c+v) - df(c))(u)$$

(obsérvese que $L[x_0, x_0+u] \subset B(x_0, \frac{\delta}{2}) \subset A$). Si escribimos c en la forma $c = x_0 + \alpha u$ con $\alpha \in (0,1)$, tenemos que

$$\begin{aligned} & f(x_0+u+v) - f(x_0+u) - (f(x_0+v) - f(x_0)) \\ = \;& df(x_0+\alpha u+v)(u) - df(x_0+\alpha u)(u) \\ = \;& D_u f(x_0+\alpha u+v) - D_u f(x_0+\alpha u). \end{aligned}$$

Ahora bien, como $D_u f$ es diferenciable en A y $L[x_0+\alpha u, x_0+\alpha u+v]$ está contenido en A, otra aplicación del teorema del valor medio, ahora a $D_u f$, nos da que existe un punto e entre $x_0+\alpha u$ y $x_0+\alpha u+v$ (distinto de ellos) tal que

$$D_u f(x_0+\alpha u+v) - D_u f(x_0+\alpha u) = d(D_u f)(e)(v) = D_v(D_u f)(e).$$

Si finalmente escribimos $e = x_0 + \alpha u + \beta v$ (con $\beta \in (0,1)$) tenemos el resultado. □

Teorema 2.6.6 ***(de Schwarz)*** *Sean A un abierto de $\mathbb{R}^n$ y $f \in C^2(A)$, entonces $D_{ij}f(x) = D_{ji}f(x)$ para todo $x \in A$ y todo $i,j \in \{1,...,n\}$.*

Demostración. Fijemos x_0 en A. Probaremos que para todo $\varepsilon > 0$ se verifica que $|D_{ij}f(x_0) - D_{ji}f(x_0)| < \varepsilon$. Sea $\varepsilon > 0$. Como $D_{ij}f$ y $D_{ji}f$ son continuas en x_0, existe $\delta > 0$ tal que $B(x_0,\delta) \subset A$ y

$$\|x - x_0\| < \delta \Longrightarrow \begin{cases} |D_{ij}f(x) - D_{ij}f(x_0)| < \varepsilon/2 \\ |D_{ji}f(x) - D_{ji}f(x_0)| < \varepsilon/2. \end{cases}$$

Tomemos $u = te_i$ y $v = se_j$ con $s,t \in \left(0, \frac{\delta}{2}\right)$. Tenemos así puntos que están en las condiciones del lema anterior, aplicando ese lema dos veces (una con $u = te_i$ y $v = se_j$ y otra con $u = se_j$ y $v = te_i$) obtenemos que existen $\alpha_1, \beta_1, \alpha_2, \beta_2 \in (0,1)$ tales que

$$\begin{aligned} & f(x_0+te_i+se_j) - f(x_0+te_i) - (f(x_0+se_j) - f(x_0)) \\ = \;& D_{se_j}(D_{te_i}f)(x_0+\alpha_1 te_i+\beta_1 se_j), \end{aligned}$$

$$\begin{aligned} & f(x_0 + se_j + te_i) - f(x_0 + se_j) - (f(x_0 + te_i) - f(x_0)) \\ = \; & D_{te_i}(D_{se_j} f)(x_0 + \alpha_2 se_j + \beta_2 te_i). \end{aligned}$$

Como en general, $D_{\lambda u} f(x_0) = \lambda D_u f(x_0)$, se tiene que

$$stD_{ji} f(x_0 + \alpha_1 te_i + \beta_1 se_j) = tsD_{ij} f(x_0 + \alpha_2 se_j + \beta_2 te_i).$$

(pues los miembros de la izquierda coinciden). Por la desigualdad triangular

$$\begin{aligned} |D_{ij} f(x_0) - D_{ji} f(x_0)| \;\leq\; & |D_{ij} f(x_0) - D_{ij} f(x_0 + \alpha_2 se_j + \beta_2 te_i)| \\ & + |D_{ji} f(x_0 + \alpha_1 te_i + \beta_1 se_j) - D_{ji} f(x_0)| \end{aligned}$$

y, tal como habíamos tomado s y t, esta suma es menor que ε pues $\|\alpha_2 se_j + \beta_2 te_i\| < s + t < \delta$ y $\|\alpha_1 te_i + \beta_1 se_j\| < s+t < \delta$. La arbitrariedad de ε nos da que $D_{ij} f(x_0) = D_{ji} f(x_0)$.
□

Observaciones:

1. Puede probarse que si f es de clase C^1 en $A \subset \mathbb{R}^2$ y existe $\frac{\partial^2 f}{\partial x \partial y}$ y es continua, entonces también existe $\frac{\partial^2 f}{\partial y \partial x}$ y coinciden (ver Marsden-Hoffman, p. 375 o el libro de Carlos Fernández Pérez y otros, p. 256).

2. También se puede obtener la igualdad entre D_{12} y D_{21} a partir de la existencia de $D_1 f$ y $D_2 f$ en un entorno de (x_0, y_0) y su diferenciabilidad (T. de Heffter-Young. Ver el libro de Carlos Fernández Pérez y otros, p. 257).

Corolario 2.6.7 *Si $f \in C^3(A)$, entonces $D_{ijk} f(x) = D_{\sigma(i)\sigma(j)\sigma(k)} f(x)$ para todo $x \in A$ y toda permutación σ de $\{i, j, k\}$.*

Demostración. Supongamos que σ está dada por $\sigma(i) = k, \sigma(j) = i$ y $\sigma(k) = j$ (el caso general es análogo).

$$\begin{aligned} D_{ijk} f(x) &= D_i(D_j(D_k f))(x) \\ &= D_i(D_{jk} f)(x) = D_i(D_{kj} f)(x) \\ &= D_i(D_k(D_j f))(x) = D_{ik}(D_j f)(x) = D_{ki}(D_j f)(x) = D_{kij} f(x). \end{aligned}$$

Nótese que las $D_p f$ son de clase C^2 en A por lo que se les puede aplicar a ellas el teorema de Schwarz. □

Obtenemos ahora una generalización, al caso de varias variables, del conocido teorema de Taylor para funciones de una variable real.

Teorema 2.6.8 *(de Taylor) Sea $f : A \subset \mathbb{R}^n \to \mathbb{R}$ una función de clase $k+1$ en el abierto A (siendo $k = 0, 1...$) y sean $x_0 \in A$ y $h \neq 0$ tales que $L[x_0, x_0 + h] \subset A$. Entonces existe*

$c \in L[x_0, x_0 + h], c \neq x_0, x_0 + h$, *tal que*

$$\begin{aligned} f(x_0+h) &= f(x_0) + \sum_{i_1=1}^{n} D_{i_1} f(x_0) h_{i_1} \\ &+ \frac{1}{2!} \sum_{i_1=1, i_2=1}^{n} D_{i_1 i_2} f(x_0) h_{i_1} h_{i_2} + \cdots \\ &+ \frac{1}{k!} \sum_{i_1=1,\dots,i_k=1}^{n} D_{i_1 i_2 \cdots i_k} f(x_0) h_{i_1} \cdots h_{i_k} \\ &+ \frac{1}{(k+1)!} \sum_{i_1=1,\dots,i_{k+1}=1}^{n} D_{i_1 i_2 \cdots i_k i_{k+1}} f(c) h_{i_1} \cdots h_{i_k} h_{i_{k+1}}. \end{aligned}$$

Los suma de los $k+1$ primeros sumandos es un polinomio de grado menor o igual que k en las variables $h_1, ..., h_n$ y se llama el **polinomio de Taylor** de grado menor o igual que k correspondiente a f y al punto x_0. El último sumando se llama el **resto de Taylor** correspondiente f y al punto x_0.

Observaciones previas a la demostración:

1. Si $k = 0$ se trata del teorema del valor medio demostrado anteriormente pues $\sum_{i_1=1}^{n} h_{i_1} D_{i_1} f(c) = df(c)(h)$.

2. El teorema de Taylor no se verifica, en general, cuando la aplicación toma valores en $\mathbb{R}^m$ con $m > 1$. Nótese que este teorema coincide, como acabamos de indicar, con el del valor medio si $k = 0$ y ya hemos comentado en su momento que esta versión del teorema del valor medio no sirve para aplicaciones con valores en $\mathbb{R}^m$, con $m > 1$. El ejemplo que pusimos era
$$f(t) = (t^2, t^3).$$

3. Si estamos en $\mathbb{R}^2$ la fórmula anterior, para $h = (x - x_0, y - y_0)$, queda en la forma:

$$\begin{aligned} f(x,y) &= f(x_0,y_0) + D_1 f(x_0,y_0)(x-x_0) + D_2 f(x_0,y_0)(y-y_0) \\ &+ \frac{1}{2!}(D_{11} f(x_0,y_0)(x-x_0)^2 + 2D_{12} f(x_0,y_0)(x-x_0)(y-y_0) \\ &+ D_{22} f(x_0,y_0)(y-y_0)^2) \\ &+ \cdots \\ &+ \frac{1}{k!}\left(D_{1\overset{k}{\cdots}1} f(x_0,y_0)(x-x_0)^k + \cdots + D_{2\overset{k}{\cdots}2} f(x_0,y_0)(y-y_0)^k\right) \\ &+ \frac{1}{(k+1)!}(D_{1\overset{k+1}{\cdots}1} f(c_1,c_2)(x-x_0)^{k+1} + \cdots \\ &+ D_{2\overset{k+1}{\cdots}2} f(c_1,c_2)(y-y_0)^{k+1}) \end{aligned}$$

(se ha usado el teorema de Schwarz).

Demostración. (del teorema de Taylor) Consideremos la aplicación $\varphi : \mathbb{R} \to \mathbb{R}^n$ definida por $\varphi(t) = x_0 + th$. Como φ continua, $B = \varphi^{-1}(A)$ es un abierto de $\mathbb{R}$. Consideremos la composición $g = f \circ \varphi : B \subset \mathbb{R} \to \mathbb{R}$. La función g es de clase $k+1$ en el abierto B y además para $j = 1, ..., k+1$ se tiene que

$$g^{j)}(t) = \sum_{i_1=1,\ldots,i_j=1}^{n} D_{i_1 \ldots i_j} f(\varphi(t)) h_{i_1} \cdots h_{i_j}.$$

En efecto, por la regla de la cadena,

$$g'(t) = (f \circ \varphi)'(t) = \langle \nabla f(\varphi(t))), h \rangle = \sum_{i_1=1}^{n} D_{i_1} f(\varphi(t)) h_{i_1} = \sum_{i_1=1}^{n} D_{i_1} f \circ \varphi(t) h_{i_1}.$$

Si suponemos ahora que la fórmula que estamos queriendo demostrar es cierta para un $r = 1, 2...$, veamos como también lo es para $r+1$:

$$\begin{aligned}
g^{r+1)}(t) &= (g^{r)})'(t) \\
&= \sum_{i_1=1,\ldots,i_r=1}^{n} (D_{i_1 \ldots i_r} f \circ \varphi)'(t) h_{i_1} \cdots h_{i_r} \\
&= \sum_{i_1=1,\ldots,i_r=1}^{n} \langle \nabla D_{i_1 \ldots i_r} f(\varphi(t)), h \rangle h_{i_1} \cdots h_{i_r} \\
&= \sum_{i_1=1,\ldots,i_r=1}^{n} \left(\sum_{i_{r+1}=1}^{n} D_{i_{r+1}} (D_{i_1 \ldots i_r} f)(\varphi(t)) h_{i_{r+1}} \right) h_{i_1} \cdots h_{i_r} \\
&= \sum_{i_1=1,\ldots,i_r,i_{r+1}=1}^{n} D_{i_1 \ldots i_r i_{r+1}} f(\varphi(t)) h_{i_1} \cdots h_{i_r} h_{i_{r+1}}.
\end{aligned}$$

Apliquemos ahora el teorema de Taylor para funciones reales de una variable real a la función $g : B \subset \mathbb{R} \to \mathbb{R}$. El intervalo $[0,1]$ está contenido en B (pues $L[x_0, x_0 + h] \subset A$), luego existe $\alpha \in (0,1)$ tal que

$$g(1) = g(0) + g'(0) + \frac{1}{2!} g^{2)}(0) + \cdots + \frac{1}{k!} g^{k)}(0) + \frac{1}{(k+1)!} g^{k+1)}(\alpha).$$

Si hacemos $c = x_0 + \alpha h$ y sustituimos los valores de $g^{j)}(0)$ que hemos calculado antes tenemos el resultado. $\square$

2.7. Extremos locales

Recordamos del curso de Análisis de Variable Real que una función $f:(a,b)\to\mathbb{R}$ tiene un máximo relativo (o local) en un punto $x_0\in(a,b)$ si existe $r>0$ tal que $(x_0-r,x_0+r)\subset(a,b)$ y

$$f(x)\le f(x_0) \text{ para todo } x\in(x_0-r,x_0+r).$$

Analogamente $f:(a,b)\to\mathbb{R}$ tiene un mínimo relativo (o local) en un punto $x_0\in(a,b)$ si existe $r>0$ tal que $(x_0-r,x_0+r)\subset(a,b)$ y

$$f(x)\ge f(x_0) \text{ para todo } x\in(x_0-r,x_0+r).$$

Recordamos también que si f es derivable en x_0 y tiene un extremo (máximo o mínimo) relativo en ese punto, entonces $f'(x_0)=0$. Veamos lo análogo en varias variables.

Definición 2.7.1 *Sean $f:A\subset\mathbb{R}^n\to\mathbb{R}$ y $x_0\in A$ (A abierto). Diremos que f tiene un **máximo relativo** en x_0 si existe $r>0$ tal que $B(x_0,r)\subset A$ y $f(x)\le f(x_0)$ para todo $x\in B(x_0,r)$. Si existe una bola centrada en x_0 en la que $f(x)\ge f(x_0)$ diremos que f tiene un **mínimo relativo** en x_0. Cuando se da alguna de esas desigualdades para todo $x\in A$ se habla de **máximo o mínimo absolutos**. Finalmente si f tiene todas las derivadas parciales de primer orden en x_0, diremos que x_0 es un **punto crítico** de f si* $\nabla f(x_0)=0$.

Teorema 2.7.2 *Si $f:A\subset\mathbb{R}^n\to\mathbb{R}$ (A abierto), tiene un extremo relativo (máximo o mínimo) en $x_0\in A$ y f tiene todas las parciales en ese punto, entonces x_0 es un punto crítico.*

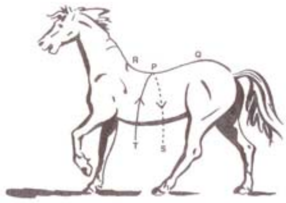
Demostración. Supongamos que f tiene un máximo relativo en x_0 y fijemos $i=1,...,n$. Entonces, como el numerador del cociente

$$\frac{f(x_0+te_i)-f(x_0)}{t}$$

es negativo (para t próximos a 0) su signo es opuesto al de t. Dado que existe el límite de ese cociente, necesariamente tiene que ser 0. Para el caso de mínimo relativo el argumento es análogo. □

Observamos que ni siquiera en el caso de funciones de una variable real tiene por qué verificarse que la anulación de la derivada en un punto implique la existencia de un extremo en ese punto. Por ejemplo $f(x)=x^3$ tiene derivada nula en el 0 y sin embargo no tiene un extremo en ese punto.

Los puntos críticos en los que no hay extremos se llaman **puntos de ensilladura** (o de silla), en referencia a las sillas de montar a caballo.



Una condición suficiente que, para funciones de una variable, garantiza la existencia de máximos o mínimos locales en un punto x_0; a saber, que la función sea de clase C^2 en un entorno de x_0, que $f'(x_0) = 0$ y que $f''(x_0)$ sea mayor que 0 para mínimo y menor que 0 para máximo, se podrá generalizar a varias variables como veremos enseguida.

Sean f una función de clase C^2 en un abierto A y sea $x_0 \in A$. Llamaremos **matriz hessiana**[2] de f en x_0 a la matriz

$$H_f(x_0) = \begin{pmatrix} D_{11}f(x_0) & D_{12}f(x_0) & \cdots & D_{1n}f(x_0) \\ D_{12}f(x_0) & D_{22}f(x_0) & \cdots & D_{2n}f(x_0) \\ \vdots & \vdots & \vdots & \vdots \\ D_{1n}f(x_0) & D_{2n}f(x_0) & \cdots & D_{nn}f(x_0) \end{pmatrix}$$

Obsérvese que ésta es una matriz simétrica y que, por el teorema de Schwarz, coincide con la matriz

$$\begin{pmatrix} D_{11}f(x_0) & D_{12}f(x_0) & \cdots & D_{1n}f(x_0) \\ D_{21}f(x_0) & D_{22}f(x_0) & \cdots & D_{2n}f(x_0) \\ \vdots & \vdots & \vdots & \vdots \\ D_{n1}f(x_0) & D_{n2}f(x_0) & \cdots & D_{nn}f(x_0) \end{pmatrix}$$

La matriz hessiana da lugar a la forma cuadrática:

$$Q(h) = \begin{pmatrix} h_1 & \cdots & h_n \end{pmatrix} \begin{pmatrix} D_{11}f(x_0) & D_{12}f(x_0) & \cdots & D_{1n}f(x_0) \\ D_{12}f(x_0) & D_{22}f(x_0) & \cdots & D_{2n}f(x_0) \\ \vdots & \vdots & \vdots & \vdots \\ D_{1n}f(x_0) & D_{2n}f(x_0) & \cdots & D_{nn}f(x_0) \end{pmatrix} \begin{pmatrix} h_1 \\ \\ \vdots \\ \\ h_n \end{pmatrix},$$

esto es:

$$Q(h) = \sum_{i,j=1}^{n} D_{ij}f(x_0)h_ih_j; \; h \in \mathbb{R}^n.$$

En las asignaturas de álgebra se usa la notación $Q(h) = hH_f(x_0)h^t$.

[2]El nombre proviene del matemático alemán Ludwig Otto Hess, 1811-1874.

Definición 2.7.3 *Se dice que una forma cuadrática $Q : \mathbb{R}^n \to \mathbb{R}$ es* ***definida positiva*** *si $Q(h) > 0$ para todo $h \in \mathbb{R}^n$, $h \neq 0$. Cuando $Q(h) \geq 0$ para todo $h \in \mathbb{R}^n$ se dice que Q es* ***semidefinida positiva****. De manera obvia se definen los conceptos de* ***definida negativa*** *y* ***semidefinida negativa****. Cuando existen h y h' tales que $Q(h) < 0$ y $Q(h') > 0$ se dice que Q es* ***indefinida****.*

Nota 2.7.4 *En el caso de $\mathbb{R}^1$, $h \mapsto f''(x_0)h^2$ es una forma cuadrática que es definida positiva si (y sólo si) $f''(x_0) > 0$ y definida negativa si (y sólo si) $f''(x_0) < 0$. Para varias variables se tiene el siguiente resultado:*

Teorema 2.7.5 *Sea f una función de clase $\mathcal{C}^2$ en un abierto de $\mathbb{R}^n$ y supongamos que un punto x_0 de él es un punto crítico para f.*

a) Si la forma cuadrática Q asociada a $H_f(x_0)$ es definida negativa, entonces f tiene en x_0 un máximo relativo.

b) Si f tiene un máximo relativo en x_0, entonces Q es semidefinida negativa.

c) Si Q es definida positiva, entonces f tiene en x_0 un mínimo relativo.

d) Si f tiene un mínimo relativo en x_0, entonces Q es semidefinida positiva.

e) Si Q es indefinida entonces la función tiene un punto de ensilladura en x_0.

Demostración. Haremos solamente las demostraciones de (a) y (b). Las de (c) y (d) se obtienen aplicando (a) y (b) respectivamente a $-f$, y la de (e) es consecuencia directa de (b) y (d).

a) Como el conjunto $K = \{x \in \mathbb{R}^n : \|x\| = 1\}$ es un compacto de $\mathbb{R}^n$, existe un punto $\widetilde{x} \in K$ tal que

$$Q(\widetilde{x}) \geq Q(x) \text{ para todo } x \in K$$

(esto es debido, a que como sabemos, las aplicaciones continuas en un compacto alcanzan en él el máximo). Por otra parte, al ser Q definida negativa, $Q(\widetilde{x}) < 0$. Sea $\varepsilon_0 = -Q(\widetilde{x})$. La continuidad de las derivadas parciales de segundo orden de f en x_0 nos da que existe un $\delta > 0$ tal que $B(x_0, \delta) \subset A$ y

$$||x - x_0|| < \delta \Rightarrow |D_{ij}f(x) - D_{ij}f(x_0)| < \frac{\varepsilon_0}{2n^2} \text{ para todo } i, j = 1, ..., n.$$

Por otra parte para $h \in \mathbb{R}^n$, $h \neq 0$, se verifica que

$$Q(h) = \|h\|^2 Q\left(\frac{h}{\|h\|}\right) \leq \|h\|^2 Q(\widetilde{x}) = -\varepsilon_0 \|h\|^2 .$$

Aplicamos ahora el teorema de Taylor a f, x_0 y $x_0 + h$, con $0 < ||h|| < \delta$ y $k = 1$. Obtenemos un $c \in L[x_0, x_0 + h]$, $c \neq x_0, x_0 + h$, tal que

$$f(x_0 + h) - f(x_0) = \frac{1}{2} \sum_{i,j=1}^{n} D_{ij}f(c)h_i h_j$$

(nótese que $\nabla f(x_0) = 0$).

Ahora
$$\sum_{i,j=1}^{n} D_{ij}f(c)h_ih_j = \sum_{i,j=1}^{n} (D_{ij}f(c) - D_{ij}f(x_0))h_ih_j + Q(h)$$

Para $0 < ||h|| < \delta$ se tiene que

$$\sum_{i,j=1}^{n} (D_{ij}f(c) - D_{ij}f(x_0))h_ih_j < \frac{\varepsilon_0}{2}||h||^2.$$

Como por otra parte, si $h \neq 0$,
$$Q(h) \leq -\varepsilon_0||h||^2,$$

tenemos que:
$$f(x_0 + h) - f(x_0) \leq \frac{1}{2}(-\varepsilon_0 + \frac{\varepsilon_0}{2})\,\|h\|^2 < 0,$$

por lo que $f(x) < f(x_0)$ para todo $x \in B(x_0, \delta)$, $x \neq x_0$, (obsérvese que todo $x \in B(x_0, \delta)$, $x \neq x_0$ es de la forma $x_0 + h$ con $0 < ||h|| < \delta$), lo que nos da que f tiene un máximo local en x_0.

b) Si existe h_0 tal que $Q(h_0) > 0$ consideremos $g(t) = -f(x_0+th_0) = -f(\varphi(t))$, siendo $\varphi(t) = x_0 + th_0$. Se verifica que g es de clase $\mathcal{C}^2$ en un intervalo centrado en 0,

$$g'(t) = -(f \circ \varphi)'(t) = -\sum_{i=1}^{n} D_if(x_0 + th_0)h_{0i} = -\sum_{i=1}^{n} (D_if \circ \varphi)(t)h_{0i}.$$

y

$$g''(t) = -\sum_{i=1}^{n} (D_if \circ \varphi))'(t)h_{0i} = -\sum_{i=1}^{n} \left(\sum_{j=1}^{n} D_j(D_if)(\varphi(t))h_{0j} \right) h_{0i}$$

por lo que $g'(0) = 0$ y $g''(0) = -Q(h_0) < 0$. El argumento que acabamos de usar para probar (a) aplicado a la forma cuadrática definida negativa, $Q_g(t) = g''(0)t^2$, nos da que $g(t) < g(0)$ para los t de un cierto intervalo $(-\delta, \delta)$, $t \neq 0$, esto es,

$$-f(x_0 + th_0) < -f(x_0) \quad \forall t \in (-\delta, \delta),\ t \neq 0$$

o, equivalentemente,

$$f(x_0 + th_0) > f(x_0) \quad \forall t \in (-\delta, \delta) \quad t \neq 0$$

por lo que f no tendría un máximo local en x_0. □

Nos centramos ahora en el caso $n = 2$. En ese caso la matriz hessiana de una función f, de clase $\mathcal{C}^2$ en un abierto A de $\mathbb{R}^2$ en (x_0, y_0), es

$$H_f(x_0,y_0) = \begin{pmatrix} D_{11}f(x_0,y_0) & D_{12}f(x_0,y_0) \\ D_{12}f(x_0,y_0) & D_{22}f(x_0,y_0) \end{pmatrix}$$

y puede escribirse en la forma

$$H_f(x_0,y_0) = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

siendo $a = D_{11}f(x_0,y_0),\ b = D_{12}f(x_0,y_0)$ y $c = D_{22}f(x_0,y_0)$.

Teorema 2.7.6 *Sean f una función de clase $\mathcal{C}^2$ en un abierto A de $\mathbb{R}^2$, $(x_0,y_0) \in A$ un punto crítico de f y sea $\Delta := \det(H_f(x_0,y_0)) = ac - b^2$.*

1) Si $\Delta > 0$ y $a > 0$, entonces f tiene un mínimo relativo en (x_0,y_0)

2) Si $\Delta > 0$ y $a < 0$, entonces f tiene un máximo relativo en (x_0,y_0)

(nótese que si $\Delta > 0$, entonces $a \neq 0$)

3) Si $\Delta < 0$, entonces f tiene un punto de silla en (x_0,y_0) independiente de como sea a

4) Si $\Delta = 0$, entonces no se sabe nada en relación con un posible extremo de f en (x_0,y_0).

Demostración. Para la forma cuadrática Q asociada a $H_f(x_0,y_0)$ se verifica que

$$Q(h_1,h_2) = ah_1^2 + bh_1h_2 + bh_2h_1 + h_2^2c =$$

$$ah_1^2 + 2bh_1h_2 + ch_2^2.$$

Si $\Delta > 0$ entonces $a \neq 0$ (pues $\Delta = ac - b^2$), por lo que

$$ah_1^2 + 2bh_1h_2 + ch_2^2 = a\left(h_1 + \frac{b}{a}h_2\right)^2 + \frac{\Delta}{a}h_2^2.$$

1) Si $a > 0$ ese valor es mayor que 0 cualesquiera que sean h_1 y h_2, $(h_1,h_2) \neq (0,0)$, lo que implica que f tiene un mínimo relativo en (x_0,y_0).

2) Si $\Delta > 0$ y $a < 0$, $Q(h_1,h_2) < 0$ cualesquiera que sean h_1 y h_2, $(h_1,h_2) \neq (0,0)$ lo que implica que f tenga un máximo relativo en (x_0,y_0).

3) Si $\Delta < 0$ y $a = 0$ entonces $Q(h_1,h_2) = 2bh_1h_2 + ch_2^2$ y necesariamente $b \neq 0$ (pues si $b = a = 0$, necesariamente $\Delta = 0$). Si $c = 0$ entonces $Q(1,1) = 2b$ y $Q(-1,1) = -2b$. Si $c \neq 0$, entonces $Q(-\frac{c}{b},1) = -c$ y $Q(-\frac{c}{4b},1) = \frac{c}{2}$ y Q toma valores de signos opuestos en puntos distintos, por lo que es indefinida.

Si $\Delta < 0$ y $a \neq 0$, entonces $Q(1,0) = a$ y $Q(-\frac{b}{a},1) = \frac{\Delta}{a}$ que de nuevo muestran que Q es indefinida pues toma valores de signos opuestos en dos puntos distintos.

4) Finalmente, si $\Delta = 0$ no podemos decir nada sobre el posible extremo que f pueda tener en (x_0,y_0). Veamos un ejemplo de dos funciones para las que $\Delta = 0$ y una tiene un extremo y la otra no.

Sean $f(x,y) = x^2 - 2xy + y^2 + x^4 + y^4$ y $g(x,y) = x^2 - 2xy + y^2 - x^4 - y^4$. En ambos casos la matriz hessiana en $(0,0)$ es:

$$\begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix}$$

y su determinante es 0. Sin embargo, como $f(0,0) = 0$ y $f(x,y) = (x-y)^2 + x^4 + y^4 \geq 0$ para todo (x,y), resulta que f tiene un mínimo, incluso absoluto, en $(0,0)$.

Por otra parte $g(x,x) = -2x^4 < 0$ para todo $x \in \mathbb{R},\ x \neq 0$ y $g(x,-x) = 2x^2(2-x^2) > 0$ para $x \in \mathbb{R}$ pequeños, $x \neq 0$. Por lo tanto $(0,0)$ es un punto de ensilladura para g. $\square$

Nota 2.7.7 *Las funciones de más de una variable puede tener un único punto crítico, y éste puede ser un extremo relativo pero no absoluto. Esto le pasa, por ejemplo, a la función*

$$f(x,y) = y^2 + x^2(1+y)^3.$$

Sucede que tiene un único punto crítico en $(0,0)$*, en él tiene un mínimo relativo, pero este mínimo no es absoluto pues en el punto* $(0,0)$ *vale* 0*, mientras que en el punto* $(1,-4)$ *vale* -9*.*

Hemos demostrado el teorema anterior para el caso particular en el que $n = 2$. Puede obtenerse un resultado análogo para n arbitrario. El resultado es el llamado **criterio de Sylvester**:

Si f es una función de clase $\mathcal{C}^2$ en un abierto A de $\mathbb{R}^n$, $x_0 \in A$ es un punto crítico de f y si denotamos por Δ_k al determinante de la matriz cuadrada de orden k superior izquierda de la matriz hessiana de f en x_0 (menor principal), entonces:

a) Si $\Delta_k > 0$ para todo $k = 1, ..., n$, f tiene un mínimo relativo en x_0.
b) Si $(-1)^k \Delta_k > 0$ para todo $k = 1, ..., n$, f tiene un máximo relativo en x_0.
c) Si $\Delta_n \neq 0$ y no estamos ni en el caso (a) ni en el (b), entonces no hay extremo.
d) Si $\Delta_n = 0$ no se sabe nada.

Nótese que las condiciones de (a) y (b) caracterizan la definición (positiva o negativa respectivamente) de la forma cuadrática asociada a la matriz hessiana de f en x_0.

Nota 2.7.8 *Pueden también usarse los autovalores de la matriz hessiana para obtener información sobre los posibles extremos. Así, si todos los autovalores de ella son positivos hay un mínimo relativo y si son todos negativos hay un máximo relativo. En los demás casos no se sabe nada. La razón para ello es la siguiente:*

Toda matriz simétrica M *se puede reducir a una matriz diagonal* N *formada por los autovalores de* M*, esto es, por las soluciones de la ecuación* $det(M - \lambda I) = 0$*, en el sentido* *de que* $M = P^t N P$ *siendo* P *una matriz no singular (con determinante no nulo). Véase por ejemplo el libro "Álgebra lineal" de Fernando, Gamboa y Ruiz, Proposición 19.13.*

Ahora bien, la forma cuadrática asociada a M *es definida positiva si lo es la asociada a* N*. En efecto, si* $n = 2$ *para simplificar la escritura,*

$$\begin{aligned}
& (h_1\ h_2)\begin{pmatrix} m_{11} & m_{12} \\ m_{12} & m_{22} \end{pmatrix}\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \\
= \;& (h_1\ h_2)\begin{pmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{pmatrix}\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}\begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \\
= \;& (p_{11}h_1 + p_{12}h_2, p_{21}h_1 + p_{22}h_2)\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}\begin{pmatrix} p_{11}h_1 + p_{12}h_2 \\ p_{21}h_1 + p_{22}h_2 \end{pmatrix}.
\end{aligned}$$

Si $(h_1\ h_2) \neq (0,0)$*, entonces* $(p_{11}h_1 + p_{12}h_2, p_{21}h_1 + p_{22}h_2) \neq (0,0)$ *pues* P *es no singular y en consecuencia el sistema homogéneo*

$$\begin{aligned}
p_{21}h_1 + p_{22}h_2 = 0 \\
p_{21}h_1 + p_{22}h_2 = 0
\end{aligned}$$

sólo tiene la solución trivial. Luego si $(h,k) \neq (0,0)$ *y la forma cuadrática asociada a* $\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$ *es definida positiva se verifica que*

$$(h_1\ h_2)\begin{pmatrix} m_{11} & m_{12} \\ m_{12} & m_{22} \end{pmatrix}\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} > 0.$$

Finalmente observamos que es fácil ver si la forma cuadrática asociada a una matriz diagonal N *es definida positiva, pues,*

$$(h^*, k^*)\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}\begin{pmatrix} h^* \\ k^* \end{pmatrix} = \lambda_1 {h^*}^2 + \lambda_2 {k^*}^2,$$

y este valor claramente es positivo si todos los λ_j *son positivos y negativo si todos los* λ_j *lo son. Si los hay positivos y negativos es fácil encontrar pares* (h^*, k^*) *en los que da positivo y otros en los que da negativo. Por ejemplo, si* $\lambda_1 > 0$ *y* $\lambda_2 < 0$*, entonces* $\lambda_1 1^2 + \lambda_2 0^2 > 0$ *y* $\lambda_1 0^2 + \lambda_2 1^2 < 0$*.*

Capítulo 3

Teoremas de la función inversa e implícita

3.1. Teorema de la función inversa

Se sabe de los cursos de álgebra que dados $y_1, ..., y_n \in \mathbb{R}$, el sistema de ecuaciones lineales

$$\left.\begin{array}{c} a_{11}x_1 + \cdots + a_{1n}x_n = y_1 \\ \cdots \\ a_{n1}x_1 + \cdots + a_{nn}x_n = y_n \end{array}\right\}$$

tiene una (única) solución $(x_1, ..., x_n)$ si la matriz de sus coeficientes,

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ & \cdots & \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$$

tiene determinante no nulo.

Nosotros aquí estaremos interesados en obtener un resultado relativo a la existencia de solución de un sistema no necesariamente lineal

$$\left.\begin{array}{c} f_1(x_1, ..., x_n) = y_1 \\ \cdots \\ f_n(x_1, ..., x_n) = y_n \end{array}\right\}$$

Por ejemplo,

$$\left.\begin{array}{c} x_1^2 + x_2 = y_1 \\ x_1 + x_2^3 = y_2 \end{array}\right\}$$

Obsérvese que si $f_1, ..., f_n$ son lineales se trata de un sistema lineal como los que se consideran en álgebra.

El papel de la matriz de los coeficientes en un sistema lineal lo desempeñará aquí la matriz jacobiana de la aplicación $f = (f_1, ..., f_n)$ que define al sistema.

Supongamos por un momento que $n = 1$ y sea $f : A \subset \mathbb{R} \to \mathbb{R}$ una función derivable en un cierto punto $x_0 \in A$ para el que se verifica que $f'(x_0) = 0$. Entonces f puede ser invertible o no en un entorno de x_0. Piénsese en lo que les pasa a las funciones $f(x) = x^3$ y $g(x) = x^2$ en el 0. Cuando f es de clase $\mathcal{C}^1$ en A y $f'(x_0) \neq 0$ se puede garantizar que f tiene una inversa local. Este hecho, que ya es conocido (para funciones de una variable), se generalizará aquí al caso de varias variables mediante un teorema que daremos un poco más adelante (teorema de la función inversa). Obtenemos previamente dos proposiciones que necesitaremos para probarlo.

Proposición 3.1.1 *Sea $f : \overline{B}(x_0, r) \subset \mathbb{R}^n \to \mathbb{R}^n$ una aplicación continua en $\overline{B}(x_0, r)$ y diferenciable en $B(x_0, r)$. Si $\det(J_f(x)) \neq 0$ para todo $x \in B(x_0, r)$ y $f(x) \neq f(x_0)$ para todo x tal que $\|x - x_0\| = r$, entonces $f(x_0)$ es un punto interior de $f(B(x_0, r))$.*

Demostración. Consideremos la función

$$\varphi : x \in \overline{B}(x_0, r) \subset \mathbb{R}^n \mapsto \varphi(x) = \|f(x) - f(x_0)\| \in \mathbb{R}.$$

φ es continua en $\overline{B}(x_0, r)$. Como $\{x : \|x - x_0\| = r\}$ es un compacto, existe un punto $x^* \in \{x : \|x - x_0\| = r\}$ tal que

$$\varphi(x^*) \leq \varphi(x) \text{ para todo } x \text{ tal que } \|x - x_0\| = r.$$

Como $f(x) \neq f(x_0)$ para todo x tal que $\|x - x_0\| = r$ resulta que $m := \varphi(x^*) > 0$.

Vamos a comprobar que

$$B(f(x_0), \tfrac{m}{2}) \subset f(B(x_0, r))$$

lo que nos da el resultado. Fijado $y \in B(f(x_0), \frac{m}{2})$ consideremos la función auxiliar

$$\psi : x \in \overline{B}(x_0, r) \subset \mathbb{R}^n \mapsto \psi(x) = \|f(x) - y\| \in \mathbb{R}.$$

Esta función ψ es continua en el compacto $\overline{B}(x_0, r)$ por lo que existe $x^{**} \in \overline{B}(x_0, r)$ tal que $\psi(x^{**}) \leq \psi(x)$ para todo $x \in \overline{B}(x_0, r)$. Comprobaremos primero que $x^{**} \in B(x_0, r)$ y después que $f(x^{**}) = y$; esto nos da que $y \in f(B(x_0, r))$.

Para ver que $x^{**} \in B(x_0, r)$ observamos que como $y \in B(f(x_0), \frac{m}{2})$ resulta que:

$$\frac{m}{2} > \|f(x_0) - y\| =: \psi(x_0) \geq \psi(x^{**}).$$

Por otra parte, si $\|x - x_0\| = r$,

$$\psi(x) = \|f(x) - y\| \geq \|f(x) - f(x_0)\| - \|f(x_0) - y\| > m - \frac{m}{2} = \frac{m}{2}.$$

Esto es,

$$\psi(x) > \frac{m}{2} > \psi(x^{**}) \text{ para todo } x \text{ tal que } \|x - x_0\| = r,$$

lo que implica que $x^{**} \in B(x_0, r)$, pues si fuese $\|x^{**} - x_0\| = r$, entonces obtendríamos que $\psi(x^{**}) > \psi(x^{**})$.

Veamos ahora como $f(x^{**}) = y$.

Al ser f diferenciable en $B(x_0, r)$ también lo es la función

$$\psi^2(x) = (f_1(x) - y_1)^2 + \cdots + (f_n(x) - y_n)^2.$$

Como ψ^2 tiene un extremo (mínimo) en $x^{**} \in B(x_0, r)$, pues lo tiene ψ y ésta es positiva, sabemos que $\nabla\psi^2(x^{**}) = 0$, y como

$$D_i\psi^2(x) = 2(f_1(x) - y_1)D_i f_1(x) + \cdots + 2(f_n(x) - y_n)D_i f_n(x) \text{ para } i = 1, ..., n.$$

resulta que

$$\left.\begin{array}{ll} (f_1(x^{**}) - y_1)D_1 f_1(x^{**}) + \cdots + (f_n(x^{**}) - y_n)D_1 f_n(x^{**}) & = 0 \\ \cdots & \\ (f_1(x^{**}) - y_1)D_n f_1(x^{**}) + \cdots + (f_n(x^{**}) - y_n)D_n f_n(x^{**}) & = 0 \end{array}\right\}$$

Como por hipótesis el $det(D_i f_j(x^{**})) \neq 0$ pues $\det(J_f(x^{**})) \neq 0$ (la matriz del sistema es la traspuesta de esa jacobiana), resulta que el anterior sistema lineal tiene solución única y ésta es necesariamente la $(0, ..., 0)$. Luego

$$f_1(x^{**}) = y_1, ..., \; f_n(x^{**}) = y_n.$$

Esto es, $f(x^{**}) = y$. □

Corolario 3.1.2 *Sea $f : A \subset \mathbb{R}^n \to \mathbb{R}^n$ una aplicación de $\mathcal{C}^1$ en el abierto A y supongamos que f es inyectiva en A y que $\det(J_f(x)) \neq 0$ para todo punto $x \in A$. Entonces f es una aplicación abierta.*

Demostración. Sea $B \subset A$ un conjunto abierto y sea $f(x) \in f(B)$. Por ser B un conjunto abierto existe $r > 0$ tal que $\overline{B}(x, r) \subset B$. Por la proposición anterior $f(x)$ es un punto interior de $f(B(x, r))$ y por lo tanto de $f(B)$. □

Proposición 3.1.3 *Sea $f : A \subset \mathbb{R}^n \to \mathbb{R}^n$ una aplicación de clase $\mathcal{C}^1$ en el abierto A y supongamos que para un punto $x_0 \in A$ se verifica que $\det(J_f(x_0)) \neq 0$. Entonces existe $r > 0$ tal que $B(x_0, r) \subset A$, f es inyectiva en $B(x_0, r)$ y $\det(J_f(x)) \neq 0$ para todo $x \in B(x_0, r)$.*

Demostración. Consideremos la aplicación

$$\phi : A \times \overset{n}{\cdots} \times A \subset \mathbb{R}^n \times \overset{n}{\cdots} \times \mathbb{R}^n \to \mathbb{R}$$

definida por

$$\phi(a_1, ..., a_n) = \det \begin{pmatrix} D_1 f_1(a_1) & \cdots & D_n f_1(a_1) \\ \cdots & \cdots & \cdots \\ D_1 f_n(a_n) & \cdots & D_n f_n(a_n) \end{pmatrix}$$

(nótese que cada a_i es un vector de $\mathbb{R}^n$). Se usará la notación $det(D_i f_j(a_j))$ para denotar este determinante. Por hipótesis $\phi(x_0, ..., x_0) \neq 0$ pues $\det(J_f(x_0)) \neq 0$. Dado que f es de clase $\mathcal{C}^1$ las $D_i f_j$ son continuas y, en consecuencia, ϕ es continua, luego existe un $r > 0$ tal que $B(x_0, r) \subset A$ y $\phi(a_1, ..., a_n) \neq 0$ si $a_i \in B(x_0, r)$ para $i = 1, ..., n$. Veamos que f es inyectiva en $B(x_0, r)$.

Sean $x, y \in B(x_0, r)$. Si suponemos que $x \neq y$ y que $f(x) = f(y)$, entonces, aplicando el teorema del valor medio a cada f_j, obtenemos un $c_j \in L[x, y] \subset B(x_0, r)$ tal que

$$0 = f_j(x) - f_j(y) = D_1 f_j(c_j)(x_1 - y_1) + \cdots + D_n f_j(c_j)(x_n - y_n).$$

Al ser $\phi(c_1, ..., c_n) \neq 0$ (pues $c_j \in B(x_0, r)$), el sistema

$$\left.\begin{array}{rl} D_1 f_1(c_1)(x_1 - y_1) + \cdots + D_n f_1(c_1)(x_n - y_n) & = 0 \\ \cdots & \\ D_1 f_n(c_n)(x_1 - y_1) + \cdots + D_n f_n(c_n)(x_n - y_n) & = 0 \end{array}\right\}$$

tiene como (única) solución la $(0, ..., 0)$. Esto es, $x_i = y_i$ para todo $i = 1., ..., n$ en contra de lo que habíamos supuesto, luego necesariamente $f(x) \neq f(y)$. $\square$

Teorema 3.1.4 (de la función inversa) *Sean A un abierto de $\mathbb{R}^n$, $f : A \to \mathbb{R}^n$, $f \in \mathcal{C}^1(A, \mathbb{R}^n)$ y supongamos que para un punto $x_0 \in A$ se verifica que $\det(J_f(x_0)) \neq 0$. Entonces existen entornos abiertos U de x_0 y V de $f(x_0)$ y una única función $g : V \subset \mathbb{R}^n \to U \subset \mathbb{R}^n$ verificando:*

a) $f(U) = V$

b) f *es inyectiva en* U

c) $g \circ f = Id$ *en* U

d) $g \in \mathcal{C}^1(V, \mathbb{R}^n)$.

Demostración. En la demostración de la proposición anterior hemos visto, que con estas hipótesis, que existe un $r > 0$ tal que $det(D_i f_j(a_j)) \neq 0$ para todo $a_j \in B(x_0, r)$ y que f es inyectiva en $B(x_0, r)$. Además por el Corolario anterior a esa proposición se tiene que $f(B(x_0, r)$ es un conjunto abierto por lo que existe $s > 0$ tal que

$$B(f(x_0), s) \subset f(B(x_0, r)).$$

Definamos $V = B(f(x_0), s)$ y $U = f^{-1}(V) \cap B(x_0, r)$. Claramente U y V son abiertos (véase la Nota 1.4.15) $x_0 \in U$ y $f(x_0) \in V$. Veamos la verificación de (a), (b), (c) y (d).

(a) $f(U) = V$.

"$\subset$" Si $x \in U$ entonces $f(x) \in V$ pues $U = f^{-1}(V) \cap B(x_0, r)$.

"$\supset$" Si $y \in V = B(f(x_0), s) \subset f(B(x_0, r))$ entonces existe $x \in B(x_0, r)$ tal que $y = f(x)$. Luego $y \in f(U)$ (nótese que $x \in f^{-1}(V) \cap B(x_0, r)$).

(b) f es inyectiva en U pues lo es en $B(x_0, r)$.

(c) Todo $y \in V$ es de la forma $f(x)$ para un único $x \in U$ (inyectividad), se define entonces

$$g : y \in V \mapsto g(y) = x.$$

Es evidente que esta g así definida verifica que $g \circ f = Id$ en U. Es también claro que esta g es única, pues si $h : V \to \mathbb{R}^n$ verifica que $h \circ f = Id$ en U, entonces $h(y) = h(f(x)) = x = g(y)$ para todo $y \in V$. Nótese que también $f \circ g = Id$ en V (se usa más adelante) pues para todo $y = f(x) \in V$, $f \circ g(y) = f \circ g(f(x)) = f(x) = y$.

(d) $g \in \mathcal{C}^1(V, \mathbb{R}^n)$. Empezaremos viendo que g es continua, cosa que usaremos en lo que sigue. Para ello consideramos un abierto B de $\mathbb{R}^n$ que supondremos contenido en U (obsérvese que $g(V) = U$) . Sucede que $g^{-1}(B) = f(B)$ que es un abierto como consecuencia del corolario anterior.

Vamos a ver ahora como las componentes g_i de g tienen todas las parciales de primer orden y éstas son continuas en V. Para el caso en el que $n = 1$ el resultado es conocido de primer curso. Veremos ahora la existencia de $D_2 g_1$. Los demás casos son análogos.

Sea $y_0 \in V = f(U)$, entonces $g(y_0) \in U$ y así existe $\varepsilon_0 > 0$ verificando que $B(g(y_0), \varepsilon_0) \subset U$. Por ser g continua en y_0 existe δ_0 tal que $g(B(y_0, \delta_0)) \subset B(g(y_0), \varepsilon_0)$. Veamos como existe el

$$\lim_{t \to 0} \frac{g_1(y_0 + te_2) - g_1(y_0)}{t}.$$

Dado que $f(g(y_0)) = y_0$ y $f(g(y_0 + te_2)) = y_0 + te_2$ resulta que:

$$\left.\begin{array}{lll} f_1(g(y_0 + te_2)) - f_1(g(y_0)) & = y_{0_1} - y_{0_1} & = 0 \\ f_2(g(y_0 + te_2)) - f_2(g(y_0)) & = y_{0_2} + t - y_{0_2} & = t \\ f_3(g(y_0 + te_2)) - f_3(g(y_0)) & = y_{0_3} - y_{0_3} & = 0 \\ \cdots & = \cdots & = 0 \\ f_n(g(y_0 + te_2)) - f_n(g(y_0)) & = y_{0_n} - y_{0_n} & = 0 \end{array}\right\}.$$

Para todo t, $|t| < \delta_0$, $L[g(y_0 + te_2), g(y_0)]$ está contenido en $U \subset B(x_0, r)$ pues $y_0 + te_2 \in B(y_0, \delta_0)$ y entonces $g(y_0 + te_2)$ y $g(y_0)$ pertenecen a la bola $B(g(y_0), \varepsilon_0)$ que está contenida en U. Aplicando el teorema del valor medio a f_1 en ese segmento, existe $c_{t,1} \in L[g(y_0 + te_2), g(y_0)]$ tal que

$$\begin{aligned} 0 &= \frac{f_1(g(y_0 + te_2)) - f_1(g(y_0))}{t} \\ &= D_1 f_1(c_{t,1}) \frac{g_1(y_0 + te_2) - g_1(y_0)}{t} + \cdots + D_n f_1(c_{t,1}) \frac{g_n(y_0 + te_2) - g_n(y_0)}{t}. \end{aligned}$$

Análogamente existe $c_{t,2} \in L[g(y_0+te_2), g(y_0)]$ tal que

$$\begin{aligned} 1 &= \frac{f_2(g(y_0+te_2)) - f_2(g(y_0))}{t} \\ &= D_1f_2(c_{t,2})\frac{g_1(y_0+te_2)-g_1(y_0)}{t} + \cdots + D_nf_2(c_{t,2})\frac{g_n(y_0+te_2)-g_n(y_0)}{t}. \end{aligned}$$

y se sigue así hasta llegar a obtener la existencia de un $c_{t,n} \in L[g(y_0+te_2), g(y)]$ tal que

$$\begin{aligned} 0 &= \frac{f_n(g(y_0+te_2)) - f_n(g(y_0))}{t} \\ &= D_1f_n(c_{t,n})\frac{g_1(y_0+te_2)-g_1(y_0)}{t} + \cdots + D_nf_n(c_{t,n})\frac{g_n(y_0+te_2)-g_n(y_0)}{t}. \end{aligned}$$

Por la demostración de la proposición anterior tenemos que $\det(D_if_j(a_j)) \neq 0$ para $a_1, ..., a_n \in B(x_0, r)$ y entonces $\det(D_if_j(c_{t,j})) \neq 0$ pues $L[g(y_0+te_2), g(y_0)] \subset B(x_0, r)$.

Si aplicamos la Regla de Cramer a nuestro anterior sistema lineal, vemos que éste tiene solución única y que

$$\frac{g_1(y_0+te_2)-g_1(y_0)}{t} = \frac{\det\begin{pmatrix} 0 & D_2f_1(c_{t,1}) & \cdots & D_nf_1(c_{t,1}) \\ 1 & D_2f_2(c_{t,2}) & \cdots & D_nf_2(c_{t,2}) \\ & & \cdots & \\ 0 & D_2f_n(c_{t,n}) & \cdots & D_nf_n(c_{t,n}) \end{pmatrix}}{\det(D_if_j(c_{t,j})}.$$

Cuando $t \to 0$, $c_{t,j} \to g(y_0)$, y entonces, por la continuidad de las D_if_j se tiene que $D_if_j(c_{t,j}) \to D_if_j(g(y_0))$, por lo que

$$\lim_{t\to 0}\frac{g_1(y_0+te_2)-g_1(y)}{t} = \frac{\det\begin{pmatrix} 0 & D_2f_1(g(y_0)) & \cdots & D_nf_1(g(y_0)) \\ 1 & D_2f_2(g(y_0)) & \cdots & D_nf_2(g(y_0)) \\ & & \cdots & \\ 0 & D_2f_n(g(y_0)) & \cdots & D_nf_n(g(y_0)) \end{pmatrix}}{\det(D_if_j(g(y_0)))}.$$

Entonces

$$D_2g_1(y_0) = \frac{\det\begin{pmatrix} 0 & D_2f_1(g(y_0)) & \cdots & D_nf_1(g(y_0)) \\ 1 & D_2f_2(g(y_0)) & \cdots & D_nf_2(g(y_0)) \\ & & \cdots & \\ 0 & D_2f_n(g(y_0)) & \cdots & D_nf_n(g(y_0)) \end{pmatrix}}{\det(D_if_j(g(y_0)))}.$$

Es claro entonces que existe $D_2g_1(y_0)$ en cada punto y_0 de V Si consideramos puntos $y \in V$ próximos a y_0, obtenemos una fórmula similar lo que nos permite afirmar que la aplicación D_2g_1 es continua en V por ser g continua, f de clase 1 en U y ser continua la operación de hallar determinante. $\square$

Observamos que se sigue de la demostración que acabamos de hacer que g es de la misma clase que f, esto es, si suponemos que f es de clase p entonces g también es de clase p. Hacemos notar también que al ser $g \circ f = Id$ resulta que

$$dg(y) = [df(x)]^{-1}$$

(siendo $f(x) = y$). En términos de matrices jacobianas esto se escribe así:

$$Jg(y) = [J_f(x)]^{-1}.$$

Nota 3.1.5 *No se verifica, en general, el teorema de la función inversa si sólo se pide que f sea diferenciable. Para verlo considérese la función $f : \mathbb{R} \to \mathbb{R}$ definida por*

$$f(x) = \begin{cases} \frac{x}{2} + x^2 \operatorname{sen} \frac{1}{x} & \textit{si } x \neq 0 \\ 0 & \textit{si } x = 0. \end{cases}$$

Esta función es derivable en todo punto de $\mathbb{R}$, $f'(0) = \frac{1}{2} \neq 0$, y sin embargo f no admite inversa en ningún entorno de 0. En efecto, recordamos del curso de Análisis de Variable Real que para que una función f continua admita inversa en un entorno de un punto, en este caso el 0, tiene que ser estrictamente monótona en él (véase R. G. Bartle, p. 196). Esto no ocurre en este caso pues

$$f'(x) = \frac{1}{2} + 2x \operatorname{sen} \frac{1}{x} - \cos \frac{1}{x} \quad \textit{para } x \neq 0,$$

y entonces, si tomamos puntos $x_k = \frac{1}{k\pi}$, con $k \in \mathbb{N}$ bastante grande para que x_k pertenezca al entorno, resulta que $f'(x_k) = \frac{1}{2} - \cos k\pi = \frac{1}{2} - (-1)^k$ y este valor es > 0 si k es impar y < 0 si k es par. Al no ser f' de signo constante la función no es estríctamente monótona (véase R. G. Bartle, p. 222).

3.2. Teorema de la función implícita

Empezaremos con una motivación del resultado que se va a obtener:

Supongamos que tenemos la relación $x^2+y^2=1$, o equivalentemente $x^2+y^2-1=0$. Podemos considerar la función $F:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ definida por $F(x,y)=x^2+y^2-1$. Entonces los puntos $(x,y)\in\mathbb{R}^2$ que satisfacen la relación $x^2+y^2=1$ son exactamente aquellos puntos $(x,y)\in\mathbb{R}^2$ que verifican que $F(x,y)=0$. Consideremos un punto $(x_0,y_0)\in\mathbb{R}^2$ tal que $F(x_0,y_0)=0$, estamos interesados en saber si podremos despejar, por ejemplo, la variable x en función de la y cuando la y varía en un entorno de y_0. Cuando se pueda se dirá que la relación $F(x,y)=0$ determina implícitamente a x como función de y en un entorno de (x_0,y_0). Cuando se pueda despejar la y en función de la x en un entorno de (x_0,y_0) se dirá que la relación $F(x,y)=0$ determina implícitamente a y como función de x en un entorno de (x_0,y_0). Por ejemplo, tomemos el punto $(1,0)$, se verifica que $F(1,0)=0$, ¿existirán un entorno Y_0 de 0 y una función G definida en él de forma que $G(0)=1$ y $F(G(y),y)=0$ para todo $y\in Y_0$?, ¿existirá un entorno X_0 de 1 y una función H definida en él de forma que $H(1)=0$ y $F(x,H(x))=0$ para todo $x\in X_0$? Veamos qué podemos decir.

Si tomamos como Y_0 el intervalo $(-1,1)$ y como G la función definida en él por $G(y)=+\sqrt{1-y^2}$ entonces ciertamente $G(0)=1$ y $F(G(y),y)=0$ para todo $y\in Y_0$, luego, x es una función implícita de y para (x,y) en el entorno $(0,+\infty)\times\mathbb{R}$ de $(1,0)$. Obsérvese que si tomamos $G(y)=-\sqrt{1-y^2}$ entonces $G(0)=-1$. ¿Podremos también despejar y en función de x, en un entorno de $(1,0)$? Observamos que si tomamos un entorno de 1, en él hay números reales mayores que 1 y por lo tanto nunca será posible que se verifique la relación $F(x,H(x))=0$ para tales x, pues siempre $x^2+H(x)^2$ será mayor que 1.

En puntos como el $\left(\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}}\right)$ la función F determina implícitamente a cada variable en función de la otra.

El teorema de la función implícita que damos a continuación nos da una condición suficiente para la existencia de funciones definidas implícitamente en un entorno de un punto por una relación del tipo $F(x,y)=0$.

Teorema 3.2.1 (de la función implícita) *Sea $F:A\subset\mathbb{R}^n\times\mathbb{R}^k\to\mathbb{R}^n$ una aplicación de clase C^1 en A (A abierto) y supongamos que para un punto $(x_0,y_0)\in A$ se verifica que $F(x_0,y_0)=0$ y*

$$\det(D_iF_j(x_0,y_0))_{i,j=1,\ldots,n}\neq 0.$$

Entonces existen, un entorno abierto $Y_0\subset\mathbb{R}^k$ de y_0 y una única función $G:Y_0\subset\mathbb{R}^k\to\mathbb{R}^n$ verificándose que:

a) $G\in C^1(Y_0,\mathbb{R}^n)$

b) $G(y_0) = x_0$

c) $F(G(y), y) = 0$ para todo $y \in Y_0$

d) Existe un entorno abierto $U \subset A$ de (x_0, y_0) tal que

$$\{(x, y) \in U : F(x, y) = 0\} = \{(G(y), y) : y \in Y_0\}.$$

Nota 3.2.2 *Se trata de despejar las n primeras variables en función de las k últimas. También puede darse un teorema análogo para despejar las últimas variables en función de las primeras, su enunciado preciso es:*

Teorema 3.2.3 (de la función implícita) *Sea $F : A \subset \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^k$ una aplicación de clase C^1 en A y supongamos que para un punto $(x_0, y_0) \in A$ se verifica que $F(x_0, y_0) = 0$ y*

$$\det(D_i F_j(x_0, y_0))_{i=n+1,\ldots,n+k;j=1,\ldots,k} \neq 0.$$

Entonces existen, un entorno abierto $X_0 \subset \mathbb{R}^n$ de x_0 y una única función $H : X_0 \subset \mathbb{R}^n \to \mathbb{R}^k$ verificándose que:

a) $H \in C^1(X_0, \mathbb{R}^k)$

b) $H(x_0) = y_0$

c) $F(x, H(x)) = 0$ para todo $x \in X_0$.

d) Existe un entorno abierto $V \subset A$ de (x_0, y_0) tal que

$$\{(x, y) \in V : F(x, y) = 0\} = \{(x, H(x)) : x \in X_0\}.$$

Demostración. (del teorema de la función implícita, primera versión). Consideremos la función

$$f : (x, y) \in A \subset \mathbb{R}^n \times \mathbb{R}^k \mapsto f(x, y) = (F(x, y), y) \in \mathbb{R}^n \times \mathbb{R}^k$$

Claramente $f \in C^1(A, \mathbb{R}^{n+k})$ y además se verifica que:

$$\det(J_f(x_0, y_0)) =$$

$$\det \begin{pmatrix} D_1F_1(x_0,y_0) & \cdot & D_nF_1(x_0,y_0) & D_{n+1}F_1(x_0,y_0) & \cdot & D_{n+k}F_1(x_0,y_0) \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ D_1F_n(x_0,y_0) & \cdot & D_nF_n(x_0,y_0) & D_{n+1}F_n(x_0,y_0) & \cdot & D_{n+k}F_n(x_0,y_0) \\ 0 & \cdot & 0 & 1 & \cdot & 0 \\ \cdot & \cdot & \cdot & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & 0 & 0 & \cdot & 1 \end{pmatrix} =$$

$$\det\begin{pmatrix} D_1F_1(x_0,y_0) & \cdots & D_nF_1(x_0,y_0) \\ \cdots & \cdots & \cdots \\ D_1F_n(x_0,y_0) & \cdots & D_nF_n(x_0,y_0) \end{pmatrix}$$

que es distinto de 0 por hipótesis. Podemos entonces aplicar el teorema de la función inversa a la función f y obtenemos abiertos $U \subset A \subset \mathbb{R}^{n+k}$ y $V \subset \mathbb{R}^{n+k}$, siendo U un entorno de (x_0, y_0) y V un entorno de $f(x_0, y_0) = (F(x_0, y_0), y_0) = (0, y_0)$ y una función $g : V \subset \mathbb{R}^{n+k} \to U \subset \mathbb{R}^{n+k}$ tales que:

a) $f(U) = V$
b) f es inyectiva en U
c) $g \circ f = Id$ en U
d) $g \in C^1(V, \mathbb{R}^{n+k})$.

Definimos,

$$Y_0 = \{y \in \mathbb{R}^k : (0, y) \in V\}$$

y

$$G : y \in Y_0 \subset \mathbb{R}^k \mapsto G(y) = (g_1(0, y), ..., g_n(0, y)) \in \mathbb{R}^n.$$

Veamos como Y_0 y G verifican las condiciones del teorema. En primer lugar $y_0 \in Y_0$ pues $(0, y_0) = (F(x_0, y_0), y_0) = f(x_0, y_0) \in V$. El conjunto Y_0 es abierto pues es igual a $\varphi^{-1}(V)$ siendo φ la aplicación continua $\varphi : y \in \mathbb{R}^k \mapsto \varphi(y) = (0, y) \in \mathbb{R}^n \times \mathbb{R}^k$. La función G es de C^1 en Y_0 pues G está formada por las n primeras componentes de la composición de φ y g, y ambas son de clase C^1, la primera en todo $\mathbb{R}^k$, y la segunda en V.

Además,

$$G(y_0) = (g_1(0, y_0), ..., g_n(0, y_0)) = (g_1(f(x_0, y_0)), ..., g_n(f(x_0, y_0))) =$$

$$((g \circ f)_1(x_0, y_0), ..., (g \circ f)_n(x_0, y_0)) = (x_{0_1}, ..., x_{0_n}) = x_0.$$

Veamos la relación $F(G(y), y) = 0$ para todo $y \in Y_0$. Sea $y \in Y_0$, por la propia definición de Y_0 se tiene que $(0, y) \in V = f(U)$, por lo que existe $(x^*, y^*) \in U$ tal que $f(x^*, y^*) = (0, y)$. Como por definición $f(x^*, y^*) = (F(x^*, y^*), y^*)$, resulta que

$$F(x^*, y^*) = 0 \quad \text{e} \quad y^* = y. \tag{*}$$

Por otra parte, como $(x^*, y^*) \in U$ y $g \circ f = Id$ en U,

$$(x^*, y^*) = g \circ f(x^*, y^*) = g(0, y).$$

Dado que $G(y) = (g_1(0, y), ..., g_n(0, y))$, que son las n primeras componentes de $g(0, y)$, resulta que $G(y) = x^*$ y por lo tanto se sigue de la relación (*) que $(G(y), y) = (x^*, y^*) \in U$ y $F(G(y), y) = 0$.

La unicidad de G puede obtenerse así: Si hubiese otra función G^* con las mismas características que G, entonces $F(G(y), y) = F(G^*(y), y)$ para todo $y \in Y_0$, y así

$$(F(G(y), y), y) = (F(G^*(y), y), y) \quad \forall y \in Y_0$$

Esto nos da que $f(G(y), y) = f(G^*(y), y) \in V$ para todo $y \in Y_0$ y entonces, como f es inyectiva en U, resulta que $G(y) = G^*(y)$ para todo $y \in Y_0$.

Veamos por último la demostración del apartado (d).

Consideremos un punto $(x, y) \in U$ tal que $F(x, y) = 0$. Como $f(x, y) = (F(x, y), y)$, resulta que $f(x, y) = (0, y)$, y como $f(U) = V$ (el V del teorema de la función inversa), necesariamente $(0, y) \in V$, como por definición $Y_0 = \varphi^{-1}(V)$ (siendo $\varphi(y) = (0, y)$) se tiene que $y \in Y_0$. Además, como $(x, y) \in U$,

$$(x, y) = g \circ f(x, y) = g(0, y) = (G(y), g_{n+1}(0, y), ..., g_{n+k}(0, y)),$$

en particular $x = G(y)$.

Recíprocamente, se acaba de probar que si $y \in Y_0$, entonces $(G(y), y) \in U$ y $F(G(y), y) = 0$. $\square$

Ejemplo 3.2.4 *Consideremos la función*

$$F : (x, y; u, v) \in \mathbb{R}^2 \times \mathbb{R}^2 \mapsto F(x, y; u, v) = (xu + yv^2, xv^3 + y^2u^6) \in \mathbb{R}^2.$$

En esta situación $n = 2$ *y* $k = 2$*. Para* $x_0^* = (x_0, y_0) = (1, -1)$ *e* $y_0^* = (u_0, v_0) = (1, -1)$*, se verifica que* $F(x_0^*, y_0^*) = (1 - 1, -1 + 1) = (0, 0)$*. Por otra parte,*

$$\begin{pmatrix} D_1F_1(x, y; u, v) & D_2F_1(x, y; u, v) \\ D_1F_2(x, y; u, v) & D_2F_2(x, y; u, v) \end{pmatrix} = \begin{pmatrix} u & v^2 \\ v^3 & 2yu^6 \end{pmatrix}$$

En consecuencia,

$$\det(D_jF_i(1, -1, 1, -1) = \det \begin{pmatrix} 1 & 1 \\ -1 & -2 \end{pmatrix} = -1 \neq 0.$$

Por lo tanto existirán un entorno Y_0^* *de* $y_0^* = (1, -1)$ *y una función* $G : Y_0^* \subset \mathbb{R}^2 \to \mathbb{R}^2$ *tal que* $G((1, -1)) = (1, -1)$ *y* $F(G(u, v), u, v) = 0$ *para todo* $(u, v) \in Y_0^*$*. Esto es,*

$$\begin{gathered} G_1(u, v)u + G_2(u, v)v^2 = 0 \\ G_1(u, v)v^3 + G_2(u, v)^2u^6 = 0 \end{gathered}$$

Como se ve, se despejan así las variables x *e* y *en función de las variables* u *y* v *en un entorno del punto* $(x_0^*, y_0^*) = (1, -1, -1, 1)$*.*

Vamos a ver ahora como también es posible conocer la diferencial de la función implícita G obtenida en el teorema de la función implícita en un entorno de y_0. Lo hacemos en un caso particular pero el resultado es válido en general.

Consideremos una función $F : A \subset \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ de clase C^1 en A, a la que se le aplica el teorema de la función implícita para un punto $(x_0, y_0) \in A$ para el cual $F(x_0, y_0) = 0$ y

$D_1F(x_0, y_0) \neq 0$, para obtener una función $G : Y_0 \to \mathbb{R}$ tal que $F(G(y), y) = 0$ para todo $y \in Y_0$, entonces la diferencial de la composición

$$y \mapsto F(G(y), y)$$

de la función $\Psi(y) = (G(y), y)$ con F, es 0 en todo punto de Y_0, y la correspondiente matriz jacobiana es:

$$(D_1F(G(y), y)\ D_2F(G(y), y)) \begin{pmatrix} G'(y) \\ 1 \end{pmatrix} = 0,$$

de donde, $D_1F(G(y), y) \cdot G'(y) + D_2F(G(y), y) = 0$, como $D_1F(G(y_0), y_0) \neq 0$ por continuidad $D_1F(G(y), y) \neq 0$ en un entorno de y_0, con lo que,

$$G'(y) = -\frac{D_2F(G(y), y)}{D_1F(G(y), y)}.$$

Apliquemos esto a un caso concreto.

Ejemplo 3.2.5 *Dada la función $F : (x, y) \in A \subset \mathbb{R} \times \mathbb{R} \longmapsto F(x, y) = tg(xy) - xy$, vamos a ver en qué subconjunto abierto A está definida y para qué puntos (x_0, y_0) de A el teorema de la función implícita nos garantiza la existencia de un abierto Y_0, entorno de y_0 y de una función $G : Y_0 \to \mathbb{R}$ de clase C^1 en Y_0 verificando que $G(y_0) = x_0$ y $F(G(y), y) = 0$ para todo $y \in Y_0$. Calcular $G'(y_0)$.*

F está definida en los puntos (x, y) tales que $\cos(xy) \neq 0$, es decir en los puntos (x, y) con $xy \neq \frac{\pi}{2} + k\pi$ con $k \in \mathbb{Z}$, este conjunto es la unión de los conjuntos abiertos $\{(x, y) : \frac{\pi}{2} + k\pi < xy < \frac{\pi}{2} + (k + 1)\pi\}$, $k \in \mathbb{Z}$.

En un punto (x_0, y_0) del conjunto de definición de F se verifica que $F(x_0, y_0) = 0$ si y sólo si $tg(x_0y_0) = x_0y_0$. En un tal punto

$$D_1F(x_0, y_0) = \frac{y_0}{\cos^2(x_0y_0)} - y_0 = y_0\left(\frac{1}{\cos^2(x_0y_0)} - 1\right) = y_0tg^2(x_0y_0) = y_0(x_0y_0)^2.$$

Entonces podremos aplicar el teorema de la función implícita a aquellos (x_0, y_0) tales que $tg(x_0y_0) = x_0y_0$ y además $x_0y_0 \neq 0$ (nótese que F es de clase C^1 en un entorno de esos puntos). En tales puntos el teorema de la función implícita nos da la existencia de un entorno abierto Y_0 de y_0 y una función $G : Y_0 \to \mathbb{R}$ tal que G es de clase C^1 en Y_0, $G(y_0) = x_0$ y $F(G(y), y) = 0$ para todo $y \in Y_0$. Sabemos además que

$$G'(y) = -\frac{D_2F(G(y), y)}{D_1F(G(y), y)} = -\frac{G(y)tg^2(G(y)y)}{ytg^2(G(y)y)} = -\frac{G(y)}{y}$$

Podemos entonces, en este caso determinar G resolviendo esa "ecuación diferencial", cuyas soluciones son

$$G(y) = \frac{C}{y}$$

siendo C cualquier constante real. Tal constante se obtiene de la condición $G(y_0) = x_0$, lo que nos da que $C = x_0 y_0$. Obsérvese que si $G(y) = \frac{C}{y}$ entonces

$$G'(y) = -\frac{C}{y^2} = -\frac{G(y)}{y}.$$

3.3. Extremos condicionados. Multiplicadores de Lagrange

En este último tema vamos a dar una introducción al concepto de variedad regular en $\mathbb{R}^n$ y a estudiar los extremos condicionados.

Definición 3.3.1 *Sea $f : A \subset \mathbb{R}^n \to \mathbb{R}^k$, siendo A un abierto. Se dice que un punto $x_0 \in A$ es un punto regular de f (o que f es regular en x_0) si f es de clase $\mathcal{C}^1$ en un entorno abierto de x_0 y la matriz jacobiana de f en x_0 tiene rango máximo (el rango de una matriz es el orden de la mayor submatriz cuadrada de ella con determinante no nulo).*

Definición 3.3.2 *Se llama **variedad** (o superficie) **regular** de dimensión m en $\mathbb{R}^n$ ($m < n$) a todo subconjunto M de $\mathbb{R}^n$ con la siguiente propiedad: Para cada $x_0 \in M$ existe un abierto $A \subset \mathbb{R}^n$ que contiene a x_0 y existe una aplicación $F : A \to \mathbb{R}^{n-m}$, regular en cada punto de A (por lo que el rango de la matriz jacobiana de F en todo punto de A es $n-m$), tal que*

$$M \cap A = \{x \in A : F(x) = 0 \in \mathbb{R}^{n-m}\}.$$

Ejemplos:

1. Sea $f : A \subset \mathbb{R}^n \to \mathbb{R}^{n-m}$ una aplicación de clase C^1 en el abierto A cuya matriz jacobiana tiene rango $n-m$ en cada punto de A. Entonces

$$M = \{x \in A : f(x) = 0\}$$

es una variedad regular de dimensión m en $\mathbb{R}^n$.

Tómese, por cada $x_0 \in M$ como A el propio A y como F la aplicación f.

Veamos algunos casos particulares:

a) $n = 2$, $m = 1$, $A = \mathbb{R}^2$ y sea $f(x, y) = y - x^2$. Entonces $M = \{(x, y) \in \mathbb{R}^2 : y - x^2 = 0\}$ es precisamente la parábola $y = x^2$. Nótese que $J_f(x, y) = (-2x, 1)$ tiene rango 1 en cualquier punto de $\mathbb{R}^2$.

b) $n = 2$, $m = 1$, $A = \mathbb{R}^2$ y sea $f(x, y) = x^2 + y^2 - 1$. Entonces $M = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 - 1 = 0\}$ es precisamente la circunferencia de centro $(0, 0)$ y radio 1. En este caso

$J_f(x,y) = (2x, 2y)$ tiene rango 1 para todo $(x,y) \in \mathbb{R}^2$ distinto del $(0,0)$, pero este punto no pertenece a M.

c) $n = 3$, $m = 1$, $B = \mathbb{R}^3$ y sea $f(x,y,z) = (y - x, z - x^2)$,

$$J_f(x,y,z) = \begin{pmatrix} -1 & 1 & 0 \\ -2x & 0 & 1 \end{pmatrix}$$

que es de rango 2 cualquiera que sea el punto (x,y,z). Luego $M = \{(x,y,z) \in \mathbb{R}^3 : (y - x, z - x^2) = (0,0)\}$ es una variedad en $\mathbb{R}^3$ de dimensión 1, es precisamente una parábola situada en el plano $x = y$.

d) $n = 3$, $m = 2$, $A = \mathbb{R}^3$ y sea $f(x;y,z) = x + y + z$. Ciertamente f es de clase C^1 en todo punto de $\mathbb{R}^3$ y para cualquier $(x,y,z) \in \mathbb{R}^3$ se tiene que

$$J_f(x,y,z) = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}.$$

Esta matriz tiene rango 1 cualquiera que sea el punto $(x,y,z) \in \mathbb{R}^3$, luego $M = \{(x,y,z) \in \mathbb{R}^3 : x + y + z = 0\}$ es una variedad regular de dimensión 2 en $\mathbb{R}^3$, es un plano.

2. Si $f : B \subset \mathbb{R}^m \to \mathbb{R}^{n-m}$ es una aplicación de clase C^1 en el abierto B, entonces

$$M = \{(x_1, ..., x_m; f_1(x_1, ..., x_m), ..., f_{n-m}(x_1, ..., x_m)) : (x_1, ..., x_m) \in B\}$$

es una variedad regular de dimensión m en $\mathbb{R}^n$. Obsérvese que esto nos dice que la gráfica de una función de clase C^1 de un abierto de $\mathbb{R}^m$ en $\mathbb{R}^{n-m}$ es una variedad regular de dimensión m en $\mathbb{R}^n$.

En efecto, consideremos para cualquier punto de M el abierto $A = B \times \mathbb{R}^{n-m}$ de $\mathbb{R}^n$ y la aplicación $F : A \to \mathbb{R}^{n-m}$ definida por

$$F : (x_1, ..., x_m; x_{m+1}, ..., x_n) \mapsto (x_{m+1} - f_1(x_1, ..., x_m), ..., x_n - f_{n-m}(x_1, ..., x_m)).$$

F es de clase C^1 en A y al ser la matriz jacobiana de F en cada $(x_1, ..., x_m; x_{m+1}, ..., x_n) \in A$ igual a

$$\begin{pmatrix} -\frac{\partial f_1}{\partial x_1}(x_1, ..., x_m) & \cdots & -\frac{\partial f_1}{\partial x_m}(x_1, ..., x_m) & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ -\frac{\partial f_{n-m}}{\partial x_1}(x_1, ..., x_m) & \cdots & -\frac{\partial f_{n-m}}{\partial x_m}(x_1, ..., x_m) & 0 & \cdots & 1 \end{pmatrix}$$

y tiene rango máximo $(n - m)$ en cada punto de A.

Los ejemplos (a), (c) y (d) anteriores también se pueden considerar como ejemplos de este caso. Para (a) se considera $f(x) = x^2$ y entonces

$$M = \{(x, x^2) : x \in \mathbb{R}\},$$

para (c), $f(x) = (x, x^2)$ y entonces

$$M = \{(x, x, x^2) : x \in \mathbb{R}\}$$

y para (d), $f(x, y) = -x - y$ con lo que

$$M = \{(x, y, -x - y) : (x, y) \in \mathbb{R}^2\}.$$

En el caso (b) no podemos hacer lo mismo pues "no se puede despejar globalmente la y en función de la x", esto es, no podemos poner $M = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 - 1 = 0\}$ en la forma $\{(x, f(x)), x \in \mathbb{D}\}$ para ninguna f. y ningún D abierto de $\mathbb{R}$.

3. Sea $\varphi : B \subset \mathbb{R}^m \to \mathbb{R}^n$, de clase $\mathcal{C}^1$ en el abierto B, una aplicación cuya matriz jacobiana tiene rango m en cada punto de B. Entonces para cada punto $b \in B$ existe un entorno abierto U de b tal que $\varphi(U)$ es una variedad regular de dimensión m en $\mathbb{R}^n$.

Ejemplo: $\varphi(t) = (\cos t, \operatorname{sen} t),\ t \in \mathbb{R}$.

Veremos la demostración sólo para $m = 1$ y $n = 2$ para simplificar la escritura (el caso general es análogo): Sea b un punto de B, Al ser φ regular en b (lo es en todo punto de B), la matriz jacobiana de φ en b tiene rango máximo, que en este caso es 1. Como esa matriz jacobiana es $\varphi'(b) = \begin{pmatrix} \varphi_1'(b) \\ \varphi_2'(b) \end{pmatrix}$, necesariamente alguno de sus térmimos tiene que ser distinto de 0. Supongamos que lo es el $\varphi_1'(b)$. Aplicamos el teorema de la función inversa a $\varphi_1 : B \subset \mathbb{R} \to \mathbb{R}$, $(\varphi_1'(b) \neq 0)$ y obtenemos la existencia de:

-Un entorno abierto U de b en el que φ_1 es inyectiva,
-un entorno abierto V de $\varphi_1(b)$ tal que $\varphi_1(U) = V$, y
-una función g_1 de clase C^1 en V, con valores en U, de forma que $g_1 \circ \varphi_1 = Id$ en U.

Veamos como $\varphi(U)$ es una variedad regular de dimensión 1 en $\mathbb{R}^2$. Dado $c \in U$, consideremos el entorno abierto $A = V \times \mathbb{R}$ de $\varphi(c)$ (nótese que $\varphi_1(c) \in V$), y la función F definida en él por:

$$F(x, y) = y - \varphi_2(g_1(x));\ (x, y) \in A.$$

$F \in C^1(A)$ y su matriz jacobiana (que es un gradiente) tiene rango 1 en todo punto de A (obsérvese que esa matriz es el vector gradiente: $(-\varphi_2'(g_1(x))g_1'(x), 1)$, y sucede que

$$\varphi(U) \cap A = \{(\varphi_1(u), \varphi_2(u));\ u \in U\} =$$

$$\{(x, y) \in A : F(x, y) = 0\}$$

La primera de las igualdades es inmediata (téngase en cuenta que $A = V \times \mathbb{R}$ y que $\varphi_1(U) = V$). Veamos la segunda: para $u \in U$ se tiene que $\varphi(u) = (\varphi_1(u), \varphi_2(u)) \in A$ y

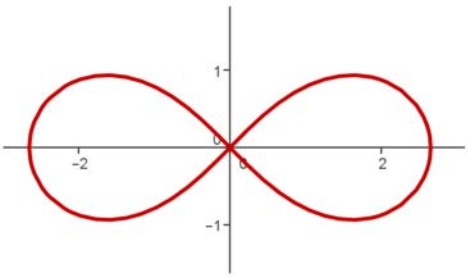
$$F(\varphi_1(u), \varphi_2(u)) = \varphi_2(u) - \varphi_2(g_1(\varphi_1(u))) = \varphi_2(u) - \varphi_2(u) = 0.$$

Recíprocamente, si $F(x, y) = y - \varphi_2(g_1(x)) = 0$ para un $(x, y) \in A$, entonces $y = \varphi_2(g_1(x))$. Como $x \in V$, para un cierto $y^* \in U$ se tiene que $x = \varphi_1(y^*)$ e $y = \varphi_2(g_1(x)) = \varphi_2(g_1(\varphi_1(y^*))) = \varphi_2(y^*)$, luego $(x, y) = (\varphi_1(y^*), \varphi_2(y^*))$ con $y^* \in U$.

Observamos que, en general no se verifica que $\varphi(B)$ sea una variedad regular, esto sucede, por ejemplo, si φ es la "lemniscata de Bernoulli" dada por

$$\varphi(t) = \left(\frac{t(1+t^2)}{1+t^4}, \frac{t(1-t^2)}{1+t^4} \right).$$

También se puede expresar como el conjunto de puntos (x, y) de $\mathbb{R}^2$ tales que $(x^2+y^2)^2 + (y^2 - x^2)^2 = 0$. Su gráfica es:



Pueden verse los detalles en el libro de F. Castillo, "Análisis Matemático II", Ed. Alhambra, 1980, pp. 141-142.

No todos los subconjuntos de $\mathbb{R}^n$ son variedades regulares; por ejemplo, $M = \{(x, y) \in \mathbb{R}^2 : x^2 = y^2\}$ no es una variedad regular de dimensión 1 en $\mathbb{R}^2$. Si lo fuese, dado $(0, 0) \in M$ deberían existir un entorno abierto A de él en $\mathbb{R}^2$ y una función de clase C^1 regular en cada punto de A tales que

$$M \cap A = \{(x, y) \in A : F(x, y) = 0\}.$$

Como F es regular en $(0, 0)$ una de sus derivadas parciales en ese punto debe ser distinta de 0, supongamos que lo es la $D_1F(0, 0)$. Por el teorema de la función implícita existen un entorno abierto Y_0 de 0 y una función $G : Y_0 \to \mathbb{R}$ de clase C^1 en Y_0 tales que $G(0) = 0$ y $F(G(y), y) = 0$ para todo $y \in Y_0$, pero por la afirmación (d) del teorema de la función implícita,

$$\{(x, y) \in U : F(x, y) = 0\} = \{(G(y), y) : y \in Y_0\}$$

(U es un cierto entorno abierto de $(0, 0)$ que se obtuvo en la demostración del teorema de la función implícita) y claramente esto no es posible pues en $M \cap A \cap U$ hay puntos de la forma $(\varepsilon, \varepsilon)$ y $(-\varepsilon, \varepsilon)$ (si ε es pequeño) pero $G(\varepsilon)$ sólo toma uno de los valores ε o $-\varepsilon$, luego $\{(x,y) \in A : F(x,y) = 0\}$ es estrictamente más pequeño que $M \cap A$.

Un argumento parecido a este es el que se usa para ver que la lemniscata de Bernoulli no da lugar a una variedad regular.

Definición 3.3.3 *Sean B es un conjunto abierto de $\mathbb{R}^n$, f una función definida de B en $\mathbb{R}$ y M un subconjunto de B. Se dice que un punto $x_0 \in M$ es un* ***máximo relativo de f condicionado por*** *M si existe $r > 0$ tal que $B(x_0, r) \subset B$ y $f(x) \leq f(x_0)$ $\forall x \in M \cap B(x_0, r)$. La definición de* ***mínimo relativo de f condicionado por*** *M es análoga.*

Con el objetivo de obtener el resultado final del curso (teorema de los multiplicadores de Lagrange), que nos ayuda a encontrar extremos condicionados, hacemos el siguiente lema.

Lema 3.3.4 *Sea $F : A \subset \mathbb{R}^{n-m} \times \mathbb{R}^m \to \mathbb{R}^{n-m}$ de clase C^1 en el abierto A y sea*

$$M = \{x = (x_1, ..., x_{n-m}, ..., x_n) \in A : F(x) = 0\}.$$

Entonces para todo $x_0 = (x_{0_1}, ..., x_{0_{n-m}}, ..., x_{0_n}) \in M$ tal que

$$\det(D_i F_j(x_0))_{i,j=1,...,n-m} \neq 0$$

existe un $Y_0 \subset \mathbb{R}^m$, entorno de $y_0 = (x_{0_{n-m+1}}, ..., x_{0_n})$ y existe $\psi : Y_0 \to A$, regular en cada punto de Y_0, tal que $\psi(y_0) = x_0$ y $F[\psi(y)] = 0$ para todo $y \in Y_0$.

Demostración. Estamos en las hipótesis para poder aplicar el teorema de la función implícita (nótese que como $x_0 \in M$ se verifica que $F(x_0) = 0$), y de él obtenemos la existencia de un Y_0, entorno de y_0 en $\mathbb{R}^m$, y de una función $G : Y_0 \to \mathbb{R}^{n-m}$ de clase C^1 en Y_0 tales que $G(y_0) = (x_{0_1}, ..., x_{0_{n-m}})$ y $F(G(y), y) = 0$ para todo $y \in Y_0$. Si definimos

$$\psi : y \in Y_0 \subset \mathbb{R}^m \mapsto \psi(y) = (G(y), y) = (G_1(y), \ldots, G_{n-m}(y), y_1, \ldots, y_m)$$

entonces ψ es regular en Y_0 (su matriz jacobiana en cada punto contiene una matriz cuadrada de orden m diagonal formada por unos en la diagonal y el resto ceros) y

$$F[\psi(y)] = F(G(y), y) = 0 \text{ para todo } y \in Y_0.$$

$\square$

Teorema 3.3.5 (de los multiplicadores de Lagrange) *Sean B un abierto de $\mathbb{R}^n$, $f : B \to \mathbb{R}$ una función de clase C^1 en B y M una variedad regular en $\mathbb{R}^n$ de dimensión m contenida en B. Si un punto $x_0 = (x_{0_1}, ..., x_{0_{n-m}}, ..., x_{0_n}) \in M$ es un extremo relativo de f condicionado por M y si A es un entorno de x_0 y $F : A \subset \mathbb{R}^{n-m} \times \mathbb{R}^m \to \mathbb{R}^{n-m}$ es una función regular en A de forma que*

$$M \cap A = \{x \in A : F(x) = 0\},$$

entonces existen $n - m$ números reales $\lambda_1, ..., \lambda_{n-m}$ de forma que

$$\nabla f(x_0) = \lambda_1 \nabla F_1(x_0) + \cdots + \lambda_{n-m} \nabla F_{n-m}(x_0).$$

Demostración. Al ser F regular en x_0 el rango de la matriz jacobiana de F en x_0 es máximo, por lo que existe un menor de orden $n - m$ de esa matriz con determinante no nulo. Supondremos por comodidad que un tal menor es el dado por la matriz $(D_i F_j(x_0))_{i,j=1,...,n-m}$. Tenemos así que $F(x_0) = 0$ y $\det(D_i F_j(x_0))_{i,j=1,...,n-m} \neq 0$. Entonces por el Lema anterior existe un entorno Y_0 de $y_0 = (x_{0_{n-m+1}}, ..., x_{0_n})$ y existe $\psi : Y_0 \to A$ de clase C^1 regular en todo punto de Y_0 tal que $\psi(y_0) = x_0$ y $F[\psi(y)] = 0$ para todo $y \in Y_0$. La relación $F(\psi(y)) = 0$ nos da que para todo $y \in Y_0$

$$\begin{pmatrix} D_1 F_1(\psi(y)) & \cdots & D_n F_1(\psi(y)) \\ \cdots & \cdots & \cdots \\ D_1 F_{n-m}(\psi(y)) & \cdots & D_n F_{n-m}(\psi(y)) \end{pmatrix} \cdot \begin{pmatrix} D_1 \psi_1(y) & \cdots & D_m \psi_1(y) \\ \cdots & \cdots & \cdots \\ D_1 \psi_n(y) & \cdots & D_m \psi_n(y) \end{pmatrix} = (0).$$

Por otra parte, f tiene un extremo en $x_0 \in M$, condicionado por M, por lo que $f \circ \psi$ tiene un extremo en y_0 (nótese que si en $M \cap B(x_0, r)$ se da una de las desigualdades que caracterizan al extremo, entonces $Y = \psi^{-1}(B(x_0, r)) \subset Y_0$ es un entorno abierto de y_0 tal que $\psi(Y) \subset M \cap B(x_0, r)$, téngase en cuenta que $\psi(Y) \subset \psi(Y_0) \subset A$ y como $F(\psi(y)) = 0$ para $y \in Y$ resulta que $\psi(y) \in A \cap M \cap B(x_0, r) \subset M \cap B(x_0, r)$). La condición necesaria para la existencia de un extremo en un abierto nos dice entonces que $d(f \circ \psi)(y_0) = 0$ y por lo tanto que

$$\begin{pmatrix} D_1 f(\psi(y_0)) & \cdots & D_n f(\psi(y_0)) \end{pmatrix} \cdot \begin{pmatrix} D_1 \psi_1(y_0) & \cdots & D_m \psi_1(y_0) \\ \cdots & \cdots & \cdots \\ D_1 \psi_n(y_0) & \cdots & D_m \psi_n(y_0) \end{pmatrix} = (0)$$

Nos encontramos así con que los vectores $(D_1 F_j)\psi(y_0)), ..., D_n F_j(\psi(y_0)))$, $j = 1, ..., n - m$ y $(D_1 f(\psi(y_0)), ..., D_n f(\psi(y_0)))$ son todos ellos ortogonales a los vectores columna de la matriz

$$\begin{pmatrix} D_1 \psi_1(y_0) & \cdots & D_m \psi_1(y_0) \\ \cdots & \cdots & \cdots \\ D_1 \psi_n(y_0) & \cdots & D_m \psi_n(y_0) \end{pmatrix}.$$

Como ψ es regular el rango de esta matriz máximo (m), el espacio que generan los vectores columna tiene dimensión m por lo que su complemento ortogonal en $\mathbb{R}^n$ tiene dimensión

$n-m$. Como nosotros tenemos $(n-m)+1$ vectores en ese espacio ortogonal, siendo los del tipo $(D_1F_i(\psi(y_0)), ..., D_nF_i(\psi(y_0)))$, $i = 1, ..., n-m$, independientes entre si, resulta que el otro es combinación lineal de éstos y por lo tanto existen $\lambda_1, ..., \lambda_{n-m}$ números reales tales que

$$(D_1f(x_0), ..., D_nf(x_0)) =$$

$$\lambda_1(D_1F_1(x_0), ..., D_nF_1(x_0)) + \cdots + \lambda_{n-m}(D_1F_{n-m}(x_0), ..., D_nF_{n-m}(x_0)).$$

$\square$

Nota 3.3.6 *Cuando se trabaja con funciones* $f : B \subset \mathbb{R}^2 \to \mathbb{R}$ *el teorema nos da la existencia de un* $\lambda \in \mathbb{R}$ *tal que* $\nabla f(x_{01}, x_{0_2}) = \lambda \nabla F(x_{0_1}, x_{0_2})$.