

Análisis Matemático I

ÍNDICE

Tema 1: El espacio euclídeo. Espacios normados y espacios métricos	1
Tema 2: Topología de un espacio métrico	13
Tema 3: Continuidad y límite funcional	27
Tema 4: Compacidad y conexión	39
Tema 5: Complicidad y continuidad uniforme	52
Tema 6: Diferenciabilidad	61
Tema 7: Vector derivada	76
Tema 8: Vector gradiente.....	84
Tema 9: Matriz jacobiana	99
Tema 10: Teorema del valor medio	113
Tema 11: Función inversa.....	121
Tema 12: Función implícita.....	133

El espacio euclídeo, espacios normados y espacios métricos

Como punto de partida para el estudio de las funciones de varias variables reales, debemos familiarizarnos con la estructura y propiedades del espacio en el que dichas funciones tendrán su conjunto de definición, el *espacio euclídeo* N -dimensional, donde N es un número natural. Al tiempo que estudiamos algunas propiedades de dicho espacio, las iremos abstrayendo, para entender ciertos conceptos generales que son importantes en Análisis Matemático.

Partimos de la definición de \mathbb{R}^N y su estructura algebraica básica, la de *espacio vectorial*. Al estudiar el *producto escalar* en \mathbb{R}^N , completamos la definición del espacio euclídeo, así llamado porque formaliza analíticamente los axiomas y resultados de la geometría de Euclides. De hecho introducimos la noción más general de *espacio pre-hilbertiano* y manejamos en este tipo de espacio la norma asociada al producto escalar, que en el caso particular de \mathbb{R}^N es la *norma euclídea*.

Abstrayendo las propiedades básicas de la norma de un espacio pre-hilbertiano, definimos lo que se entiende en general por una *norma* en un espacio vectorial, llegando así a la noción de *espacio normado*. Veremos ejemplos de normas en \mathbb{R}^N distintas de la euclídea, así como algunos espacios normados de dimensión infinita. Por último veremos que, en todo espacio normado, se puede definir de manera coherente la distancia entre dos puntos, motivando así las nociones generales de *distancia* y *espacio métrico*.

1.1. El espacio vectorial \mathbb{R}^N

Repasamos brevemente la estructura de espacio vectorial de \mathbb{R}^N , aprovechando para fijar la notación que vamos a usar. En todo lo que sigue, N será un número natural fijo y escribiremos

$$\Delta_N = \{k \in \mathbb{N} : k \leq N\} = \{1, 2, \dots, N\}$$

Recordamos que \mathbb{R}^N es el producto cartesiano de N copias de \mathbb{R} , es decir, el conjunto de todas las posibles N -uplas de números reales:

$$\mathbb{R}^N = \mathbb{R} \times \mathbb{R} \times \dots \times \mathbb{R} = \{(x_1, x_2, \dots, x_N) : x_1, x_2, \dots, x_N \in \mathbb{R}\}$$

Dado $x = (x_1, x_2, \dots, x_N) \in \mathbb{R}^N$, los números reales x_1, x_2, \dots, x_N son las **componentes** de la N -upla x , más concretamente, x_k será la k -ésima componente de x , para cada $k \in \Delta_N$. Esta es una notación muy habitual: cuando usamos una determinada letra para denotar una N -upla, la misma letra con los subíndices $1, 2, \dots, N$ nos sirve para denotar sus componentes.

Sin embargo, no siempre es conveniente usar subíndices para denotar las componentes de los elementos de \mathbb{R}^N , pues podemos necesitar los subíndices para otra finalidad. Es lo que ocurre, por ejemplo, cuando consideramos sucesiones de elementos de \mathbb{R}^N . Conviene por tanto disponer de una notación alternativa. Para valores concretos de N , podemos denotar las componentes con letras diferentes, siendo habitual escribir

$$\mathbb{R}^2 = \{(x, y) : x, y \in \mathbb{R}\} \quad \text{y} \quad \mathbb{R}^3 = \{(x, y, z) : x, y, z \in \mathbb{R}\}$$

En general, cabe recordar que una N -upla de números reales es una aplicación de Δ_N en \mathbb{R} . De hecho, la N -upla $(x_1, x_2, \dots, x_N) \in \mathbb{R}^N$ es la aplicación $x : \Delta_N \rightarrow \mathbb{R}$ dada por $x(k) = x_k$ para todo $k \in \Delta_N$. Por tanto, aunque pocas veces pensaremos en x como una función, tiene perfecto sentido denotar por $x(k)$ a la k -ésima componente de la N -upla x . Tenemos así una notación que evita los subíndices, permitiendo usarlos para otros fines.

En \mathbb{R}^N disponemos de las operaciones de **suma** y **producto por escalares**, que vienen definidas, para $x = (x_1, x_2, \dots, x_N) \in \mathbb{R}^N$, $y = (y_1, y_2, \dots, y_N) \in \mathbb{R}^N$ y $\lambda \in \mathbb{R}$, por

$$\begin{aligned} x + y &= (x_1 + y_1, x_2 + y_2, \dots, x_N + y_N) \\ \lambda x &= (\lambda x_1, \lambda x_2, \dots, \lambda x_N) \end{aligned}$$

Obsérvese que estas operaciones son caso particular de las que usamos para funciones reales de variable real, pues para cualesquiera $x, y \in \mathbb{R}^N$, $\lambda \in \mathbb{R}$ y $k \in \Delta_N$, tenemos claramente

$$(x + y)(k) = x(k) + y(k) \quad \text{y} \quad (\lambda x)(k) = \lambda x(k)$$

Sabemos que, con estas dos operaciones, \mathbb{R}^N tiene estructura de *espacio vectorial*. Aquí y en lo sucesivo, *cundo hablemos de un espacio vectorial, se entenderá siempre construido sobre el cuerpo \mathbb{R} de los números reales*, único cuerpo escalar que vamos a manejar.

El espacio vectorial \mathbb{R}^N tiene *dimensión* N , es decir, todas sus *bases* constan de N vectores. Destacamos la que llamaremos *base usual*, $\Phi = \{e_1, e_2, \dots, e_N\}$ donde, para cada $k \in \Delta_N$, llamamos e_k a la N -upla cuya k -ésima componente es 1 y las demás se anulan, es decir:

$$e_k(k) = 1 \quad \text{y} \quad e_k(j) = 0 \quad \forall j \in \Delta_N \setminus \{k\}$$

He aquí un ejemplo en el que resulta cómoda la notación antes explicada. Usamos subíndices para numerar los elementos de Φ , así que, para cada $k \in \Delta_N$, tenemos una N -upla $e_k \in \mathbb{R}^N$, cuyas componentes ya no conviene denotar con subíndices. Para $j, k \in \Delta_N$, es más cómodo denotar por $e_k(j)$ a la j -ésima componente de la N -upla e_k , como hemos hecho.

Para cada $x = (x_1, x_2, \dots, x_N) \in \mathbb{R}^N$, la única expresión de x como combinación lineal de elementos de Φ viene dada por

$$x = \sum_{k=1}^N x_k e_k = \sum_{k=1}^N x(k) e_k$$

Así pues, las *coordenadas* de cada vector $x \in \mathbb{R}^N$, con respecto a la base usual de \mathbb{R}^N , son precisamente las componentes de la N -upla x .

Recordemos que \mathbb{R}^N es, salvo isomorfismos, el *único* espacio vectorial de dimensión N . Se dice que dos espacios vectoriales son *isomorfos* cuando existe entre ellos un *isomorfismo*, esto es, una biyección lineal. Si Ψ es una base de un espacio vectorial X , y Ψ consta de N vectores, la aplicación $f: X \rightarrow \mathbb{R}^N$, que a cada vector de X hace corresponder la N -upla formada por sus coordenadas en la base Ψ , es lineal y biyectiva. Nótese que $f(\Psi) = \Phi$, es decir, f transforma la base Ψ que hemos usado para definirla, en la base usual de \mathbb{R}^N . Queda bien claro que *todo espacio vectorial de dimensión N es isomorfo a \mathbb{R}^N* .

Damos por conocida la interpretación geométrica de los elementos de \mathbb{R}^N como los puntos, o los vectores libres, en un espacio de N dimensiones, en el que hemos fijado un sistema de referencia cartesiano. Para $N = 2$ tenemos un plano y para $N = 3$ el espacio tridimensional que somos capaces de percibir, pues para $N \geq 4$ perdemos la intuición geométrica. Como ocurre en el caso $N = 1$, al interpretar \mathbb{R} como una recta, esta visión geométrica sólo se usa para guiar la intuición, nunca como argumento válido en las demostraciones.

1.2. Producto escalar

El *producto escalar de dos vectores* $x = (x_1, x_2, \dots, x_N) \in \mathbb{R}^N$ e $y = (y_1, y_2, \dots, y_N) \in \mathbb{R}^N$ es, por definición, el número real $(x|y)$ dado por

$$(x|y) = \sum_{k=1}^N x_k y_k = \sum_{k=1}^N x(k) y(k) \quad (1)$$

y decimos también que la aplicación $(\cdot|\cdot) : \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}$, dada por $(x, y) \mapsto (x|y)$ para cualesquiera $x, y \in \mathbb{R}^N$, es el **producto escalar** en \mathbb{R}^N . Para $N = 1$, el producto escalar en \mathbb{R} es el producto usual de números reales.

En general, el producto escalar en \mathbb{R}^N tiene tres propiedades clave, todas ellas evidentes:

- (P.1) $(\lambda u + \mu v|y) = \lambda(u|y) + \mu(v|y) \quad \forall u, v, y \in \mathbb{R}^N, \forall \lambda, \mu \in \mathbb{R}$
- (P.2) $(x|y) = (y|x) \quad \forall x, y \in \mathbb{R}^N$
- (P.3) $(x|x) > 0 \quad \forall x \in \mathbb{R}^N \setminus \{0\}$

Vamos a recordar la nomenclatura que suele usarse en relación con estas tres propiedades, trabajando en un espacio vectorial arbitrario X , con una *forma en dos variables*, esto es, una aplicación $\varphi : X \times X \rightarrow \mathbb{R}$.

Se dice que φ es una *forma bilineal* en X , cuando es lineal en cada variable, es decir, cuando fijado cualquier $z \in X$, tanto $x \mapsto \varphi(x, z)$ como $y \mapsto \varphi(z, y)$ son aplicaciones lineales de X en \mathbb{R} . Se dice que la forma bilineal φ es *simétrica* cuando verifica que $\varphi(x, y) = \varphi(y, x)$ para cualesquiera $x, y \in X$. Entonces, la *forma cuadrática* asociada a φ es la aplicación $Q : X \rightarrow \mathbb{R}$ definida por $Q(x) = \varphi(x, x)$ para todo $x \in X$, y se dice que Q es *definida positiva*, cuando verifica que $Q(x) > 0$ para todo $x \in X \setminus \{0\}$.

La propiedad **(P.1)** nos dice que el producto escalar en \mathbb{R}^N es lineal en la primera variable, pero entonces **(P.2)** nos da también la linealidad en la segunda variable, así como la simetría. Por último, **(P.3)** nos dice que la forma cuadrática asociada es definida positiva. Queda así motivada la definición que sigue.

Un **producto escalar** en un espacio vectorial X es una forma bilineal simétrica en X , cuya forma cuadrática asociada es definida positiva. Un **espacio pre-hilbertiano** es, por definición, un espacio vectorial dotado de un producto escalar. Por supuesto, \mathbb{R}^N con el producto escalar definido como en (1):

$$\varphi(x, y) = (x|y) = \sum_{k=1}^N x(k)y(k) \quad \forall x, y \in \mathbb{R}^N$$

es un espacio pre-hilbertiano, que se conoce como el **espacio euclídeo N -dimensional**.

Como ejemplo de dimensión infinita, el espacio vectorial $C[0, 1]$ de todas las funciones continuas del intervalo $[0, 1]$ en \mathbb{R} , se convierte en espacio pre-hilbertiano definiendo

$$\varphi(x, y) = \int_0^1 x(t)y(t) dt \quad \forall x, y \in C[0, 1] \quad (2)$$

En efecto, la linealidad de la integral nos dice que φ es una forma bilineal en $C[0, 1]$, que obviamente es simétrica. Pero del crecimiento estricto de la integral deducimos que la forma cuadrática asociada es definida positiva, puesto que si $x \in C[0, 1] \setminus \{0\}$, tenemos $x(t)^2 \geq 0$ para todo $t \in [0, 1]$ y esta desigualdad ha de ser estricta para algún $t \in [0, 1]$, luego

$$\varphi(x, x) = (x|x) = \int_0^1 x(t)^2 dt > 0$$

1.3. Norma euclídea

A cada $x \in \mathbb{R}^N$ vamos ahora a asociar un número real no negativo que, como veremos, se interpreta geoméricamente como la longitud del vector x , o la longitud de cualquier segmento que usemos para representarlo. De hecho, lo hacemos en cualquier espacio pre-hilbertiano X . Igual que en \mathbb{R}^N , denotaremos por $(x, y) \mapsto (x|y)$ al producto escalar de X .

Se define la **norma de un vector** $x \in X$ como la raíz cuadrada del producto escalar de x por sí mismo, es decir, el número real no negativo $\|x\|$ dado por

$$\|x\| = (x|x)^{1/2} \quad (3)$$

Se dice también que la aplicación $\|\cdot\| : X \rightarrow \mathbb{R}$, definida por $x \mapsto \|x\|$ para todo $x \in X$, es la **norma** del espacio pre-hilbertiano X , o la norma asociada al producto escalar de X .

Destacamos las tres propiedades básicas de la norma recién definida:

$$(N.1) \quad \|x + y\| \leq \|x\| + \|y\| \quad \forall x, y \in X$$

$$(N.2) \quad \|\lambda x\| = |\lambda| \|x\| \quad \forall x \in X, \forall \lambda \in \mathbb{R}$$

$$(N.3) \quad x \in X, \|x\| = 0 \implies x = 0$$

La tercera es evidente: si fuese $x \neq 0$ se tendría $\|x\|^2 = (x|x) > 0$. La igualdad (N.2) se comprueba también inmediatamente:

$$\|\lambda x\|^2 = (\lambda x|\lambda x) = \lambda^2 (x|x) = \lambda^2 \|x\|^2$$

y basta tomar raíces cuadradas. Para probar (N.1) necesitamos la siguiente relación clave entre norma y producto escalar.

Desigualdad de Cauchy-Schwartz. *En todo espacio pre-hilbertiano X , se tiene:*

$$|(x|y)| \leq \|x\| \|y\| \quad \forall x, y \in X \quad (4)$$

Además, se verifica la igualdad si, y sólo si, x e y son linealmente dependientes.

Demostración. Si $x, y \in X$ son linealmente dependientes, se tendrá $y = 0$, o bien, $x = \lambda y$ con $\lambda \in \mathbb{R}$. En el primer caso la igualdad buscada es trivial y en el segundo basta usar (N.2):

$$|(x|y)| = |(\lambda y|y)| = |\lambda| \|y\|^2 = \|x\| \|y\|$$

Supongamos pues que $x, y \in X$ son linealmente independientes, y en particular no nulos, para probar la desigualdad estricta en (4). Para todo $\lambda \in \mathbb{R}$, tenemos

$$0 < (x - \lambda y|x - \lambda y) = \|x\|^2 - 2\lambda(x|y) + \lambda^2 \|y\|^2$$

y en particular, tomando $\lambda = (x|y) / \|y\|^2$ obtenemos

$$0 < \|x\|^2 - \frac{(x|y)^2}{\|y\|^2} \quad \text{es decir,} \quad \frac{(x|y)^2}{\|y\|^2} < \|x\|^2$$

Basta ahora multiplicar ambos miembros por $\|y\|^2 > 0$ y tomar raíces cuadradas. ■

La propiedad (N.1) se comprueba ya fácilmente. Para $x, y \in X$, basta pensar que

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

donde hemos usado la desigualdad de Cauchy-Schwartz. Abstrayendo las tres propiedades que tiene la norma de un espacio pre-hilbertiano, llegaremos más adelante a la noción general de norma en un espacio vectorial.

Por supuesto, todo lo dicho sobre la norma de un espacio pre-hilbertiano, es válido en el espacio euclídeo N -dimensional, cuyo producto escalar se definió en (1). La norma asociada a dicho producto escalar, definida como caso particular de (3), es la **norma euclídea** en \mathbb{R}^N . Así pues, la norma euclídea de un vector $x = (x_1, x_2, \dots, x_N) \in \mathbb{R}^N$ viene dada por

$$\|x\| = (x|x)^{1/2} = \left(\sum_{k=1}^N x_k^2 \right)^{1/2} = \left(\sum_{k=1}^N x(k)^2 \right)^{1/2} \quad (3')$$

Usando el teorema de Pitágoras, confirmamos claramente lo que habíamos anunciado: $\|x\|$ se interpreta como la longitud del vector x , la distancia del origen al punto x , o la longitud de cualquier otro segmento que usemos para representar al vector x . Nótese que en el caso $N = 1$ se tiene $\|x\| = |x|$ para todo $x \in \mathbb{R}$. Podemos y debemos entender la norma euclídea como una generalización natural del valor absoluto, disponible para cualquier dimensión N .

Resaltamos la desigualdad de Cauchy-Schwartz, que en \mathbb{R}^N toma la forma

$$\left| \sum_{k=1}^N x(k)y(k) \right| \leq \left(\sum_{k=1}^N x(k)^2 \right)^{1/2} \left(\sum_{k=1}^N y(k)^2 \right)^{1/2} \quad \forall x, y \in \mathbb{R}^N \quad (4')$$

La igualdad sólo se da cuando x e y son linealmente dependientes, es decir, cuando los puntos x e y están alineados con el origen. En el caso $N = 1$ tenemos una igualdad bien conocida, $|xy| = |x| |y|$ para cualesquiera $x, y \in \mathbb{R}$. Recíprocamente, la igualdad en (4') sólo se verifica para cualesquiera $x, y \in \mathbb{R}^N$, cuando $N = 1$.

Comentemos la interpretación geométrica de las propiedades básicas de la norma euclídea. La propiedad (N.1) recibe el nombre de *desigualdad triangular*, pues afirma que cada lado de un triángulo es menor o igual que la suma de los otros dos. La propiedad (N.2) se suele conocer como *homogeneidad por homotecias*. Equivale claramente a que, para todo $x \in \mathbb{R}^N$, se verifiquen tres igualdades:

$$\|0\| = 0, \quad \|-x\| = \|x\| \quad \text{y} \quad \|\lambda x\| = \lambda \|x\| \quad \forall \lambda \in \mathbb{R}^+$$

La primera no merece mucho comentario, la longitud de un segmento degenerado, cuyo origen y extremo coinciden, debe ser 0. Igual ocurre con la segunda, tampoco es una sorpresa que la longitud de un vector coincida con la de su opuesto, o que la longitud de un segmento no dependa de su orientación. Al hablar de homogeneidad por homotecias, ponemos el énfasis en la tercera igualdad, como vamos a explicar.

En cualquier espacio vectorial X , fijado $\lambda \in \mathbb{R}^+$, la aplicación $x \mapsto \lambda x$, de X en sí mismo, es la *homotecia* de razón λ . Intuitivamente, podríamos decir que consiste en aplicar un zoom, un cambio de escala en el espacio X . Pues bien, la propiedad (N.2) nos dice que, al aplicar a un segmento una homotecia, su longitud resulta multiplicada por la razón de homotecia.

Finalmente, (N.3) es una propiedad de *no degeneración*: la longitud de un segmento sólo se anula cuando se trata de un segmento degenerado.

1.4. Espacios normados

Una **norma** en un espacio vectorial X es una aplicación $\|\cdot\| : X \rightarrow \mathbb{R}$, que a cada $x \in X$ hace corresponder un número real $\|x\|$, verificando las tres condiciones siguientes:

- (N.1) *Desigualdad triangular:* $\|x + y\| \leq \|x\| + \|y\| \quad \forall x, y \in X$
- (N.2) *Homogeneidad por homotecias:* $\|\lambda x\| = |\lambda| \|x\| \quad \forall x \in X, \forall \lambda \in \mathbb{R}$
- (N.3) *No degeneración:* $x \in X, \|x\| = 0 \implies x = 0$

Un **espacio normado** es un espacio vectorial X , en el que hemos fijado una norma $\|\cdot\|$. Para cada $x \in X$, se dice también que el número real $\|x\|$ es la *norma del vector* x . No debemos confundir la norma del espacio X , que es una aplicación de X en \mathbb{R} , con la norma de cada vector $x \in X$, que es un número real.

Enseguida veremos que, en un mismo espacio vectorial X , podemos tener varias normas distintas. Debe quedar claro que, con cada una de esas normas, tendremos un espacio normado diferente.

Antes de ver abundantes ejemplos de espacios normados, comentemos algunas propiedades inmediatas que todas las normas tienen.

Si $\|\cdot\|$ es una norma en un espacio vectorial X , usando la homogeneidad por homotecias con $\lambda = 0$ obtenemos $\|0\| = 0$, mientras que tomando $\lambda = -1$, vemos que $\|-x\| = \|x\|$ para todo $x \in X$. Pero entonces, la desigualdad triangular implica que

$$0 = \|x + (-x)\| \leq \|x\| + \|-x\| = 2\|x\| \quad \forall x \in X$$

así que una norma nunca puede tomar valores negativos.

Por otra parte, dados $x, y \in X$, tenemos $\|x\| - \|y\| = \|x - y + y\| - \|y\| \leq \|x - y\|$, pero podemos intercambiar x e y para concluir que $|\|x\| - \|y\|| \leq \|x - y\|$. Además, tanto esta desigualdad como la triangular pueden aplicarse usando $-y$ en lugar de y . Disponemos así de estimaciones por defecto y por exceso para la norma de cualquier suma o diferencia.

Comentemos finalmente que, mediante una obvia inducción, la desigualdad triangular puede aplicarse a sumas finitas de vectores con cualquier número de sumandos. Si además usamos la homogeneidad por homotecias, estimamos por exceso la norma de cualquier combinación lineal. En resumen, toda norma tiene siempre las propiedades que siguen, algunas de las cuales generalizan propiedades bien conocidas del valor absoluto en \mathbb{R} :

■ En todo espacio normado X , se tiene:

- (i) $\|x\| \geq 0$ para todo $x \in X$, siendo $\|x\| = 0$ si, y sólo si, $x = 0$.
- (ii) $|\|x\| - \|y\|| \leq \|x \pm y\| \leq \|x\| + \|y\|$ para cualesquiera $x, y \in X$.
- (iii) Si $n \in \mathbb{N}$, $x_1, x_2, \dots, x_n \in X$ y $\lambda_1, \lambda_2, \dots, \lambda_n \in \mathbb{R}$, entonces:

$$\left\| \sum_{k=1}^n \lambda_k x_k \right\| \leq \sum_{k=1}^n \|\lambda_k x_k\| = \sum_{k=1}^n |\lambda_k| \|x_k\|$$

1.5. Ejemplos

Claramente, el primer ejemplo de espacio normado es el que hemos usado como motivación: *todo espacio pre-hilbertiano es un espacio normado, con la norma asociada a su producto escalar*. En particular, el espacio euclídeo N -dimensional es un espacio normado, \mathbb{R}^N con la norma euclídea.

Antes de ver otros ejemplos, conviene comentar que, en todo espacio normado X , la norma de cada vector $x \in X$ se interpreta siempre como la longitud que asignamos al vector x . En particular, cuando consideremos en \mathbb{R}^N una norma distinta de la euclídea, habremos cambiado de criterio sobre cual es la longitud de cada vector de \mathbb{R}^N .

La norma euclídea en \mathbb{R} coincide con el valor absoluto, y cualquier otra norma $\|\cdot\|$ en \mathbb{R} es proporcional a ella, pues tomando $K = \|1\| > 0$, tenemos: $\|x\| = \|x \cdot 1\| = |x| \|1\| = K|x|$, para todo $x \in \mathbb{R}$. En cualquier espacio vectorial, dos normas proporcionales son en esencia idénticas, pues podemos entender que al sustituir una por otra, simplemente hemos cambiado la unidad de longitud. Por tanto, salvo cambios de escala, el valor absoluto es la única norma que cabe considerar en \mathbb{R} . *Siempre que hablemos de \mathbb{R} como espacio normado, se entenderá que su norma es el valor absoluto*.

Para $N > 1$, es bien fácil definir en \mathbb{R}^N normas que no son proporcionales a la euclídea. Concretamente, podemos definir la **norma del máximo**, $\|\cdot\|_\infty$, y la **norma de la suma**, $\|\cdot\|_1$, escribiendo, para $x = (x_1, x_2, \dots, x_N) \in \mathbb{R}^N$,

$$\|x\|_\infty = \max \{|x_k| : k \in \Delta_N\} \quad \text{y} \quad \|x\|_1 = \sum_{k=1}^N |x_k| \quad (5)$$

Se comprueba fácilmente que se trata de dos nuevas normas en \mathbb{R}^N , incluso la desigualdad triangular es inmediata. También es fácil ver que estas normas no proceden de ningún producto escalar en \mathbb{R}^N . Veremos sin embargo que, para ciertos problemas, la norma de la suma o la del máximo pueden ser más útiles que la euclídea.

Consideremos de nuevo el espacio vectorial $C[0, 1]$ formado por las funciones continuas del intervalo $[0, 1]$ en \mathbb{R} . Conocemos en este espacio un producto escalar que lo convierte en un espacio pre-hilbertiano, luego es un espacio normado con la norma dada por

$$\|x\| = (x|x)^{1/2} = \left(\int_0^1 x(t)^2 dt \right)^{1/2} \quad \forall x \in C[0, 1]$$

La analogía con la norma euclídea de \mathbb{R}^N puede continuarse, definiendo dos normas en $C[0, 1]$ claramente inspiradas en las normas del máximo y de la suma. Para $x \in C[0, 1]$, basta escribir

$$\|x\|_\infty = \max \{|x(t)| : t \in [0, 1]\} \quad \text{y} \quad \|x\|_1 = \int_0^1 |x(t)| dt \quad (6)$$

Es fácil ver que así se obtienen dos nuevas normas en $C[0, 1]$, que no pueden proceder de un producto escalar. Tenemos pues dos ejemplos de espacios normados de dimensión infinita, que no son espacios pre-hilbertianos. Mencionamos, aunque sin explicar el motivo, que para trabajar en $C[0, 1]$, la norma más adecuada es $\|\cdot\|_\infty$, y no la asociada al producto escalar, como podría pensarse.

1.6. Espacios métricos

Dados dos puntos $x, y \in \mathbb{R}^N$, el segmento que va de x a y se obtiene trasladando el que va del origen a $y - x$, luego la longitud del vector $y - x$ nos da la distancia de x a y . Por tanto, cada norma que usemos en \mathbb{R}^N nos va a dar una posible definición de la distancia entre dos puntos de \mathbb{R}^N . De hecho, vamos a hacer lo mismo en cualquier espacio normado.

Si X es un espacio normado, se define la *distancia entre dos puntos* de X por

$$d(x, y) = \|y - x\| \quad \forall x, y \in X \quad (7)$$

Obtenemos así una función de dos variables $d : X \times X \rightarrow \mathbb{R}$, a partir de la cual podemos recuperar la norma de X , puesto que evidentemente

$$\|x\| = d(0, x) \quad \forall x \in X \quad (8)$$

Destacamos tres propiedades casi evidentes de la función d , que enseguida usaremos para definir lo que se entiende en general por una distancia en un conjunto cualquiera, que ya no tiene por qué ser un espacio vectorial. Para $x, y, z \in X$ tenemos

$$\begin{aligned} d(x, z) &= \|z - x\| \leq \|y - x\| + \|z - y\| = d(x, y) + d(y, z) \\ d(x, y) &= \|y - x\| = \|x - y\| = d(y, x) \\ d(x, y) &= 0 \iff \|y - x\| = 0 \iff x = y \end{aligned} \quad (9)$$

Nótese que las tres propiedades obtenidas son consecuencias de las condiciones que ha de cumplir una norma, pero se han escrito de forma que en ellas sólo interviene la función d , no aparece la norma ni las operaciones de X . Ello nos lleva a la siguiente definición:

Una **distancia** en un conjunto no vacío E es una función $d : E \times E \rightarrow \mathbb{R}$ verificando las siguientes condiciones:

- (D.1) *Desigualdad triangular* $d(x, z) \leq d(x, y) + d(y, z) \quad \forall x, y, z \in E$
- (D.2) *Simetría:* $d(x, y) = d(y, x) \quad \forall x, y \in E$
- (D.3) *No degeneración.* Para $x, y \in E$, se tiene: $d(x, y) = 0 \iff x = y$

Como hicimos con las normas, destacamos algunas propiedades de todas las distancias:

■ Si $d : E \times E \rightarrow \mathbb{R}$ es una distancia en un conjunto no vacío E , se tiene:

- (i) $d(x, y) \geq 0 \quad \forall x, y \in E$.
- (ii) $|d(x, z) - d(y, z)| \leq d(x, y) \quad \forall x, y, z \in E$.
- (iii) $n \in \mathbb{N}, x_0, x_1, \dots, x_n \in E \implies d(x_0, x_n) \leq \sum_{k=1}^n d(x_{k-1}, x_k)$.

Para (i) basta pensar que $0 = d(x, x) \leq d(x, y) + d(y, x) = 2d(x, y)$. La desigualdad triangular nos da $d(x, z) - d(y, z) \leq d(x, y)$, e intercambiando x con y obtenemos (ii). Finalmente (iii) se prueba por inducción sobre n , de manera obvia. ■

Cuando para un conjunto no vacío E disponemos de una distancia $d : E \times E \rightarrow \mathbb{R}$, decimos que E es un **espacio métrico**. El primer ejemplo debe ser el que hemos usado como motivación. En vista de (9), si $\|\cdot\|$ es una norma en un espacio vectorial X , la función $d : X \times X \rightarrow \mathbb{R}$, definida en (7), es una distancia en X , que llamaremos *distancia asociada a la norma* $\|\cdot\|$ y es la que siempre usaremos en un espacio normado. Dicho de forma más solemne: *todo espacio normado se considera siempre como espacio métrico, con la distancia asociada a su norma*. Nótese que, en virtud de (8), si las distancias asociadas a dos normas en un mismo espacio vectorial son iguales, dichas normas también coinciden.

En \mathbb{R}^N disponemos ya de tantas distancias diferentes como normas hemos estudiado. Para cualesquiera $x, y \in \mathbb{R}^N$, podemos usar la *distancia euclídea*

$$d(x, y) = \left(\sum_{k=1}^N (y(k) - x(k))^2 \right)^{1/2}$$

pero también la *distancia del máximo* d_∞ y la *distancia de la suma* d_1 , asociadas a las normas que llevan el mismo nombre:

$$d_\infty(x, y) = \max \{ |y(k) - x(k)| : k \in \Delta_N \} \quad \text{y} \quad d_1(x, y) = \sum_{k=1}^N |y(k) - x(k)|$$

En el caso $N = 1$ las tres coinciden, tenemos la que llamaremos **distancia usual** de \mathbb{R} , por ser la que siempre consideramos cuando hablamos de \mathbb{R} como espacio métrico, salvo que se diga expresamente lo contrario. Viene dada por:

$$d(x, y) = |y - x| \quad \forall x, y \in \mathbb{R}$$

Conviene resaltar que la noción de espacio métrico es mucho más general que la de espacio normado, cosa que pondremos de manifiesto de varias formas. En primer lugar pensemos que normas y distancias se pueden restringir, para usarlas en contextos más reducidos.

Más concretamente, si X es un espacio normado, e Y es un subespacio vectorial de X , la norma $\|\cdot\|$ de X , como aplicación de X en \mathbb{R} que es, se puede restringir a Y , obteniendo una aplicación $\|\cdot\|_Y : Y \rightarrow \mathbb{R}$, que evidentemente es una norma en el espacio vectorial Y . Suele decirse que $\|\cdot\|_Y$ es la **norma inducida** en Y por la norma de X . También es habitual decir que Y , con la norma $\|\cdot\|_Y$, es un **subespacio normado** de X .

Lo mismo puede hacerse con una distancia, aunque ahora se trate de una función de dos variables. Si d es una distancia en un conjunto E , y A es un subconjunto no vacío de E , la función $d : E \times E \rightarrow \mathbb{R}$ puede restringirse al conjunto $A \times A \subset E \times E$, con lo que se obtiene una función $d_A : A \times A \rightarrow \mathbb{R}$, que evidentemente es una distancia en A . Decimos que d_A es la **distancia inducida** por d en A y también que A con la distancia d_A es un **subespacio métrico** de E . Por ejemplo, un subespacio métrico de \mathbb{R} no es más que un subconjunto no vacío \mathbb{R} con la distancia inducida por la usual de \mathbb{R} .

Resaltemos la mayor generalidad de este segundo proceso. Si X es un espacio normado, luego un espacio métrico, y A es un subconjunto no vacío de X que no sea un subespacio vectorial, no podemos ver A como subespacio normado de X , pero sí como subespacio métrico.

Pero veamos un ejemplo de una distancia, que se puede definir en cualquier conjunto y no guarda relación con norma alguna. Sea E un conjunto no vacío arbitrario y $\delta : E \times E \rightarrow \mathbb{R}$ la función definida, para $x, y \in E$, de la siguiente forma:

$$\delta(x, y) = \begin{cases} 0 & \text{si } x = y \\ 1 & \text{si } x \neq y \end{cases}$$

Se comprueba fácilmente que δ es una distancia, conocida como la **distancia discreta** en el conjunto E . Así pues, cualquier conjunto no vacío puede considerarse como un espacio métrico con la distancia discreta. Por supuesto, el conjunto E puede ser un espacio vectorial y, salvo en el caso trivial $E = \{0\}$, no puede haber una norma en E cuya distancia asociada sea la discreta. Más adelante iremos viendo que la distancia discreta no es tan exótica como a primera vista podría parecer.

1.7. Ejercicios

1. Probar que, en cualquier espacio pre-hilbertiano X , el producto escalar se obtiene a partir de la norma mediante la llamada *identidad de polarización*:

$$4(x|y) = \|x+y\|^2 - \|x-y\|^2 \quad \forall x, y \in X$$

2. Si X e Y son espacios pre-hilbertianos, es una sana costumbre denotar ambos productos escalares por $(\cdot|\cdot)$ y ambas normas asociadas por $\|\cdot\|$. Sea $f : X \rightarrow Y$ una aplicación lineal que preserve la norma, es decir,

$$\|f(x)\| = \|x\| \quad \forall x \in X$$

Probar que entonces f también preserva el producto escalar:

$$(f(u)|f(v)) = (u|v) \quad \forall u, v \in X$$

3. Probar que todo espacio pre-hilbertiano X de dimensión $N \in \mathbb{N}$, se identifica totalmente con el espacio euclídeo N -dimensional, es decir, existe una biyección lineal $f : X \rightarrow \mathbb{R}^N$ que preserve el producto escalar:

$$(f(x)|f(y)) = (x|y) \quad \forall x, y \in X$$

En este sentido podemos decir que el espacio euclídeo N -dimensional es el único espacio pre-hilbertiano de dimensión N .

4. Probar que, en todo espacio pre-hilbertiano X , se verifica la *identidad del paralelogramo*:

$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2 \quad \forall x, y \in X$$

Interpretar geométricamente el resultado.

5. Para cualquier espacio pre-hilbertiano X , discutir la posibilidad de que la desigualdad triangular sea una igualdad, es decir, encontrar la condición necesaria y suficiente que deben cumplir dos vectores $x, y \in X$ para verificar que $\|x + y\| = \|x\| + \|y\|$.
6. Discutir la posibilidad de que la desigualdad triangular para la norma de la suma en \mathbb{R}^N sea una igualdad, es decir, encontrar la condición necesaria y suficiente que deben cumplir dos vectores $x, y \in \mathbb{R}^N$ para verificar que $\|x + y\|_1 = \|x\|_1 + \|y\|_1$.
7. Probar que, para $N > 1$, no existe un producto escalar en \mathbb{R}^N cuya norma asociada sea la de la suma, y que lo mismo le ocurre a la norma del máximo. Probar también que, en el espacio vectorial $C[0, 1]$, las normas $\|\cdot\|_1$ y $\|\cdot\|_\infty$ definidas en las igualdades (6) no son las asociadas a ningún producto escalar.
8. Sea X un espacio vectorial y sean $\mu, \nu : X \rightarrow \mathbb{R}$ dos normas en X . En cada uno de los siguientes casos, probar que la función $\|\cdot\| : X \rightarrow \mathbb{R}$, definida para todo $x \in X$ en la forma que se indica, es una norma en X :

- (a) $\|x\| = \mu(x) + \nu(x)$
- (b) $\|x\| = \max\{\mu(x), \nu(x)\}$
- (c) $\|x\| = (\mu(x)^2 + \nu(x)^2)^{1/2}$

9. Probar que la función $\rho : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ definida por

$$\rho(x, y) = |y - x|^{1/2} \quad \forall x, y \in \mathbb{R}$$

es una distancia en \mathbb{R} .

10. Sean X un espacio normado, Y un espacio vectorial y $f : Y \rightarrow X$ una aplicación lineal e inyectiva. Probar que, definiendo

$$\|y\| = \|f(y)\| \quad \forall y \in Y$$

se obtiene una norma en Y . Establecer un resultado análogo para espacios métricos.

Topología de un espacio métrico

Nuestro próximo objetivo es estudiar ciertas *propiedades topológicas* de un espacio métrico, así llamadas porque sólo dependen de una familia de subconjuntos del espacio que recibe el nombre de *topología*. Empezamos definiendo los elementos de dicha topología, que son los *conjuntos abiertos*, y mencionamos la noción de *espacio topológico*.

Ocurre frecuentemente que dos distancias diferentes en un conjunto dan lugar a la misma topología, en cuyo caso diremos que dichas distancias son *equivalentes*. Prestaremos especial atención a la posible equivalencia entre las distancias asociadas a dos normas en un mismo espacio vectorial, lo que nos llevará a la definición de la *topología usual* de \mathbb{R}^N .

Estudiamos, en un espacio métrico arbitrario, varios conceptos topológicos elementales, que describen la posición relativa de un punto con respecto a un subconjunto del espacio. El *interior* de un conjunto y los *entornos* de un punto, son nociones directamente relacionadas con los conjuntos abiertos. Después definimos los conjuntos *cerrados* y, ligadas a ellos, las nociones de *cierre*, *punto adherente*, *punto de acumulación*, y *frontera* de un conjunto.

Por otra parte, estudiamos la *convergencia de sucesiones*, generalizando la que conocemos en \mathbb{R} . Veremos que la topología de un espacio métrico queda caracterizada por las sucesiones convergentes. En \mathbb{R}^N con la topología usual, ocurre que la convergencia de una sucesión de vectores equivale a la de sus sucesiones de componentes, lo que servirá de motivación para un breve estudio del *producto* de espacios normados, o de espacios métricos.

2.1. Bolas abiertas

En todo lo que sigue trabajaremos en un espacio métrico arbitrario E , cuya distancia se denotará por d , teniendo presente el caso particular que más nos interesa: un espacio normado.

Dados $x \in E$ y $r \in \mathbb{R}^+$, la **bola abierta** de centro x y radio r , que se denota por $B(x, r)$, es el conjunto de puntos de E cuya distancia al centro es estrictamente menor que el radio:

$$B(x, r) = \{y \in E : d(x, y) < r\}$$

Es obvio que $B(x, r) \neq \emptyset$, pues al menos $x \in B(x, r)$, así como que la bola crece cuando, manteniendo su centro, aumentamos el radio: si $0 < r < s$, entonces $B(x, r) \subset B(x, s)$.

Las bolas abiertas en \mathbb{R} son intervalos abiertos acotados:

$$B(x, r) = \{y \in \mathbb{R} : |x - y| < r\} =]x - r, x + r[\quad \forall x \in \mathbb{R}, \forall r \in \mathbb{R}^+$$

Recíprocamente, para $\alpha, \beta \in \mathbb{R}$ con $\alpha < \beta$, tomando $x = (\alpha + \beta)/2$ y $r = (\beta - \alpha)/2$, se tendrá $] \alpha, \beta[= B(x, r)$. Así pues, al definir las bolas abiertas en espacios métricos, estamos generalizando la noción de intervalo abierto acotado, que tan útil resulta para trabajar en \mathbb{R} .

Cuando $E = X$ es un espacio normado, la bola abierta de centro $x \in X$ y radio $r \in \mathbb{R}^+$ viene dada por $B(x, r) = \{y \in X : \|y - x\| < r\}$. Tiene especial interés la *bola abierta unidad*, que es la centrada en el origen, con radio 1: $B(0, 1) = \{u \in X : \|u\| < 1\}$. A partir de ella, mediante homotecias y traslaciones, obtenemos las demás, pues dados $x \in X$ y $r \in \mathbb{R}^+$, vemos fácilmente que $B(x, r) = \{x + ru : u \in B(0, 1)\}$, así que $B(x, r)$ se obtiene a partir de $B(0, 1)$ usando dos aplicaciones de X en sí mismo, primero $u \mapsto ru$, que es la homotecia de razón r , y luego, $y \mapsto x + y$ que es la *traslación* mediante x . Intuitivamente hablando, la forma geométrica de un conjunto no cambia cuando le aplicamos una homotecia o una traslación, luego todas las bolas abiertas de un espacio normado tienen el mismo aspecto.

En \mathbb{R}^2 con la norma euclídea, tenemos que $B(0, 1) = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$ es el conjunto de los puntos rodeados por una circunferencia. En \mathbb{R}^3 tendríamos una esfera, y esto explica que usemos el término “bola” en cualquier espacio métrico. Conviene observar que el aspecto de las bolas en \mathbb{R}^N cambia totalmente cuando usamos una norma distinta de la euclídea. Por ejemplo, con la norma del máximo, para $x \in \mathbb{R}^N$ y $r \in \mathbb{R}^+$, tenemos que

$$B(x, r) = \{y \in \mathbb{R}^N : |y(k) - x(k)| < r \quad \forall k \in \Delta_N\} = \prod_{k=1}^N]x(k) - r, x(k) + r[$$

producto cartesiano de N bolas abiertas en \mathbb{R} , todas con el mismo radio. En \mathbb{R}^2 tenemos el conjunto de puntos situados dentro de un cuadrado de lados paralelos a los ejes, y en \mathbb{R}^3 los situados dentro de un cubo de aristas paralelas a los ejes. Pero más allá de la forma geométrica de las bolas, vemos aquí una ventaja de la norma del máximo frente a la euclídea o cualquier otra: del mismo modo que \mathbb{R}^N es producto cartesiano de rectas, las bolas abiertas de \mathbb{R}^N con la norma del máximo son productos cartesianos de intervalos.

Como un ejemplo más, pensemos en un conjunto no vacío E con la distancia discreta. Para cualesquiera $x \in E$ y $r \in \mathbb{R}^+$ se tiene $B(x, r) = \{x\}$ si $r \leq 1$, y $B(x, r) = E$ si $r > 1$.

Comprobamos finalmente una sencilla propiedad de las bolas abiertas en cualquier espacio métrico. Es geoméricamente muy intuitiva y su utilidad se verá enseguida.

- Sea E un espacio métrico, $a \in E$ y $r \in \mathbb{R}^+$. Entonces, para cada $x \in B(a, r)$ puede encontrarse un $\varepsilon \in \mathbb{R}^+$ tal que $B(x, \varepsilon) \subset B(a, r)$.

Como $d(x, a) < r$, podemos tomar $\varepsilon = r - d(x, a) > 0$, y la desigualdad triangular nos dice que, para todo $y \in B(x, \varepsilon)$, se tiene: $d(y, a) \leq d(y, x) + d(x, a) < \varepsilon + d(x, a) = r$. ■

2.2. Conjuntos abiertos

Sea E un espacio métrico y $U \subset E$. Decimos que U es un **subconjunto abierto** de E , o simplemente un abierto de E , cuando U contiene una bola abierta centrada en cada uno de sus puntos, es decir,

$$\forall x \in U \exists \varepsilon \in \mathbb{R}^+ : B(x, \varepsilon) \subset U$$

Si no es necesario enfatizar el espacio métrico E , podemos decir que U es un conjunto abierto, o simplemente *un abierto*. Es obvio, por ejemplo, que el conjunto vacío y el propio E son abiertos, y la propiedad de las bolas abiertas comprobada anteriormente nos dice que, como su nombre indica, *las bolas abiertas son conjuntos abiertos*.

Recordemos el interior de un conjunto $A \subset \mathbb{R}$. Para $x \in \mathbb{R}$ sabemos que $x \in A^\circ$ cuando existe un $\varepsilon \in \mathbb{R}^+$ tal que $]x - \varepsilon, x + \varepsilon[\subset A$. Es obvio que $A^\circ \subset A$, inclusión que puede ser estricta o ser una igualdad. Vemos ahora que, en el espacio métrico \mathbb{R} , con la distancia usual, un subconjunto U es abierto si, y sólo si, todos sus puntos son interiores, es decir, $U = U^\circ$.

Es claro que existen subconjuntos de \mathbb{R} que no son abiertos. Lo mismo ocurre en la mayoría de los espacios métricos, pero no en todos. Si E es un conjunto no vacío con la distancia discreta, todos los subconjuntos de E son abiertos. En efecto, dado $U \subset E$, para todo $x \in U$ se tiene obviamente $B(x, 1) = \{x\} \subset U$. Queda muy claro que, si cambiamos la distancia de E , algunos abiertos pueden dejar de serlo, y también pueden aparecer nuevos abiertos. Esto no nos debe extrañar, pues si cambiamos la distancia de E , estaremos en un espacio métrico diferente.

Se define la **topología del espacio métrico** E como la familia \mathcal{T} formada por todos los subconjuntos abiertos de E . Si conviene enfatizar la distancia d que estamos usando en E , decimos que \mathcal{T} es la **topología generada por la distancia** d . Tenemos $\mathcal{T} \subset \mathcal{P}(E)$, donde $\mathcal{P}(E)$ es el conjunto de todos los subconjuntos de E . La familia \mathcal{T} tiene tres propiedades básicas:

(A.1) *El conjunto vacío y el conjunto E son abiertos: $\emptyset, E \in \mathcal{T}$*

(A.2) *La unión de cualquier familia de abiertos es un abierto: $\mathcal{S} \subset \mathcal{T} \implies \bigcup \mathcal{S} \in \mathcal{T}$*

(A.3) *La intersección de dos abiertos es siempre un abierto: $U, V \in \mathcal{T} \implies U \cap V \in \mathcal{T}$*

Se ha comentado ya que (A.1) es evidente. Para (A.2) basta pensar que, dado $x \in \bigcup \mathcal{S}$, existirá $S \in \mathcal{S}$ tal que $x \in S$, pero por ser S abierto, existe $\varepsilon > 0$ tal que $B(x, \varepsilon) \subset S \subset \bigcup \mathcal{S}$. Finalmente comprobamos (A.3): si $x \in U \cap V$, existen $\varepsilon_1, \varepsilon_2 \in \mathbb{R}^+$ tales que $B(x, \varepsilon_1) \subset U$ y $B(x, \varepsilon_2) \subset V$, con lo que tomando $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$, tenemos $B(x, \varepsilon) \subset U \cap V$. ■

De (A.2) se deduce una descripción alternativa de los abiertos. Como las bolas abiertas son conjuntos abiertos, (A.2) nos dice que la unión de cualquier familia de bolas abiertas es un abierto. Pero recíprocamente, si $U \neq \emptyset$ es abierto, para cada $x \in U$ tenemos un $\varepsilon_x \in \mathbb{R}^+$ tal que $B(x, \varepsilon_x) \subset U$, y es evidente que $U = \bigcup_{x \in U} B(x, \varepsilon_x)$, una unión de bolas abiertas. Para que no sea una excepción, podemos ver el conjunto vacío como unión de una familia vacía. Así pues, *los abiertos de un espacio métrico son las uniones de bolas abiertas*.

A partir de (A.3), una obvia inducción permite deducir que *la intersección de cualquier familia finita de abiertos es un abierto*:

$$n \in \mathbb{N}, U_1, U_2, \dots, U_n \in \mathcal{T} \implies U_1 \cap U_2 \cap \dots \cap U_n \in \mathcal{T}$$

pero, a diferencia de la unión, en general no podemos asegurar que la intersección de cualquier familia de abiertos sea un conjunto abierto.

Abstrayendo las tres propiedades anteriores, llegamos a la noción de espacio topológico. Una **topología** en un conjunto no vacío Ω es un conjunto $\mathcal{T} \subset \mathcal{P}(\Omega)$, es decir, una familia de subconjuntos de Ω , que verifica las tres condiciones siguientes:

$$(T.1) \quad \emptyset, \Omega \in \mathcal{T}$$

$$(T.2) \quad \mathcal{S} \subset \mathcal{T} \implies \cup \mathcal{S} \in \mathcal{T}$$

$$(T.3) \quad U, V \in \mathcal{T} \implies U \cap V \in \mathcal{T}$$

Resaltamos que en (T.2), la familia \mathcal{S} es arbitraria, así que una topología es estable por uniones arbitrarias, mientras que de (T.3) sólo se deduce que la intersección de una familia finita de elementos de \mathcal{T} pertenece a \mathcal{T} . Pues bien, un **espacio topológico** es un conjunto no vacío Ω en el que hemos fijado una topología \mathcal{T} .

Ejemplo claro de espacio topológico, de hecho el único que vamos a manejar, es el usado como motivación: un espacio métrico E , con la topología generada por su distancia d . Nos interesa especialmente la distancia d asociada a una norma $\|\cdot\|$ en un espacio vectorial X . Decimos entonces que la topología generada por d es la **topología de la norma** $\|\cdot\|$.

En \mathbb{R} tenemos la topología del valor absoluto, o **topología usual** de \mathbb{R} , por ser la topología generada por la distancia usual de \mathbb{R} . Los elementos de esta topología se llaman simplemente *abiertos* de \mathbb{R} , y sabemos que son las uniones de intervalos abiertos acotados.

En cualquier conjunto no vacío Ω , siempre disponemos de la **topología discreta**, $\mathcal{P}(\Omega)$, la más grande posible, que es la generada por la distancia discreta.

Como hemos dicho, no vamos a trabajar con espacios topológicos cualesquiera, ese es el terreno de la Topología General, en el que no vamos a entrar. Pero de un espacio métrico, nos interesan sus muy abundantes *propiedades topológicas*, las que sólo involucran su topología, y no la distancia que la genera. Son las propiedades que podemos definir, o caracterizar, usando sólo los conjuntos abiertos y olvidado la distancia que dio lugar a ellos, por lo que podríamos estudiarlas en cualquier espacio topológico, aunque no lo hagamos.

2.3. Normas equivalentes

Se dice que dos distancias en un conjunto E son **equivalentes**, cuando generan la misma topología, es decir, los conjuntos abiertos para ambas distancias son los mismos. Tenemos así, valga la redundancia, una relación de equivalencia en el conjunto de todas las distancias en E . Decimos que dos normas en un mismo espacio vectorial X son *equivalentes* cuando lo son las distancias asociadas, esto es, cuando las topologías de ambas normas coinciden. La equivalencia entre dos normas se caracteriza de forma muy sencilla, como vamos a ver.

- Para dos normas $\|\cdot\|_1$ y $\|\cdot\|_2$ definidas en un mismo espacio vectorial X , las siguientes afirmaciones son equivalentes:

- (i) Existe una constante $\rho \in \mathbb{R}^+$ tal que $\|x\|_2 \leq \rho \|x\|_1$ para todo $x \in X$.
- (ii) La topología de la norma $\|\cdot\|_2$ está incluida en la de $\|\cdot\|_1$.

Para la demostración, dados $x \in X$ y $r \in \mathbb{R}^+$, denotamos por $B_1(x, r)$ y $B_2(x, r)$ a las bolas abiertas de centro x y radio r para las normas $\|\cdot\|_1$ y $\|\cdot\|_2$, respectivamente.

(i) \Rightarrow (ii). Si U es un conjunto abierto para la norma $\|\cdot\|_2$, para cada $x \in U$ existe $\varepsilon > 0$ tal que $B_2(x, \varepsilon) \subset U$. De (i) deducimos entonces claramente que $B_1(x, \varepsilon/\rho) \subset B_2(x, \varepsilon) \subset U$, luego U es abierto para la norma $\|\cdot\|_1$, como queríamos.

(ii) \Rightarrow (i). Como $B_2(0, 1)$ es abierto para $\|\cdot\|_2$, también lo es para $\|\cdot\|_1$, luego existe $\delta > 0$ tal que $B_1(0, \delta) \subset B_2(0, 1)$. Tomando $\rho = 1/\delta > 0$ conseguimos la desigualdad buscada. En efecto, si $x \in X$ verificase que $\|x\|_2 > \rho \|x\|_1$, tomando $y = x/\|x\|_2$ tendríamos

$$\|y\|_1 = \frac{\|x\|_1}{\|x\|_2} < \frac{1}{\rho} = \delta$$

de donde $\|y\|_2 < 1$, lo cual es una contradicción, puesto que claramente $\|y\|_2 = 1$. Así pues, tenemos $\|x\|_2 \leq \rho \|x\|_1$ para todo $x \in X$, como queríamos. ■

Está claro que dos normas serán equivalentes cuando verifiquen la desigualdad que aparece en la condición (i) anterior, junto con la que se obtendría intercambiando los papeles de ambas normas. Esta otra desigualdad se puede escribir para que enlace con la que ya teníamos, sin más que dividir ambos miembros por la constante que en ella aparezca. Obtenemos así:

- Dos normas $\|\cdot\|_1$ y $\|\cdot\|_2$ en un espacio vectorial X son equivalentes si, y sólo si, existen constantes $\lambda, \rho \in \mathbb{R}^+$ tales que

$$\lambda \|x\|_1 \leq \|x\|_2 \leq \rho \|x\|_1 \quad \forall x \in X$$

Nótese que dos normas proporcionales verifican las desigualdades anteriores con $\lambda = \rho$, luego son equivalentes, como cabía esperar.

2.4. La topología usual de \mathbb{R}^N

Aplicando el criterio antes obtenido, vemos fácilmente que las tres normas que hasta ahora hemos considerado en \mathbb{R}^N son equivalentes:

- En \mathbb{R}^N , la norma euclídea, la de la suma y la del máximo, son equivalentes.

La relación entre la norma del máximo $\|\cdot\|_\infty$ y la de la suma $\|\cdot\|_1$ es evidente:

$$\|x\|_\infty \leq \|x\|_1 \leq N \|x\|_\infty \quad \forall x \in \mathbb{R}^N$$

Para la norma euclídea $\|\cdot\|$, el razonamiento es también evidente, incluso mejorando la segunda desigualdad:

$$\|x\|_\infty \leq \|x\| \leq N^{1/2} \|x\|_\infty \quad \forall x \in \mathbb{R}^N$$

Por supuesto, la norma euclídea y la de la suma son equivalentes, pues hemos visto que ambas son equivalentes a la del máximo. ■

La topología común a las tres normas cuya equivalencia acabamos de comprobar, se conoce como **topología usual** de \mathbb{R}^N , por ser la que siempre se usa en \mathbb{R}^N . A sus elementos se les llama simplemente **abierto** de \mathbb{R}^N . Usando la norma del máximo obtenemos ahora una útil descripción de los mismos.

- Si U_1, U_2, \dots, U_N son abiertos de \mathbb{R} , entonces el producto cartesiano $U = \prod_{k=1}^N U_k$ es un abierto de \mathbb{R}^N . De hecho, todo abierto de \mathbb{R}^N se puede expresar como unión de una familia de productos cartesianos de abiertos de \mathbb{R} .

Para ver que U es abierto fijamos $x = (x_1, x_2, \dots, x_N) \in U$. Para cada $k \in \Delta_N$ tenemos $x_k \in U_k$, luego por ser U_k un abierto de \mathbb{R} , existe $r_k \in \mathbb{R}^+$ tal que $]x_k - r_k, x_k + r_k[\subset U_k$. Tomando entonces $r = \min\{r_k : k \in \Delta_N\}$ y usando en \mathbb{R}^N la norma del máximo, es claro que $B(x, r) \subset U$, y esto prueba que U es abierto. Además, todo abierto de \mathbb{R}^N es unión de bolas abiertas para la norma del máximo que, como sabemos, son productos cartesianos de abiertos de \mathbb{R} . ■

Para definir y manejar cómodamente la que llamaremos topología usual de un subconjunto de \mathbb{R}^N , conviene discutir en general la relación entre la topología de un espacio métrico y la de un subespacio suyo. Volvamos pues a nuestro espacio métrico E cuya distancia d genera una topología \mathcal{T} . Si A es un subespacio métrico de E , tenemos en A la distancia inducida d_A , que genera una topología \mathcal{T}_A , y nos preguntamos por la relación entre \mathcal{T} y \mathcal{T}_A . Nótese que $A \in \mathcal{T}_A$ pero puede ocurrir que $A \notin \mathcal{T}$. Para buscar la relación entre los abiertos de E y los de A , es natural empezar por la relación entre las bolas abiertas en uno y otro espacio.

Dados $a \in A$ y $r \in \mathbb{R}^+$, denotando por $B(a, r)$ y $B_A(a, r)$ a las bolas abiertas de centro a y radio r en los espacios métricos E y A respectivamente, es claro que $B_A(a, r) = B(a, r) \cap A$. A partir de aquí, la relación entre las topologías \mathcal{T} y \mathcal{T}_A se adivina fácilmente:

- Si \mathcal{T} es la topología de un espacio métrico E y \mathcal{T}_A la de un subespacio métrico $A \subset E$, se tiene:

$$\mathcal{T}_A = \{U \cap A : U \in \mathcal{T}\}$$

En particular, cuando $A \in \mathcal{T}$, para $V \subset A$ se tiene que $V \in \mathcal{T}_A$ si, y sólo si, $V \in \mathcal{T}$.

Si $U \in \mathcal{T}$, para cada $a \in U \cap A$ existe $r \in \mathbb{R}^+$ tal que $B(a, r) \subset U$, luego $B_A(a, r) \subset U \cap A$ y esto prueba que $U \cap A \in \mathcal{T}_A$. Recíprocamente, si $V \in \mathcal{T}_A$, para cada $a \in V$ existe $r_a \in \mathbb{R}^+$ tal que $B_A(a, r_a) \subset V$. Tomando entonces $U = \bigcup \{B(a, r_a) : a \in V\}$, tenemos que $U \in \mathcal{T}$, por ser una unión de bolas abiertas, pero es claro que $U \cap A = V$.

Supongamos finalmente que $A \in \mathcal{T}$ y sea $V \subset A$. De $V \in \mathcal{T}$ deducimos $V = V \cap A \in \mathcal{T}_A$. Pero recíprocamente, si $V \in \mathcal{T}_A$, hemos visto que $V = U \cap A$ con $U \in \mathcal{T}$, luego $V \in \mathcal{T}$, por ser la intersección de dos elementos de \mathcal{T} . ■

Obtenemos pues directamente la topología \mathcal{T}_A a partir de \mathcal{T} , olvidando la distancia de E y la inducida en A . Por ello decimos que \mathcal{T}_A es la **topología inducida** por \mathcal{T} en A , lo cual es coherente, pues la distancia inducida genera la topología inducida. Si dos distancias en E son equivalentes, las inducidas en cualquier subespacio también lo son, pues ambas generan la topología inducida.

Destaquemos el caso particular que nos interesa. Si A es un subconjunto no vacío de \mathbb{R}^N , llamaremos **topología usual** de A a la inducida en A por la usual de \mathbb{R}^N . Como hemos visto, los abiertos de A son las intersecciones con A de los abiertos de \mathbb{R}^N , que no tienen por qué ser abiertos de \mathbb{R}^N . Para enfatizar esto, se suele decir que los abiertos de A son **abiertos relativos**. Hemos visto también que, cuando A es un abierto de \mathbb{R}^N , los abiertos relativos no son ni más ni menos que los abiertos de \mathbb{R}^N que estén contenidos en A .

2.5. Interior y entornos

En todo lo que sigue, E es un espacio métrico con distancia d , y \mathcal{T} es la topología generada por d . Las nociones que vamos a estudiar son topológicas, sólo dependen de la topología \mathcal{T} y no se alteran al sustituir la distancia d por otra equivalente.

Todo conjunto $A \subset E$ contiene un abierto, pues al menos $\emptyset \subset A$. Se define el **interior** de A , que se denota por A° , como la unión de todos los abiertos incluidos en A :

$$A^\circ = \bigcup \{U \in \mathcal{T} : U \subset A\}$$

Claramente, A° es abierto y $A^\circ \subset A$. De hecho A° es el *máximo abierto incluido* en A , pues si $U \in \mathcal{T}$ y $U \subset A$, se tiene obviamente $U \subset A^\circ$. Por tanto, A es abierto si, y sólo si, $A = A^\circ$. Cuando $x \in A^\circ$, decimos que x es un **punto interior** de A , o que A es un **entorno** de x , y denotamos por $\mathcal{U}(x)$ al conjunto de todos los entornos de x .

Como A° es abierto, para $x \in A^\circ$ existe $\varepsilon > 0$ tal que $B(x, \varepsilon) \subset A^\circ \subset A$. Recíprocamente, si existe $\varepsilon > 0$ tal que $B(x, \varepsilon) \subset A$, tenemos $x \in A^\circ$, pues $B(x, \varepsilon) \in \mathcal{T}$ y $B(x, \varepsilon) \subset A$. Quedan así caracterizados el interior de un conjunto y los entornos de un punto, en términos de bolas abiertas: $x \in A^\circ \iff A \in \mathcal{U}(x) \iff \exists \varepsilon > 0 : B(x, \varepsilon) \subset A$.

Cuando $E = \mathbb{R}$, tenemos $x \in A^\circ$ si, y sólo si, existe $\varepsilon > 0$ tal que $]x - \varepsilon, x + \varepsilon[\subset A$, luego el concepto de interior en espacios métricos generaliza el que ya conocíamos en \mathbb{R} .

Volviendo al caso general, destacamos dos propiedades de la familia $\mathcal{U}(x)$ de todos los entornos de un punto $x \in E$. Por una parte es obvio que, si $A \in \mathcal{U}(x)$ y $A \subset C \subset E$, entonces se tiene $C \in \mathcal{U}(x)$: contener un entorno de x es lo mismo que ser entorno de x . Por otra, dados dos entornos $A_1, A_2 \in \mathcal{U}(x)$, existen $U_1, U_2 \in \mathcal{T}$ tales que $x \in U_1 \subset A_1$ y $x \in U_2 \subset A_2$, pero entonces tenemos $U_1 \cap U_2 \in \mathcal{T}$ y $x \in U_1 \cap U_2 \subset A_1 \cap A_2$, luego $A_1 \cap A_2 \in \mathcal{U}(x)$. Por inducción, la intersección de cualquier familia finita de entornos de x es un entorno de x . Finalmente, conocer los entornos de todos los puntos de E , equivale a conocer la topología, pues un conjunto es abierto si, y sólo si, es entorno de todos sus puntos.

2.6. Conjuntos cerrados

Dado $C \subset E$, decimos que C es un **subconjunto cerrado** de E , o simplemente un cerrado de E , cuando su complemento $E \setminus C$ es abierto. Si no es preciso enfatizar el espacio métrico E , podemos decir que C es un conjunto cerrado, o simplemente, un cerrado. Las propiedades de la topología se traducen equivalentemente en términos de conjuntos cerrados. Denotando por \mathcal{C} a la familia de todos los subconjuntos cerrados de E , tenemos

$$(1) \emptyset, E \in \mathcal{C}; \quad (2) \mathcal{D} \subset \mathcal{C} \implies \bigcap \mathcal{D} \in \mathcal{C}; \quad (3) C, D \in \mathcal{C} \implies C \cup D \in \mathcal{C}$$

Resaltamos que (2) asegura la estabilidad de \mathcal{C} por intersecciones arbitrarias, mientras de (3) sólo podemos deducir que la unión de cualquier familia *finita* de cerrados es un cerrado.

Todo $A \in \mathcal{P}(E)$ está incluido en un cerrado, al menos $A \subset E$. Se define el **cierre** de A , que se denota por \overline{A} , como la intersección de todos los cerrados en los que A está incluido:

$$\overline{A} = \bigcap \{C \in \mathcal{C} : A \subset C\}$$

Vemos claramente que \overline{A} es cerrado y $A \subset \overline{A}$. De hecho \overline{A} es el *mínimo cerrado que contiene* al conjunto A , pues si C es cerrado y $A \subset C$, se tiene obviamente $\overline{A} \subset C$. Por tanto, A es cerrado si, y sólo si, $A = \overline{A}$. Las operaciones de cierre e interior están claramente relacionadas:

- Para todo subconjunto A de un espacio métrico E , se tiene:

$$E \setminus \overline{A} = (E \setminus A)^\circ \quad y \quad E \setminus A^\circ = \overline{E \setminus A}$$

La comprobación de la primera igualdad es inmediata:

$$\begin{aligned} E \setminus \overline{A} &= E \setminus \left(\bigcap \{C \in \mathcal{C} : A \subset C\} \right) = \bigcup \{E \setminus C : C \in \mathcal{C}, A \subset C\} \\ &= \bigcup \{U \in \mathcal{T} : U \subset E \setminus A\} = (E \setminus A)^\circ \end{aligned}$$

Para la segunda igualdad, basta aplicar la primera al conjunto $E \setminus A$ en lugar de A :

$$E \setminus \overline{E \setminus A} = (E \setminus (E \setminus A))^\circ = A^\circ, \quad \text{luego} \quad \overline{E \setminus A} = E \setminus A^\circ \quad \blacksquare$$

Usando el resultado anterior podemos caracterizar los puntos del cierre de un conjunto A . Para $x \in E$ tenemos $x \in \overline{A}$ si, y sólo si, $E \setminus A$ no es entorno de x . Esto equivale a que $E \setminus A$ no contenga ningún entorno de x , es decir, a que todo entorno de x contenga puntos de A . A su vez, esto equivale a que toda bola abierta de centro x contenga puntos de A . En resumen:

$$x \in \overline{A} \iff U \cap A \neq \emptyset \quad \forall U \in \mathcal{U}(x) \iff B(x, \varepsilon) \cap A \neq \emptyset \quad \forall \varepsilon \in \mathbb{R}^+$$

Cuando esto ocurre decimos que x es un **punto adherente** al conjunto A , así que \overline{A} es el conjunto de todos los puntos adherentes al conjunto A . Esta denominación tiene un significado intuitivo muy claro: x es un punto adherente al conjunto A cuando existen puntos de A tan cerca de x como se quiera.

Veamos algunos ejemplos sencillos de conjuntos cerrados. Es claro que si A tiene un sólo punto, $A = \{x\}$ con $x \in E$, entonces A es cerrado: para $y \in E \setminus \{x\}$, tomando $\varepsilon = d(x, y) > 0$ se tiene $B(y, \varepsilon) \cap A = \emptyset$, luego $y \notin \overline{A}$. En vista de las propiedades de los conjuntos cerrados, deducimos lo siguiente.

- En cualquier espacio métrico E , todo subconjunto finito de E es cerrado.

Como otro ejemplo típico, para $x \in E$ y $r \in \mathbb{R}^+$, la **bola cerrada** de centro x y radio r viene dada por

$$\overline{B}(x, r) = \{y \in E : d(y, x) \leq r\}$$

que ciertamente es un conjunto cerrado, como vamos a comprobar. Si z pertenece a su cierre, para todo $\varepsilon > 0$ existe $y \in B(z, \varepsilon) \cap \overline{B}(x, r)$, con lo que $d(z, x) \leq d(z, y) + d(y, x) < \varepsilon + r$. Deducimos claramente que $d(z, x) \leq r$, como se quería.

Como consecuencia, el conjunto

$$S(x, r) = \{y \in E : d(y, x) = r\} = \overline{B}(x, r) \setminus B(x, r)$$

es cerrado, por ser la intersección de dos cerrados. Se dice que $S(x, r)$ es la **esfera** de centro x y radio r , nomenclatura inspirada en el caso particular de \mathbb{R}^3 con la distancia euclídea.

Veamos otra noción muy intuitiva, relacionada con el cierre. Definimos la **frontera** de un conjunto $A \subset E$, que se denota por $\text{Fr}(A)$, como el conjunto de todos los puntos adherentes al conjunto A que no sean interiores. Tenemos por tanto

$$\text{Fr}(A) = \overline{A} \setminus A^\circ = \overline{A} \cap (E \setminus A^\circ) = \overline{A} \cap \overline{E \setminus A}$$

donde, para la última igualdad, hemos usado la relación entre interior y cierre, ya conocida. Como consecuencia, $\text{Fr}(A)$ es un conjunto cerrado y $\text{Fr}(A) = \text{Fr}(E \setminus A)$. Observamos también que $\overline{A} = A \cup \text{Fr}(A)$ y $A^\circ = A \setminus \text{Fr}(A)$. Por tanto, la topología de E queda determinada cuando conocemos la frontera de cada subconjunto, ya que A es abierto si, y sólo si, $A \cap \text{Fr}(A) = \emptyset$, mientras que A es cerrado si, y sólo si, $\text{Fr}(A) \subset A$. Así pues, A es abierto y cerrado si, y sólo si, $\text{Fr}(A) = \emptyset$, como ocurre cuando $A = \emptyset$ o $A = E$. En general, cada conjunto $A \subset E$ da lugar a una partición del espacio E , más concretamente:

$$E = A^\circ \cup \text{Fr}(A) \cup (E \setminus A)^\circ \quad \text{con} \quad A^\circ \cap \text{Fr}(A) = A^\circ \cap (E \setminus A)^\circ = \text{Fr}(A) \cap (E \setminus A)^\circ = \emptyset$$

2.7. Puntos de acumulación y puntos aislados

Los puntos adherentes a un conjunto A pueden ser de dos tipos excluyentes, que ahora vamos a distinguir. En primer lugar, decimos que $x \in E$ es un **punto de acumulación** de A , cuando x es adherente al conjunto $A \setminus \{x\}$, esto es, $x \in \overline{A \setminus \{x\}}$. Esto significa que todo entorno de x , o toda bola abierta de centro x , contiene puntos de A distintos de x . Denotamos por A' al conjunto de todos los puntos de acumulación de A :

$$x \in A' \iff U \cap (A \setminus \{x\}) \neq \emptyset \quad \forall U \in \mathcal{U}(x) \iff B(x, \varepsilon) \cap (A \setminus \{x\}) \neq \emptyset \quad \forall \varepsilon \in \mathbb{R}^+$$

Pensemos ahora en los puntos adherentes a un conjunto que no sean puntos de acumulación. Tenemos $x \in \overline{A} \setminus A'$ si, y sólo si, existe $U \in \mathcal{U}(x)$ tal que $U \cap A = \{x\}$, o lo que es lo mismo, existe $\varepsilon > 0$ tal que $B(x, \varepsilon) \cap A = \{x\}$, en cuyo caso decimos que x es un **punto aislado** de A .

Resaltamos que los puntos aislados de un conjunto han de pertenecer a dicho conjunto. Por tanto, el conjunto de los puntos aislados de un conjunto A es $\overline{A} \setminus A' = A \setminus A'$. Observamos también que $\overline{A} = A' \cup A$, luego A es cerrado si, y sólo si, $A' \subset A$.

En el caso del espacio métrico $E = \mathbb{R}$, para $A \subset \mathbb{R}$ y $x \in \mathbb{R}$, tenemos $x \in A'$ si, y sólo si, $]x - \varepsilon, x + \varepsilon[\cap (A \setminus \{x\}) \neq \emptyset$ para todo $\varepsilon \in \mathbb{R}^+$, mientras que x es punto aislado de A cuando existe $\varepsilon > 0$ tal que $]x - \varepsilon, x + \varepsilon[\cap A = \{x\}$. Vemos así que en \mathbb{R} con la topología usual, los puntos de acumulación y los puntos aislados de un subconjunto de \mathbb{R} son los que ya conocíamos, como ocurrió con los puntos interiores. Son las tres nociones topológicas que ya habíamos usado en \mathbb{R} y que ahora hemos generalizado para cualquier espacio métrico.

2.8. Sucesiones convergentes

Sabemos que una **sucesión** de elementos de un conjunto $E \neq \emptyset$ es una aplicación $\varphi: \mathbb{N} \rightarrow E$, que se denota por $\{x_n\}$, donde $x_n = \varphi(n)$ para todo $n \in \mathbb{N}$. Las **sucesiones parciales** de $\{x_n\}$ son las de la forma $\{x_{\sigma(n)}\}$ donde $\sigma: \mathbb{N} \rightarrow \mathbb{N}$ es estrictamente creciente. Volviendo a nuestro espacio métrico E , definimos ahora la convergencia de una sucesión $\{x_n\}$ de puntos de E .

Decimos que la sucesión $\{x_n\}$ **converge** a un punto $x \in E$, y escribimos $\{x_n\} \rightarrow x$, cuando cada entorno de x contiene a todos los términos de la sucesión, a partir de uno en adelante:

$$\{x_n\} \rightarrow x \iff [\forall U \in \mathcal{U}(x) \exists m \in \mathbb{N} : n \geq m \Rightarrow x_n \in U] \quad (1)$$

Tenemos claramente una noción topológica, sólo involucra los entornos de x . Por tanto, al hablar de convergencia de sucesiones en un espacio métrico E , no es necesario especificar la distancia concreta d que estemos usando en E , sino solamente la topología que genera. En particular, como en \mathbb{R}^N usamos siempre la topología usual, para una sucesión de vectores de \mathbb{R}^N , podemos hablar de su convergencia, sin ninguna ambigüedad.

Por otra parte, es claro que en (1), en vez de entornos, podemos usar sólo bolas abiertas,

$$\{x_n\} \rightarrow x \iff [\forall \varepsilon > 0 \exists m \in \mathbb{N} : n \geq m \Rightarrow d(x_n, x) < \varepsilon] \quad (2)$$

y si $E = X$ es un espacio normado tendremos

$$\{x_n\} \rightarrow x \iff [\forall \varepsilon > 0 \exists m \in \mathbb{N} : n \geq m \Rightarrow \|x_n - x\| < \varepsilon]$$

Esto se aplicará a \mathbb{R}^N con cualquier norma cuya topología sea la usual. Para $N = 1$ será

$$\{x_n\} \rightarrow x \iff [\forall \varepsilon > 0 \exists m \in \mathbb{N} : n \geq m \Rightarrow |x_n - x| < \varepsilon]$$

luego la única noción de convergencia que vamos a usar en \mathbb{R} es la que ya conocíamos.

Volviendo al caso general, la afirmación que aparece a la derecha de (2) significa, lisa y llanamente, que la sucesión $\{d(x_n, x)\}$ converge a cero. Por tanto, la convergencia de una sucesión de puntos de un espacio métrico, equivale siempre a la convergencia a cero de una concreta sucesión de números reales no negativos:

$$\{x_n\} \rightarrow x \iff \{d(x_n, x)\} \rightarrow 0 \quad (3)$$

Esta equivalencia permite trasladar a cualquier espacio métrico propiedades conocidas de la convergencia en \mathbb{R} , como haremos ahora. Decimos que la sucesión $\{x_n\}$ es **convergente**, cuando existe $x \in E$ tal que $\{x_n\} \rightarrow x$, en cuyo caso x es único. En efecto, si también se tiene que $\{x_n\} \rightarrow y$, puesto que $d(x, y) \leq d(x, x_n) + d(x_n, y)$ para todo $n \in \mathbb{N}$, de (3) deducimos que $d(x, y) = 0$, luego $x = y$. Decimos entonces que x es el **límite** de la sucesión $\{x_n\}$ y escribimos $x = \lim_{n \rightarrow \infty} \{x_n\}$.

Usando también (3), relacionamos la convergencia de una sucesión con la de sus sucesiones parciales. De $\{x_n\} \rightarrow x$ se deduce que toda sucesión parcial de $\{x_n\}$ también converge a x . Vemos igualmente que, fijado $k \in \mathbb{N}$, se tiene

$$\{x_{k+n}\} \rightarrow x \iff \{x_n\} \rightarrow x \iff \{x_{2n-1}\} \rightarrow x \text{ y } \{x_{2n}\} \rightarrow x$$

Por ejemplo, si existen $x \in E$ y $k \in \mathbb{N}$ tales que para $n > k$ se tiene $x_n = x$, entonces $\{x_n\} \rightarrow x$.

2.9. Caracterización secuencial de la topología

Es muy importante observar que la convergencia de sucesiones determina la topología de cualquier espacio métrico:

- *En todo espacio métrico E , un punto $x \in E$ es adherente a un conjunto $A \subset E$ si, y sólo si, existe una sucesión de puntos de A que converge a x .*

Si $x \in \overline{A}$, para cada $n \in \mathbb{N}$ podemos tomar $x_n \in B(x, 1/n) \cap A$, obteniendo una sucesión $\{x_n\}$ de puntos de A tal que $\{d(x_n, x)\} \rightarrow 0$, luego $\{x_n\} \rightarrow x$. Pero recíprocamente, si $\{x_n\} \rightarrow x$ con $x_n \in A$ para todo $n \in \mathbb{N}$, es obvio que $U \cap A \neq \emptyset$ para todo $U \in \mathcal{U}(x)$, luego $x \in \overline{A}$. ■

Deducimos que un conjunto $A \subset E$ es cerrado si, y sólo si, A contiene a los límites de todas las sucesiones de puntos de A que sean convergentes. Así pues, la topología de un espacio métrico queda caracterizada por la convergencia de sucesiones: si conocemos la convergencia de sucesiones, conocemos los conjuntos cerrados, luego conocemos la topología.

Aprovechamos esta idea para conseguir una cómoda caracterización de la equivalencia entre dos distancias:

- *Si d_1 y d_2 son dos distancias en un conjunto E , equivalen las afirmaciones siguientes:*
 - (i) *La topología generada por d_1 está incluida en la generada por d_2 .*
 - (ii) *Toda sucesión convergente para la distancia d_2 es convergente para d_1 .*

Por tanto, d_1 y d_2 son equivalentes si, y sólo si, dan lugar a las mismas sucesiones convergentes.

(i) \Rightarrow (ii). Sea $\{x_n\}$ una sucesión de puntos de E y supongamos que $\{x_n\} \rightarrow x \in E$ para la distancia d_2 . Si U es un entorno de x para la distancia d_1 , aplicando (i) tenemos que U también es entorno de x para d_2 . Por tanto, existe $m \in \mathbb{N}$ tal que $x_n \in U$ para $n \geq m$, y esto nos dice que $\{x_n\} \rightarrow x$ para la distancia d_1 .

(ii) \Rightarrow (i). Si $A \subset E$ es cerrado para d_1 , bastará ver que también lo es para d_2 . Sea pues x un punto adherente al conjunto A para d_2 y veamos que $x \in A$. Por el resultado anterior, existe una sucesión $\{x_n\}$ de puntos de A tal que $\{d_2(x_n, x)\} \rightarrow 0$. Tomamos $y_{2n-1} = x_n$ e $y_{2n} = x$ para todo $n \in \mathbb{N}$, con lo que también tenemos $\{d_2(y_n, x)\} \rightarrow 0$. Aplicando (ii) sabemos que la sucesión $\{y_n\}$ es convergente para la distancia d_1 , pero su límite no puede ser otro que x , puesto que $y_{2n} = x$ para todo $n \in \mathbb{N}$. Así pues, tenemos $\{d_1(y_n, x)\} \rightarrow 0$, de donde deducimos que $\{d_1(x_n, x)\} = \{d_1(y_{2n-1}, x)\} \rightarrow 0$. Por ser A cerrado para d_1 , concluimos que $x \in A$, como se quería. ■

2.10. Convergencia en \mathbb{R}^N

Ha quedado claro que, para conocer la topología usual de \mathbb{R}^N , basta conocer la convergencia de sucesiones de vectores de \mathbb{R}^N , cuyo estudio se puede reducir al de la convergencia en \mathbb{R} . Para ello, ni siquiera necesitamos usar una norma o distancia concreta en \mathbb{R}^N , basta mirar a las componentes de los términos de la sucesión:

- Para toda sucesión $\{x_n\}$ de vectores de \mathbb{R}^N y todo $x \in \mathbb{R}^N$, se tiene:

$$\{x_n\} \rightarrow x \quad \Longleftrightarrow \quad \{x_n(k)\} \rightarrow x(k) \quad \forall k \in \Delta_N$$

Por tanto, $\{x_n\}$ es convergente si, y sólo si, $\{x_n(k)\}$ es convergente para todo $k \in \Delta_N$.

En efecto, basta pensar, por ejemplo, que para $n \in \mathbb{N}$ y $k \in \Delta_N$ se tiene

$$|x_n(k) - x(k)| \leq \|x_n - x\|_\infty \leq \sum_{j=1}^N |x_n(j) - x(j)|$$

Si $\{x_n\} \rightarrow x$, tenemos $\{\|x_n - x\|_\infty\} \rightarrow 0$ y la primera desigualdad nos da $\{x_n(k)\} \rightarrow x(k)$ para todo $k \in \Delta_N$. Recíprocamente, si $\{x_n(k)\} \rightarrow x(k)$ para todo $k \in \Delta_N$, de la segunda igualdad deducimos que $\{\|x_n - x\|_\infty\} \rightarrow 0$, es decir, $\{x_n\} \rightarrow x$. ■

Usaremos a menudo este resultado para trasladar de \mathbb{R} a \mathbb{R}^N diversos resultados acerca de la convergencia de sucesiones. Pero de hecho, podemos usar la misma idea en un contexto más general, como enseguida veremos.

2.11. Producto de espacios normados o métricos

Los mismos procedimientos que, a partir del valor absoluto, nos permitieron definir tres normas equivalentes en \mathbb{R}^N , sirven para definir otras tantas normas en cualquier producto de espacios normados. Por su relación más directa con el producto cartesiano, casi siempre es preferible usar la norma del máximo. Cuando se trabaja con varios espacios normados, es una sana costumbre denotar por $\|\cdot\|$ a las normas de todos ellos. No hay peligro de confusión, pues según cual sea el vector cuya norma usemos en cada momento, estará claro en qué espacio calculamos dicha norma.

Supongamos pues que, para cada $k \in \Delta_N$ tenemos un espacio normado X_k , y consideremos el espacio vectorial producto $X = X_1 \times X_2 \times \dots \times X_N$. Igual que en \mathbb{R}^N , para $x \in X$ y $k \in \Delta_N$ denotamos por $x(k) \in X_k$ a la k -ésima componente de x . Las operaciones de X tienen entonces el mismo aspecto que las de \mathbb{R}^N : para $x, y \in X$, $\lambda \in \mathbb{R}$ y $k \in \Delta_N$, se tiene

$$(x + y)(k) = x(k) + y(k) \quad \text{y} \quad (\lambda x)(k) = \lambda x(k)$$

Para convertir X en un espacio normado, definimos

$$\|x\|_\infty = \max \{ \|x(k)\| : k \in \Delta_N \} \quad \forall x \in X$$

Se comprueba sin ninguna dificultad que $\|\cdot\|_\infty$ es una norma en X . Decimos que X con la norma $\|\cdot\|_\infty$ es el **espacio normado producto** de X_1, X_2, \dots, X_N . También decimos que la topología de la norma $\|\cdot\|_\infty$ es la **topología producto** de las topologías de la norma en los espacios X_1, X_2, \dots, X_N .

Obviamente, la topología usual de \mathbb{R}^N es un ejemplo de topología producto, en el que todos los factores coinciden con \mathbb{R} . Si $M \in \mathbb{N}$ y pensamos que $\mathbb{R}^{N+M} = \mathbb{R}^N \times \mathbb{R}^M$, es claro que la topología usual de \mathbb{R}^{N+M} es el producto de la usual de \mathbb{R}^N por la usual de \mathbb{R}^M . En efecto, si tanto en \mathbb{R}^N como en \mathbb{R}^M consideramos la norma del máximo, la norma del producto $\mathbb{R}^N \times \mathbb{R}^M$ no es otra que la norma del máximo en \mathbb{R}^{N+M} .

Para trabajar con la topología producto, es muy cómodo usar la convergencia de sucesiones, que sabemos caracteriza a dicha topología. El mismo razonamiento que hemos usado en \mathbb{R}^N nos dice que, conocer la convergencia en el producto, equivale a conocerla en los factores:

- Si $X = X_1 \times X_2 \times \dots \times X_N$ es un producto de espacios normados, $\{x_n\}$ una sucesión de vectores de X y $x \in X$, se tiene:

$$\{x_n\} \rightarrow x \quad \Longleftrightarrow \quad \{x_n(k)\} \rightarrow x(k) \quad \forall k \in \Delta_N$$

En efecto, basta tener en cuenta que, para cualesquiera $n \in \mathbb{N}$ y $k \in \Delta_N$, se tiene

$$\|x_n(k) - x(k)\| \leq \|x_n - x\|_\infty \leq \sum_{j=1}^N \|x_n(j) - x(j)\| \quad \blacksquare$$

Todo lo hecho con un producto de espacios normados puede hacerse, de forma análoga, con un producto de espacios métricos, lo repasamos brevemente. Sean E_1, E_2, \dots, E_N espacios métricos cuyas distancias denotamos todas por d . El producto $E = E_1 \times E_2 \times \dots \times E_N$ se convierte en un espacio métrico sin más que definir

$$d_\infty(x, y) = \max \{ d(x(k), y(k)) : k \in I_N \} \quad \forall x, y \in E$$

pues se comprueba sin ninguna dificultad que d_∞ es una distancia en E . Decimos que E es el **espacio métrico producto** de E_1, E_2, \dots, E_N , y que la topología generada por d_∞ es la **topología producto** de las generadas por las distancias de E_1, E_2, \dots, E_N . Si $\{x_n\}$ es una sucesión de puntos de E , y $x \in E$, vemos que $\{x_n\} \rightarrow x$ si, y sólo si, $\{x_n(k)\} \rightarrow x(k)$ para todo $k \in \Delta_N$.

2.12. Ejercicios

1. Probar que, en todo espacio métrico, la distancia queda determinada cuando se conocen las bolas abiertas. En el caso particular de un espacio normado, probar que la norma queda determinada cuando se conoce la bola abierta unidad.
2. Sea X un espacio normado, $x, y \in X$ y $r, \rho \in \mathbb{R}^+$. Probar que:

$$\begin{aligned} \text{(a)} \quad B(x, r) \cap B(y, \rho) \neq \emptyset &\iff \|y - x\| < r + \rho \\ \text{(b)} \quad B(y, \rho) \subset B(x, r) &\iff \|y - x\| \leq r - \rho \end{aligned}$$

¿Son ciertos los resultados análogos en un espacio métrico cualquiera?

3. Dar un ejemplo de una familia numerable de abiertos de \mathbb{R} cuya intersección no sea un conjunto abierto.
4. Si A es un subconjunto no vacío de un espacio métrico E con distancia d , se define la *distancia* de un punto $x \in E$ al conjunto A por

$$d(x, A) = \inf \{ d(x, a) : a \in A \}$$

Probar que $\overline{A} = \{ x \in E : d(x, A) = 0 \}$.

5. Si X un espacio normado, $x \in X$ y $r \in \mathbb{R}^+$, probar que

$$\overline{B(x, r)} = \overline{B}(x, r) \quad \text{y} \quad B(x, r) = [\overline{B}(x, r)]^\circ$$

Deducir que $\text{Fr}(B(x, r)) = \text{Fr}(\overline{B}(x, r)) = S(x, r)$. ¿Son ciertos estos resultados en un espacio métrico cualquiera?

6. Para un intervalo $J \subset \mathbb{R}$, calcular los conjuntos $J^\circ, \overline{J}, J'$ y $\text{Fr } J$.
7. En el espacio métrico \mathbb{R} y para cada uno de los conjuntos $\mathbb{N}, \mathbb{Z}, \mathbb{Q}$ y $\mathbb{R} \setminus \mathbb{Q}$, calcular su interior y su cierre, sus puntos de acumulación, sus puntos aislados y su frontera.
8. Si un subconjunto A de un espacio métrico E verifica que $A' = \emptyset$, probar que la topología inducida por E en A es la discreta. ¿Es cierto el recíproco?
9. Sean $\{x_n\}$ e $\{y_n\}$ sucesiones convergentes en un espacio métrico E con distancia d . Probar que la sucesión $\{d(x_n, y_n)\}$ es convergente y calcular su límite.
10. Sea $E = \prod_{k=1}^N E_k$ un producto de espacios métricos y $A = \prod_{k=1}^N A_k \subset E$, donde $A_k \subset E_k$ para todo $k \in I_N$. Probar que $A^\circ = \prod_{k=1}^N A_k^\circ$ y que $\overline{A} = \prod_{k=1}^N \overline{A_k}$. Deducir que A es un abierto de E si, y sólo si, A_k es un abierto de E_k para todo $k \in \Delta_N$, mientras que A es un cerrado de E si, y sólo si, A_k es un cerrado de E_k para todo $k \in \Delta_N$.

Continuidad y límite funcional

Generalizando lo que conocemos para funciones reales de variable real, vamos a estudiar las nociones de *límite* y *continuidad* para funciones entre dos espacios métricos cualesquiera. Las definimos de forma que quede claro que se trata de nociones topológicas. Analizamos con detalle el *carácter local* de ambas nociones, aclaramos la relación entre ellas y comprobamos que la composición de aplicaciones preserva la continuidad. Al considerar el límite de una composición de funciones, obtenemos una regla de cambio de variable, útil en la práctica para el cálculo de límites. Prestamos especial atención al caso particular de funciones definidas en un subconjunto de \mathbb{R}^N y con valores en \mathbb{R}^M donde $M \in \mathbb{N}$, que se denominan *campos escalares* cuando $M = 1$, o *campos vectoriales* cuando $M > 1$.

3.1. Continuidad en un punto

Para motivar la definición de continuidad, recordemos el caso conocido de una función real de variable real, es decir, una función $f : E \rightarrow \mathbb{R}$, donde E es un subconjunto no vacío de \mathbb{R} . Sabemos que f es continua en un punto $x \in E$, cuando verifica la siguiente condición:

$$\forall \varepsilon > 0 \exists \delta > 0 : y \in E, |y - x| < \delta \Rightarrow |f(y) - f(x)| < \varepsilon \quad (1)$$

Pensemos en \mathbb{R} como espacio métrico con la distancia usual, y en E como subespacio métrico de \mathbb{R} , en el que tenemos la distancia inducida. Usando bolas abiertas en ambos espacios métricos, (1) toma la forma:

$$\forall \varepsilon > 0 \exists \delta > 0 : f(B(x, \delta)) \subset B(f(x), \varepsilon) \quad (2)$$

Como ha ocurrido otras veces, podemos sustituir las bolas abiertas por entornos arbitrarios. Más concretamente, usando los entornos de $f(x)$ en \mathbb{R} y los de x en E , es claro que (2) equivale a

$$\forall V \in \mathcal{U}(f(x)) \exists U \in \mathcal{U}(x) : f(U) \subset V \quad (3)$$

La última inclusión equivale a $U \subset f^{-1}(V)$, con la notación habitual para la imagen inversa de un conjunto por una función: $f^{-1}(V) = \{y \in E : f(y) \in V\}$. Finalmente, que $f^{-1}(V)$ contenga un entorno de x , equivale a que $f^{-1}(V)$ sea entorno de x .

En resumen, considerando tanto en E como en \mathbb{R} la topología usual, hemos visto que f es continua en x si, y sólo si, la imagen inversa por f de cada entorno de $f(x)$ en \mathbb{R} es un entorno de x en E . Tenemos así expresada la continuidad de una forma que sólo involucra entornos en los espacios métricos de partida y llegada de nuestra función. Esta condición es la que tomaremos como definición de continuidad para una función entre dos espacios métricos cualesquiera, definición que podríamos usar también para espacios topológicos. En todo lo que sigue, E y F serán espacios métricos arbitrarios, cuyas distancias se denotan ambas por d .

Decimos que una función $f : E \rightarrow F$ es **continua en un punto** $x \in E$ cuando la imagen inversa por f de cada entorno de $f(x)$ en el espacio F es un entorno de x en E :

$$V \in \mathcal{U}(f(x)) \implies f^{-1}(V) \in \mathcal{U}(x)$$

Puede llamar la atención que sólo consideremos funciones definidas en todo el espacio E y no sólo en un subconjunto suyo. No perdemos generalidad, pues cualquier subconjunto de E es a su vez un espacio métrico con la distancia inducida.

Ha quedado muy claro que la continuidad de una función en un punto es una propiedad topológica. Sin embargo, sustituyendo como siempre entornos por bolas abiertas, tenemos una caracterización de la continuidad en términos de las distancias que estemos usando. Además, como no podía ser de otra forma, tenemos una caracterización *secuencial* de la continuidad, es decir, en términos de convergencia de sucesiones:

■ Para $f : E \rightarrow F$ y $x \in E$, las siguientes afirmaciones son equivalentes:

- (i) f es continua en el punto x
- (ii) $\forall \varepsilon > 0 \exists \delta > 0 : y \in E, d(y, x) < \delta \implies d(f(y), f(x)) < \varepsilon$
- (iii) $x_n \in E \forall n \in \mathbb{N}, \{x_n\} \rightarrow x \implies \{f(x_n)\} \rightarrow f(x)$

(i) \implies (ii). Dado $\varepsilon > 0$, $B(f(x), \varepsilon)$ es entorno de $f(x)$ en F , luego su imagen inversa por f será entorno de x en E , es decir, existe $\delta > 0$ tal que $B(x, \delta) \subset f^{-1}[B(f(x), \varepsilon)]$. Para $y \in E$ con $d(y, x) < \delta$ se tiene entonces $f(y) \in B(f(x), \varepsilon)$, es decir, $d(f(y), f(x)) < \varepsilon$.

(ii) \implies (iii). Para $\varepsilon > 0$, tenemos $\delta > 0$ dado por (ii). Por ser $\{x_n\} \rightarrow x$, existe $m \in \mathbb{N}$ tal que, para $n \geq m$ se tiene $d(x_n, x) < \delta$, luego $d(f(x_n), f(x)) < \varepsilon$. Esto prueba que $\{f(x_n)\} \rightarrow f(x)$.

(iii) \implies (i). Si f no es continua en x , veremos que no se verifica (iii). Existe $V \in \mathcal{U}(f(x))$ tal que $f^{-1}(V) \notin \mathcal{U}(x)$, luego para cada $n \in \mathbb{N}$, $f^{-1}(V)$ no puede contener a la bola abierta de centro x y radio $1/n$, así que existe $x_n \in E$ tal que $d(x_n, x) < 1/n$ pero $f(x_n) \notin V$. Está claro entonces que $\{x_n\} \rightarrow x$ pero $\{f(x_n)\}$ no converge a $f(x)$. ■

Conviene aclarar una cuestión sencilla, que se refiere al espacio de llegada de una función. Si F es subespacio métrico de otro espacio G , una función $f : E \rightarrow F$ puede verse también como función de E en G . Pues bien, la continuidad de f en un punto $x \in E$ no depende de que consideremos F o G como espacio métrico de llegada. Ello es obvio con cualquiera de las condiciones equivalentes del enunciado anterior. Por ejemplo, la segunda sólo involucra las distancias $d(f(y), f(x))$ para ciertos puntos $y \in E$, que son las mismas en F que en G , precisamente porque F es subespacio métrico de G .

Cuestión diferente se plantea cuando cambiamos el espacio métrico de partida, al restringir nuestra función. La relación entre la continuidad de f y la de sus restricciones es la que sigue.

- Sea $f : E \rightarrow F$ una función y sea A un subconjunto no vacío de E , que consideramos como espacio métrico con la distancia inducida. Para $x \in A$ se tiene:

- (i) Si f es continua en x , entonces $f|_A$ es continua en x .
- (ii) Si $f|_A$ es continua en x y A es entorno de x en E , entonces f es continua en x .

(i). Si $V \in \mathcal{U}(f(x))$ sabemos que $f^{-1}(V)$ es un entorno de x en el espacio métrico E , de donde deducimos que $(f|_A)^{-1}(V) = f^{-1}(V) \cap A$ es entorno de x en el espacio métrico A .

(ii). Si $V \in \mathcal{U}(f(x))$, sabemos ahora que $(f|_A)^{-1}(V) = f^{-1}(V) \cap A$ es entorno de x en A , luego $f^{-1}(V) \cap A \supset U \cap A$ donde U es un abierto de E tal que $x \in U$. Entonces $U \cap A$ es entorno de x en E , luego igual le ocurre a $f^{-1}(V)$, pues $U \cap A \subset f^{-1}(V)$. ■

Queda claro que, para decidir si una función es continua en un punto, basta conocerla en un entorno de dicho punto, tan pequeño como se quiera. Por ello decimos que el resultado anterior pone de manifiesto el *carácter local* de la continuidad, en el que enseguida insistiremos.

3.2. Continuidad global

Decimos que una función $f : E \rightarrow F$ es **continua en un conjunto** no vacío $A \subset E$ cuando es continua en todo punto $x \in A$. Si f es continua en E decimos simplemente que f es **continua**. Reunimos en un sólo enunciado tres caracterizaciones de esta propiedad:

- Para cualquier función $f : E \rightarrow F$ las siguientes afirmaciones son equivalentes:
 - (i) f es continua
 - (ii) Para todo abierto $V \subset F$, se tiene que $f^{-1}(V)$ es un abierto de E
 - (iii) Para todo cerrado $C \subset F$, se tiene que $f^{-1}(C)$ es un cerrado de E
 - (iv) f preserva la convergencia de sucesiones: para toda sucesión convergente $\{x_n\}$ de puntos de E , la sucesión $\{f(x_n)\}$ es convergente.

(i) \Rightarrow (ii). Si $V = V^\circ \subset F$ y $x \in f^{-1}(V)$, como $V \in \mathcal{U}(f(x))$, la continuidad de f en x nos dice que $f^{-1}(V) \in \mathcal{U}(x)$, luego $f^{-1}(V)$ es entorno de todos sus puntos, es decir, es abierto.

(ii) \Rightarrow (i). Si $x \in E$ y $W \in \mathcal{U}(f(x))$, existe un abierto V de F , tal que $f(x) \in V \subset W$, con lo que $x \in f^{-1}(V) \subset f^{-1}(W)$. Por (ii) sabemos que $f^{-1}(V)$ es abierto, luego $f^{-1}(W) \in \mathcal{U}(x)$.

(ii) \Rightarrow (iii). Si $C = \overline{C} \subset F$, como $F \setminus C$ es abierto, (ii) nos dice que $f^{-1}(F \setminus C)$ es abierto, pero $f^{-1}(F \setminus C) = E \setminus f^{-1}(C)$, luego $f^{-1}(C)$ es cerrado.

(iii) \Rightarrow (ii). Es enteramente análoga: si $V = V^\circ \subset F$, (iii) nos dice que $f^{-1}(F \setminus V)$ es cerrado, luego $f^{-1}(V)$ es abierto.

(i) \Rightarrow (iv). Es evidente.

(iv) \Rightarrow (i). Sea $x \in E$ y $\{x_n\} \rightarrow x$ con $x_n \in E$ para todo $n \in \mathbb{N}$, para ver que $\{f(x_n)\} \rightarrow f(x)$. Si para cada $n \in \mathbb{N}$ tomamos $y_{2n-1} = x_n$, e $y_{2n} = x$, es claro que $\{y_n\} \rightarrow x$, y (iv) nos dice que $\{f(y_n)\}$ es convergente. Como $f(y_{2n}) = f(x)$ para todo $n \in \mathbb{N}$, se tiene $\{f(y_n)\} \rightarrow f(x)$, luego $\{f(x_n)\} = \{f(y_{2n-1})\} \rightarrow f(x)$, como queríamos. ■

Las caracterizaciones de la continuidad, dadas por las condiciones (ii) y (iii), se usan a menudo para probar que ciertos subconjuntos de E son abiertos o cerrados. Por ejemplo, tomando $F = \mathbb{R}$ observamos que, si $f : E \rightarrow \mathbb{R}$ es una función continua, entonces los conjuntos

$$\{x \in E : f(x) > 0\} = f^{-1}(\mathbb{R}^+) \quad \text{y} \quad \{x \in E : f(x) < 0\} = f^{-1}(\mathbb{R}^-)$$

son abiertos, mientras que el conjunto $\{x \in E : 0 \leq f(x) \leq 1\} = f^{-1}([0, 1])$ es cerrado.

Volvamos a la relación entre la continuidad de una función y la de sus restricciones, pero pensando en la continuidad en un conjunto no vacío $A \subset E$. Es obvio que, si $f : E \rightarrow F$ es continua en A , entonces $f|_A$ es continua. En general, el recíproco es falso, pero sí se verifica cuando A es abierto. En efecto, para cada $x \in A$ tenemos que A es entorno de x , y de ser $f|_A$ continua en x se deduce que f es continua en x . Así pues:

- Si A es un subconjunto abierto no vacío de E , una función $f : E \rightarrow F$ es continua en A si, y sólo si, $f|_A$ es continua.

En la práctica, esta observación se usa frecuentemente para estudiar la continuidad de una función que viene definida mediante una disyuntiva. Un ejemplo sencillo es el siguiente:

- Supongamos que $E = U \cup V$ donde U y V son subconjuntos abiertos de E . Entonces, una función $f : E \rightarrow F$ es continua si, y sólo si, $f|_U$ y $f|_V$ son continuas.

Resaltamos que la continuidad de f , no sólo en un punto, sino en todo el espacio E , tiene **carácter local**, en el sentido de que se puede comprobar con sólo conocer f *localmente*: f es continua si, y sólo si, para cada $x \in E$ existe $U \in \mathcal{U}(x)$ tal que $f|_U$ es continua.

3.3. Límite funcional y su relación con la continuidad

Recordemos la definición de límite en un punto para una función real de variable real. Dada una función $f : A \rightarrow \mathbb{R}$ donde $\emptyset \neq A \subset \mathbb{R}$, y dados $\alpha \in A'$ y $L \in \mathbb{R}$, tenemos

$$\lim_{x \rightarrow \alpha} f(x) = L \iff [\forall \varepsilon > 0 \exists \delta > 0 : x \in A, 0 < |x - \alpha| < \delta \Rightarrow |f(x) - L| < \varepsilon]$$

Está muy claro cómo podemos reformular la afirmación anterior de forma que tenga sentido para una función entre espacios métricos cualesquiera. Seguimos trabajando con dos espacios métricos E y F cuyas distancias se denotan por d , pero ahora tendremos una función $f : A \rightarrow F$ donde A es un subconjunto no vacío de E . No nos limitamos al caso $A = E$, para poder hablar de límite de f en puntos de $E \setminus A$, que deberán ser puntos de acumulación de A .

Así pues, dado $\alpha \in A'$, decimos que f **tiene límite** en el punto α cuando existe $L \in F$ verificando la siguiente condición:

$$\forall \varepsilon > 0 \exists \delta > 0 : x \in A, 0 < d(x, \alpha) < \delta \Rightarrow d(f(x), L) < \varepsilon \quad (4)$$

Comprobaremos enseguida que entonces L es único, le llamamos **límite** de f en el punto α y escribimos $\lim_{x \rightarrow \alpha} f(x) = L$.

En efecto, si $L_1, L_2 \in F$ verifican (4), dado $\varepsilon > 0$ podemos claramente encontrar $\delta > 0$ tal que, para $x \in A$ con $0 < d(x, \alpha) < \delta$ se tiene $d(f(x), L_1) < \varepsilon$ y también $d(f(x), L_2) < \varepsilon$. Como $\alpha \in A'$, existe efectivamente $x \in A$ con $0 < d(x, \alpha) < \delta$, y usando un tal x , deducimos que $d(L_1, L_2) \leq d(L_1, f(x)) + d(f(x), L_2) < 2\varepsilon$, desigualdad que es válida para todo $\varepsilon > 0$. Tenemos por tanto $d(L_1, L_2) = 0$, es decir, $L_1 = L_2$. Nótese que la condición $\alpha \in A'$ es la que permite asegurar la unicidad del límite.

La primera observación sobre el concepto de límite funcional es muy clara: para que tenga sentido hablar de límite de la función f en un punto $\alpha \in A'$ no es necesario que f esté definida en el punto α y, aún cuando $\alpha \in A \cap A'$, el valor que tome f en α no afecta para nada a la existencia del límite, ni al valor de dicho límite, caso de que exista.

En segundo lugar, el límite funcional es una *propiedad topológica*. La definición utiliza las distancias de E y F , pero se puede reformular fácilmente para que sólo aparezcan sus topologías. Basta sustituir en (4), como hemos hecho otras veces, la bola abierta $B(L, \varepsilon)$ del espacio F por un entorno V de L , y la bola abierta $B(\alpha, \delta)$ por un entorno U de α en el espacio E . Por otra parte, el límite funcional se puede caracterizar en términos de convergencia de sucesiones. En resumen, tenemos las siguientes equivalencias, cuya demostración es muy similar a la que se hizo para la continuidad.

$$\begin{aligned} \lim_{x \rightarrow \alpha} f(x) = L &\iff \forall V \in \mathcal{U}(L) \exists U \in \mathcal{U}(\alpha) : f(U \cap (A \setminus \{\alpha\})) \subset V \\ &\iff [x_n \in A \setminus \{\alpha\} \forall n \in \mathbb{N}, \{x_n\} \rightarrow \alpha \Rightarrow \{f(x_n)\} \rightarrow L] \end{aligned}$$

Resaltemos el *carácter local* del límite funcional: para saber si nuestra función $f : A \rightarrow F$ tiene límite en un punto $\alpha \in A'$, basta conocerla “cerca” de α . Concretamente, fijemos $r \in \mathbb{R}^+$ arbitrario y consideremos el conjunto $B = \{x \in A : 0 < d(x, \alpha) < r\}$, que verifica $\alpha \in B'$. Es claro que, para $L \in F$, se tiene $\lim_{x \rightarrow \alpha} f(x) = L$ si, y sólo si, $\lim_{x \rightarrow \alpha} f|_B(x) = L$.

Las nociones de límite y continuidad guardan entre sí la misma relación que conocíamos para funciones reales de variable real. En primer lugar, cuando a es punto aislado de A , es decir, $a \in A \setminus A'$, no tiene sentido hablar de límite de f en a , pero f siempre es continua en el punto a . Basta observar que el conjunto $\{a\}$ es entorno del punto a en el espacio métrico A . En lo que sigue estudiamos las otras dos situaciones posibles.

- Para $a \in A \cap A'$ se tiene que f es continua en a si, y sólo si, $\lim_{x \rightarrow a} f(x) = f(a)$.

El límite indicado consiste en que, fijado $\varepsilon > 0$, exista $\delta > 0$ tal que se tenga $d(f(x), f(a)) < \varepsilon$ siempre que $x \in A$ verifique $0 < d(x, a) < \delta$. Para la continuidad, la única diferencia es que no se excluye el caso $x = a$, pero en ese caso, la desigualdad $d(f(x), f(a)) < \varepsilon$ es obvia. ■

- Si $\alpha \in A' \setminus A$, entonces f tiene límite en el punto α si, y sólo si, se puede definir una función $g : A \cup \{\alpha\} \rightarrow F$ que es continua en el punto α y verifica que $g(x) = f(x)$ para todo $x \in A$. En tal caso se tiene $g(\alpha) = \lim_{x \rightarrow \alpha} f(x)$, y en particular g es única.

Si existe tal función g se tiene claramente $\lim_{x \rightarrow \alpha} f(x) = \lim_{x \rightarrow \alpha} g(x) = g(\alpha)$, donde hemos usado el resultado anterior, pues g es continua en α . Esto prueba una implicación y la última afirmación del enunciado. Recíprocamente, si f tiene límite en el punto α , basta definir $g(\alpha) = \lim_{x \rightarrow \alpha} f(x)$ y $g(x) = f(x)$ para todo $x \in A$, para tener la extensión de f que es continua en el punto α . ■

3.4. Composición de funciones: cambio de variable

Como otra propiedad básica de la continuidad, comprobamos enseguida que se conserva al componer dos funciones:

- Sean G , E y F espacios métricos, consideremos dos funciones $\varphi : G \rightarrow E$ y $f : E \rightarrow F$, y su composición $f \circ \varphi : G \rightarrow F$. Si φ es continua en un punto $z \in G$ y f es continua en el punto $x = \varphi(z)$, entonces $f \circ \varphi$ es continua en z . Por tanto, si φ y f son continuas, entonces $f \circ \varphi$ es continua.

La comprobación es evidente: para $W \in \mathcal{U}[(f \circ \varphi)(z)] = \mathcal{U}(f(x))$, la continuidad de f en el punto x nos dice que $f^{-1}(W) \in \mathcal{U}(x) = \mathcal{U}(\varphi(z))$, con lo que la continuidad de φ en z nos dice que $\varphi^{-1}(f^{-1}(W)) = (f \circ \varphi)^{-1}(W) \in \mathcal{U}(z)$, como queríamos. ■

Usaremos muy frecuentemente el resultado anterior, pues para funciones de varias variables, la composición tiene más utilidad si cabe, que en el caso de una variable. El resultado análogo para el límite de una composición de funciones es una útil regla para el cálculo de límites.

- Sean E y F espacios métricos, A un subconjunto no vacío de E , $f : A \rightarrow F$ una función y $\alpha \in E$. Sea ahora T un subconjunto no vacío de otro espacio métrico G , $\varphi : T \rightarrow E$ una función y $z \in T'$. Supongamos que se cumplen las siguientes dos condiciones:

$$\lim_{t \rightarrow z} \varphi(t) = \alpha \quad \text{y} \quad \varphi(t) \in A \setminus \{\alpha\} \quad \forall t \in T \setminus \{z\} \quad (5)$$

Entonces $\alpha \in A'$ y se verifica la siguiente implicación:

$$\lim_{x \rightarrow \alpha} f(x) = L \in F \implies \lim_{t \rightarrow z} f(\varphi(t)) = L \quad (6)$$

Sea $\{t_n\}$ una sucesión de puntos de $T \setminus \{z\}$ tal que $\{t_n\} \rightarrow z$. Tenemos entonces por hipótesis que $\{\varphi(t_n)\}$ es una sucesión de puntos de $A \setminus \{\alpha\}$ tal que $\{\varphi(t_n)\} \rightarrow \alpha$. Esto nos dice de entrada que $\alpha \in A'$, pero además, de $\lim_{x \rightarrow \alpha} f(x) = L$, deducimos que $\{f(\varphi(t_n))\} \rightarrow L$. Esto prueba que $\lim_{t \rightarrow z} f(\varphi(t)) = L$, como queríamos. ■

La principal utilidad del resultado anterior consiste en que nos permite estudiar la existencia del límite de una función, usándolo como una *regla de cambio de variable*, pensando que el límite que aparece a la derecha de la implicación (6) se deduce del que aparece a la izquierda, mediante el cambio de variable $x = \varphi(t)$, siempre que φ verifique las condiciones (5). Para concretar, supongamos que queremos estudiar la existencia del límite de f en un punto $\alpha \in A'$. Podemos entonces elegir libremente la función φ , cuidando que se verifique (5), y estudiar la existencia del límite en z de la función $f \circ \varphi$. Dada la libertad que tenemos para elegir φ , lo haremos de forma que la función $f \circ \varphi$ sea más fácil de estudiar que f . Lo más habitual es tomar $G = \mathbb{R}$ y T un intervalo, con lo que $f \circ \varphi$ es una función de variable real, típicamente más sencilla que f . Veamos entonces la forma de usar el resultado anterior:

- Si φ verifica (5) y $f \circ \varphi$ no tiene límite en z , concluimos que f no tiene límite en α , pues en otro caso (6) nos llevaría a contradicción.

- Si por el contrario tenemos $\lim_{t \rightarrow z} f(\varphi(t)) = L \in F$, no podemos concluir que L sea el límite de f en α , pues (6) es sólo una implicación, no una equivalencia. Sin embargo, si sabemos que L es el *único posible límite* de f en α , pues si se tuviera $\lim_{x \rightarrow \alpha} f(x) = L' \in F$ con $L' \neq L$, (6) nos llevaría de nuevo a contradicción.
- En particular, si encontramos dos funciones φ_1 y φ_2 , ambas verificando (5), tales que tanto $f \circ \varphi_1$ como $f \circ \varphi_2$ tienen límite en el punto z , pero dichos límites no coinciden, está claro que f tampoco tiene límite en α .

Siempre que usamos (6), solemos decir que hacemos un **cambio de variable**. Deberemos concretar la función φ que estamos usando y comprobar (5). Todo ello se puede indicar de forma breve diciendo que *usamos el cambio de variable $x = \varphi(t) \in A$ con $t \in T$, teniendo en cuenta que $x \rightarrow \alpha$ cuando $t \rightarrow z$ y que $x \neq \alpha$ para $t \neq z$* . Casi nunca son necesarias tantas precisiones, podemos omitir todo aquello que se deduzca claramente del contexto.

3.5. Primeros ejemplos de funciones continuas

Para dos espacios métricos arbitrarios E y F , ejemplo obvio de funciones continuas son las *constantes*. Si $y_0 \in F$ y $f(x) = y_0$ para todo $x \in E$, la continuidad de f es obvia.

En general, puede no haber más ejemplos, pues tanto E como F pueden reducirse a un punto, pero aún excluyendo estos casos triviales, puede ocurrir que toda función continua de E en F sea constante.

Por ejemplo, supongamos que F es un subconjunto de \mathbb{R} tal que $F^\circ = \emptyset$, de forma que F no contiene intervalos no triviales. Considerando en \mathbb{R} la distancia usual y en F la inducida, toda función continua $f: \mathbb{R} \rightarrow F$ es constante. En efecto, viendo f como función de \mathbb{R} en \mathbb{R} , podemos aplicar el teorema del valor intermedio, obteniendo que $f(\mathbb{R})$ es un intervalo. Pero como $f(\mathbb{R}) \subset F$, deducimos que $f(\mathbb{R})$ se reduce a un punto, es decir, f es constante. Así pues, viendo a \mathbb{Q} como subespacio métrico de \mathbb{R} , toda función continua $f: \mathbb{R} \rightarrow \mathbb{Q}$ es constante.

Supongamos que E es un subespacio métrico de F . Entonces la *inclusión* de E en F , definida por $I(x) = x$ para todo $x \in E$, es continua. En efecto, para todo abierto $V \subset F$, se tiene que $I^{-1}(V) = V \cap E$ es un abierto de E . En el caso particular $E = F$, vemos que la función *identidad* en cualquier espacio métrico E es continua.

La *distancia* d , de todo espacio métrico E , es un ejemplo importante de función continua, de $E \times E$ en \mathbb{R} , entendiendo que $E \times E$ es el espacio métrico producto. Para comprobarlo, observamos que si $x, y, u, v \in E$, se tiene

$$|d(u, v) - d(x, y)| \leq |d(u, v) - d(x, v)| + |d(x, v) - d(x, y)| \leq d(u, x) + d(v, y)$$

Dado $\varepsilon > 0$, y teniendo en cuenta la definición de la distancia d_∞ en $E \times E$, vemos que basta tomar $d_\infty((u, v), (x, y)) < \varepsilon/2$ para tener $|d(u, v) - d(x, y)| < \varepsilon$.

Para cualquier espacio normado X , conviene destacar tres funciones continuas, ligadas a su estructura. En primer lugar, de la desigualdad

$$|\|y\| - \|x\|| \leq \|y - x\| \quad \forall x, y \in X$$

se deduce claramente que la *norma* de X es una función continua $\|\cdot\| : X \rightarrow \mathbb{R}$.

En segundo lugar, la función *suma*, $(x, y) \mapsto x + y$ es una función continua, definida en el espacio normado producto $X \times X$ y con valores en X . Ello se deduce claramente de la siguiente desigualdad, válida para cualesquiera $x, y, u, v \in X$:

$$\|(u + v) - (x + y)\| \leq \|u - x\| + \|v - y\| \leq 2\|(u, v) - (x, y)\|_\infty$$

Lo mismo ocurre con el *producto por escalares*, $(\lambda, x) \mapsto \lambda x$, que es una función continua en el espacio normado producto $\mathbb{R} \times X$ con valores en X . Lo comprobamos usando ahora la caracterización secuencial de la continuidad. Si $\{(\lambda_n, x_n)\}$ es una sucesión en $\mathbb{R} \times X$ que converge a $(\lambda, x) \in \mathbb{R} \times X$, es decir, con $\{\lambda_n\} \rightarrow \lambda$ y $\{x_n\} \rightarrow x$, debemos ver que $\{\lambda_n x_n\} \rightarrow \lambda x$. Ello se deduce claramente de la siguiente desigualdad, válida para todo $n \in \mathbb{N}$:

$$\begin{aligned} \|\lambda_n x_n - \lambda x\| &\leq |\lambda_n| \|x_n - x\| + |\lambda_n - \lambda| \|x\| \\ &\leq |\lambda_n - \lambda| \|x_n - x\| + |\lambda| \|x_n - x\| + |\lambda_n - \lambda| \|x\| \end{aligned}$$

3.6. Funciones con valores en un producto

Encontraremos con frecuencia funciones que toman valores en un producto de espacios métricos o de espacios normados, por lo que conviene aclarar algunas cuestiones básicas sobre este tipo de funciones.

Si $M \in \mathbb{N}$ y $F = F_1 \times F_2 \times \dots \times F_M$ es un producto cartesiano de conjuntos no vacíos, para cada $k \in \Delta_M$ tenemos una aplicación sobreyectiva $\pi_k : F \rightarrow F_k$, que a cada M -upla $y \in F$ hace corresponder su k -ésima componente, esto es, $\pi_k(y) = y(k)$ para todo $y \in F$. Suele decirse que estas funciones son las **proyecciones coordenadas** del producto F sobre cada uno de sus factores, o más concretamente, π_k es la k -ésima proyección coordenada, para todo $k \in \Delta_M$.

Si E es otro conjunto no vacío y $f : E \rightarrow F$ una función, para cada $k \in \Delta_M$ podemos considerar la composición $f_k = \pi_k \circ f$, y decimos que $f_k : E \rightarrow F_k$ es la k -ésima **componente** de f , de modo que f tiene M componentes f_1, f_2, \dots, f_M que la determinan mediante la obvia igualdad siguiente:

$$f(x) = (f_1(x), f_2(x), \dots, f_M(x)) \quad \forall x \in E$$

Es por ello que escribimos $f = (f_1, f_2, \dots, f_M)$ para indicar las componentes de una función que toma valores en un producto cartesiano.

Pues bien, si F_1, F_2, \dots, F_M son espacios métricos cuyas distancias se denotan todas por d y consideramos en F la distancia d_∞ del espacio métrico producto, es claro que las proyecciones coordenadas son funciones continuas, pues para todo $k \in \Delta_M$ tenemos de hecho la desigualdad

$$d(\pi_k(y), \pi_k(z)) \leq d_\infty(y, z) \quad \forall y, z \in F$$

Si E es otro espacio métrico y una función $f: E \rightarrow F$ es continua en un punto $x \in E$, la regla sobre la continuidad de una composición nos dice que la k -ésima componente f_k también es continua en x , para todo $k \in \Delta_M$. Recíprocamente, si todas las componentes de f son continuas en x , dada una sucesión $\{x_n\}$ de puntos de E tal que $\{x_n\} \rightarrow x$, la sucesión $\{y_n\} = \{f(x_n)\}$, de puntos de F , verifica que $\{y_n(k)\} = \{f_k(x_n)\} \rightarrow f_k(x)$ para todo $k \in \Delta_M$, luego $\{y_n\} \rightarrow f(x)$ y concluimos que f es continua en el punto x . Así pues, la continuidad de f equivale a la de sus componentes:

- Sea E un espacio métrico y $F = F_1 \times F_2 \times \dots \times F_M$ un producto de espacios métricos. Una función $f = (f_1, f_2, \dots, f_M): E \rightarrow F$ es continua en un punto $x \in E$ si, y sólo si, f_k es continua en x , para todo $k \in \Delta_M$.

Enunciamos el resultado análogo para el límite funcional, cuya demostración es muy similar a la que hemos hecho para la continuidad.

- Sea E un espacio métrico y $F = F_1 \times F_2 \times \dots \times F_M$ un producto de espacios métricos. Sea A un subconjunto no vacío de E , $f = (f_1, f_2, \dots, f_M): A \rightarrow F$ una función y $\alpha \in A'$. Entonces, para $y \in F$ se tiene:

$$\lim_{x \rightarrow \alpha} f(x) = y \quad \Longleftrightarrow \quad \lim_{x \rightarrow \alpha} f_k(x) = y(k) \quad \forall k \in \Delta_M$$

3.7. Operaciones con funciones continuas

Vamos a comentar algunas operaciones algebraicas que podemos hacer con funciones y a comprobar que tales operaciones preservan la continuidad. Si E, Y son conjuntos no vacíos, denotamos por $\mathcal{F}(E, Y)$ al conjunto de todas las funciones de E en Y . Si $Y = \mathbb{R}$, escribimos simplemente $\mathcal{F}(E)$ en lugar de $\mathcal{F}(E, \mathbb{R})$.

Cuando Y es un espacio vectorial, $\mathcal{F}(E, Y)$ también lo es, con la **suma y producto por escalares** definidos de manera natural:

$$\begin{aligned} (f+g)(x) &= f(x) + g(x) & \forall x \in E & \quad \forall f, g \in \mathcal{F}(E, Y) \\ (\lambda g)(x) &= \lambda g(x) & \forall x \in E & \quad \forall \lambda \in \mathbb{R} \quad \forall g \in \mathcal{F}(E, Y) \end{aligned}$$

De hecho tenemos una operación más general que el producto por escalares. En vez del escalar $\lambda \in \mathbb{R}$ podemos usar una función $\Lambda \in \mathcal{F}(E)$. Entonces, para $g \in \mathcal{F}(E, Y)$ podemos considerar la función **producto** $\Lambda g \in \mathcal{F}(E, Y)$ dada por

$$(\Lambda g)(x) = \Lambda(x)g(x) \quad \forall x \in E$$

En particular, cuando $Y = \mathbb{R}$ hemos definido el producto de dos funciones $f, g \in \mathcal{F}(E)$ que nos da una función $fg \in \mathcal{F}(E)$. Tenemos por tanto un producto en $\mathcal{F}(E)$ que, junto con la suma antes definida, convierte a $\mathcal{F}(E)$ en un anillo conmutativo con unidad. Podemos finalmente considerar el **cociente** de dos funciones $f, g \in \mathcal{F}(E)$, siempre que $g(x) \neq 0$ para todo $x \in E$. Naturalmente, dicho cociente es la función $f/g \in \mathcal{F}(E)$ dada por

$$\left(\frac{f}{g}\right)(x) = \frac{f(x)}{g(x)} \quad \forall x \in E$$

En el contexto adecuado, todas las operaciones anteriores preservan la continuidad:

- Sea E un espacio métrico e Y un espacio normado. Si $f, g \in \mathcal{F}(E, Y)$ y $\Lambda \in \mathcal{F}(E)$ son funciones continuas en un punto $x \in E$, entonces $f + g$ y Λg son continuas en x . En el caso $Y = \mathbb{R}$, si $g(E) \subset \mathbb{R}^*$, entonces f/g es continua en x .

Todo se comprueba sin dificultad usando la convergencia de sucesiones, pero merece la pena comentar un razonamiento alternativo, más elegante.

Consideramos las funciones $\Phi = (f, g) : E \rightarrow Y \times Y$ y $\Psi = (\Lambda, g) : E \rightarrow \mathbb{R} \times Y$, es decir,

$$\Phi(x) = (f(x), g(x)) \quad \text{y} \quad \Psi(x) = (\Lambda(x), g(x)) \quad \forall x \in E$$

Denotemos por $\sigma : Y \times Y \rightarrow Y$ a la suma, y $\tau : \mathbb{R} \times Y \rightarrow Y$ al producto por escalares, del espacio vectorial Y , es decir, $\sigma(y, z) = y + z$ y $\tau(\lambda, y) = \lambda y$ para cualesquiera $y, z \in Y$, $\lambda \in \mathbb{R}$. Es claro entonces que $f + g = \sigma \circ \Phi$ y $\Lambda g = \tau \circ \Psi$.

Como f , g y Λ son continuas en x , también lo son Φ y Ψ , pero sabemos que σ y τ son funciones continuas, luego las composiciones $f + g$ y Λg son continuas en x .

En el caso $Y = \mathbb{R}$, para el cociente f/g , observamos que $\Phi(E) \subset \mathbb{R} \times \mathbb{R}^*$ y podemos escribir $f/g = q \circ \Phi$ donde $q : \mathbb{R} \times \mathbb{R}^* \rightarrow \mathbb{R}$ viene dada por $q(y, z) = y/z$ para $(y, z) \in \mathbb{R} \times \mathbb{R}^*$. Es claro que q es continua, luego f/g es continua en x , por serlo Φ . ■

Si E es un espacio métrico e Y un espacio normado, denotamos por $\mathcal{C}(E, Y)$ al subconjunto de $\mathcal{F}(E, Y)$ formado por las funciones continuas de E en Y . De nuevo escribimos $\mathcal{C}(E)$ en lugar de $\mathcal{C}(E, \mathbb{R})$. Por el resultado anterior, tenemos:

- $\mathcal{C}(E, Y)$ es un subespacio vectorial de $\mathcal{F}(E, Y)$. Además, $\mathcal{C}(E)$ es un subanillo de $\mathcal{F}(E)$. Si $f, g \in \mathcal{C}(E)$ y $g(E) \subset \mathbb{R}^*$, entonces $f/g \in \mathcal{C}(E)$.

El resultado anterior sobre operaciones con funciones continuas tiene una versión análoga para el límite funcional, que nos da las reglas básicas para calcular límites de funciones. Lo enunciamos brevemente y omitimos su demostración, que no tiene dificultad.

- Sea E un espacio métrico, $A \subset E$ y $\alpha \in A'$. Sea Y un espacio normado y consideremos tres funciones $f, g : A \rightarrow Y$ y $\Lambda : A \rightarrow \mathbb{R}$ que tengan límite en el punto α , es decir,

$$\lim_{x \rightarrow \alpha} f(x) = y \in Y, \quad \lim_{x \rightarrow \alpha} g(x) = z \in Y \quad \text{y} \quad \lim_{x \rightarrow \alpha} \Lambda(x) = \lambda \in \mathbb{R}$$

Se tiene entonces que:

$$\lim_{x \rightarrow \alpha} (f + g)(x) = y + z \quad \text{y} \quad \lim_{x \rightarrow \alpha} (\Lambda f)(x) = \lambda y$$

En particular, cuando $Y = \mathbb{R}$ se tiene que $\lim_{x \rightarrow \alpha} (fg)(x) = yz$. Finalmente, también en el caso $Y = \mathbb{R}$, si $g(A) \subset \mathbb{R}^*$ y $z \in \mathbb{R}^*$ se tiene:

$$\lim_{x \rightarrow \alpha} \left(\frac{f}{g} \right)(x) = \frac{y}{z}$$

3.8. Campos escalares y vectoriales

Prestemos ahora atención a las funciones que más nos interesan. Un **campo escalar** es una función real de N -variables reales, es decir, una función $f : A \rightarrow \mathbb{R}$, donde A es un subconjunto no vacío de \mathbb{R}^N . Una función de N -variables reales que tome valores vectoriales, es decir, una función $f : A \rightarrow \mathbb{R}^M$ con $M > 1$, recibe el nombre de **campo vectorial**. Naturalmente, a la hora de discutir límites y continuidad de un campo escalar o vectorial, se trabaja siempre con la topología usual, tanto en el conjunto de definición A como en el espacio de llegada, \mathbb{R} o \mathbb{R}^M .

Como caso particular de las operaciones antes estudiadas vemos que, fijados $M \in \mathbb{N}$ y un conjunto no vacío $A \subset \mathbb{R}^N$, el conjunto $\mathcal{F}(A, \mathbb{R}^M)$ de todos los campos vectoriales definidos en A y con valores en \mathbb{R}^M , es un espacio vectorial, del que los campos vectoriales continuos forman el subespacio vectorial $\mathcal{C}(A, \mathbb{R}^M)$.

Cada campo vectorial $f \in \mathcal{F}(A, \mathbb{R}^M)$ tiene M componentes, $f = (f_1, f_2, \dots, f_M)$, que son campos escalares. Recordemos que, para cada $k \in \Delta_M$, se tiene $f_k = \pi_k \circ f$, donde $\pi_k : \mathbb{R}^M \rightarrow \mathbb{R}$ es la k -ésima proyección coordenada, es decir, $\pi_k(y) = y(k)$ para todo $y \in \mathbb{R}^M$. Sabemos que, para estudiar límites y continuidad de un campo vectorial, basta trabajar con sus componentes. Concretamente, f es continuo en un punto $a \in A$, o tiene límite en un punto $\alpha \in A'$ si, y sólo si, lo mismo le ocurre a f_k para todo $k \in \Delta_M$. Esta equivalencia se repetirá con todas las propiedades de los campos vectoriales que trataremos más adelante.

Recordemos que el conjunto $\mathcal{F}(A)$ de los campos escalares definidos en A no es sólo un espacio vectorial, sino también un anillo, en el que los campos escalares continuos forman el subanillo y subespacio vectorial $\mathcal{C}(A)$.

Tras las funciones constantes, los ejemplos más sencillos de campos escalares continuos son las restricciones al conjunto A de las proyecciones coordenadas en \mathbb{R}^N . Podemos ahora considerar el subanillo de $\mathcal{F}(A)$ engendrado por las constantes y esas restricciones, que se denota por $\mathcal{P}(A)$. Cuando $f \in \mathcal{P}(A)$ decimos que f es una **función polinómica** en A . Como hemos incluido en $\mathcal{P}(A)$ a las funciones constantes, vemos que $\mathcal{P}(A)$ también es subespacio vectorial de $\mathcal{F}(A)$. Además, como $\mathcal{C}(A)$ es un subanillo de $\mathcal{F}(A)$, tenemos que $\mathcal{P}(A) \subset \mathcal{C}(A)$. Hemos obtenido así abundantes ejemplos de campos escalares continuos, que conviene describir con más detalle.

Es claro que cada función polinómica $f \in \mathcal{P}(A)$ ha de venir definida por

$$f(x) = \sum_{j_1, j_2, \dots, j_N=0}^p \lambda_{j_1 j_2 \dots j_N} x_1^{j_1} x_2^{j_2} \dots x_N^{j_N} \quad \forall x = (x_1, x_2, \dots, x_N) \in A$$

donde $p \in \mathbb{N} \cup \{0\}$ y, para cualesquiera $j_1, j_2, \dots, j_N \in \{0, 1, 2, \dots, p\}$, el coeficiente $\lambda_{j_1 j_2 \dots j_N}$ es un número real.

De manera más general podemos considerar cocientes de funciones polinómicas. Se dice que $h : A \rightarrow \mathbb{R}$ es una **función racional** en A , si existen funciones polinómicas $f, g \in \mathcal{P}(A)$ tales que $g(x) \neq 0$ y $h(x) = f(x)/g(x)$ para todo $x \in A$. Denotamos por $\mathcal{R}(A)$ al conjunto de las funciones racionales en A y tenemos claramente:

$$\mathcal{P}(A) \subset \mathcal{R}(A) \subset \mathcal{C}(A) \subset \mathcal{F}(A)$$

donde cada conjunto de funciones es subanillo y subespacio vectorial de los que le siguen.

3.9. Ejercicios

1. Sean E y F espacios métricos y $f : E \rightarrow F$ una función. Probar que f es continua si, y sólo si, $f(\overline{A}) \subset \overline{f(A)}$ para todo conjunto $A \subset E$.
2. Dado un subconjunto A de un espacio métrico E , la *función característica* de A es la función $\chi_A : E \rightarrow \mathbb{R}$ definida por:

$$\chi_A(x) = 1 \quad \forall x \in A \quad \text{y} \quad \chi_A(x) = 0 \quad \forall x \in E \setminus A$$

Probar que χ_A es continua en un punto $x \in E$ si, y sólo si, $x \in A^\circ \cup (E \setminus A)^\circ$. Deducir que χ_A es continua si, y sólo si, A es a la vez abierto y cerrado.

3. Si E y F son espacios métricos, se dice que una función $f : E \rightarrow F$ es *localmente constante* cuando, para cada $x \in E$, existe $U \in \mathcal{U}(x)$ tal que $f|_U$ es constante. Probar que entonces f es continua. Dar un ejemplo de un conjunto $A \subset \mathbb{R}$ y una función localmente constante $f : A \rightarrow \mathbb{R}$, cuya imagen $f(A)$ sea un conjunto infinito.
4. Sea E un espacio métrico con distancia d y A un subconjunto no vacío de E . Probar la continuidad de la función $f : E \rightarrow \mathbb{R}$ dada por

$$f(x) = d(x, A) \stackrel{\text{def}}{=} \inf \{d(x, a) : a \in A\} \quad \forall x \in E$$

5. Sea E un espacio métrico con distancia d , y consideremos el espacio producto $E \times E$. Probar que, para todo $r \in \mathbb{R}_0^+$, el conjunto $\{(x, y) \in E \times E : d(x, y) < r\}$ es abierto, mientras que $\{(x, y) \in E \times E : d(x, y) \leq r\}$ es cerrado. En particular se tiene que la *diagonal* $\Delta(E) = \{(x, x) : x \in E\}$ es un conjunto cerrado. Deducir que, si F es otro espacio métrico y $f, g : E \rightarrow F$ son funciones continuas, entonces $\{x \in E : f(x) = g(x)\}$ es un subconjunto cerrado de E .
6. Sean E, F espacios métricos y $f : E \rightarrow F$ una función continua. Probar que su *gráfica*, es decir, el conjunto $\text{Gr } f = \{(x, f(x)) : x \in E\}$, es un subconjunto cerrado del espacio métrico producto $E \times F$.
7. Sea E un espacio métrico e Y un espacio pre-hilbertiano. Para $f, g \in \mathcal{F}(E, Y)$, se define una función $h \in \mathcal{F}(E)$ por $h(x) = (f(x) | g(x))$ para todo $x \in E$. Probar que, si f y g son continuas en un punto $a \in E$, entonces h también lo es.
8. Sea E un espacio métrico y $f, g : E \rightarrow \mathbb{R}$ funciones continuas en un punto $a \in E$. Probar que la función $h : E \rightarrow \mathbb{R}$ definida por $h(x) = \max \{f(x), g(x)\}$ para todo $x \in E$, también es continua en a .
9. Probar que si Y es un espacio normado, E un espacio métrico y $f : E \rightarrow Y$ una función continua en un punto $a \in E$, entonces la función $g : E \rightarrow \mathbb{R}$ definida por $g(x) = \|f(x)\|$ para todo $x \in E$, también es continua en el punto a .
10. Probar las siguientes igualdades:

$$(a) \quad \lim_{(x,y) \rightarrow (0,0)} \frac{\sin(x^2 + y^2)}{x^2 + y^2} = 1 \quad (b) \quad \lim_{(x,y) \rightarrow (0,0)} \frac{\log(1 + x^4 + y^4)}{x^4 + y^4} = 1$$

Compacidad y conexión

Estudiamos ahora dos propiedades clave de las funciones continuas entre espacios métricos, que generalizan sendos teoremas bien conocidos para funciones reales de variable real: el de Bolzano o del valor intermedio, acerca de la imagen de un intervalo por una función continua, y el de Weierstrass para el caso particular de un intervalo cerrado y acotado.

Para generalizar el segundo de estos teoremas, empezamos por estudiar, para subconjuntos de un espacio métrico, la noción de *acotación*. A diferencia de otras que hemos manejado hasta ahora, la acotación no es una propiedad topológica, no se conserva al cambiar la distancia del espacio por otra equivalente. Sin embargo, sí se conserva en un espacio normado, al cambiar la norma por otra equivalente, lo que permite manejarla cómodamente en \mathbb{R}^N . Probamos sin dificultad la versión para \mathbb{R}^N del teorema de Bolzano-Weierstrass, afirmando que toda sucesión acotada de vectores de \mathbb{R}^N admite una sucesión parcial convergente.

Para llegar a una visión más abstracta y general del teorema de Weierstrass, estudiamos la noción de *compacidad*, más fuerte que la acotación. Esta vez sí se trata de una propiedad topológica, pero no usaremos la noción general de compacidad en espacios topológicos, sino otra propiedad mucho más sencilla que, para espacios métricos, es equivalente a la compacidad. La preservación de la compacidad por funciones continuas es la versión abstracta del teorema de Weierstrass que vamos buscando.

Como consecuencia fundamental de dicha versión, probaremos un teorema de Hausdorff, afirmando que todas las normas en un espacio vectorial de dimensión finita son equivalentes. Esto clarifica definitivamente la definición de la topología usual de \mathbb{R}^N , no sólo es la topología de la norma euclídea, la de la suma o la del máximo, es la topología de cualquier norma que queramos usar en \mathbb{R}^N . Más aún, se puede hablar sin ninguna ambigüedad de dicha topología en cualquier espacio vectorial de dimensión N , pues no depende de la base que usemos para identificar tal espacio con \mathbb{R}^N .

Si el análisis general del teorema de Weierstrass nos ha llevado al estudio de la compacidad, el teorema del valor intermedio nos lleva directamente a otra importante propiedad topológica, la *conexión*, que también estudiamos brevemente. Como una condición suficiente, muy intuitiva, para que un subconjunto de un espacio normado sea conexo, usamos una propiedad puramente algebraica como es la *convexidad*.

4.1. Acotación en espacios métricos

La acotación, un concepto muy útil para trabajar en \mathbb{R} , nace ligada al orden de \mathbb{R} , y de hecho se puede hablar de acotación en cualquier conjunto ordenado. Pero también podemos caracterizar la acotación mediante la distancia usual de \mathbb{R} , lo que permite hablar de acotación en cualquier espacio métrico. Concretamente, las bolas abiertas en \mathbb{R} son los intervalos abiertos acotados, luego un subconjunto de \mathbb{R} está acotado si, y sólo si, está incluido en una bola abierta, o lo que es lo mismo, en una bola cerrada. Queda así motivada la definición que sigue.

Se dice que un subconjunto de un espacio métrico está **acotado**, cuando está incluido en una bola. Observamos enseguida que el centro de dicha bola puede elegirse a voluntad:

- Si A es un subconjunto acotado de un espacio métrico E , entonces, para todo $x \in E$ existe $r \in \mathbb{R}^+$ tal que $A \subset B(x, r)$.

Sean $x_0 \in E$ y $r_0 \in \mathbb{R}^+$ tales que $A \subset B(x_0, r_0)$. Si d es la distancia de E , dado $x \in E$, para todo $a \in A$ tenemos

$$d(a, x) \leq d(a, x_0) + d(x_0, x) < r_0 + d(x_0, x)$$

luego $A \subset B(x, r)$ sin más que tomar $r = r_0 + d(x_0, x)$. ■

Es obvio que todo subconjunto finito de un espacio métrico E , está acotado. El siguiente paso es pensar en un subconjunto numerable de E , o lo que es lo mismo, en el conjunto de los términos de una sucesión de puntos de E . Naturalmente, si $x_n \in E$ para todo $n \in \mathbb{N}$, decimos que $\{x_n\}$ es una **sucesión acotada** cuando el conjunto $\{x_n : n \in \mathbb{N}\}$ está acotado. Esto equivale claramente a que exista $z \in E$ tal que la sucesión $\{d(x_n, z)\}$ esté acotada. Ahora es obvia la relación entre sucesiones convergentes y acotadas: si $\{x_n\} \rightarrow x \in E$, de $\{d(x_n, x)\} \rightarrow 0$ deducimos que la sucesión $\{d(x_n, x)\}$ está acotada. Así pues:

- En cualquier espacio métrico, toda sucesión convergente está acotada.

Por supuesto, el recíproco es falso: en cualquier espacio métrico que contenga al menos dos puntos, es fácil dar ejemplos de sucesiones acotadas que no son convergentes.

Es importante resaltar que la acotación no es una propiedad topológica: cuando sustituimos la distancia de un espacio métrico por otra equivalente, un subconjunto acotado puede dejar de serlo, y viceversa, como vamos a ver.

Ejemplo. Dos distancias equivalentes que no dan lugar a los mismos conjuntos acotados.

Si d es una distancia cualquiera en un conjunto no vacío E , podemos definir otra distancia ρ en E de la siguiente forma:

$$\rho(x, y) = \frac{d(x, y)}{1 + d(x, y)} \quad \forall x, y \in E \quad (1)$$

Comprobemos que ρ es una distancia en E , equivalente a d . De entrada, para $x, y \in E$ es evidente que $\rho(x, y) = \rho(y, x)$, mientras que $\rho(x, y) = 0$ si, y sólo si, $x = y$.

Para la desigualdad triangular usamos que la función $t \mapsto t/(1+t)$, de \mathbb{R}_0^+ es sí mismo, es creciente. Entonces, para $x, y, z \in E$ tenemos

$$\begin{aligned} \rho(x, z) &= \frac{d(x, z)}{1 + d(x, z)} \leq \frac{d(x, y) + d(y, z)}{1 + d(x, y) + d(y, z)} \\ &\leq \frac{d(x, y)}{1 + d(x, y)} + \frac{d(y, z)}{1 + d(y, z)} = \rho(x, y) + \rho(y, z) \end{aligned}$$

Observamos además que

$$\rho(x, y) < 1 \quad \text{y} \quad d(x, y) = \frac{\rho(x, y)}{1 - \rho(x, y)} \quad \forall x, y \in E \quad (2)$$

En vista de (1) y (2) vemos claramente que si $\{x_n\}$ es una sucesión de puntos de E y $x \in E$, entonces $\{d(x_n, x)\} \rightarrow 0$ si, y sólo si, $\{\rho(x_n, x)\} \rightarrow 0$. Así pues, las distancias d y ρ dan lugar a las mismas sucesiones convergentes, luego son equivalentes.

Finalmente, es claro que todo subconjunto de E está acotado para la distancia ρ , cosa que no tiene por qué ocurrir para d . Por ejemplo, d podría ser la distancia usual de \mathbb{R} . ■

4.2. Teorema de Bolzano-Weierstrass

En un espacio normado X usamos siempre, como ya hemos dicho, la distancia asociada a su norma. Entonces, un conjunto $A \subset X$ está acotado si, y sólo si, está contenido en una bola de centro 0, lo que equivale a que A esté **acotado en norma**, es decir, a que el conjunto de números reales $\{\|x\| : x \in A\}$ esté mayorado:

$$A \text{ acotado} \iff \exists M > 0 : \|x\| \leq M \quad \forall x \in A$$

Si $\|\cdot\|'$ es otra norma en X , y existe $\rho > 0$ tal que $\|x\|' \leq \rho \|x\|$ para todo $x \in X$, está claro que todo conjunto acotado para la norma $\|\cdot\|$ lo estará también para $\|\cdot\|'$. Cuando ambas normas son equivalentes, podemos intercambiarlas, y obtenemos:

- *Dos normas equivalentes en un espacio vectorial dan lugar a los mismos subconjuntos acotados.*

En particular, en \mathbb{R}^N podemos hablar sin ambigüedad de subconjuntos acotados, son los acotados para cualquier norma cuya topología sea la usual de \mathbb{R}^N . Podemos averiguar si un subconjunto de \mathbb{R}^N está acotado, viendo lo que ocurre en cada coordenada. De hecho, lo mismo es cierto en cualquier producto de espacios normados:

- Sean X_1, X_2, \dots, X_N espacios normados y $X = \prod_{k=1}^N X_k$ el espacio normado producto. Un subconjunto $A \subset X$ está acotado si, y sólo si, el conjunto $A_k = \{x(k) : x \in A\}$ está acotado, para todo $k \in \Delta_N$.

Si A está acotado tenemos $M \in \mathbb{R}^+$ tal que $\|x\|_\infty \leq M$ para todo $x \in A$. Fijado un $k \in \Delta_N$, para todo $x \in A$ se tiene $\|x(k)\| \leq \|x\|_\infty \leq M$, luego A_k está acotado. Recíprocamente, supongamos que, para cada $k \in \Delta_N$, el conjunto A_k está acotado y sea $M_k \in \mathbb{R}^+$ tal que $\|x(k)\| \leq M_k$ para todo $x \in A$. Tomando $M = \max \{M_k : k \in \Delta_N\}$, es claro que $\|x\|_\infty \leq M$ para todo $x \in A$, con lo que A está acotado. ■

Como caso particular del último resultado, para una sucesión $\{x_n\}$ de vectores de \mathbb{R}^N , tenemos:

$$\{x_n\} \text{ está acotada} \iff \{x_n(k)\} \text{ está acotada } \forall k \in \Delta_N$$

Para sucesiones convergentes teníamos la equivalencia análoga, luego podemos ya adivinar que el principal resultado sobre convergencia de sucesiones de números reales también será cierto en \mathbb{R}^N , pero debemos salvar un pequeño obstáculo.

Si $\{x_n\}$ es una sucesión acotada de vectores de \mathbb{R}^N , para $k \in \Delta_N$ la sucesión de números reales $\{x_n(k)\}$ admite una sucesión parcial convergente. El problema es que, en principio, dicha sucesión parcial se obtiene mediante una aplicación $\sigma_k : \mathbb{N} \rightarrow \mathbb{N}$ que depende de k , y no tenemos una sola sucesión parcial $\{x_{\sigma(n)}\}$ tal que $\{x_{\sigma(n)}(k)\}$ sea convergente *para todo* $k \in \Delta_N$, para concluir que $\{x_{\sigma(n)}\}$ es convergente. La forma de resolver este problema se puede adivinar: extraer sucesiones parciales en un proceso iterativo, de forma que en cada paso se consigue la convergencia en una nueva componente, manteniendo la de las anteriores. Este proceso se entiende mejor si razonamos por inducción.

Teorema de Bolzano-Weierstrass. *Toda sucesión acotada de vectores de \mathbb{R}^N admite una sucesión parcial convergente.*

Demostración. Razonamos por inducción sobre N . La etapa base está clara, sabemos que el teorema es cierto para $N = 1$. Suponiendo que es cierto en \mathbb{R}^N , lo demostramos en \mathbb{R}^{N+1} . Sea pues $\{x_n\}$ una sucesión acotada de vectores de \mathbb{R}^{N+1} .

Para cada $n \in \mathbb{N}$, escribiendo $y_n(k) = x_n(k)$ para todo $k \in \Delta_N$ tenemos $y_n \in \mathbb{R}^N$, y hemos conseguido así una sucesión $\{y_n\}$ de vectores de \mathbb{R}^N , que evidentemente está acotada. Por la hipótesis de inducción $\{y_n\}$ admite una sucesión parcial convergente $\{y_{\varphi(n)}\}$. De esta forma tenemos que $\{x_{\varphi(n)}(k)\} = \{y_{\varphi(n)}(k)\}$ converge para todo $k \in \Delta_N$.

Ahora $\{x_{\varphi(n)}(N+1)\}$ es una sucesión acotada de números reales, que a su vez tendrá una sucesión parcial convergente $\{x_{\varphi(\tau(n))}(N+1)\}$. Definiendo $\sigma = \varphi \circ \tau$ es claro que $\sigma : \mathbb{N} \rightarrow \mathbb{N}$ es estrictamente creciente, luego $\{x_{\sigma(n)}\}$ es una sucesión parcial de $\{x_n\}$. Por una parte sabemos que $\{x_{\sigma(n)}(N+1)\}$ es convergente, pero por otra, para todo $k \in \Delta_N$, la sucesión $\{x_{\sigma(n)}(k)\}$ también converge, pues se trata de una sucesión parcial de $\{x_{\varphi(n)}(k)\}$, que era convergente. En resumen, $\{x_{\sigma(n)}(k)\}$ es convergente para todo $k \in \Delta_{N+1}$, luego $\{x_{\sigma(n)}\}$ converge. ■

Es un gran error pensar que el teorema de Bolzano-Weierstrass pueda ser cierto en cualquier espacio normado, no digamos ya en cualquier espacio métrico. De hecho se puede probar que *en todo espacio normado de dimensión infinita existe una sucesión acotada que no admite ninguna sucesión parcial convergente.*

4.3. Compacidad

El teorema de Bolzano-Weierstrass motiva una propiedad muy deseable para un espacio métrico, que ahora vamos a estudiar. Se dice que un espacio métrico E es **compacto** cuando toda sucesión de puntos de E admite una sucesión parcial convergente. Se trata claramente de una propiedad topológica, pues se define mediante la convergencia de sucesiones, pero conviene aclarar que ésta no es la noción de compacidad que se usa en espacios topológicos generales, sino otra que no vamos a explicar. Ambas nociones no son equivalentes en general, pero sí lo son para espacios métricos.

Si E es un espacio métrico y $A \subset E$, diremos que A es un subconjunto compacto de E cuando A sea un espacio métrico compacto con la distancia inducida por la de E . Claramente, esto significa que toda sucesión de puntos de A admite una sucesión parcial que converge a un punto de A . Enseguida encontramos dos condiciones necesarias para que A sea compacto:

- *Sea A un subconjunto compacto de un espacio métrico E . Entonces A está acotado y es un subconjunto cerrado de E .*

Suponiendo que A no está contenido en ninguna bola, llegaremos a contradicción. Fijado un punto cualquiera $x \in E$, para cada $n \in \mathbb{N}$ existe $x_n \in A$ tal que $d(x_n, x) > n$. Entonces $\{x_n\}$ admite una sucesión parcial $\{x_{\sigma(n)}\}$ que converge a un punto $a \in A$, es decir, $\{d(x_{\sigma(n)}, a)\} \rightarrow 0$. Pero $d(x_{\sigma(n)}, x) \leq d(x_{\sigma(n)}, a) + d(a, x)$ para todo $n \in \mathbb{N}$, luego la sucesión $\{d(x_{\sigma(n)}, x)\}$ está mayorada, lo cual es una contradicción, ya que $d(x_{\sigma(n)}, x) \geq \sigma(n) \geq n$ para todo $n \in \mathbb{N}$.

Sea ahora $x \in \overline{A}$ y $\{a_n\}$ una sucesión de puntos de A con $\{a_n\} \rightarrow x$. Por ser A compacto, tenemos una sucesión parcial $\{a_{\sigma(n)}\}$ que converge a un $a \in A$, pero también $\{a_{\sigma(n)}\} \rightarrow x$, luego $x = a \in A$. Esto prueba que $\overline{A} \subset A$, es decir, A es cerrado. ■

En general, las dos condiciones necesarias para la compacidad que acabamos de encontrar están muy lejos de ser suficientes. Sin embargo, para subconjuntos de \mathbb{R}^N , usando el teorema de Bolzano-Weierstrass, sí conseguimos el recíproco del resultado anterior:

- *Un subconjunto de \mathbb{R}^N es compacto si, y sólo si, es cerrado y acotado.*

Ya hemos visto que una implicación es válida, no sólo en \mathbb{R}^N , sino en cualquier espacio métrico. Para el recíproco, sea A un subconjunto cerrado y acotado de \mathbb{R}^N . Toda sucesión $\{a_n\}$ de puntos de A es una sucesión acotada de vectores de \mathbb{R}^N , y el teorema de Bolzano-Weierstrass nos dice que $\{a_n\}$ admite una sucesión parcial $\{a_{\sigma(n)}\}$ que converge a un vector $x \in \mathbb{R}^N$. Como A es cerrado, tenemos $x \in A$, lo que prueba que toda sucesión de puntos de A admite una sucesión parcial que converge a un punto de A , es decir, A es compacto. ■

Recordemos el teorema de Weierstrass que conocemos para funciones reales de variable real. Teniendo en cuenta lo que acabamos de probar, dicho teorema afirmaba que la imagen de un intervalo compacto $[a, b] \subset \mathbb{R}$ por una función continua $f : [a, b] \rightarrow \mathbb{R}$ también es un subconjunto compacto de \mathbb{R} . Conseguiremos ahora muy fácilmente un resultado mucho más general.

Teorema. Sean E y F dos espacios métricos y $f : E \rightarrow F$ una función continua. Si E es compacto, entonces $f(E)$ es compacto.

Demostración. Dada una sucesión $\{y_n\}$ de puntos de $f(E)$, deberemos probar que $\{y_n\}$ admite una sucesión parcial que converge a un punto de $f(E)$. Para cada $n \in \mathbb{N}$, existirá un punto $x_n \in E$ tal que $f(x_n) = y_n$. Como, por hipótesis, E es un espacio métrico compacto, la sucesión $\{x_n\}$ admitirá una sucesión parcial $\{x_{\sigma(n)}\}$ que converge a un punto $x \in E$. Por ser f continua, deducimos que $\{f(x_{\sigma(n)})\} \rightarrow f(x)$, es decir, $\{y_{\sigma(n)}\} \rightarrow f(x) \in f(E)$. ■

Merece la pena destacar una consecuencia importante, que se obtiene tomando $F = \mathbb{R}$. Observamos que todo conjunto compacto no vacío $K \subset \mathbb{R}$ tiene máximo y mínimo. En efecto, sabemos que K está acotado y entonces, tanto el supremo como el ínfimo de K son puntos adherentes a K , pero K es cerrado, luego ambos pertenecen a K , así que K tiene máximo y mínimo. Si ahora E es un espacio métrico compacto y $f : E \rightarrow \mathbb{R}$ es continua, el teorema anterior nos dice que $f(E)$ es compacto, luego tiene máximo y mínimo. Esto es tanto como decir que f tiene un máximo absoluto y un mínimo absoluto en sendos puntos de E :

- Si E es un espacio métrico compacto y $f : E \rightarrow \mathbb{R}$ una función continua, existen $u, v \in E$ tales que $f(u) \leq f(x) \leq f(v)$ para todo $x \in E$.

El resultado anterior se aplica obviamente al caso en que E es un subconjunto compacto, es decir, cerrado y acotado, de \mathbb{R}^N .

4.4. Teorema de Hausdorff

La equivalencia entre la norma euclídea, la del máximo y la de la suma, comprobada en su momento, no es más que un caso muy particular del siguiente resultado:

Teorema (Hausdorff, 1932). Todas las normas en \mathbb{R}^N son equivalentes.

Demostración. Bastará ver que toda norma $\|\cdot\|$ en \mathbb{R}^N es equivalente a la norma de la suma $\|\cdot\|_1$. Sea $\{e_k : k \in \Delta_N\}$ la base usual de \mathbb{R}^N y tomemos $\rho = \max \{\|e_k\| : k \in \Delta_N\}$. Por ser $\|\cdot\|$ una norma, para todo $x \in \mathbb{R}^N$ tenemos

$$\|x\| = \left\| \sum_{k=1}^N x(k) e_k \right\| \leq \sum_{k=1}^N |x(k)| \|e_k\| \leq \rho \sum_{k=1}^N |x(k)| = \rho \|x\|_1$$

Hemos conseguido, así de fácilmente, una de las dos desigualdades que buscamos:

$$\exists \rho \in \mathbb{R}^+ : \|x\| \leq \rho \|x\|_1 \quad \forall x \in \mathbb{R}^N \quad (3)$$

Debemos ahora encontrar $\lambda \in \mathbb{R}^+$ que verifique $\lambda \|x\|_1 \leq \|x\|$, también para todo $x \in \mathbb{R}^N$. En particular, si $\|x\|_1 = 1$ se deberá tener $\|x\| \geq \lambda$, lo que nos indica cómo encontrar λ .

Consideramos por tanto el conjunto $S = \{x \in \mathbb{R}^N : \|x\|_1 = 1\}$, que es cerrado y acotado, luego compacto, para la topología usual de \mathbb{R}^N . Además, la función $\|\cdot\| : \mathbb{R}^N \rightarrow \mathbb{R}$ es continua, pues de hecho, usando (3) tenemos:

$$|\|x\| - \|y\|| \leq \|x - y\| \leq \rho \|x - y\|_1 \quad \forall x, y \in \mathbb{R}^N$$

Deducimos que la función continua $\|\cdot\|$ tiene mínimo en el conjunto compacto S , lo que nos permite tomar $\lambda = \min\{\|x\| : x \in S\}$. Como $\lambda = \|x_0\|$ para algún $x_0 \in S$, será $\lambda \in \mathbb{R}^+$. Además, para $x \in \mathbb{R}^N \setminus \{0\}$ tenemos:

$$\frac{x}{\|x\|_1} \in S, \quad \text{luego} \quad \lambda \leq \left\| \frac{x}{\|x\|_1} \right\| = \frac{\|x\|}{\|x\|_1} \quad \text{es decir,} \quad \lambda \|x\|_1 \leq \|x\|$$

Esta desigualdad es obvia para $x = 0$ y hemos probado que

$$\exists \lambda \in \mathbb{R}^+ : \lambda \|x\|_1 \leq \|x\| \quad \forall x \in \mathbb{R}^N \quad (4)$$

En vista de (3) y (4) las normas $\|\cdot\|$ y $\|\cdot\|_1$ son equivalentes, como queríamos. ■

En algunos resultados ya estudiados anteriormente, se hablaba de una norma en \mathbb{R}^N cuya topología sea la usual de \mathbb{R}^N . El teorema anterior deja claro que dicha condición es automática, la topología de cualquier norma en \mathbb{R}^N es siempre la usual de \mathbb{R}^N . Así pues, por ejemplo, los subconjuntos acotados de \mathbb{R}^N son los mismos para todas las normas en \mathbb{R}^N . Para resaltar mejor el contenido del teorema de Hausdorff, es conveniente hacer un enunciado que, formalmente, es más general:

Teorema. *Todas las normas en un espacio vectorial de dimensión finita son equivalentes.*

Demostración. Si X es un espacio vectorial de dimensión $N \in \mathbb{N}$, existe una biyección lineal $\Phi : \mathbb{R}^N \rightarrow X$. Cualesquiera dos normas $\|\cdot\|_1$ y $\|\cdot\|_2$ en X , se “trasladan” a \mathbb{R}^N mediante la aplicación Φ . Más concretamente, basta definir, para todo $y \in \mathbb{R}^N$,

$$\|y\|'_1 = \|\Phi(y)\|_1 \quad \text{y} \quad \|y\|'_2 = \|\Phi(y)\|_2$$

Es rutinario comprobar que de esta forma se obtienen dos normas en \mathbb{R}^N que, por el teorema anterior, son equivalentes, es decir, existen constantes $\lambda, \rho \in \mathbb{R}^+$ tales que

$$\lambda \|y\|'_1 \leq \|y\|'_2 \leq \rho \|y\|'_1 \quad \forall y \in \mathbb{R}^N$$

Entonces, para cada $x \in X$ podemos tomar $y = \Phi^{-1}(x)$ para obtener

$$\lambda \|x\|_1 = \lambda \|y\|'_1 \leq \|y\|'_2 = \|x\|_2 = \|y\|'_2 \leq \rho \|y\|'_1 = \rho \|x\|_1$$

luego las normas de partida en X también son equivalentes, como queríamos. ■

Explicamos el interés de este enunciado que, como hemos visto, es equivalente al teorema de Hausdorff. La clave está en que, al hablar de \mathbb{R}^N , inevitablemente tenemos en mente su base usual, mientras que al hablar de un espacio vectorial de dimensión N , digamos X , está claro que no estamos pensando en ninguna base concreta de X .

El primer enunciado nos dice que, si fijamos una base de X e identificamos X con \mathbb{R}^N , todas las normas en X tienen la misma topología, pero esa topología podría depender de la forma en que hemos identificado X con \mathbb{R}^N , es decir, de la base fijada en X . El segundo enunciado deja claro que no hay tal dependencia, en el espacio X hay una topología común a todas las normas, que no depende de ninguna base que podamos fijar en X para identificarlo con \mathbb{R}^N .

La situación cambia drásticamente cuando consideramos espacios vectoriales de dimensión infinita. Aunque no vamos a demostrarlo, porque tampoco vamos a tener ocasión de usarlo, conviene conocer el siguiente resultado: *en todo espacio vectorial de dimensión infinita, existen dos normas que no son equivalentes*.

4.5. Conexión

Estudiamos ahora la propiedad topológica que nos va a permitir obtener una versión general para espacios métricos del teorema del valor intermedio que conocemos para funciones reales de variable real, afirmando que la imagen por una función continua de un intervalo es un intervalo. Es natural preguntarse qué condición debe cumplir un espacio métrico E para poder asegurar que la imagen de toda función continua de E en \mathbb{R} sea un intervalo.

Para encontrar esta condición, supongamos que $f : E \rightarrow \mathbb{R}$ es continua, pero su imagen no es un intervalo. Por la caracterización de los intervalos, existen $\alpha, \beta \in f(E)$, con $\alpha < \beta$, y existe también un $\lambda \in]\alpha, \beta[$ tal que $\lambda \notin f(E)$. Consideramos entonces los conjuntos

$$U = \{x \in E : f(x) < \lambda\} \quad \text{y} \quad V = \{x \in E : f(x) > \lambda\}$$

Como f es continua, U y V son subconjuntos abiertos de E , ya que son las imágenes inversas por f de dos semirrectas abiertas de \mathbb{R} . Por ser $\alpha \in f(E)$, existe $a \in E$ tal que $f(a) = \alpha < \lambda$ luego $a \in U$ y $U \neq \emptyset$. Análogamente, de $\beta \in f(E)$ deducimos $V \neq \emptyset$. Además, como $\lambda \notin f(E)$, para $x \in E$ se tiene, o bien $f(x) < \lambda$ y $x \in U$, o bien $f(x) > \lambda$ y $x \in V$, así que $E = U \cup V$. Finalmente, es evidente que $U \cap V = \emptyset$. En resumen, si existe una función continua $f : E \rightarrow \mathbb{R}$ tal que $f(E)$ no es un intervalo, entonces E se puede expresar como unión de dos subconjuntos abiertos, no vacíos y disjuntos. Enseguida veremos que el recíproco también es cierto, lo que motiva la siguiente definición.

Decimos que un espacio métrico E es **conexo**, cuando no se puede expresar como unión de dos subconjuntos abiertos, no vacíos y disjuntos. Conviene reformular esta condición negativa como una implicación. Decir que dos subconjuntos abiertos de E , digamos U y V , no pueden cumplir las cuatro condiciones $U \cup V = E$, $U \cap V = \emptyset$, $U \neq \emptyset$ y $V \neq \emptyset$, equivale obviamente a decir que, si U y V cumplen algunas de esas condiciones, no pueden cumplir las restantes. Esto deja varias posibilidades entre las que destacamos por ejemplo la siguiente:

$$U = U^\circ, \quad V = V^\circ, \quad U \cup V = E, \quad U \cap V = \emptyset \quad \implies \quad U = \emptyset \quad \text{o} \quad V = \emptyset$$

A su vez, esta implicación puede expresarse de manera que sólo aparezca uno de los dos conjuntos, digamos U . En efecto, las condiciones $U \cup V = E$ y $U \cap V = \emptyset$ equivalen a que se tenga $V = E \setminus U$, y entonces, decir que V es abierto equivale a decir que U es cerrado. En cuanto a la conclusión, es claro que $V = \emptyset$ equivale a $U = E$.

Por tanto, la implicación anterior equivale a

$$U \subset E, \quad U^\circ = U = \overline{U} \quad \implies \quad U = \emptyset \text{ o } U = E$$

Así pues, vemos que *un espacio métrico E es conexo si, y sólo si, \emptyset y E son los únicos subconjuntos de E que son a la vez abiertos y cerrados.*

Resaltamos que la conexión es una propiedad topológica, se expresa usando sólo conjuntos abiertos y se puede definir de la misma forma para espacios topológicos cualesquiera. Veamos ya que esta propiedad responde adecuadamente a la pregunta planteada como motivación:

■ *Para un espacio métrico E , las siguientes afirmaciones son equivalentes:*

- (i) *E es conexo.*
- (ii) *La imagen de toda función continua de E en \mathbb{R} es un intervalo.*
- (iii) *Toda función continua de E en $\{0, 1\}$ es constante.*

(i) \implies (ii). Hemos visto anteriormente que si E no verifica (ii), no puede ser conexo, pero merece la pena repetir el razonamiento de forma más directa. Si $f : E \rightarrow \mathbb{R}$ es una función continua, para probar que $f(E)$ es un intervalo, tomamos $\alpha, \beta \in f(E)$ con $\alpha < \beta$, y $a, b \in E$, tales que $f(a) = \alpha$ y $f(b) = \beta$. Fijado $\lambda \in]\alpha, \beta[$, debemos probar que $\lambda \in f(E)$. Para ello consideramos los conjuntos abiertos

$$U = f^{-1}(]-\infty, \lambda[) \quad \text{y} \quad V = f^{-1}(]\lambda, +\infty[)$$

Tenemos $U \neq \emptyset$ porque $a \in U$, y $V \neq \emptyset$ porque $b \in V$. También es obvio que $U \cap V = \emptyset$. Por ser E conexo, no podrá ser $E = U \cup V$, luego ha de existir $x \in E \setminus (U \cup V)$. Entonces $x \notin U$, luego $f(x) \geq \lambda$, y $x \notin V$, luego $f(x) \leq \lambda$. Por tanto, $\lambda = f(x) \in f(E)$ como queríamos.

(ii) \implies (iii). Toda función continua $f : E \rightarrow \{0, 1\}$ es una función continua de E en \mathbb{R} , luego tenemos que $f(E)$ es un intervalo contenido en $\{0, 1\}$, lo que implica que f es constante.

(iii) \implies (i). Si U es un subconjunto abierto y cerrado de E , podemos considerar la función característica $\chi_U : E \rightarrow \mathbb{R}$ definida por

$$\chi_U(x) = 1 \quad \forall x \in U \quad \text{y} \quad \chi_U(x) = 0 \quad \forall x \in E \setminus U$$

Puesto que U y $E \setminus U$ son abiertos y χ_U es constante en cada uno de esos conjuntos, el carácter local de la continuidad nos dice que χ_U es continua. De (iii) deducimos que χ_U es constante, luego $U = E$ o $U = \emptyset$. ■

Lógicamente, decimos que un subconjunto A de un espacio métrico E es conexo, cuando A es un espacio métrico conexo con la distancia inducida por la de E . Es importante resaltar que entonces, en la definición de espacio métrico conexo debemos obviamente usar subconjuntos abiertos de A , que no tienen por qué ser abiertos de E .

Por el teorema del valor intermedio, todo intervalo es un subconjunto conexo de \mathbb{R} , puesto que verifica la afirmación (ii) anterior. Recíprocamente, si A es un subconjunto conexo de \mathbb{R} , como la inclusión $I : A \rightarrow \mathbb{R}$, definida por $I(x) = x$ para todo $x \in A$, es continua, deducimos que $I(A) = A$ es un intervalo. Tenemos así una caracterización *topológica* de los intervalos.

- *Un subconjunto de \mathbb{R} es conexo si, y sólo si, es un intervalo.*

Queda claro ahora que el siguiente resultado es la generalización, para funciones continuas entre espacios métricos cualesquiera, del teorema del valor intermedio:

Teorema. *Sean E y F espacios métricos y $f : E \rightarrow F$ una función continua. Si E es conexo, entonces $f(E)$ es un subconjunto conexo de F .*

Demostración. Para toda función continua $g : f(E) \rightarrow \mathbb{R}$ tenemos que $g \circ f$ es continua, luego de ser E conexo deducimos que $(g \circ f)(E)$ es un intervalo. Así pues, la imagen de toda función continua de $f(E)$ en \mathbb{R} es un intervalo, luego $f(E)$ es conexo.

Merece la pena mostrar un razonamiento alternativo. Si V es un subconjunto abierto y cerrado de $f(E)$, la continuidad de f nos dice que $U = f^{-1}(V)$ es un subconjunto abierto y cerrado de E . Como E es conexo, tenemos $U = \emptyset$, o bien $U = E$. Por ser $V \subset f(E)$ tenemos que $V = f(U)$, luego $V = \emptyset$ o $V = f(E)$. ■

Uniendo las dos propiedades básicas de las funciones continuas que hemos estudiado, en el caso particular de funciones con valores reales, tenemos la siguiente conclusión:

- *Si E es un espacio métrico compacto y conexo, y $f : E \rightarrow \mathbb{R}$ es una función continua, entonces $f(E)$ es un intervalo cerrado y acotado.*

4.6. Conjuntos convexos

A poco que se piense, comprobar en la práctica que un espacio métrico es conexo no es fácil. Por ejemplo, sabemos que \mathbb{R} es conexo, pero no está nada claro que \mathbb{R}^N lo sea, para $N > 1$. Si queremos sacar provecho de la versión recién probada del teorema del valor intermedio, debemos disponer de condiciones suficientes para que un espacio métrico sea conexo, que sean fáciles de comprobar. Para conseguirlas, empezamos caracterizando de nuevo la conexión:

- *Un espacio métrico E es conexo si, y sólo si, para cualesquiera dos puntos $x, y \in E$ existe un conjunto conexo $C \subset E$, tal que $x, y \in C$.*

Obviamente, si E es conexo, basta tomar $C = E$ para cualesquiera $x, y \in E$. Para probar el recíproco, suponiendo $E = U \cup V$ con U y V abiertos no vacíos, debemos ver que $U \cap V \neq \emptyset$. Tomados $x \in U$ e $y \in V$, existe un conjunto conexo $C \subset E$ tal que $x, y \in C$. Vemos que $U \cap C$ y $V \cap C$ son subconjuntos abiertos de C , que no son vacíos, porque $x \in U \cap C$ e $y \in V \cap C$, y es claro que $C = (E \cap C) = (U \cup V) \cap C = (U \cap C) \cup (V \cap C)$. Por tanto, al ser C conexo, deducimos que $(U \cap C) \cap (V \cap C) \neq \emptyset$, luego $U \cap V \neq \emptyset$ como se quería. ■

Así pues, podemos decir que un espacio métrico E es conexo cuando cualesquiera dos puntos de E están *conectados*, en el sentido de que existe un subconjunto conexo de E que los contiene. Es bien fácil adivinar cómo podemos *conectar* dos puntos de un espacio normado, basta usar el segmento que los une.

Un subconjunto E de un espacio vectorial X es **convexo** cuando, para cualesquiera dos puntos de E , el segmento que los une está contenido en E , es decir:

$$x, y \in E \implies \{ (1-t)x + ty : t \in [0, 1] \} \subset E$$

Nótese que la convexidad es una propiedad algebraica, tiene sentido en cualquier espacio vectorial. Es claro que un subconjunto de \mathbb{R} es convexo si, y sólo si, es un intervalo, pues para cualesquiera $x, y \in \mathbb{R}$ con $x < y$, se tiene que $\{ (1-t)x + ty : t \in [0, 1] \} = [x, y]$. Tenemos pues una caracterización algebraica de los intervalos y, para un subconjunto de \mathbb{R} , ser convexo equivale a ser conexo. En todo espacio normado se tiene una implicación:

- *Todo subconjunto convexo de un espacio normado es conexo.*

Sea E un subconjunto convexo de un espacio normado y fijemos $x, y \in E$. Consideramos la función $f : [0, 1] \rightarrow E$ dada por $f(t) = (1-t)x + ty$ para todo $t \in [0, 1]$ y comprobamos sin dificultad que f es continua. De hecho, para $s, t \in [0, 1]$ se tiene claramente

$$\|f(s) - f(t)\| = \|(t-s)(x-y)\| = |t-s| \|x-y\|$$

Como $[0, 1]$ es conexo y f es continua, deducimos que la imagen de f es un subconjunto conexo de E , que claramente contiene a los puntos $f(0) = x$ y $f(1) = y$. Puesto que x e y eran puntos arbitrarios de E , el resultado anterior nos dice que E es conexo. ■

Por ejemplo, vemos que el propio espacio normado X es conexo, e igual le ocurre a toda bola en X , abierta o cerrada, que siempre es un conjunto convexo. En efecto, si $a \in X$ y $r \in \mathbb{R}^+$, para cualesquiera $x, y \in B(a, r)$ y $t \in [0, 1]$ se tiene

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

Para una bola cerrada el razonamiento es análogo. Por supuesto, todo lo dicho para un espacio normado, se aplica en particular a \mathbb{R}^N con cualquier norma.

Con el fin de poner ejemplos de subconjuntos conexos de un espacio normado, que no sean convexos, conviene hacer la siguiente observación:

- *Sean C y D subconjuntos conexos de un espacio métrico. Si $C \cap D \neq \emptyset$, entonces $C \cup D$ es conexo.*

Dada una función continua $f : C \cup D \rightarrow \{0, 1\}$, probaremos que f es constante. Puesto que las restricciones de f a C y D son continuas, de ser C y D conexos deducimos que f toma un solo valor en C y un solo valor en D . Pero ambos valores han de ser iguales, ya que $C \cap D \neq \emptyset$, luego f es constante como se quería. ■

Consideremos por ejemplo en \mathbb{R}^2 los semiplanos $C = \mathbb{R}^+ \times \mathbb{R}$ y $D = \mathbb{R} \times \mathbb{R}^+$, que son convexos, luego conexos. Como $C \cap D \neq \emptyset$ el resultado anterior nos dice que $C \cup D$ es un subconjunto conexo de \mathbb{R}^2 , que claramente no es convexo.

4.7. Ejercicios

1. Probar que dos normas definidas en un mismo espacio vectorial, que den lugar a los mismos conjuntos acotados, han de ser equivalentes.
2. Dado un subconjunto A de un espacio normado, probar que las siguientes afirmaciones son equivalentes:
 - (i) A está acotado
 - (ii) Si $\{a_n\}$ es una sucesión de puntos de A y $\{\lambda_n\}$ una sucesión de números reales tal que $\{\lambda_n\} \rightarrow 0$, entonces $\{\lambda_n a_n\} \rightarrow 0$.
 - (iii) Para toda sucesión $\{a_n\}$ de puntos de A se tiene que $\{a_n/n\} \rightarrow 0$.
 ¿Significa esto que si $\{a_n/n\} \rightarrow 0$, entonces la sucesión $\{a_n\}$ está acotada?
3. Probar que todo espacio métrico finito es compacto. Probar también que, en un conjunto no vacío E con la distancia discreta, todo subconjunto compacto de E es finito.
4. Sea $\{x_n\}$ una sucesión convergente de puntos de un espacio métrico y $x = \lim_{n \rightarrow \infty} x_n$. Probar que el conjunto $A = \{x_n : n \in \mathbb{N}\} \cup \{x\}$ es compacto.
5. Probar que, si E es un espacio métrico compacto, y A es un subconjunto infinito de E , entonces $A' \neq \emptyset$.
6. Sea E un espacio métrico con distancia d y K un subconjunto compacto de E . Probar que, para cada $x \in E$, existe un punto $k_x \in K$ tal que $d(x, k_x) \leq d(x, k)$ para todo $k \in K$.
7. Sea A un subconjunto cerrado de \mathbb{R}^N y consideremos en \mathbb{R}^N cualquier distancia d que genere la topología usual. Probar que, para todo $x \in \mathbb{R}^N$, se puede encontrar un $a_x \in A$, tal que $d(x, a_x) \leq d(x, a)$ para todo $a \in A$.
8. Sea F un conjunto no vacío con la distancia discreta. Probar que si E es un espacio métrico conexo, toda función continua de E en F es constante.
9. Probar que, en todo espacio métrico, el cierre de un conjunto conexo es conexo.
10. Probar que el conjunto $A = \{(x, y) \in \mathbb{R}^2 : xy \geq 0\}$ es conexo pero A° no lo es.
11. Sea X un espacio normado. Probar que para cualesquiera $x, y \in X \setminus \{0\}$ se tiene

$$\left\| \frac{x}{\|x\|} - \frac{y}{\|y\|} \right\| \leq \frac{2\|x - y\|}{\|x\|}$$

Deducir que la función $\varphi : X \setminus \{0\} \rightarrow X$, dada por $\varphi(x) = x/\|x\|$ para todo $x \in X \setminus \{0\}$, es continua.

12. Probar que, si X es un espacio normado de dimensión mayor que 1, el conjunto $X \setminus \{0\}$ es conexo. Deducir que la *esfera unidad* $S(0, 1) = \{x \in X : \|x\| = 1\}$ es un conjunto conexo.

Complitud y continuidad uniforme

Como ocurría con la acotación, las nociones que ahora vamos a estudiar no son topológicas, pero se conservan en espacios normados al cambiar las normas por otras equivalentes. Gracias al teorema de Hausdorff, para estudiarlas en \mathbb{R}^N podemos usar cualquier norma.

El concepto de *sucesión de Cauchy* se generaliza fácilmente, para manejarlo en un espacio métrico arbitrario, lo que nos llevará a la definición de *espacio métrico completo*. Generalizando el teorema de complitud de \mathbb{R} , probaremos el resultado análogo en \mathbb{R}^N .

En cuanto a la *continuidad uniforme*, probamos la versión general del *teorema de Heine*, completando así el análisis de los principales teoremas que conocíamos sobre continuidad de funciones reales de variable real. Entre las funciones uniformemente continuas destacamos las *lipschitzianas*, y entre ellas las llamadas *contractivas*. Sobre estas últimas trata un importante resultado, el *teorema del punto fijo de Banach*, que es útil en muy diversos campos y aquí servirá para el estudio del cálculo diferencial. Finalmente veremos una caracterización, también muy útil, de las aplicaciones lineales y continuas entre espacios normados.

5.1. Sucesiones de Cauchy

Si recordamos las sucesiones de Cauchy de números reales, la siguiente definición no será ninguna sorpresa. Si E es un espacio métrico con distancia d , y $x_n \in E$ para todo $n \in \mathbb{N}$, se dice que $\{x_n\}$ es una **sucesión de Cauchy** en E , cuando:

$$\forall \varepsilon > 0 \quad \exists m \in \mathbb{N} : p, q \geq m \Rightarrow d(x_p, x_q) < \varepsilon$$

Es obvio que, en \mathbb{R} con la distancia usual, las sucesiones de Cauchy son las que ya conocíamos, que coinciden con las convergentes. En general, tenemos siempre una implicación:

- *En cualquier espacio métrico, toda sucesión convergente es una sucesión de Cauchy.*

Para comprobarlo, sea $\{x_n\}$ una sucesión de puntos de un espacio métrico E , y supongamos que $\{x_n\} \rightarrow x \in E$. Dado $\varepsilon > 0$, existe $m \in \mathbb{N}$ tal que, para $n \geq m$ se tiene $d(x_n, x) < \varepsilon/2$. Para $p, q \geq m$, concluimos entonces que $d(x_p, x_q) \leq d(x_p, x) + d(x, x_q) < \varepsilon$. ■

Es fácil ver que, en general, el recíproco del resultado anterior no es cierto. En el espacio métrico \mathbb{Q} , subespacio métrico de \mathbb{R} , existen sucesiones de Cauchy que no son convergentes. Si $r_n \in \mathbb{Q}$ para todo $n \in \mathbb{N}$ y, en el espacio métrico \mathbb{R} , tenemos $\{r_n\} \rightarrow \alpha \in \mathbb{R} \setminus \mathbb{Q}$, es claro que $\{r_n\}$ es una sucesión de Cauchy en el espacio métrico \mathbb{Q} que no es convergente, pues si fuese $\{r_n\} \rightarrow r \in \mathbb{Q}$, entonces $\{r_n\}$ también convergería a r en el espacio métrico \mathbb{R} , luego se tendría $\alpha = r$, lo cual es una contradicción.

Aclaremos ahora que la noción de sucesión de Cauchy no es topológica:

Ejemplo. *Dos distancias equivalentes que no dan lugar a las mismas sucesiones de Cauchy.* Sea d la distancia usual de \mathbb{R} y definamos otra distancia ρ de la siguiente forma:

$$\rho(x, y) = |e^x - e^y| \quad \forall x, y \in \mathbb{R}$$

Es obvio que $\rho(x, y) = \rho(y, x)$ para cualesquiera $x, y \in \mathbb{R}$, así como que $\rho(x, y) = 0$ si, y sólo si, $x = y$. La desigualdad triangular también es evidente, pues para $x, y, z \in \mathbb{R}$ se tiene:

$$\rho(x, z) = |e^x - e^z| \leq |e^x - e^y| + |e^y - e^z| = \rho(x, y) + \rho(y, z)$$

Por tanto ρ es una distancia en \mathbb{R} y, para ver que es equivalente a d , comprobamos que ambas dan lugar a las mismas sucesiones convergentes. Si $x_n \in \mathbb{R}$ para todo $n \in \mathbb{N}$ y $x \in \mathbb{R}$, usando la continuidad de la exponencial y del logaritmo, tenemos

$$\{\rho(x_n, x)\} \rightarrow 0 \Leftrightarrow \{|e^{x_n} - e^x|\} \rightarrow 0 \Leftrightarrow \{|x_n - x|\} \rightarrow 0 \Leftrightarrow \{d(x_n, x)\} \rightarrow 0$$

Es claro que una sucesión $\{x_n\}$ es de Cauchy para la distancia ρ si, y sólo si, $\{e^{x_n}\}$ es de Cauchy para la distancia usual, es decir, $\{e^{x_n}\}$ es convergente. Puesto que $\{e^{-n}\} \rightarrow 0$, vemos que $\{-n\}$ es una sucesión de Cauchy para la distancia ρ , pero obviamente no lo es para d . Está claro que, en \mathbb{R} con la distancia ρ , existen sucesiones de Cauchy que no son convergentes. ■

5.2. Complitud

Dado un espacio métrico E con distancia d , la siguiente definición está ya motivada. Se dice que d es una **distancia completa**, o también que E es un **espacio métrico completo**, cuando toda sucesión de Cauchy de puntos de E es convergente.

El teorema de complitud de \mathbb{R} , como su nombre indica, afirma que la distancia usual de \mathbb{R} es completa, o que \mathbb{R} con la distancia usual es un espacio métrico completo. Pero en el ejemplo anterior hemos visto una distancia ρ en \mathbb{R} , equivalente a la usual, que no es completa. Por tanto, la complitud de un espacio métrico no es una propiedad topológica.

Diremos ahora que una norma $\|\cdot\|$ en un espacio vectorial X es una **norma completa**, cuando es completa la distancia d asociada, definida por $d(x, y) = \|y - x\|$ para $x, y \in X$. Un espacio normado cuya norma es completa recibe el nombre de **espacio de Banach**, en honor del matemático polaco Stefan Banach (1892-1945). Así pues, un espacio de Banach es un espacio normado que, con la distancia asociada a su norma, es un espacio métrico completo. Un espacio pre-hilbertiano cuya norma, la asociada a su producto escalar, es completa, recibe el nombre de **espacio de Hilbert**, esta vez en honor del matemático alemán David Hilbert (1862-1943).

En el ambiente particular de los espacios normados tenemos una ventaja que no teníamos en espacios métricos cualesquiera:

- *Dos normas equivalentes en un mismo espacio vectorial dan lugar a la mismas sucesiones de Cauchy. Por tanto, toda norma equivalente a una norma completa es completa.*

Si $\|\cdot\|_1$ y $\|\cdot\|_2$ son dos normas equivalentes en un espacio vectorial X , sabemos que existen constantes $\lambda, \rho \in \mathbb{R}^+$ tales que $\lambda\|x\|_1 \leq \|x\|_2 \leq \rho\|x\|_1$. Entonces, para toda sucesión $\{x_n\}$ de vectores de X tendremos

$$\lambda\|x_p - x_q\|_1 \leq \|x_p - x_q\|_2 \leq \rho\|x_p - x_q\|_1 \quad \forall p, q \in \mathbb{N}$$

Es ya evidente que toda sucesión de Cauchy para $\|\cdot\|_1$ es de Cauchy para $\|\cdot\|_2$ y viceversa. Como ambas normas también dan lugar a las mismas sucesiones convergentes, deducimos que una de ellas es completa si, y sólo si, lo es la otra. ■

Así pues, un espacio de Banach lo sigue siendo al sustituir su norma por otra equivalente. Gracias al teorema de Hausdorff, todas las normas en \mathbb{R}^N dan lugar a las mismas sucesiones de Cauchy, así que podemos hablar sin ambigüedad de sucesiones de Cauchy de vectores de \mathbb{R}^N . Pero es que, en realidad tales sucesiones no son otras que las convergentes:

Teorema. *Todo espacio normado de dimensión finita es un espacio de Banach. Por tanto, el espacio euclídeo N -dimensional es un espacio de Hilbert.*

Demostración. Basta, por ejemplo, probar que la norma del máximo en \mathbb{R}^N es completa. Si $\{x_n\}$ es una sucesión de Cauchy de vectores de \mathbb{R}^N , le aplicamos una desigualdad que ya hemos usado varias veces:

$$|x_p(k) - x_q(k)| \leq \|x_p - x_q\|_\infty \quad \forall p, q \in \mathbb{N} \quad \forall k \in \Delta_N$$

Deducimos que, para cada $k \in \Delta_N$, la sucesión de números reales $\{x_n(k)\}$ es de Cauchy luego, por el teorema de complitud de \mathbb{R} , es convergente. Por tanto, $\{x_n\}$ es convergente. ■

Procede preguntar ahora qué subespacios métricos de \mathbb{R}^N son espacios métricos completos, pero es preferible analizar esta pregunta en general, para subconjuntos de un espacio métrico cualquiera.

- *Sea E un espacio métrico y A un subespacio métrico de E . Se tiene:*

- (i) *Si A es completo, entonces A es un subconjunto cerrado de E .*
- (ii) *Si E es completo y A es un subconjunto cerrado de E , entonces A es completo.*

(i). Si $\{a_n\}$ es una sucesión de puntos de A tal que $\{a_n\} \rightarrow x \in E$, bastará ver que $x \in A$. Pero está claro que $\{a_n\}$ es una sucesión de Cauchy en A , luego es convergente: $\{a_n\} \rightarrow a \in A$. Deducimos claramente que $x = a$, luego $x \in A$ como queríamos.

(ii). Sea $\{x_n\}$ una sucesión de Cauchy en el espacio métrico A , luego también en E . Como ahora E es completo, tenemos $\{x_n\} \rightarrow x \in E$ y, al ser A cerrado, deducimos que $x \in A$. Por tanto, $\{x_n\}$ es convergente en el espacio métrico A , así que A es completo. ■

Deducimos que, cuando el espacio métrico E es completo, un subespacio A es completo si, y sólo si, A es un subconjunto cerrado de E . Esto se aplica en particular al caso de \mathbb{R}^N con cualquier norma: los subconjuntos completos de \mathbb{R}^N no son otros que los cerrados.

5.3. Funciones uniformemente continuas

La definición de continuidad uniforme no precisa ninguna motivación, basta observar que la definición que conocemos para funciones reales de variable real viene expresada en términos de la distancia usual de \mathbb{R} . En lo que sigue fijamos dos espacios métricos E y F , cuyas distancias se denotan ambas por d , y una función $f : E \rightarrow F$. La continuidad de f se expresa en la forma

$$\forall x \in E \quad \forall \varepsilon > 0 \quad \exists \delta > 0 : y \in E, \quad d(x, y) < \delta \Rightarrow d(f(x), f(y)) < \varepsilon$$

donde sabemos que δ puede depender tanto de ε como del punto $x \in E$ considerado. Pues bien, tendremos continuidad uniforme cuando podamos conseguir que δ sólo dependa de ε .

Por tanto, decimos que f es **uniformemente continua** cuando:

$$\forall \varepsilon > 0 \quad \exists \delta > 0 : x, y \in E, \quad d(x, y) < \delta \Rightarrow d(f(x), f(y)) < \varepsilon$$

Está claro que así generalizamos la definición de continuidad uniforme que conocemos para funciones reales de variable real. También es obvio que toda función uniformemente continua es continua, pero sabemos que el recíproco es falso. Caracterizamos fácilmente la continuidad uniforme en términos de sucesiones:

- Si f es uniformemente continua y $\{x_n\}, \{y_n\}$ son sucesiones de puntos de E verificando que $\{d(x_n, y_n)\} \rightarrow 0$, entonces $\{d(f(x_n), f(y_n))\} \rightarrow 0$. El recíproco también es cierto, más aún: si f no es uniformemente continua, existen dos sucesiones $\{x_n\}$ e $\{y_n\}$ de puntos de E y existe un $\varepsilon > 0$ tales que $d(x_n, y_n) < 1/n$ para todo $n \in \mathbb{N}$, pero también se tiene $d(f(x_n), f(y_n)) \geq \varepsilon$ para todo $n \in \mathbb{N}$.

Dado $\varepsilon > 0$, sea $\delta > 0$ dado por la continuidad uniforme de f . Existe $m \in \mathbb{N}$ tal que, para todo $n \geq m$, se tiene $d(x_n, y_n) < \delta$, luego $d(f(x_n), f(y_n)) < \varepsilon$. Recíprocamente, si f no es uniformemente continua, existe $\varepsilon > 0$ tal que, para cada $\delta > 0$, podemos encontrar $x, y \in E$ tales que $d(x, y) < \delta$, pero $d(f(x), f(y)) \geq \varepsilon$. Para cada $n \in \mathbb{N}$ tomamos entonces $\delta = 1/n$ para encontrar $x_n, y_n \in E$ verificando que $d(x_n, y_n) < 1/n$ y $d(f(x_n), f(y_n)) \geq \varepsilon$. En particular tenemos $\{d(x_n, y_n)\} \rightarrow 0$ pero la sucesión $\{d(f(x_n), f(y_n))\}$ no converge a cero. ■

Probamos ya el principal resultado acerca de la continuidad uniforme:

Teorema de Heine. Sean E y F dos espacios métricos y $f : E \rightarrow F$ una función continua. Si E es compacto, entonces f es uniformemente continua.

Demostración. Por reducción al absurdo, suponemos que f no es uniformemente continua. Existen sucesiones $\{x_n\}$ e $\{y_n\}$ de puntos de E , y un $\varepsilon > 0$, tales que, para todo $n \in \mathbb{N}$ se tiene

$$d(x_n, y_n) < 1/n \quad \text{y} \quad d(f(x_n), f(y_n)) \geq \varepsilon$$

Por ser E compacto, tenemos una sucesión parcial $\{x_{\sigma(n)}\}$ que converge a un punto $x \in E$. Puesto que $\{d(x_{\sigma(n)}, y_{\sigma(n)})\} \rightarrow 0$, deducimos que también $\{y_{\sigma(n)}\} \rightarrow x$. Como f es continua, tenemos $\{f(x_{\sigma(n)})\} \rightarrow f(x)$ y $\{f(y_{\sigma(n)})\} \rightarrow f(x)$, luego $\{d(f(x_{\sigma(n)}), f(y_{\sigma(n)}))\} \rightarrow 0$, lo cual es una contradicción, ya que $d(f(x_{\sigma(n)}), f(y_{\sigma(n)})) \geq \varepsilon$ para todo $n \in \mathbb{N}$. ■

El teorema anterior pone de manifiesto que la continuidad uniforme no es una propiedad local: si $f : \mathbb{R} \rightarrow \mathbb{R}$ es continua, para cada $x \in \mathbb{R}$ existe un entorno U de x tal que $f|_U$ es uniformemente continua, pero sabemos que esto no implica que f sea uniformemente continua.

Observemos ahora que la continuidad uniforme no es una propiedad topológica. Usamos para ello una distancia ρ en \mathbb{R} , equivalente a la usual, que ya conocemos:

$$\rho(x, y) = |e^x - e^y| \quad \forall x, y \in \mathbb{R}$$

Si como espacio métrico E tomamos \mathbb{R} con la distancia usual, mientras que F es \mathbb{R} con la distancia ρ , entonces la función identidad $f : E \rightarrow F$ no es uniformemente continua. Si lo fuese tendríamos que

$$\forall \varepsilon > 0 \quad \exists \delta > 0 : x, y \in \mathbb{R}, |x - y| < \delta \Rightarrow |e^x - e^y| < \varepsilon$$

Esto sería tanto como decir que la función exponencial es uniformemente continua, vista como función de \mathbb{R} con la distancia usual en sí mismo, cosa que sabemos que es falsa. Sin embargo, es obvio que si en F cambiamos su distancia por la usual de \mathbb{R} tendremos la continuidad uniforme de f . Lo mismo ocurre si, tanto en E como en F usamos la distancia ρ . Tenemos así ejemplos en los que una función uniformemente continua deja de serlo cuando sustituimos la distancia del espacio de partida, o del de llegada, por otra equivalente.

Ocorre sin embargo con la continuidad uniforme lo mismo que ocurrió con la acotación y con la complitud. Si X e Y son espacios normados y E es un subconjunto de X , la continuidad uniforme de una función $f : E \rightarrow Y$ se mantiene cuando sustituimos las normas, tanto la de X como la de Y , por otras equivalentes. La comprobación de este hecho es un sencillo ejercicio. En particular, si E es un subconjunto no vacío de \mathbb{R}^N , para estudiar la continuidad uniforme de una función $f : E \rightarrow \mathbb{R}^M$ podemos usar cualquier norma, tanto en \mathbb{R}^N como en \mathbb{R}^M .

5.4. Funciones lipschitzianas

Si E y F son espacios métricos, una función $f : E \rightarrow F$ es **lipschitziana** cuando existe una constante $M \in \mathbb{R}_0^+$ tal que:

$$d(f(x), f(y)) \leq M d(x, y) \quad \forall x, y \in E \quad (1)$$

Cuando convenga resaltar la constante M que aparece en esta desigualdad, podemos decir que la función f es *lipschitziana con constante M* . Es evidente que toda función lipschitziana es uniformemente continua, y sabemos que el recíproco es falso, incluso en el caso $E = F = \mathbb{R}$.

La mínima constante M_0 que verifica (1) es la **constante de Lipschitz** de f , que viene dada por

$$M_0 = \sup \left\{ \frac{d(f(x), f(y))}{d(x, y)} : x, y \in E, x \neq y \right\}$$

Cuando $M_0 \leq 1$ se dice que f es **no expansiva**. Cuando se tiene de hecho $M_0 < 1$ decimos que f es **contractiva**.

Por ejemplo, si X es un espacio normado, su norma es una función no expansiva de X en \mathbb{R} , puesto que $|\|x\| - \|y\|| \leq \|x - y\|$ para cualesquiera $x, y \in X$. Análogamente, en cualquier espacio métrico E , fijado $z \in E$, la aplicación $x \mapsto d(x, z)$ es no expansiva, puesto que $|d(x, z) - d(y, z)| \leq d(x, y)$ para cualesquiera $x, y \in E$. Como función de dos variables, definida en el espacio métrico producto $E \times E$, la distancia es una función lipschitziana, pues para cualesquiera $x, y, u, v \in E$ se tiene

$$\begin{aligned} |d(x, y) - d(u, v)| &\leq |d(x, y) - d(y, u)| + |d(y, u) - d(u, v)| \leq d(x, u) + d(y, v) \\ &\leq 2 \max\{d(x, u), d(y, v)\} = 2 d_\infty((x, y), (u, v)) \end{aligned}$$

Por tanto, d es uniformemente continua, y en particular es continua, cosa que ya sabíamos.

El hecho de que una función sea lipschitziana, y no digamos su constante de Lipschitz, es muy inestable cuando cambiamos de distancias. En el ejemplo usado para mostrar que la continuidad uniforme no es una propiedad topológica, teníamos una aplicación no expansiva que dejaba de ser uniformemente continua al sustituir la distancia del espacio métrico de partida o de llegada por otra equivalente. En el caso de dos espacios normados, es fácil ver que una función lipschitziana lo sigue siendo al cambiar las normas por otras equivalentes, pero su constante de Lipschitz puede claramente cambiar.

5.5. Teorema del punto fijo de Banach

El siguiente resultado muestra una propiedad importante de las aplicaciones contractivas y, de paso, pone muy de manifiesto la utilidad de la complitud de un espacio métrico.

Teorema. *Sea E un espacio métrico completo y $f : E \rightarrow E$ una aplicación contractiva. Entonces f tiene un único punto fijo, es decir, existe un único punto $x \in E$ tal que $f(x) = x$.*

Demostración. Fijamos $x_0 \in E$ arbitrario y definimos por inducción una sucesión $\{x_n\}$ de puntos de E , tomando $x_1 = f(x_0)$ y $x_{n+1} = f(x_n)$ para todo $n \in \mathbb{N}$. Probaremos que esta sucesión es convergente y su límite será el punto fijo que buscamos.

Si $\alpha < 1$ es la constante de Lipschitz de f y $\rho = d(x_0, x_1)$, comprobamos por inducción que

$$d(x_n, x_{n+1}) \leq \alpha^n \rho \quad \forall n \in \mathbb{N} \quad (2)$$

En efecto, tenemos $d(x_1, x_2) = d(f(x_0), f(x_1)) \leq \alpha d(x_0, x_1) = \alpha \rho$ y, suponiendo que (2) se verifica para un $n \in \mathbb{N}$, deducimos que

$$d(x_{n+1}, x_{n+2}) = d(f(x_n), f(x_{n+1})) \leq \alpha d(x_n, x_{n+1}) \leq \alpha \alpha^n \rho = \alpha^{n+1} \rho$$

Ahora, para cualesquiera $n, k \in \mathbb{N}$ tenemos

$$d(x_n, x_{n+k}) \leq \sum_{j=0}^{k-1} d(x_{n+j}, x_{n+j+1}) \leq \rho \sum_{j=0}^{k-1} \alpha^{n+j} \leq \rho \alpha^n \sum_{j=0}^{\infty} \alpha^j = \frac{\rho \alpha^n}{1 - \alpha} \quad (3)$$

De la desigualdad (3) deduciremos fácilmente que $\{x_n\}$ es una sucesión de Cauchy. En efecto, dado $\varepsilon > 0$, como $\{\alpha^n\} \rightarrow 0$, existe $m \in \mathbb{N}$ tal que, para $n \geq m$ se tiene $\rho \alpha^n < \varepsilon(1 - \alpha)$. Entonces, para $p, q \geq m$, suponiendo sin perder generalidad que $p < q$, usamos (3) con $n = p$ y $k = q - p$ para obtener

$$d(x_p, x_q) = d(x_n, x_{n+k}) \leq \frac{\rho \alpha^n}{1 - \alpha} < \varepsilon$$

Como por hipótesis E es completo, tenemos $\{x_n\} \rightarrow x \in E$, luego $\{f(x_n)\} = \{x_{n+1}\} \rightarrow x$. Pero f continua, luego $\{f(x_n)\} \rightarrow f(x)$, y concluimos que $f(x) = x$. Finalmente, si $y \in E$ es otro punto fijo de f , se tiene $d(x, y) = d(f(x), f(y)) \leq \alpha d(x, y)$, luego $(1 - \alpha)d(x, y) \leq 0$. Como $\alpha < 1$, deducimos que $d(x, y) \leq 0$, es decir, $x = y$. ■

El teorema anterior se aplica en contextos muy variados, pues se trata de una herramienta muy potente para resolver ecuaciones. Al fin y al cabo nos asegura que, bajo ciertas condiciones sobre el espacio métrico E y la función $f : E \rightarrow E$, la ecuación $f(x) = x$ tiene solución única.

5.6. Continuidad de aplicaciones lineales

Vamos a obtener ahora una útil caracterización de la continuidad de una aplicación lineal entre dos espacios normados cualesquiera:

- Sean X, Y dos espacios normados y sea $T : X \rightarrow Y$ una aplicación lineal. Las siguientes afirmaciones son equivalentes:

- (i) T es continua
- (ii) Existe una constante $M \in \mathbb{R}_0^+$ tal que $\|T(x)\| \leq M\|x\|$ para todo $x \in X$

(i) \Rightarrow (ii). Teniendo en cuenta que $T(0) = 0$, la continuidad de T en 0 nos dice que

$$\exists \delta > 0 : z \in X, \|z\| < \delta \Rightarrow \|T(z)\| < 1$$

Dado $x \in X \setminus \{0\}$, tomando $z = \frac{\delta x}{2\|x\|}$ tenemos claramente $\|z\| = \delta/2 < \delta$, luego

$$\|T(x)\| = \frac{2\|x\|}{\delta} \|T(z)\| \leq \frac{2}{\delta} \|x\|$$

Como esta desigualdad es obvia cuando $x = 0$, hemos probado (ii) con $M = 2/\delta$.

(ii) \Rightarrow (i). Para cualesquiera $u, v \in X$, de (ii) deducimos claramente que

$$\|T(u) - T(v)\| = \|T(u - v)\| \leq M\|u - v\|$$

lo que prueba que T es lipschitziana, luego continua. ■

Merece la pena comentar dos ideas importantes que han aparecido en la demostración del resultado anterior.

En primer lugar, para probar que $(i) \Rightarrow (ii)$ sólo hemos usado que T es continua en el origen. Por tanto, de ser T continua en el origen se deduce que es continua en todo punto de X . De hecho el origen no tiene nada de especial: usando la traslación adecuada, vemos fácilmente que si T es continua en un punto cualquiera de X , entonces es continua en el origen y, por tanto, en todo punto de X . Dicho de manera equivalente: si T no es continua, entonces no puede ser continua en ningún punto de X .

Por otra parte, al probar $(ii) \Rightarrow (i)$ no sólo hemos visto que T es continua, sino que T es lipschitziana y, en particular, uniformemente continua. Así pues, para una aplicación lineal entre espacios normados, ser continua equivale a ser uniformemente continua, lo que a su vez equivale a ser lipschitziana.

En el estudio del cálculo diferencial trabajaremos sobre todo con espacios normados de dimensión finita, en cuyo caso la continuidad de una aplicación lineal es automática, como muestra el siguiente resultado, importante consecuencia del teorema de Hausdorff:

- *Si X es un espacio normado de dimensión finita, toda aplicación lineal de X en cualquier otro espacio normado, es continua.*

Sea Y otro espacio normado y $T : X \rightarrow Y$ una aplicación lineal. Definimos entonces una nueva norma $\|\cdot\|_T$ en X , de la siguiente forma

$$\|x\|_T = \|x\| + \|T(x)\| \quad \forall x \in X$$

Se comprueba rutinariamente que $\|\cdot\|_T$ es efectivamente una norma en X . Como X tiene dimensión finita, el teorema de Hausdorff nos dice que $\|\cdot\|_T$ es equivalente a la norma de partida en X , luego existe una constante $\rho \in \mathbb{R}^+$ tal que $\|x\|_T \leq \rho \|x\|$ para todo $x \in X$. Pero entonces está bien claro que

$$\|T(x)\| \leq \|x\|_T \leq \rho \|x\| \quad \forall x \in X$$

y esto prueba que T es continua. ■

5.7. Norma de una aplicación lineal continua

Si X e Y son espacios normados arbitrarios, denotaremos siempre por $L(X, Y)$ al espacio vectorial de todas las aplicaciones lineales y continuas de X en Y , cuya suma y producto por escalares vienen definidos, para $T, S \in L(X, Y)$ y $\lambda \in \mathbb{R}$, por

$$(T + S)(x) = T(x) + S(x) \quad \forall x \in X \quad \text{y} \quad (\lambda T)_x = \lambda T(x) \quad \forall x \in X$$

De hecho, $L(X, Y)$ es un subespacio vectorial de $\mathcal{C}(X, Y)$, el espacio vectorial de todas las funciones continuas de X en Y , que ya conocíamos. Pues bien, nuestro próximo objetivo es convertir a $L(X, Y)$ en un espacio normado.

Ya se ha comentado que toda aplicación lineal y continua $T \in L(X, Y)$ es lipschitziana. Esto permite definir la **norma** de T , que naturalmente se denota por $\|T\|$, como la constante de Lipschitz de T , es decir, la mínima constante $M \in \mathbb{R}_0^+$ tal que $\|T(u) - T(v)\| \leq M \|u - v\|$ para cualesquiera $u, v \in X$. Se obtiene una descripción más sencilla de esta constante viendo que, para $M \in \mathbb{R}_0^+$, se tiene:

$$\|T(u) - T(v)\| \leq M \|u - v\| \quad \forall u, v \in X \iff \|T(x)\| \leq M \|x\| \quad \forall x \in X$$

En efecto, para la implicación hacia la derecha, dado $x \in X$, se toma $u = x$ y $v = 0$. Hacia la izquierda, dados $u, v \in X$, basta tomar $x = u - v$. Por tanto, la constante de Lipschitz de T también es la mínima constante que verifica la desigualdad de la derecha, es decir,

$$\|T\| = \min \{ M \in \mathbb{R}_0^+ : \|T(x)\| \leq M \|x\| \quad \forall x \in X \} \quad (4)$$

Antes de comprobar que efectivamente hemos definido una norma en $L(X, Y)$, explicamos la forma en que dicha norma suele usarse. La continuidad de una aplicación lineal $T : X \rightarrow Y$ se comprueba hallando $M \in \mathbb{R}_0^+$ tal que $\|Tx\| \leq M \|x\|$ para todo $x \in X$, con lo que (4) nos dice que $\|T\| \leq M$. Si por el contrario ya sabemos que $T \in L(X, Y)$, podemos escribir

$$\|T(x)\| \leq \|T\| \|x\| \quad \forall x \in X \quad (5)$$

y esta desigualdad es óptima, en el sentido de que $\|T\|$ es la mínima constante que puede aparecer en ella. Esto es lo que ocurrirá siempre que X tenga dimensión finita.

En general, comprobar que tenemos una norma en $L(X, Y)$ es bien fácil. Para $T, S \in L(X, Y)$, usando (5) se tiene:

$$\|(T + S)(x)\| \leq \|T(x)\| + \|S(x)\| \leq (\|T\| + \|S\|) \|x\| \quad \forall x \in X$$

y de (4) deducimos la desigualdad triangular: $\|T + S\| \leq \|T\| + \|S\|$.

Para la homogeneidad por homotecias preferimos usar otra descripción de la norma de una aplicación lineal y continua, que pasamos a explicar. La igualdad (4) nos dice que $\|T\|$ es el mínimo mayorante, es decir, el supremo, de un cierto conjunto de números reales:

$$\|T\| = \sup \left\{ \frac{\|T(x)\|}{\|x\|} : x \in X \setminus \{0\} \right\}$$

Esta expresión se simplifica aún más si pensamos que $\|T(x)\|/\|x\| = \|T(x/\|x\|)\|$. Además, es claro que $u = x/\|x\|$ verifica $\|u\| = 1$. Pero recíprocamente, dado $u \in X$ con $\|u\| = 1$, tenemos $u = x/\|x\|$ sin más que tomar $x = u$. Por tanto:

$$\|T\| = \sup \{ \|T(u)\| : u \in X, \|u\| = 1 \}$$

Tenemos así otra expresión muy útil de la norma de una aplicación lineal y continua.

Ahora, para $T \in L(X, Y)$ y $\lambda \in \mathbb{R}$ se tiene claramente:

$$\begin{aligned} \|\lambda T\| &= \sup \{ \|\lambda T(u)\| : u \in X, \|u\| = 1 \} \\ &= \sup \{ |\lambda| \|T(u)\| : u \in X, \|u\| = 1 \} = |\lambda| \|T\| \end{aligned}$$

Finalmente es obvio que de $\|T\| = 0$ se deduce $T = 0$. Queda así comprobado que tenemos una norma en $L(X, Y)$ como queríamos.

Conviene finalmente resaltar que, aunque el espacio vectorial $L(X, Y)$ sólo depende de las topologías de X e Y , la norma que hemos definido en $L(X, Y)$ sí depende esencialmente de las normas concretas que tengamos en X e Y . Si sustituimos dichas normas por otras equivalentes, la norma de $L(X, Y)$ también cambiará, aunque es fácil ver que la nueva norma en $L(X, Y)$ también será equivalente a la de partida. En el caso que más nos interesa, esto ocurre de nuevo automáticamente, pues cuando X e Y tienen dimensión finita, $L(X, Y)$ también la tiene, luego todas las normas en $L(X, Y)$ son equivalentes.

5.8. Ejercicios

1. Probar que, en cualquier espacio métrico, toda sucesión de Cauchy está acotada.
2. Probar que todo espacio métrico compacto es completo.
3. Sea $f : \mathbb{R} \rightarrow \mathbb{R}$ una función continua e inyectiva. Probar que definiendo

$$\rho(x, y) = |f(x) - f(y)| \quad \forall x, y \in \mathbb{R}$$

se obtiene una distancia en \mathbb{R} , equivalente a la usual. ¿Cuándo es ρ completa?

4. ¿Qué se puede afirmar sobre la composición de dos funciones uniformemente continuas?
5. Dado un espacio normado $X \neq \{0\}$, probar que la función $f : X \setminus \{0\} \rightarrow X$, dada por

$$f(x) = \frac{x}{\|x\|} \quad \forall x \in X \setminus \{0\}$$

no es uniformemente continua. Sin embargo, probar también que, para cada $\delta \in \mathbb{R}^+$, la restricción de f al conjunto $\{x \in X : \|x\| \geq \delta\}$ es una función lipschitziana.

6. Sea A un subconjunto no vacío de un espacio métrico E y $f : E \rightarrow \mathbb{R}$ la función definida por $f(x) = \inf \{d(x, a) : a \in A\}$ para todo $x \in E$. Probar que f es no expansiva.
7. Dado $y \in \mathbb{R}^N$, se define $T_y \in L(\mathbb{R}^N, \mathbb{R})$ usando el producto escalar en \mathbb{R}^N :

$$T_y(x) = (x|y) \quad \forall x \in \mathbb{R}^N$$

Calcular la norma de la aplicación lineal T_y , considerando en \mathbb{R}^N

- a) la norma euclídea
 - b) la norma del máximo
 - c) la norma de la suma
8. Consideremos los espacios normados $X = \mathbb{R}^N$ con la norma de la suma e $Y = \mathbb{R}^N$ con la norma del máximo. Denotando como siempre por $\{e_k : k \in \Delta_N\}$ a la base usual de \mathbb{R}^N , probar que, para toda $T \in L(X, Y)$ se tiene:

$$\|T\| = \max \{ |(T(e_j)|e_k)| : j, k \in \Delta_N \}$$

Tema 6

Diferenciabilidad

Iniciamos el estudio del cálculo diferencial para funciones de varias variables en un contexto muy general, definiendo la diferenciabilidad de funciones entre espacios normados arbitrarios. La noción conocida para funciones reales de variable real que se generaliza directamente, no es la de derivada, sino la de *diferencial*, equivalente a ella. Veremos que la diferenciabilidad es una propiedad local, que implica la continuidad. Entre las reglas de diferenciación de funciones entre espacios normados, destacaremos la *regla de la cadena*.

6.1. Motivación

Recordemos el concepto de derivada para funciones reales de variable real, y su significado analítico, intentando ver la forma de generalizarlo al caso de una función entre dos espacios normados arbitrarios, digamos X e Y . Si A es un subconjunto no vacío de \mathbb{R} , sabemos que una función $f : A \rightarrow \mathbb{R}$ es derivable en un punto $a \in A \cap A'$, cuando existe $\lambda \in \mathbb{R}$ tal que

$$\lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a} = \lambda \quad (1)$$

en cuyo caso $\lambda = f'(a)$ es la *derivada* de f en a . De entrada, cuando f esté definida en un subconjunto del espacio normado X , el cociente de incrementos que aparece en (1) no tiene sentido, simplemente no podemos dividir por el vector $x - a$. Para resolver este problema con el denominador, reescribimos (1) en la forma

$$\lim_{x \rightarrow a} \frac{f(x) - f(a) - \lambda(x - a)}{x - a} = 0$$

lo que a su vez se puede expresar de dos formas equivalentes:

$$\lim_{x \rightarrow a} \frac{|f(x) - f(a) - \lambda(x - a)|}{|x - a|} = 0, \quad \text{o bien,} \quad \lim_{x \rightarrow a} \frac{f(x) - f(a) - \lambda(x - a)}{\|x - a\|} = 0 \quad (2)$$

Esto resuelve el problema que teníamos con el denominador de (1): para $a, x \in X$, bastará escribir $\|x - a\|$ en lugar de $|x - a|$.

Recordemos que las igualdades (2) son las permiten entender intuitivamente el significado analítico de la derivada: la función $P: \mathbb{R} \rightarrow \mathbb{R}$ dada por $P(x) = f(a) + \lambda(x-a)$ para todo $x \in \mathbb{R}$, que viene definida por un polinomio de primer orden, es una buena aproximación de f cerca del punto a , puesto que la diferencia $f(x) - P(x)$ tiende a cero en el punto a , “mucho más rápidamente” que la diferencia $x - a$.

Observemos que el numerador de cualquiera de los límites que aparecen en (2) nos plantea un nuevo problema. En el caso general, tendremos $f(x) - f(a) \in Y$ mientras que $\lambda(x-a) \in X$. La solución está en considerar, en vez de las nociones de derivabilidad y derivada de f en a , las nociones equivalentes de diferenciabilidad y diferencial, que enseguida recordaremos. Lo que subyace a esta equivalencia es la identificación total del espacio normado \mathbb{R} con el espacio normado $L(\mathbb{R}, \mathbb{R})$ de todas las aplicaciones lineales continuas de \mathbb{R} en \mathbb{R} , así que aclaremos dicha identificación.

Para cada $\alpha \in \mathbb{R}$ consideramos la aplicación lineal de \mathbb{R} en \mathbb{R} que consiste en multiplicar por α , esto es, la aplicación $T_\alpha \in L(\mathbb{R}, \mathbb{R})$ dada por $T_\alpha(x) = \alpha x$ para todo $x \in \mathbb{R}$. Es claro que, definiendo $\Phi(\alpha) = T_\alpha$ para todo $\alpha \in \mathbb{R}$, obtenemos una aplicación lineal $\Phi: \mathbb{R} \rightarrow L(\mathbb{R}, \mathbb{R})$. Recordando la norma de $L(\mathbb{R}, \mathbb{R})$, es evidente que $\|T_\alpha\| = |\alpha|$ para todo $\alpha \in \mathbb{R}$, así que Φ conserva la norma: $\|\Phi(\alpha)\| = |\alpha|$ para todo $\alpha \in \mathbb{R}$. En particular, Φ es inyectiva, pero también es sobreyectiva, pues para toda $T \in L(\mathbb{R}, \mathbb{R})$, basta tomar $\alpha = T(1)$ para tener $T = T_\alpha = \Phi(\alpha)$. Por tanto, Φ nos permite identificar totalmente los espacios normados \mathbb{R} y $L(\mathbb{R}, \mathbb{R})$.

Volviendo a nuestra función f derivable en el punto a , cuando la anterior identificación se aplica a la derivada $\lambda = f'(a)$, se obtiene una aplicación $T_\lambda \in L(\mathbb{R}, \mathbb{R})$, que será la diferencial de f en a , y diremos que f es diferenciable en el punto a . Es claro que las igualdades (2) se expresan equivalentemente en términos de la diferencial, sin más que escribir $T_\lambda(x-a)$ en lugar de $\lambda(x-a)$. Más formalmente, decimos que $f: A \rightarrow \mathbb{R}$ es *diferenciable* en el punto $a \in A \cap A'$ cuando existe $T \in L(\mathbb{R}, \mathbb{R})$ verificando que

$$\lim_{x \rightarrow a} \frac{|f(x) - f(a) - T(x-a)|}{|x-a|} = 0, \quad \text{o bien,} \quad \lim_{x \rightarrow a} \frac{f(x) - f(a) - T(x-a)}{|x-a|} = 0 \quad (3)$$

en cuyo caso T es única, y la llamamos *diferencial* de f en a , que se denota por $Df(a)$. Está claro que si $\lambda \in \mathbb{R}$ verifica (1), o equivalentemente (2), entonces tomando $T = T_\lambda \in L(\mathbb{R}, \mathbb{R})$, tenemos (3). Recíprocamente, si $T \in L(\mathbb{R}, \mathbb{R})$ verifica (3), basta tomar $\lambda = T(1) \in \mathbb{R}$ para obtener (2), luego también (1). Así pues, f es diferenciable en a si, y sólo si, es derivable en a , en cuyo caso se tiene $f'(a) = Df(a)(1)$, o lo que es lo mismo, $Df(a)x = f'(a)x$ para todo $x \in \mathbb{R}$. Queda claro que, para funciones reales de variable real, la distinción entre derivada y diferencial es sólo una cuestión de matiz, hablamos de la derivada cuando pensamos en el número real $f'(a)$, mientras que hablamos de la diferencial cuando pensamos en la aplicación lineal $Df(a)$, pero la identificación total de \mathbb{R} con $L(\mathbb{R}, \mathbb{R})$ hace que la derivada y la diferencial sean conceptos equivalentes.

Sin embargo, en el caso general, la distinción entre derivada y diferencial se vuelve crucial. Las igualdades (2) no tenían sentido, pero las de (3) sí lo tienen, con sólo sustituir $T \in L(\mathbb{R}, \mathbb{R})$ por $T \in L(X, Y)$. Queda así motivada la noción general de función diferenciable.

6.2. Funciones diferenciables

En todo lo que sigue, X e Y serán espacios normados arbitrarios, si bien se descartan siempre los casos triviales $X = \{0\}$ e $Y = \{0\}$. Fijamos una función $f : A \rightarrow Y$, donde A es un subconjunto no vacío de X , y un punto $a \in A^\circ$.

Pues bien, se dice que f es **diferenciable** en el punto a cuando existe una aplicación lineal y continua $T \in L(X, Y)$ verificando las siguientes condiciones, que son equivalentes:

$$\lim_{x \rightarrow a} \frac{\|f(x) - f(a) - T(x-a)\|}{\|x-a\|} = 0, \quad \text{o bien,} \quad \lim_{x \rightarrow a} \frac{f(x) - f(a) - T(x-a)}{\|x-a\|} = 0 \quad (4)$$

En ambas igualdades tenemos el límite en el punto a de una función definida en $A \setminus \{a\}$, que tiene sentido, puesto que a es punto de acumulación de $A \setminus \{a\}$. La primera igualdad es conceptualmente más sencilla, pues dicha función toma valores en \mathbb{R} , mientras en la segunda tenemos el límite de una función con valores en Y . Conviene disponer de ambas, para usar una u otra según convenga.

También podemos trasladar el cálculo al origen, usando el conjunto $B = \{x-a : x \in A\}$, que verifica $0 \in B^\circ$. Mediante el cambio de variable $x = a+h \in A$ con $h \in B$, teniendo en cuenta que $x \rightarrow a$ cuando $h \rightarrow 0$ y que $x \neq a$ para $h \neq 0$, deducimos de (4) que

$$\lim_{h \rightarrow 0} \frac{\|f(a+h) - f(a) - T(h)\|}{\|h\|} = 0, \quad \text{o bien,} \quad \lim_{h \rightarrow 0} \frac{f(a+h) - f(a) - T(h)}{\|h\|} = 0 \quad (5)$$

Recíprocamente, de (5) se deduce (4) usando el cambio de variable $h = x-a$. Por tanto, f es diferenciable en a si, y sólo si, existe $T \in L(X, Y)$ verificando (5).

Está claro que, salvo la hipótesis $a \in A^\circ$, más restrictiva que $a \in A \cap A'$, hemos generalizado la noción conocida para el caso $X = Y = \mathbb{R}$. Poco a poco, iremos analizando si los resultados que conocemos en dicho caso particular son o no válidos a plena generalidad. De momento, haremos cinco observaciones clave sobre la definición de diferenciabilidad en un punto.

6.2.1. Unicidad de la diferencial

- Si f es diferenciable en a , la aplicación $T \in L(X, Y)$ que aparece en (4) es única.

Por ser $a \in A^\circ$, existe $\delta > 0$ tal que $B(a, \delta) \subset A$. Para $v \in X \setminus \{0\}$, usamos en (4) el cambio de variable $x = a+tv \in A$ con $t \in \mathbb{R}$ y $|t| < \delta/\|v\|$, teniendo en cuenta que $x \rightarrow a$ cuando $t \rightarrow 0$ y que $x \neq a$ para $t \neq 0$. Obtenemos que

$$0 = \lim_{t \rightarrow 0} \frac{\|f(a+tv) - f(a) - tT(v)\|}{|t| \|v\|} = \frac{1}{\|v\|} \lim_{t \rightarrow 0} \left\| \frac{f(a+tv) - f(a)}{t} - T(v) \right\|$$

donde hemos usado la linealidad de T . Así pues,

$$T(v) = \lim_{t \rightarrow 0} \frac{f(a+tv) - f(a)}{t} \quad (6)$$

Como esto es válido para todo $v \in X \setminus \{0\}$, queda claro que T es única: si dos aplicaciones lineales $T_1, T_2 : X \rightarrow Y$ verifican (4), también deberán verificar (6), de donde $T_1 = T_2$. ■

Nótese que, para comprobar la unicidad de T se ha usado que T es lineal, pero no que sea continua; la razón para exigir su continuidad se verá enseguida. También conviene resaltar que en el razonamiento anterior se usa que $a \in A^\circ$. Suponiendo sólo $a \in A \cap A'$, los límites que aparecen en (4) tendrían sentido, pero en general no podríamos asegurar la unicidad de la diferencial. De hecho, la condición $a \in A^\circ$ no es imprescindible para que T sea única, pero sí es la hipótesis más cómoda para asegurarse dicha unicidad.

Si f es diferenciable en a , a la única $T \in L(X, Y)$ que verifica (4), la llamamos **diferencial** de f en a , y se denota por $Df(a)$. Por tanto, $Df(a) \in L(X, Y)$ se caracteriza por

$$\lim_{x \rightarrow a} \frac{f(x) - f(a) - Df(a)(x - a)}{\|x - a\|} = 0 \quad (7)$$

y como consecuencia verifica también que

$$Df(a)(v) = \lim_{t \rightarrow 0} \frac{f(a + tv) - f(a)}{t} \quad \forall v \in X \quad (8)$$

Conviene dejar claro que, en general, de (8) no se deduce (7).

6.2.2. Relación con la continuidad

■ Si f es diferenciable en a , entonces f es continua en a .

Para todo $x \in A \setminus \{a\}$ podemos claramente escribir

$$f(x) = f(a) + Df(a)(x - a) + \|x - a\| \frac{f(x) - f(a) - Df(a)(x - a)}{\|x - a\|}$$

La igualdad (7) nos dice sobradamente que el último sumando tiene límite cero en el punto a . Por otra parte, como $Df(a)$ es continua, también tenemos $\lim_{x \rightarrow a} Df(a)(x - a) = 0$, y concluimos que $\lim_{x \rightarrow a} f(x) = f(a)$. ■

Resaltamos que la continuidad de $Df(a)$ ha sido esencial para probar que f es continua. Esto explica que al definir la diferenciabilidad hayamos exigido $T \in L(X, Y)$. En los casos que nos interesan, X tendrá dimensión finita, luego toda aplicación lineal de X en Y será continua.

6.2.3. Significado analítico

Podemos interpretar la diferenciabilidad exactamente igual que para funciones reales de variable real: que f sea diferenciable en a significa que, “cerca” del punto a , f admite una “buena aproximación” mediante una función sencilla: la función $g : X \rightarrow Y$ dada por

$$g(x) = f(a) + Df(a)(x - a) = f(a) - Df(a)(a) + Df(a)(x) \quad \forall x \in X$$

Nótese que g es la composición de $Df(a) \in L(X, Y)$ con una traslación, luego es **afín** y continua. Basta pensar en el caso particular $X = Y = \mathbb{R}$, para darse cuenta de que las funciones afines y continuas de X en Y son la generalización natural de los polinomios de primer orden en \mathbb{R} . Tras las constantes, y las lineales continuas, las afines continuas son las funciones de X en Y más sencillas que podemos imaginar. Pues bien, en virtud de (7), tenemos:

$$\lim_{x \rightarrow a} \frac{f(x) - g(x)}{\|x - a\|} = 0$$

Esto significa que, “cerca” del punto a , la función g es una “buena aproximación” de f , pues la diferencia $f(x) - g(x)$ tiende a cero en el punto a “más rápidamente” que $\|x - a\|$. Cabe esperar que, cerca del punto a , la función f se comporte de forma similar a como lo haga g . En el fondo, todos los resultados del cálculo diferencial consistirán en aprovechar esta idea.

Alternativamente, pensemos que f describe la forma en que la variable $y \in Y$ “depende” de otra variable “independiente” $x \in A$, mediante la “relación funcional” $y = f(x)$, con lo que cada variación $\Delta x = x - a$, da lugar a una variación $\Delta y = f(x) - f(a)$. Cuando f es diferenciable en el punto a , la igualdad (7) significa que, si Δx es “pequeña”, Δy es “aproximadamente” igual a $Df(a)\Delta x$. Esta interpretación explica el nombre de la diferencial: controla la relación entre diferencias de las variables, permitiendo pensar que Δy depende de Δx de manera lineal y continua. Para variables con valores vectoriales, este es el tipo de dependencia más sencillo que podemos imaginar. En casos particulares que iremos considerando, este *significado analítico* de la diferencial nos llevará a sus interpretaciones geométrica y física.

6.2.4. Carácter local

Está claro que la diferenciabilidad de una función en un punto es una propiedad local: se expresa mediante un límite, que obviamente sólo involucra los valores de la función en un entorno de dicho punto. Además, como sólo trabajamos en puntos interiores del conjunto en el que nuestra función está definida, este carácter local se expresa de forma muy sencilla:

- Si $U \subset A$ y $a \in U^\circ$, entonces f es diferenciable en a si, y sólo si, $f|_U$ es diferenciable en a , en cuyo caso se tiene $Df(a) = D(f|_U)(a)$.

En el resultado anterior siempre podemos tomar $U = A^\circ$, luego f es diferenciable en a si, y sólo si, lo es $f|_{A^\circ}$, que es una función definida en el conjunto abierto A° . Puesto que sólo estudiamos la diferenciabilidad de f en puntos de A° y, para dichos puntos, podemos trabajar con $f|_{A^\circ}$ en lugar de f , está claro que no perdemos generalidad trabajando solamente con funciones definidas en conjuntos abiertos, como haremos casi siempre a partir de ahora.

6.2.5. Independencia de las normas

- La diferenciabilidad de f en a , así como su diferencial $Df(a)$, se conservan al sustituir las normas de X e Y por otras equivalentes a ellas.

Denotando por $\|\cdot\|_1$ a las normas de partida en X e Y , y por $\|\cdot\|_2$ otras equivalentes a ellas, existen constantes $\lambda, \rho \in \mathbb{R}^+$ tales que $\lambda\|x\|_1 \leq \|x\|_2$ para todo $x \in X$ y $\|y\|_2 \leq \rho\|y\|_1$ para todo $y \in Y$. Si f es diferenciable en a con las normas $\|\cdot\|_1$, para $x \in A \setminus \{a\}$ tenemos

$$0 \leq \frac{\|f(x) - f(a) - Df(a)(x-a)\|_2}{\|x-a\|_2} \leq \frac{\rho}{\lambda} \frac{\|f(x) - f(a) - Df(a)(x-a)\|_1}{\|x-a\|_1}$$

y deducimos que $Df(a)$ sigue siendo la diferencial de f en a para las normas $\|\cdot\|_2$. ■

En el caso que nos interesa, los espacios normados X e Y tendrán dimensión finita, luego para estudiar la diferenciabilidad de una función f podemos considerar cualquier norma en dichos espacios.

El tipo de invariancia recién obtenido se añade a otro, que es completamente evidente: la diferenciabilidad de una función no depende de las bases que podamos fijar en los espacios vectoriales X e Y , simplemente porque la hemos definido sin usar base alguna. Aunque sólo vamos a trabajar con la diferenciabilidad de funciones definidas en un abierto de \mathbb{R}^N y con valores en \mathbb{R}^M , al haber definido la diferenciabilidad en abstracto, para espacios normados cualesquiera, tenemos a priori una noción invariante por cambios de base.

6.2.6. Diferenciabilidad global

Antes de presentar los primeros ejemplos de funciones diferenciables, pensemos en el tipo de función que nos va a interesar, que no será diferenciable en un solo punto, sino en todos los puntos del conjunto abierto en el que está definida.

Sean pues X e Y espacios normados, Ω un subconjunto abierto de X y $f: \Omega \rightarrow Y$ una función. Cuando f sea diferenciable en todo punto $x \in \Omega$, diremos simplemente que f es **diferenciable**, y $D(\Omega, Y)$ será el conjunto de todas las funciones diferenciables de Ω en Y , contenido en el conjunto $\mathcal{C}(\Omega, Y)$ de todas las funciones continuas de Ω en Y .

Para $f \in D(\Omega, Y)$ podemos considerar la función $Df: \Omega \rightarrow L(X, Y)$ que a cada punto $x \in \Omega$ hace corresponder la diferencial $Df(x) \in L(X, Y)$, y decimos que Df es la función diferencial de f , o simplemente la **diferencial** de f . Esta nomenclatura no debe causar confusión: por el contexto se sabe siempre si al hablar de la diferencial de f , nos referimos a la función Df o a la diferencial de f en un punto concreto $x \in \Omega$, que también es una función: $Df(x) \in L(X, Y)$. Ocurre aquí lo mismo que para funciones reales de variable real, no es lo mismo la derivada de una función en un punto que la función derivada.

Como $L(X, Y)$ es a su vez un espacio normado, tiene sentido plantearse la continuidad de la función Df . Decimos que $f \in D(\Omega, Y)$ es una **función de clase C^1** cuando Df es continua, y denotamos por $C^1(\Omega, Y)$ al conjunto de todas las funciones de clase C^1 de Ω en Y . También tendría sentido plantearse la diferenciabilidad de la función Df , pero esto es un asunto más delicado, que abordaremos en su momento, para el estudio de las diferenciales sucesivas de una función. Por ahora resaltamos la relación entre los conjuntos de funciones recién definidos:

$$C^1(\Omega, Y) \subset D(\Omega, Y) \subset \mathcal{C}(\Omega, Y)$$

6.3. Primeros ejemplos

Comentemos los ejemplos más obvios de funciones diferenciables entre espacios normados arbitrarios X e Y . Si $f : X \rightarrow Y$ es constante, cualquiera que sea $a \in X$, tenemos $f(x) = f(a)$ para todo $x \in X$, luego se cumple obviamente (4) sin más que tomar $T = 0$. Por tanto:

- Si $f : X \rightarrow Y$ es constante, entonces f es diferenciable, con $Df(a) = 0$ para todo $a \in X$. Por tanto $f \in C^1(X, Y)$.

Si $f : X \rightarrow Y$ es lineal y continua, fijado $a \in X$, tenemos entonces $f(x) - f(a) = f(x - a)$ para todo $x \in X$, luego se cumple (4) con $T = f$. Por tanto:

- Si $f \in L(X, Y)$, entonces f es diferenciable con $Df(a) = f$ para todo $a \in X$. Vemos por tanto que Df es constante, luego $f \in C^1(X, Y)$.

Está claro que la noción de diferenciabilidad no está pensada para trabajar con ejemplos tan obvios como los anteriores. Tan pronto como descendamos a casos particulares, es decir, concretemos los espacios normados X e Y , tendremos ejemplos menos obvios de funciones diferenciables.

Para funciones con valores en un espacio normado, las operaciones disponibles son la suma y el producto por escalares. Veamos pues una primera regla para el cálculo de diferenciales, que muestra la diferenciación como una operación *lineal*.

- Si Ω es un abierto no vacío de X , $f, g : \Omega \rightarrow Y$ son diferenciables en $a \in \Omega$ y $\alpha, \beta \in \mathbb{R}$, entonces $\alpha f + \beta g$ es diferenciable en a con:

$$D(\alpha f + \beta g)(a) = \alpha Df(a) + \beta Dg(a)$$

Si $f, g \in D(\Omega, Y)$, tendremos $\alpha f + \beta g \in D(\Omega, Y)$ con $D(\alpha f + \beta g) = \alpha Df + \beta Dg$. Por tanto, si $f, g \in C^1(\Omega, Y)$, se tiene $\alpha f + \beta g \in C^1(\Omega, Y)$. Así pues, $D(\Omega, Y)$ y $C^1(\Omega, Y)$ son subespacios vectoriales de $\mathcal{C}(\Omega, Y)$.

Basta pensar que, para todo $x \in \Omega \setminus \{a\}$, se tiene claramente

$$\begin{aligned} & \frac{(\alpha f + \beta g)(x) - (\alpha f + \beta g)(a) - (\alpha Df(a) + \beta Dg(a))(x - a)}{\|x - a\|} \\ &= \alpha \frac{f(x) - f(a) - Df(a)(x - a)}{\|x - a\|} + \beta \frac{g(x) - g(a) - Dg(a)(x - a)}{\|x - a\|} \end{aligned}$$

Por ser f y g diferenciables en a , los dos sumandos del segundo miembro tienen límite 0 en dicho punto, luego lo mismo le ocurre a la función que aparece en el primer miembro. Puesto que $\alpha Df(a) + \beta Dg(a) \in L(X, Y)$, tenemos el resultado deseado. ■

6.4. Regla de la cadena

Comprobamos ahora que la composición de aplicaciones preserva la diferenciabilidad. Esta es, sin duda, la regla más útil para el cálculo de diferenciales.

Teorema. Sean X, Y, Z tres espacios normados, Ω y U abiertos no vacíos de X e Y respectivamente, y sean $f : \Omega \rightarrow U$ y $g : U \rightarrow Z$ dos funciones. Si f es diferenciable en un punto $a \in \Omega$ y g es diferenciable en $b = f(a)$, entonces $g \circ f$ es diferenciable en a , con

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

Por tanto, si $f \in D(\Omega, Y)$ y $g \in D(U, Z)$, entonces $g \circ f \in D(\Omega, Z)$.

Demostración. Para mayor claridad, la dividimos en tres etapas.

(a). Primero traducimos la diferenciabilidad de f y g , con una notación que haga más fáciles los cálculos. Por una parte, definimos una función $\Phi : \Omega \rightarrow \mathbb{R}_0^+$ de la siguiente forma:

$$\Phi(x) = \frac{\|f(x) - f(a) - Df(a)(x-a)\|}{\|x-a\|} \quad \forall x \in \Omega \setminus \{a\} \quad \text{y} \quad \Phi(a) = 0$$

Por ser f diferenciable en a , tenemos que Φ es continua en a , y verifica:

$$\|f(x) - f(a) - Df(a)(x-a)\| = \Phi(x) \|x-a\| \quad \forall x \in \Omega \quad (9)$$

Análogamente, definimos $\Psi : U \rightarrow \mathbb{R}_0^+$ por

$$\Psi(y) = \frac{\|g(y) - g(b) - Dg(b)(y-b)\|}{\|y-b\|} \quad \forall y \in U \setminus \{b\} \quad \text{y} \quad \Psi(b) = 0$$

Entonces Ψ es continua en b y verifica:

$$\|g(y) - g(b) - Dg(b)(y-b)\| = \Psi(y) \|y-b\| \quad \forall y \in U \quad (10)$$

Por último, sólo para abreviar la notación, definimos también $\Lambda : \Omega \rightarrow \mathbb{R}_0^+$ por

$$\Lambda(x) = \|(g \circ f)(x) - (g \circ f)(a) - (Dg(b) \circ Df(a))(x-a)\| \quad \forall x \in \Omega$$

Puesto que $Dg(b) \circ Df(a) \in L(X, Z)$, bastará comprobar que $\lim_{x \rightarrow a} \frac{\Lambda(x)}{\|x-a\|} = 0$.

(b). En una segunda fase, obtenemos una desigualdad clave, que relaciona Λ con Φ y Ψ .

Para $x \in \Omega$, escribimos $y = f(x) \in U$, y usamos en Z la desigualdad triangular:

$$0 \leq \Lambda(x) \leq \|g(y) - g(b) - Dg(b)(y-b)\| + \|Dg(b)[y-b - Df(a)(x-a)]\| \quad (11)$$

Trabajamos ahora por separado con los dos sumandos que han aparecido.

El último es el más sencillo, usamos la definición de la norma en $L(Y, Z)$ junto con (9):

$$\begin{aligned} \|Dg(b)[y-b - Df(a)(x-a)]\| &\leq \|Dg(b)\| \|y-b - Df(a)(x-a)\| \\ &= \|Dg(b)\| \|f(x) - f(a) - Df(a)(x-a)\| \\ &= \|Dg(b)\| \Phi(x) \|x-a\| \end{aligned} \quad (12)$$

En vista de (10), el otro sumando es:

$$\|g(y) - g(b) - Dg(b)(y - b)\| = \Psi(y) \|y - b\| = \Psi(f(x)) \|f(x) - f(a)\|$$

La desigualdad triangular en Y , junto con (9) y la definición de la norma en $L(X, Y)$, nos dan

$$\begin{aligned} \|f(x) - f(a)\| &\leq \|f(x) - f(a) - Df(a)(x - a)\| + \|Df(a)(x - a)\| \\ &\leq [\Phi(x) + \|Df(a)\|] \|x - a\| \end{aligned}$$

de donde deducimos que

$$\|g(y) - g(b) - Dg(b)(y - b)\| \leq \Psi(f(x)) [\Phi(x) + \|Df(a)\|] \|x - a\| \quad (13)$$

Usando (12) y (13), deducimos de (11) la desigualdad que buscábamos:

$$0 \leq \Lambda(x) \leq \Psi(f(x)) [\Phi(x) + \|Df(a)\|] \|x - a\| + \|Dg(b)\| \Phi(x) \|x - a\| \quad (14)$$

(c). Usando la desigualdad (14), junto con las propiedades conocidas de Φ y Ψ , concluimos fácilmente la demostración. Para $x \in \Omega \setminus \{a\}$, de (14) deducimos que

$$0 \leq \frac{\Lambda(x)}{\|x - a\|} \leq \Psi(f(x)) [\Phi(x) + \|Df(a)\|] + \|Dg(b)\| \Phi(x) \quad (15)$$

luego bastará probar que el último miembro de esta desigualdad tiene límite 0 en el punto a .

Como f es diferenciable, luego continua, en el punto a , y Ψ es continua en $b = f(a)$, vemos que $\Psi \circ f$ es continua en a . Además, Φ es continua en a , luego tenemos

$$\lim_{x \rightarrow a} \Psi(f(x)) = \Psi(f(a)) = \Psi(b) = 0 \quad \text{y} \quad \lim_{x \rightarrow a} \Phi(x) = \Phi(a) = 0$$

Así pues, concluimos claramente que

$$\lim_{x \rightarrow a} \left[\Psi(f(x)) [\Phi(x) + \|Df(a)\|] + \|Dg(b)\| \Phi(x) \right] = 0$$

y (15) nos dice que $\lim_{x \rightarrow a} \frac{\Lambda(x)}{\|x - a\|} = 0$, como queríamos demostrar. ■

Para poder estudiar la composición de dos funciones de clase C^1 necesitamos una sencilla observación sobre la norma de una composición de aplicaciones lineales y continuas:

■ Si X, Y, Z son espacios normados, $T \in L(X, Y)$ y $S \in L(Y, Z)$, entonces:

$$\|S \circ T\| \leq \|S\| \|T\| \quad (16)$$

La comprobación de esta desigualdad es inmediata. Para $x \in X$ tenemos:

$$\|(S \circ T)x\| = \|S(Tx)\| \leq \|S\| \|Tx\| \leq \|S\| \|T\| \|x\|$$

donde hemos usado la definición de la norma en $L(Y, Z)$ y en $L(X, Y)$. Pero usándola también en $L(X, Z)$, la desigualdad obtenida nos dice que $\|S \circ T\| \leq \|S\| \|T\|$, como se quería. ■

Podemos ya completar la regla de la cadena en la forma que sigue.

- Sean X, Y, Z espacios normados y Ω, U subconjuntos abiertos de X, Y respectivamente. Si $f \in C^1(\Omega, Y)$ verifica que $f(\Omega) \subset U$ y $g \in C^1(U, Z)$, entonces $g \circ f \in C^1(\Omega, Z)$.

Para $x, a \in \Omega$, escribiendo $y = f(x) \in U$ y $b = f(a) \in U$ tenemos claramente

$$\begin{aligned} D(g \circ f)(x) - D(g \circ f)(a) &= Dg(y) \circ Df(x) - Dg(b) \circ Df(a) \\ &= Dg(y) \circ (Df(x) - Df(a)) + (Dg(y) - Dg(b)) \circ Df(a) \end{aligned}$$

La desigualdad triangular, junto con (16), nos dan:

$$\begin{aligned} \|D(g \circ f)(x) - D(g \circ f)(a)\| &\leq \|Dg(y)\| \|Df(x) - Df(a)\| + \|Dg(y) - Dg(b)\| \|Df(a)\| \\ &= \|Dg(f(x))\| \|Df(x) - Df(a)\| + \|Dg(f(x)) - Dg(f(a))\| \|Df(a)\| \end{aligned}$$

Por ser f continua en a y Dg continua en $f(a)$, la función $Dg \circ f$ es continua en a . Por otra parte, Df también es continua en a , luego tenemos

$$\lim_{x \rightarrow a} Dg(f(x)) = Dg(f(a)) \quad \text{y} \quad \lim_{x \rightarrow a} Df(x) = Df(a)$$

Usando ahora que, en cualquier espacio normado, la norma es una función continua, de la desigualdad anterior deducimos claramente que

$$\lim_{x \rightarrow a} \|D(g \circ f)(x) - D(g \circ f)(a)\| = 0$$

luego $D(g \circ f)$ es continua en a , y esto es válido para todo $a \in \Omega$, como queríamos. ■

6.5. Funciones con valores en un producto

Como ocurrió con la continuidad, vamos a ver que la diferenciabilidad de una función que tome valores en un producto de espacios normados equivale a la de sus componentes. Esto permitirá reducir el estudio de la diferenciabilidad de campos vectoriales al caso de campos escalares. En lo que sigue Y_1, Y_2, \dots, Y_M son espacios normados y consideramos el espacio

normado producto $Y = \prod_{j=1}^M Y_j$. Para cada $j \in \Delta_M$ recordemos la j -ésima proyección coordenada

en Y , que es la aplicación sobreyectiva $\pi_j : Y \rightarrow Y_j$ dada por $\pi_j(y) = y(j)$ para todo $y \in Y$. Es claro que $\pi_j \in L(Y, Y_j)$ con $\|\pi_j\| = 1$, luego $\pi_j \in C^1(Y, Y_j)$ con $D\pi_j(y) = \pi_j$ para todo $y \in Y$.

En sentido contrario, usaremos la *inyección natural* de Y_j en Y , es decir, la aplicación inyectiva $I_j : Y_j \rightarrow Y$, definida como sigue: para cada $u \in Y_j$, $I_j(u)$ es el vector de Y cuyas componentes son todas nulas, salvo la j -ésima, que es u :

$$[I_j(u)](j) = u \quad \text{y} \quad [I_j(u)](i) = 0 \quad \forall i \in \Delta_M \setminus \{j\}$$

De nuevo es evidente que $I_j \in L(Y_j, Y)$ con $\|I_j\| = 1$, luego $I_j \in C^1(Y_j, Y)$ con $DI_j(u) = I_j$ para todo $u \in Y_j$. Nótese también que $\pi_j \circ I_j$ es la aplicación identidad en Y_j .

Si ahora Ω es un abierto no vacío de un espacio normado X , dada una función $f : \Omega \rightarrow Y$, escribimos $f = (f_1, f_2, \dots, f_M)$ para indicar sus componentes. Esto significa que, para $x \in \Omega$ y $j \in \Delta_M$, la j -ésima componente del vector $f(x) \in Y$ es $f_j(x)$. Usando las proyecciones e inyecciones antes definidas, la relación de ida y vuelta entre f y sus componentes se expresa con gran comodidad:

$$f_j = \pi_j \circ f \quad \forall j \in \Delta_M \quad \text{y} \quad f = \sum_{j=1}^M I_j \circ f_j \quad (17)$$

Fijado $j \in \Delta_M$, la regla de la cadena y la primera igualdad de (17), nos dicen que, si f es diferenciable en $a \in \Omega$, entonces f_j también lo es, con $Df_j(a) = \pi_j \circ Df(a)$.

Recíprocamente, supongamos que f_j es diferenciable en a para todo $j \in \Delta_M$. Usando la regla de la cadena, la de diferenciación de una suma y la segunda igualdad de (17), deducimos que f es diferenciable en a con $Df(a) = \sum_{j=1}^M I_j \circ Df_j(a)$. En resumen, tal como habíamos anunciado, la diferenciabilidad de f en cada punto equivale a la de todas sus componentes, y tenemos la relación explícita entre las correspondientes diferenciales:

- Sea Ω un abierto no vacío de un espacio normado X y sea $f = (f_1, f_2, \dots, f_M) : \Omega \rightarrow Y$. Entonces f es diferenciable en un punto $a \in \Omega$ si, y sólo si, f_j es diferenciable en a para todo $j \in \Delta_M$, en cuyo caso se tiene:

$$Df_j(a) = \pi_j \circ Df(a) \quad \forall j \in \Delta_M \quad \text{y} \quad Df(a) = \sum_{j=1}^M I_j \circ Df_j(a) \quad (18)$$

Por tanto, $f \in D(\Omega, Y)$ si, y sólo si, $f_j \in D(\Omega, Y_j)$ para todo $j \in \Delta_M$. Análogamente se tiene $f \in C^1(\Omega, Y)$ si, y sólo si, $f_j \in C^1(\Omega, Y_j)$ para todo $j \in \Delta_M$.

Sólo queda comprobar la última afirmación. Para $a, x \in \Omega$ y $j \in \Delta_M$ tenemos:

$$\|Df_j(x) - Df_j(a)\| = \|\pi_j \circ (Df(x) - Df(a))\| \leq \|Df(x) - Df(a)\|$$

Por tanto Df_j es continua siempre que Df lo sea. Recíprocamente, tenemos

$$\|Df(x) - Df(a)\| \leq \sum_{j=1}^M \|I_j \circ (Df_j(x) - Df_j(a))\| \leq \sum_{j=1}^M \|Df_j(x) - Df_j(a)\| \quad \forall x, a \in \Omega$$

y deducimos que Df será continua siempre que lo sea Df_j para todo $j \in \Delta_M$. ■

Resaltamos el significado de las igualdades (18), análogas a las de (17), con $Df(a)$ en lugar de f : para cada $j \in \Delta_M$, la j -ésima componente de la diferencial $Df(a)$ coincide la diferencial $Df_j(a)$ de la j -ésima componente de f . Equivalentemente, podemos escribir:

$$Df(a) = (Df_1(a), Df_2(a), \dots, Df_M(a))$$

6.6. Productos y cocientes de funciones diferenciables

Para dos nuevas reglas de diferenciación manejamos funciones con valores en \mathbb{R} . Dado un abierto no vacío Ω de un espacio normado X , como ya hicimos con $\mathcal{C}(\Omega)$ y $\mathcal{C}(\Omega, \mathbb{R})$, abreviamos escribiendo $D(\Omega)$ y $C^1(\Omega)$ en lugar de $D(\Omega, \mathbb{R})$ y $C^1(\Omega, \mathbb{R})$, respectivamente. Pretendemos probar que $D(\Omega)$ y $C^1(\Omega)$ son subanillos de $\mathcal{C}(\Omega)$. Para ello nos basamos en el siguiente resultado:

- La función $P : \mathbb{R}^2 \rightarrow \mathbb{R}$, definida por $P(x, y) = xy$ para todo $(x, y) \in \mathbb{R}^2$, es de clase C^1 en \mathbb{R}^2 y para todo $(x, y) \in \mathbb{R}^2$ se tiene:

$$DP(x, y)(h, k) = yh + xk \quad \forall (h, k) \in \mathbb{R}^2$$

Fijado $(x, y) \in \mathbb{R}^2$, definimos $T \in L(\mathbb{R}^2, \mathbb{R})$ por $T(h, k) = yh + xk$ para todo $(h, k) \in \mathbb{R}^2$. Usando en \mathbb{R}^2 la norma euclídea, para todo $(h, k) \in \mathbb{R}^2$ tenemos claramente

$$|P(x+h, y+k) - P(x, y) - T(h, k)| = |(x+h)(y+k) - xy - yh - xk| = |hk| \leq \|(h, k)\|^2 / 2$$

y deducimos que

$$\lim_{(h, k) \rightarrow (0, 0)} \frac{|P(x+h, y+k) - P(x, y) - T(h, k)|}{\|(h, k)\|} = 0$$

es decir, P es diferenciable en el punto (x, y) con $DP(x, y) = T$, como se quería.

Para ver que DP es continua, dados $(x, y), (a, b), (h, k) \in \mathbb{R}^2$, se tiene

$$|(DP(x, y) - DP(a, b))(h, k)| = |(y-b)h + (x-a)k| \leq \|(y-b, x-a)\| \|(h, k)\|$$

donde hemos usado la desigualdad de Cauchy-Schwartz. Por definición de la norma en $L(\mathbb{R}^2, \mathbb{R})$ deducimos que

$$\|DP(x, y) - DP(a, b)\| \leq \|(y-b, x-a)\| = \|(x, y) - (a, b)\| \quad \forall (x, y), (a, b) \in \mathbb{R}^2$$

Esto prueba que DP no sólo es continua, sino que es una aplicación no expansiva. ■

Podemos ya probar la regla para la diferenciación de un producto de funciones:

- Sea Ω un abierto de un espacio normado X y $f, g : \Omega \rightarrow \mathbb{R}$ funciones diferenciables en un punto $a \in \Omega$. Entonces fg es diferenciable en a con

$$D(fg)(a) = g(a)Df(a) + f(a)Dg(a) \quad (19)$$

Si $f, g \in D(\Omega)$ se tendrá $fg \in D(\Omega)$ con $D(fg) = gDf + fDg$, luego $fg \in C^1(\Omega)$ siempre que $f, g \in C^1(\Omega)$. Así pues, $D(\Omega)$ y $C^1(\Omega)$ son subanillos de $\mathcal{C}(\Omega)$.

La función $h = (f, g) : \Omega \rightarrow \mathbb{R}^2$ es diferenciable en a por serlo sus componentes f y g , pero es claro que $fg = P \circ h$ donde $P : \mathbb{R}^2 \rightarrow \mathbb{R}$ es la función estudiada en el resultado anterior.

Puesto que P es diferenciable en todo punto de \mathbb{R}^2 , la regla de la cadena nos dice que $f \circ g$ es diferenciable en a . Para calcular su diferencial basta observar que, para todo $x \in X$ tenemos claramente:

$$\begin{aligned} D(f \circ g)(a)(x) &= [DP(h(a)) \circ Dh(a)](x) = DP(f(a), g(a))(Df(a)(x), Dg(a)(x)) \\ &= g(a)Df(a)(x) + f(a)Dg(a)(x) \end{aligned} \quad \blacksquare$$

Nótese que la igualdad (19) es la generalización natural de la bien conocida regla para la derivada de un producto de dos funciones reales de variable real.

Podemos ya dar nuevos ejemplos de funciones de clase C^1 en un abierto no vacío $\Omega \subset \mathbb{R}^N$. Recordemos que el conjunto $\mathcal{P}(\Omega)$ de las funciones polinómicas en Ω es el subanillo de $\mathcal{C}(\Omega)$ engendrado por las funciones constantes, junto con las restricciones a Ω de las proyecciones coordenadas. Como todas estas funciones son de clase C^1 en Ω , el subanillo que engendran está contenido en $C^1(\Omega)$ que, según hemos visto, es un subanillo de $\mathcal{C}(\Omega)$. Así pues:

- Si Ω es un abierto no vacío de \mathbb{R}^N , se tiene $\mathcal{P}(\Omega) \subset C^1(\Omega)$, es decir, toda función polinómica en Ω es de clase C^1 .

Pasamos a estudiar ahora la diferenciabilidad de un cociente de funciones diferenciables:

- Sea Ω un abierto no vacío de un espacio normado X y sean $f, g : \Omega \rightarrow \mathbb{R}$ diferenciables en un punto $a \in \Omega$, con $g(x) \neq 0$ para todo $x \in \Omega$. Entonces la función cociente f/g es diferenciable en a con

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

Por tanto, si $f, g \in D(\Omega)$ se tiene $f/g \in D(\Omega)$, y $f/g \in C^1(\Omega)$ siempre que $f, g \in C^1(\Omega)$.

Tenemos $1/g = \varphi \circ g$ donde $\varphi : \mathbb{R}^* \rightarrow \mathbb{R}$ viene dada por $\varphi(x) = 1/x$ para todo $x \in \mathbb{R}^*$, y es claro que $\varphi \in C^1(\mathbb{R}^*)$. Como g es diferenciable en a , la regla de la cadena nos dice que $1/g$ también lo es, y la regla para la diferenciabilidad de un producto nos da que $f/g = f(1/g)$ es diferenciable en a . Si $f, g \in C^1(\Omega)$, se tendrá también $1/g \in C^1(\Omega)$, luego $f/g \in C^1(\Omega)$.

Por último, para calcular la diferencial de f/g en a , la regla de la cadena nos da

$$D(1/g)(a)(x) = D\varphi(g(a))(Dg(a)(x)) = \frac{-1}{g(a)^2} Dg(a)(x) \quad \forall x \in X$$

y ahora basta usar la diferencial de un producto:

$$D(f/g)(a) = \frac{1}{g(a)} Df(a) + f(a)D(1/g)(a) = \frac{1}{g(a)^2} (g(a)Df(a) - f(a)Dg(a)) \quad \blacksquare$$

Recordemos el conjunto $\mathcal{R}(\Omega)$ de las funciones racionales en un abierto no vacío $\Omega \subset \mathbb{R}^N$, es decir $f \in \mathcal{R}(\Omega)$ cuando $f = P/Q$ donde $P, Q \in \mathcal{P}(\Omega)$ y $Q(x) \neq 0$ para todo $x \in \Omega$. Del resultado anterior deducimos claramente:

- Si Ω es un subconjunto abierto de \mathbb{R}^N , se tiene $\mathcal{R}(\Omega) \subset C^1(\Omega)$, es decir, toda función racional en Ω es de clase C^1 .

6.7. Ejercicios

1. Probar que la norma de un espacio normado $X \neq \{0\}$ nunca es diferenciable en 0.
2. Probar que la norma euclídea es diferenciable en todo punto de $\mathbb{R}^N \setminus \{0\}$ y calcular su diferencial. ¿En qué puntos de \mathbb{R}^2 es diferenciable la norma de la suma?
3. Probar que, tanto la diferenciabilidad de una función, como su diferencial, se conservan por traslaciones. Más concretamente, sean X, Y espacios normados, Ω un abierto no vacío de X y $f : \Omega \rightarrow Y$ diferenciable en un punto $a \in \Omega$. Fijados $x_0 \in X$ e $y_0 \in Y$, se define $\widehat{\Omega} = \{x \in X : x + x_0 \in \Omega\}$ y la función $\widehat{f} : \widehat{\Omega} \rightarrow Y$ dada por

$$\widehat{f}(x) = f(x + x_0) + y_0 \quad \forall x \in \widehat{\Omega}$$

Probar que \widehat{f} es diferenciable en $a - x_0$ con $D\widehat{f}(a - x_0) = Df(a)$.

4. Sea Ω un abierto no vacío de un espacio normado X y $f : \Omega \rightarrow \mathbb{R}$ diferenciable en un punto $a \in \Omega$. Sea J un intervalo abierto tal que $f(\Omega) \subset J$ y $g : J \rightarrow \mathbb{R}$ una función derivable en el punto $b = f(a)$. Probar que $g \circ f$ es diferenciable en a . ¿Cuál es la relación entre $D(g \circ f)(a)$ y $Df(a)$?
5. Dar un ejemplo de una función $f \in D(\mathbb{R}^N)$ tal que $f \notin C^1(\mathbb{R}^N)$.

Vector derivada

La noción de diferenciabilidad, que hemos estudiado en un contexto general o abstracto, para una función de un espacio normado en otro, se irá concretando en los casos particulares que más nos interesan. Consideramos en primer lugar el caso en que el espacio de partida es \mathbb{R} , es decir, tenemos una función de una sola variable real. Su diferencial en un punto se identifica entonces con un vector del espacio de llegada, que llamaremos *vector derivada*.

Nos interesa sobre todo el caso particular en que el espacio de llegada es \mathbb{R}^M con $M > 1$. Sabemos que la diferenciabilidad de nuestra función equivale entonces a la de sus componentes, que son funciones reales de variable real. Interpretamos geoméricamente el vector derivada como el vector de dirección de la *recta tangente* a una curva paramétrica, y físicamente, como el *vector velocidad* de un móvil.

7.1. Definición de vector derivada

Vamos a trabajar con funciones de una variable real y valores vectoriales, es decir, funciones definidas en un abierto de \mathbb{R} , que toman valores en un espacio normado Y . En realidad, no es necesario suponer que el conjunto de definición sea abierto, pero esto es poco relevante. Para describir la diferencial en un punto de una tal función, usamos una sencilla observación, que generaliza lo conocido para $Y = \mathbb{R}$.

- Si Y es un espacio normado, la aplicación $\Phi: L(\mathbb{R}, Y) \rightarrow Y$, definida por $\Phi(T) = T(1)$ para toda $T \in L(\mathbb{R}, Y)$, es una biyección lineal que preserva la norma, luego permite identificar totalmente el espacio normado $L(\mathbb{R}, Y)$ con el espacio normado Y .

Es obvio que Φ es lineal y, para $T \in L(\mathbb{R}, Y)$ se tiene que $T(t) = t T(1)$ para todo $t \in \mathbb{R}$, luego

$$\|T\| = \sup \{ |t| \|T(1)\| : t \in \mathbb{R}, |t| = 1 \} = \|T(1)\|$$

En particular Φ es inyectiva, pero también es sobreyectiva, puesto que, dado un $y \in Y$, basta definir $T(t) = ty$ para todo $t \in \mathbb{R}$, para obtener $T \in L(\mathbb{R}, Y)$ tal que $\Phi(T) = T(1) = y$. ■

Sea ahora Ω un abierto de \mathbb{R} y $f : \Omega \rightarrow Y$ una función diferenciable en un punto $a \in \Omega$. Su diferencial $Df(a) \in L(\mathbb{R}, Y)$ se identifica con el vector $y = Df(a)(1) \in Y$, lo cual significa que $Df(a)(t) = ty$ para todo $t \in \mathbb{R}$. Para calcular explícitamente el vector y , la definición de diferencial nos dice que

$$0 = \lim_{t \rightarrow a} \frac{\|f(t) - f(a) - (t-a)y\|}{|t-a|} = \lim_{t \rightarrow a} \left\| \frac{f(t) - f(a)}{t-a} - y \right\| \quad (1)$$

lo cual es equivalente a

$$y = \lim_{t \rightarrow a} \frac{f(t) - f(a)}{t-a} \quad (2)$$

El límite que ha aparecido es exactamente el mismo que usamos para definir la derivada de una función real de variable real, sólo que ahora tenemos el límite de una función con valores en Y , que es un vector de Y . Esto motiva la definición que sigue.

Decimos que f es **derivable** en $a \in \Omega$, cuando la función $t \mapsto \frac{f(t) - f(a)}{t-a}$, de $\Omega \setminus \{a\}$ en Y , tiene límite en el punto a . Dicho límite es entonces el **vector derivada** de f en a , que se denota por $f'(a)$, es decir, $f'(a) = \lim_{t \rightarrow a} \frac{f(t) - f(a)}{t-a} \in Y$.

Decimos simplemente que f es **derivable**, cuando es derivable en todo punto de Ω , en cuyo caso podemos considerar la función $f' : \Omega \rightarrow Y$ que a cada punto $x \in \Omega$ hace corresponder el vector derivada $f'(x)$, y decimos que f' es la **función derivada** de f .

Hemos visto que, si f es diferenciable en $a \in \Omega$, entonces f es derivable en a y se verifica que $f'(a) = Df(a)(1)$. Pero recíprocamente, si f es derivable en a , basta tomar $y = f'(a)$ para tener (2), luego también (1), así que f es diferenciable en a con $Df(a)(t) = tf'(a)$ para todo $t \in \mathbb{R}$. En resumen:

- Sea Y un espacio normado y Ω un subconjunto abierto de \mathbb{R} . Una función $f : \Omega \rightarrow Y$ es diferenciable en un punto $a \in \Omega$ si, y sólo si, f es derivable en a , en cuyo caso, la diferencial y el vector derivada de f en a se determinan mutuamente por:

$$f'(a) = Df(a)(1) \quad \text{y} \quad Df(a)(t) = tf'(a) \quad \forall t \in \mathbb{R}$$

Por tanto f es diferenciable si, y sólo si, es derivable. Por último, $f \in C^1(\Omega, Y)$ si, y sólo si, f es derivable y su función derivada f' es continua.

Sólo queda comprobar la última afirmación, que se deduce de la total identificación de $L(\mathbb{R}, Y)$ con Y antes comprobada: para $x, a \in \Omega$ tenemos

$$\|Df(x) - Df(a)\| = \|f'(x) - f'(a)\|$$

luego Df es continua si, y sólo si, f' es continua. ■

Tenemos pues la misma situación que en el caso $Y = \mathbb{R}$. Para funciones de una variable real con valores en un espacio normado arbitrario Y , la identificación de $L(\mathbb{R}, Y)$ con Y hace que la distinción entre diferencial y derivada sea sólo cuestión de matiz.

Comprobamos ahora una propiedad del vector derivada que será útil para su interpretación geométrica y física.

- Sea Ω un abierto de \mathbb{R} , Y un espacio normado y $f : \Omega \rightarrow Y$ una función derivable en un punto $a \in \Omega$. Entonces, para todo $\varepsilon > 0$, existe un $\delta > 0$ verificando:

$$\left. \begin{array}{l} t_1, t_2 \in \Omega, \quad t_1 \neq t_2 \\ a - \delta < t_1 \leq a \leq t_2 < a + \delta \end{array} \right\} \implies \left\| \frac{f(t_2) - f(t_1)}{t_2 - t_1} - f'(a) \right\| \leq \varepsilon \quad (3)$$

Dado $\varepsilon > 0$, por definición del vector derivada, existe $\delta > 0$ tal que:

$$t \in \Omega, \quad |t - a| < \delta \implies \|f(t) - f(a) - (t - a)f'(a)\| \leq \varepsilon |t - a|$$

Entonces, si $t_1, t_2 \in \Omega$ verifican que $a - \delta < t_1 \leq a \leq t_2 < a + \delta$, usamos la desigualdad anterior para $t = t_1$, y también para $t = t_2$, obteniendo:

$$\begin{aligned} & \|f(t_2) - f(t_1) - (t_2 - t_1)f'(a)\| \\ & \leq \|f(t_2) - f(a) - (t_2 - a)f'(a)\| + \|f(a) - f(t_1) - (a - t_1)f'(a)\| \\ & \leq \varepsilon (|t_2 - a| + |a - t_1|) = \varepsilon (t_2 - t_1) \end{aligned}$$

Cuando $t_1 \neq t_2$, basta dividir ambos miembros por $t_2 - t_1 > 0$ para tener (3). ■

Nótese que lo interesante de la observación anterior es la posibilidad de tomar $t_1 < a < t_2$, pues para $t_1 = a$ o $t_2 = a$ el resultado no es más que la definición del vector derivada.

Pensemos en el vector derivada en el caso $Y = \mathbb{R}^M$, con cualquier norma. Entonces cada función $f : \Omega \rightarrow \mathbb{R}^M$ tiene M componentes, $f = (f_1, f_2, \dots, f_M)$, que son funciones reales de variable real. Sabemos que f es diferenciable en un punto $a \in \Omega$ si, y sólo si, lo son todas sus componentes, en cuyo caso, para $j \in \Delta_M$ se tiene $Df_j(a) = \pi_j \circ Df(a)$ donde π_j es la j -ésima proyección coordenada en \mathbb{R}^M . Como consecuencia,

$$f'_j(a) = Df_j(a)(1) = (\pi_j \circ Df(a))(1) = \pi_j(f'(a))$$

Para cada $j \in \Delta_M$, vemos que $f'_j(a) \in \mathbb{R}$ es la j -ésima coordenada del vector derivada $f'(a)$ en la base usual de \mathbb{R}^M , que como siempre denotamos por $\{e_1, e_2, \dots, e_M\}$. Así pues, tenemos:

- Sea Ω un abierto de \mathbb{R} y $f = (f_1, f_2, \dots, f_M) : \Omega \rightarrow \mathbb{R}^M$ una función. Entonces f es derivable en un punto $a \in \Omega$ si, y sólo si, f_j es derivable en a para todo $j \in \Delta_M$, en cuyo caso se tiene $f'(a) = (f'_1(a), f'_2(a), \dots, f'_M(a))$ es decir:

$$f'_j(a) = \pi_j(f'(a)) \quad \forall j \in \Delta_M \quad \text{o bien,} \quad f'(a) = \sum_{j=1}^M f'_j(a) e_j \quad (4)$$

7.2. Recta tangente a una curva paramétrica

Para hacer una interpretación geométrica del vector derivada, fijamos un intervalo abierto no vacío $J \subset \mathbb{R}$ y una función $\gamma : J \rightarrow \mathbb{R}^M$, que suponemos continua. La imagen de γ , es decir, el conjunto $C = \gamma(J) = \{\gamma(t) : t \in J\} \subset \mathbb{R}^M$, es lo que en Geometría se conoce como una curva definida en forma paramétrica, o más brevemente, una **curva paramétrica**, en \mathbb{R}^M .

Los casos más interesantes se presentan cuando $M = 2$ y tenemos una *curva plana*, o bien cuando $M = 3$ y tenemos una *curva alabeada*. Aunque sólo sea para tratar simultáneamente ambos casos, merece la pena trabajar en general con $M \in \mathbb{N}$, $M > 1$.

La denominación de curva paramétrica se refiere a que entendemos t como un *parámetro*, y a cada valor $t \in J$ de dicha variable, corresponde un único punto $\gamma(t) \in C$. Se dice que la función γ parametriza la curva C , o que γ es una **parametrización** de C . Esto es sólo una forma de hablar, cualquier función parametriza siempre a su imagen.

Conviene resaltar que γ puede no ser inyectiva, distintos valores del parámetro pueden dar lugar al mismo punto de la curva C . Por otra parte, es claro que la curva C no determina a la función γ , una misma curva tiene siempre parametrizaciones muy diversas.

Pues bien, si γ es derivable en un punto $a \in J$ con $\gamma'(a) \neq 0$, podemos considerar la recta

$$R = \{ \gamma(a) + t\gamma'(a) : t \in \mathbb{R} \}$$

es decir, la única recta en \mathbb{R}^M que pasa por el punto $\gamma(a)$ y tiene vector de dirección $\gamma'(a)$. Se dice que R es la **recta tangente** a la curva C en el punto $x = \gamma(a)$. Así pues, el vector derivada $\gamma'(a) \neq 0$ es un *vector de dirección de la recta tangente* a la curva $C = \gamma(J)$ en el punto $x = \gamma(a) \in C$.

Nótese que, para otro valor $b \in J$ del parámetro, podemos tener también $x = \gamma(b)$, pero γ podría no ser derivable en b , o ser derivable, pero con $\gamma'(b) \neq \gamma'(a)$, luego es ambiguo hablar de la recta tangente en un punto $x \in C$. Indicando que $x = \gamma(a)$ evitamos la ambigüedad, al precisar el valor del parámetro al que nos estamos refiriendo.

La denominación de la recta tangente tiene una clara explicación geométrica, como vamos a ver. Fijada cualquier norma en \mathbb{R}^M , y dado $\varepsilon > 0$ con $\varepsilon < \|\gamma'(a)\|$, podemos conseguir $\delta > 0$ con $]a - \delta, a + \delta[\subset J$, de forma que, igual que en (3), se tenga:

$$a - \delta < t_1 \leq a \leq t_2 < a + \delta, \quad t_1 \neq t_2 \implies \left\| \frac{\gamma(t_2) - \gamma(t_1)}{t_2 - t_1} - \gamma'(a) \right\| \leq \varepsilon$$

Entonces $\frac{\gamma(t_2) - \gamma(t_1)}{t_2 - t_1} \neq 0$ es un vector de dirección de la recta (secante a la curva C), que pasa por los puntos $\gamma(t_1)$ y $\gamma(t_2)$, que pueden ser ambos distintos de $\gamma(a)$. Pues bien, dicho vector tiende a ser $\gamma'(a)$ cuando ambos valores t_1, t_2 del parámetro tienden a coincidir con a , siempre que se mantenga la condición $t_1 \leq a \leq t_2$. Esto concuerda con la visión geométrica de la recta tangente como “límite” de rectas secantes.

Resaltamos que, para todo lo dicho anteriormente, la hipótesis $\gamma'(a) \neq 0$ es esencial. Si γ es derivable en a con $\gamma'(a) \neq 0$ se dice que $x = \gamma(a)$ es un *punto regular* de la curva $C = \gamma(J)$ y, si esto ocurre para todo $a \in J$ decimos que C es una *curva regular*. Si, por el contrario, γ no es derivable en un punto $a \in J$, o bien es derivable en a pero $\gamma'(a) = 0$, entonces $x = \gamma(a)$ es un *punto singular* de la curva $C = \gamma(J)$. Debe quedar claro que estas nociones dependen de la parametrización γ que estamos usando para la curva C . Con otra parametrización distinta, los puntos regulares pueden dejar de serlo, y viceversa, por eso hablamos de la curva $C = \gamma(J)$, para resaltar la parametrización γ a la que nos referimos. Cuando $C = \gamma(J)$ es una curva regular, también podemos decir que γ es una *parametrización regular* de la curva C .

Pensemos finalmente en las componentes de γ . Para cada $j \in \Delta_M$, la función $\pi_j \circ \gamma: J \rightarrow \mathbb{R}$, suele denotarse por x_j , en vez de γ_j . Se dice entonces que las M igualdades

$$x_j = x_j(t) \quad (t \in J) \quad \text{con } j \in \Delta_M \quad (5)$$

son las *ecuaciones paramétricas* de la curva C . Debemos saber entender bien esta notación, que es intuitiva y cómoda, aunque formalmente cuestionable. En el segundo miembro, $x_j = \pi_j \circ \gamma$ es la j -ésima componente de γ , una función real de variable real definida en el intervalo J , mientras que en el primer miembro usamos x_j como una variable, cuyos valores son los que toma dicha función. Obviamente las ecuaciones paramétricas dependen de la función γ que estamos usando para parametrizar la curva C .

Sabemos que γ es derivable en un punto $a \in J$ si, y sólo si, lo es x_j para todo $j \in \Delta_M$, en cuyo caso (4) nos dice que

$$x'_j(a) = \pi_j(\gamma'(a)) \quad \forall j \in \Delta_M, \quad \text{o bien,} \quad \gamma'(a) = \sum_{j=1}^M x'_j(a) e_j$$

Pues bien, cuando γ es derivable en el punto a con $\gamma'(a) \neq 0$, la recta tangente R , como curva paramétrica que también es, tiene sus ecuaciones paramétricas, dadas por:

$$x_j = x_j(a) + t x'_j(a) \quad (t \in \mathbb{R}) \quad \text{con } j \in \Delta_M$$

7.3. Vector velocidad

Para hacer una interpretación física del vector derivada, podemos pensar que una función continua $\gamma: J \rightarrow \mathbb{R}^M$ describe un movimiento en el espacio M -dimensional, de forma que J es un intervalo de *tiempo* y, en cada instante $t \in J$, el móvil ocupa la posición $\gamma(t)$, por lo que se dice que $\gamma(t)$ es el *vector de posición* del móvil en el instante t . La curva paramétrica $C = \gamma(J)$ es la *trayectoria* del movimiento y sus ecuaciones paramétricas (5) son las *ecuaciones del movimiento*. La situación más intuitiva se presenta cuando $M = 2$ y tenemos un movimiento en el plano, o cuando $M = 3$ y se trata de un movimiento en el espacio tridimensional.

En este planteamiento físico, es natural suponer que γ es derivable en todo punto de J . Fijados $t_1, t_2 \in J$ con $t_1 \leq t \leq t_2$ y $t_1 \neq t_2$, el vector $\gamma(t_2) - \gamma(t_1)$ indica el *desplazamiento* del móvil durante el intervalo de tiempo $[t_1, t_2]$ luego el vector $\frac{\gamma(t_2) - \gamma(t_1)}{t_2 - t_1}$ nos da la *velocidad media* del móvil en dicho intervalo. La afirmación (3) nos dice que dicha velocidad media tiende a coincidir con $\gamma'(t)$ cuando $t_2 - t_1$ tiende a ser cero, luego $\gamma'(t)$ debe entenderse como una velocidad “instantánea”. Por ello, para todo $t \in J$, se dice que $\gamma'(t)$ es el **vector velocidad** del móvil en el instante t . La norma euclídea $\|\gamma'(t)\|$ del vector velocidad se conoce como *celeridad* del móvil en el instante t y nos informa de la rapidez con la que el móvil se está desplazando, independientemente de la dirección en que se produzca dicho desplazamiento.

Nótese que no hay inconveniente en admitir que la celeridad instantánea pueda anularse, es decir, que $\gamma'(t) = 0$ para algún $t \in J$. Ahora bien, cuando $\gamma'(t) \neq 0$, el vector velocidad nos da la dirección de la recta tangente a la trayectoria en el instante t , luego nos informa sobre la dirección en la que se produce el movimiento, información que no tenemos cuando $\gamma'(t) = 0$.

7.4. Curvas planas

En el caso $M = 2$, una curva paramétrica en \mathbb{R}^2 , o *curva plana*, es la imagen $C = \gamma(J)$ de una función continua $\gamma: J \rightarrow \mathbb{R}^2$ definida en un intervalo abierto $J \subset \mathbb{R}$. Para denotar sus componentes, evitamos los subíndices, escribiendo $x = \pi_1 \circ \gamma$ e $y = \pi_2 \circ \gamma$, con lo que las ecuaciones paramétricas de $C = \gamma(J)$ son

$$x = x(t) \quad \text{e} \quad y = y(t) \quad (t \in J)$$

Sabemos que γ es derivable en un punto $a \in J$ si, y sólo si, lo son las funciones x e y , en cuyo caso tenemos $\gamma'(a) = (x'(a), y'(a))$. Cuando $\gamma'(a) \neq 0$, la recta tangente a la curva C en el punto $\gamma(a) = (x(a), y(a))$ tiene ecuaciones paramétricas

$$x = x(a) + t x'(a) \quad \text{e} \quad y = y(a) + t y'(a) \quad (t \in \mathbb{R})$$

Por ejemplo, la elipse C de centro $(\alpha, \beta) \in \mathbb{R}^2$ con semiejes $a, b \in \mathbb{R}^+$ puede describirse mediante las ecuaciones paramétricas

$$x = \alpha + a \cos t \quad \text{e} \quad y = \beta + b \sin t \quad (t \in \mathbb{R})$$

que nos dan una parametrización regular de dicha elipse. Es fácil ver que la recta tangente a la elipse en un punto $(x_0, y_0) \in C$ tiene ecuaciones paramétricas

$$bx = bx_0 - ta(y_0 - \beta) \quad \text{y} \quad ay = ay_0 + tb(x_0 - \alpha) \quad (t \in \mathbb{R})$$

Conviene ahora aclarar la relación con el tipo de curva que mejor conocemos: la gráfica de una función continua $\varphi: J \rightarrow \mathbb{R}$. Se dice que el conjunto

$$\text{Gr } \varphi = \{ (x, \varphi(x)) : x \in J \} \subset \mathbb{R}^2 \quad (6)$$

es una **curva explícita**, resaltando que la ordenada de cada punto de este tipo de curva se obtiene *explícitamente* como función de su abscisa. Concretamente, de $(x, y_1), (x, y_2) \in \text{Gr } \varphi$ se deduce que $x \in J$ y $\varphi(x) = y_1 = y_2$, luego está claro que en la curva $\text{Gr } \varphi$, la ordenada y es función de la abscisa x , por lo que se dice que la igualdad

$$y = \varphi(x) \quad (x \in J) \quad (7)$$

es la *ecuación explícita* de la curva $\text{Gr } \varphi$. Geométricamente, poco importa intercambiar los ejes de coordenadas, así que también se considera como curva explícita a todo conjunto que sea de la forma $\{ (\psi(y), y) : y \in J \} \subset \mathbb{R}^2$, donde $\psi: J \rightarrow \mathbb{R}$ es continua. Ahora la abscisa de cada punto de la curva es función de su ordenada, y la ecuación explícita de la curva es $x = \psi(y)$ con $y \in J$. En lo que sigue consideramos solamente curvas de la forma (6), pero cualquier afirmación que hagamos para ellas tiene su análoga para el otro tipo de curvas explícitas.

Está claro que *toda curva explícita es una curva paramétrica*. Concretamente $\text{Gr } \varphi = \gamma(J)$ donde $\gamma: J \rightarrow \mathbb{R}^2$ es la función continua definida por $\gamma(t) = (t, \varphi(t))$ para todo $t \in J$.

Por tanto, las ecuaciones paramétricas de la curva explícita $\text{Gr } \varphi$, pueden ser

$$x = t \quad \text{e} \quad y = \varphi(t) \quad (t \in J)$$

Observamos ahora que γ es derivable en un punto $a \in J$ si, y sólo si, lo es φ , en cuyo caso tenemos $\gamma'(a) = (1, \varphi'(a))$, y la condición $\gamma'(a) \neq 0$ para que exista la recta tangente, se cumple automáticamente. Dicha recta tiene ecuaciones paramétricas

$$x = a + t \quad \text{e} \quad y = \varphi(a) + t \varphi'(a) \quad \forall t \in \mathbb{R}$$

y nunca es una recta vertical, responde a la ecuación explícita

$$y - \varphi(a) = \varphi'(a)(x - a) \quad (x \in \mathbb{R})$$

Recuperamos así la interpretación geométrica de la derivada $\varphi'(a)$ como la pendiente de la recta tangente a la gráfica de φ en el punto $(a, \varphi(a))$. Queda claro que la parametrización usada para la curva explícita $\text{Gr } \varphi$ es regular si, y sólo si, φ es derivable en J .

Veamos un ejemplo de curva explícita que sirve para ilustrar la noción de punto singular: la gráfica de la función valor absoluto, $C = \text{Gr } \varphi$ donde $\varphi(x) = |x|$ para todo $x \in \mathbb{R}$. El origen es un punto singular, ya que φ no es derivable en 0. En lugar de la parametrización más obvia, dada por $\gamma(t) = (t, |t|)$ para todo $t \in \mathbb{R}$, que tampoco es derivable en 0, podemos considerar la función $\chi: \mathbb{R} \rightarrow \mathbb{R}^2$ definida por $\chi(t) = (t|t|, t^2)$ para todo $t \in \mathbb{R}$, para la que también tenemos $C = \chi(\mathbb{R})$, es decir,

$$\{(x, |x|) : x \in \mathbb{R}\} = \{(t|t|, t^2) : t \in \mathbb{R}\}$$

Así pues, podemos ver C como la curva paramétrica de ecuaciones

$$x = t|t| \quad \text{e} \quad y = t^2 \quad (t \in \mathbb{R})$$

La ventaja es que χ es derivable en \mathbb{R} , con $\chi'(t) = (2|t|, 2t)$ para todo $t \in \mathbb{R}$. Físicamente, vemos que χ describe un movimiento cuya trayectoria es la gráfica de la función valor absoluto, y su velocidad está bien definida en todo instante. Sin embargo, como $\chi'(0) = 0$, el origen de coordenadas $(0, 0) = \chi(0)$ sigue siendo un punto singular de la curva $\chi(\mathbb{R})$. Geométricamente está muy claro que la curva C no puede admitir una recta tangente en el origen, de ahí que, en cualquier parametrización que hagamos, el origen sea siempre un punto singular.

Por otra parte, es fácil dar un ejemplo de un punto de una curva, que pasa de ser regular a ser singular al cambiar la parametrización. Basta considerar una curva explícita $C = \text{Gr } \psi$, para una función $\psi: \mathbb{R} \rightarrow \mathbb{R}$ que sea derivable. Obviamente C es una curva regular para su parametrización natural: $\gamma(x) = (x, \psi(x))$ para todo $x \in \mathbb{R}$. Pero también podemos considerar la parametrización $\chi(t) = (t^3, \psi(t^3))$ para todo $t \in \mathbb{R}$, pues está bien claro que $C = \chi(\mathbb{R})$. Ahora $(0, 0) = \chi(0)$ es un punto singular, ya que $\chi'(0) = 0$.

Los ejemplos anteriores dejan bien claro que toda curva explícita se puede ver como curva paramétrica, incluso usando parametrizaciones muy diversas. Resaltamos que el recíproco está muy lejos de ser cierto, la noción de curva paramétrica es mucho más general que la de curva explícita. Por ejemplo, es claro que la circunferencia $C = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ no es una curva explícita, pues basta observar que $(0, 1), (0, -1) \in C$, luego C no puede ser la gráfica de una función. Como $(1, 0), (-1, 0) \in C$, tampoco es una curva explícita del segundo tipo antes descrito: no existe una función $\psi: J \rightarrow \mathbb{R}$ con $J \subset \mathbb{R}$ tal que $\{(\psi(y), y) : y \in J\} = C$. Sin embargo, C sí es una curva paramétrica, es una de las elipses antes consideradas.

7.5. Curvas alabeadas

Aunque resulte repetitivo, como repaso pueden venir bien algunos comentarios sobre curvas paramétricas en \mathbb{R}^3 o *curvas alabeadas*. Ahora tenemos $C = \gamma(J) \subset \mathbb{R}^3$ donde J es un intervalo abierto y $\gamma: J \rightarrow \mathbb{R}^3$ es continua. Sus componentes son $x = \pi_1 \circ \gamma$, $y = \pi_2 \circ \gamma$ y $z = \pi_3 \circ \gamma$, con lo que tenemos tres ecuaciones paramétricas

$$x = x(t), \quad y = y(t) \quad y \quad z = z(t) \quad (t \in J)$$

Como ejemplo, podemos considerar la hélice circular $H \subset \mathbb{R}^3$ de ecuaciones paramétricas

$$x = \cos t, \quad y = \sin t \quad y \quad z = t \quad (t \in \mathbb{R})$$

Volviendo al caso general, sabemos que γ es derivable en un punto $a \in J$ si y sólo si, lo son las funciones x, y, z , en cuyo caso $\gamma'(a) = (x'(a), y'(a), z'(a))$. Cuando $\gamma'(a) \neq 0$ tenemos la recta tangente a la curva C en el punto $\gamma(a)$ cuyas ecuaciones paramétricas son

$$x = x(a) + tx'(a), \quad y = y(a) + ty'(a) \quad y \quad z = z(a) + tz'(a) \quad (t \in \mathbb{R})$$

Por ejemplo, la recta tangente a la hélice circular H , en cualquier punto $(x_0, y_0, z_0) \in H$, tiene la siguientes ecuaciones paramétricas:

$$x = x_0 - ty_0, \quad y = y_0 + tx_0 \quad y \quad z = z_0 + t \quad (t \in \mathbb{R})$$

7.6. Ejercicios

1. Sea Y un espacio pre-hilbertiano, Ω un abierto de \mathbb{R} y $f, g: \Omega \rightarrow Y$ dos funciones diferenciables en un punto $a \in \Omega$. Probar que la función $\varphi: \Omega \rightarrow \mathbb{R}$ definida por

$$\varphi(t) = (f(t) | g(t)) \quad \forall t \in \Omega$$

es derivable en el punto a y calcular su derivada $\varphi'(a)$.

2. Sean Y, Z espacios normados y consideremos dos funciones $f: \Omega \rightarrow U$ y $g: U \rightarrow Z$, donde $\Omega = \Omega^\circ \subset \mathbb{R}$ y $U = U^\circ \subset Y$. Supongamos que f es derivable en un punto $a \in \Omega$ y que g es diferenciable en el punto $b = f(a)$. Calcular el vector derivada de $g \circ f$ en el punto a .
3. Sean X, Z espacios normados, $\Omega = \Omega^\circ \subset X$ y $U = U^\circ \subset \mathbb{R}$. Sea $f: \Omega \rightarrow U$ una función diferenciable en un punto $a \in \Omega$ y $g: U \rightarrow Z$ una función derivable en el punto $b = f(a)$. Calcular la diferencial de $g \circ f$ en a .
4. Probar que, dados $a, b \in \mathbb{R}^+$, la hipérbola $\{(x, y) \in \mathbb{R}^2 : b^2 x^2 - a^2 y^2 = a^2 b^2\}$ no es una curva paramétrica, pero cada una de sus ramas es una curva explícita.
5. Describir geoméricamente las curvas cuyas ecuaciones paramétricas son:

$$a) \quad x = e^t, \quad y = e^{-t} \quad \forall t \in \mathbb{R}$$

$$b) \quad x = \cosh t, \quad y = \sinh t \quad \forall t \in \mathbb{R}$$

Vector gradiente

Como segundo caso particular de la noción de diferenciabilidad, estudiamos ahora lo que ocurre cuando el espacio normado de partida es \mathbb{R}^N con $N > 1$, y el de llegada es \mathbb{R} . Tenemos pues una función real de N variables reales, es decir, un campo escalar en \mathbb{R}^N . Su diferencial en un punto, cuando existe, se describe como el producto escalar por un vector de \mathbb{R}^N , que será el *vector gradiente*. Las coordenadas de este vector son las *derivadas parciales* de nuestra función, con respecto a cada una de las variables, en el punto considerado. Veremos también la interpretación geométrica y física del gradiente de un campo escalar.

8.1. Derivadas direccionales

Volvamos por un momento a la noción general de función diferenciable. Sean pues X, Y espacios normados, Ω un subconjunto abierto de X y $f : \Omega \rightarrow Y$ una función diferenciable en un punto $a \in \Omega$. Al probar la unicidad de la diferencial, obtuvimos de hecho que

$$Df(a)(u) = \lim_{t \rightarrow 0} \frac{f(a+tu) - f(a)}{t} \quad \forall u \in X \quad (1)$$

Por tanto, para estudiar la diferenciabilidad de f en a , es natural empezar por la existencia de los límites que aparecen en el segundo miembro de (1). No es necesario trabajar con todos los vectores $u \in X$. Por una parte, el caso $u = 0$ es trivial; por otra, si $v = \lambda u$ con $\lambda \in \mathbb{R}^*$, los cambios de variable $t = \lambda s$ y $s = t/\lambda$ nos dicen que

$$\lim_{t \rightarrow 0} \frac{f(a+tu) - f(a)}{t} = y \in Y \iff \lim_{s \rightarrow 0} \frac{f(a+sv) - f(a)}{s} = \lambda y \quad (2)$$

Esto permite normalizar cada vector $u \in X \setminus \{0\}$, es decir, suponer que $\|u\| = 1$ sin perder generalidad. Una **dirección** en el espacio normado X es un vector $u \in X$ con $\|u\| = 1$, y denotaremos por S al conjunto de todas las direcciones en X , es decir, $S = \{u \in X : \|u\| = 1\}$. Observamos que, para cada $u \in S$, el límite que aparece a la derecha de (1), cuando existe, es el vector derivada de una función de variable real, con valores en el espacio normado Y .

Más concretamente, fijamos $r \in \mathbb{R}^+$ tal que $B(a, r) \subset \Omega$ y, para cada $u \in S$, consideramos la función φ_u definida de la siguiente forma:

$$\varphi_u :]-r, r[\rightarrow Y, \quad \varphi_u(t) = f(a + tu) \quad \forall t \in]-r, r[$$

Obsérvese que φ_u describe el comportamiento de f en la recta que pasa por el punto a y tiene a u como vector de dirección, suficientemente cerca del punto a .

Pues bien, dado $u \in S$, decimos que f es **derivable en la dirección u** , en el punto a , cuando la función φ_u es derivable en 0, en cuyo caso, al vector derivada $\varphi'_u(0)$ lo llamamos **derivada direccional** de f en a , según la dirección u , y lo denotamos por $f'_u(a)$. Así pues,

$$f'_u(a) = \varphi'_u(0) = \lim_{t \rightarrow 0} \frac{\varphi_u(t) - \varphi_u(0)}{t} = \lim_{t \rightarrow 0} \frac{f(a + tu) - f(a)}{t}$$

Usando la equivalencia (2), con $\lambda = -1$, vemos que, siempre en el punto a , f es derivable en la dirección u si, y sólo si, lo es en la dirección $-u$, en cuyo caso se tiene $f'_{-u}(a) = -f'_u(a)$.

Decimos que f es **direccionalmente derivable** en el punto a cuando es derivable en todas las direcciones $u \in S$. En vista de (1), tenemos la siguiente relación entre la diferenciabilidad y la derivabilidad direccional:

- Sean X, Y espacios normados, Ω un abierto de X y $f : \Omega \rightarrow Y$ una función. Si f es diferenciable en un punto $a \in \Omega$, entonces f es direccionalmente derivable en a con

$$f'_u(a) = Df(a)(u) \quad \forall u \in S$$

8.2. Derivadas parciales

8.2.1. Caso general

A partir de ahora trabajamos en el caso $X = \mathbb{R}^N$, mientras que de momento, Y sigue siendo un espacio normado arbitrario. Fijamos un abierto Ω de \mathbb{R}^N , una función $f : \Omega \rightarrow Y$ y un punto $a \in \Omega$. A todos los efectos, de entrada para que las direcciones en \mathbb{R}^N estén definidas sin ambigüedad, usaremos en \mathbb{R}^N la norma euclídea, escribimos $S = \{u \in \mathbb{R}^N : \|u\| = 1\}$ y fijamos $r \in \mathbb{R}^+$ de forma que $B(a, r) \subset \Omega$. La definición de derivada direccional se aplica a las direcciones de los ejes de coordenadas, es decir, las de la base usual $\{e_1, e_2, \dots, e_N\}$. A partir de este momento, todo lo que hacemos está ligado a dicha base.

Pues bien, dado $k \in \Delta_N$, cuando f es derivable en el punto a , en la dirección e_k , decimos que f es **parcialmente derivable con respecto a la k -ésima variable** en el punto a . Entonces, la derivada direccional de f en a , según la dirección e_k , se denomina **derivada parcial de f con respecto a la k -ésima variable** en el punto a , y se denota por $\frac{\partial f}{\partial x_k}(a)$, es decir,

$$\frac{\partial f}{\partial x_k}(a) = f'_{e_k}(a) = \lim_{t \rightarrow 0} \frac{f(a + te_k) - f(a)}{t} \quad (3)$$

Cuando esto ocurre para todo $k \in \Delta_N$, decimos que f es **parcialmente derivable** en a , y entonces tenemos N derivadas parciales, $\frac{\partial f}{\partial x_k}(a)$ con $k \in \Delta_N$, definidas por la igualdad anterior. Como esta noción es más débil que la derivabilidad direccional, se tiene:

- Si f es diferenciable en a , entonces f es parcialmente derivable en a con:

$$\frac{\partial f}{\partial x_k}(a) = Df(a)(e_k) \quad \forall k \in \Delta_N \quad (4)$$

8.2.2. Caso de un campo vectorial

Pasemos al caso particular en que $Y = \mathbb{R}^M$, es decir, tenemos un campo vectorial. Como ocurrió con la diferenciabilidad, su derivabilidad parcial equivale a la de todas sus componentes:

- Sea Ω un abierto de \mathbb{R}^N , $f = (f_1, f_2, \dots, f_M) : \Omega \rightarrow \mathbb{R}^M$ un campo vectorial y $k \in \Delta_N$. Se tiene entonces que f es parcialmente derivable con respecto a la k -ésima variable en un punto $a \in \Omega$, si y sólo si, lo es f_j para todo $j \in \Delta_M$, en cuyo caso:

$$\frac{\partial f}{\partial x_k}(a) = \left(\frac{\partial f_1}{\partial x_k}(a), \frac{\partial f_2}{\partial x_k}(a), \dots, \frac{\partial f_M}{\partial x_k}(a) \right)$$

En efecto, basta usar la definición de derivada parcial, caso particular de (3), pues para $a \in \Omega$ y $k \in \Delta_M$ se tiene evidentemente

$$\lim_{t \rightarrow 0} \frac{f(a + te_k) - f(a)}{t} = y \in \mathbb{R}^M \iff \lim_{t \rightarrow 0} \frac{f_j(a + te_k) - f_j(a)}{t} = y(j) \quad \forall j \in \Delta_M \quad \blacksquare$$

Así pues, el estudio de la derivabilidad parcial de un campo vectorial se reduce al caso de un campo escalar, en el que nos centramos a partir de ahora.

8.2.3. Caso de un campo escalar

Sea Ω un abierto de \mathbb{R}^N y $f : \Omega \rightarrow \mathbb{R}$ un campo escalar. Fijado un punto $a \in \Omega$ y $k \in \Delta_N$, examinemos con más detalle la derivada parcial de f con respecto a la k -ésima variable en el punto a . Usamos para ello las coordenadas de a , es decir, $a = (a_1, a_2, \dots, a_N)$.

Para $t \in \mathbb{R}$, la k -ésima coordenada de $a + te_k$ es $a_k + t$, mientras las restantes coinciden con las de a , por lo que escribiremos $a + te_k = (a_1, \dots, a_{k-1}, a_k + t, a_{k+1}, \dots, a_N)$. Esta notación no excluye los casos $k = 1$ y $k = N$, se entiende que $a + te_1 = (a_1 + t, a_2, \dots, a_N)$ mientras que $a + te_N = (a_1, \dots, a_{N-1}, a_N + t)$. Por definición, la derivada parcial que nos interesa, caso de que exista, viene dada por

$$\frac{\partial f}{\partial x_k}(a) = \lim_{t \rightarrow 0} \frac{f(a_1, \dots, a_{k-1}, a_k + t, a_{k+1}, \dots, a_N) - f(a)}{t}$$

Fijado $r \in \mathbb{R}^+$ tal que $B(a, r) \subset \Omega$, los cambios de variable $t = x_k - a_k$ y $x_k = t + a_k$, teniendo en cuenta que $-r < t < r$ si, y sólo si, $a_k - r < x_k < a_k + r$, mientras que $t \rightarrow 0$ si, y sólo si, $x_k \rightarrow a_k$, nos dicen que, para cualquier $\alpha \in \mathbb{R}$, se tiene

$$\lim_{t \rightarrow 0} \frac{f(a + te_k) - f(a)}{t} = \alpha \iff \lim_{x_k \rightarrow a_k} \frac{f(a_1, \dots, a_{k-1}, x_k, a_{k+1}, \dots, a_N) - f(a)}{x_k - a_k} = \alpha$$

Considerando entonces la función $\psi :]a_k - r, a_k + r[\rightarrow \mathbb{R}$ definida por

$$\psi(x_k) = f(a_1, \dots, a_{k-1}, x_k, a_{k+1}, \dots, a_N) \quad \forall x_k \in]a_k - r, a_k + r[$$

la equivalencia anterior nos dice que f es parcialmente derivable con respecto a la k -ésima variable en el punto a si, y sólo si, ψ es derivable en el punto a_k , en cuyo caso se tiene

$$\frac{\partial f}{\partial x_k}(a) = \psi'(a_k) = \lim_{x_k \rightarrow a_k} \frac{f(a_1, \dots, a_{k-1}, x_k, a_{k+1}, \dots, a_N) - f(a)}{x_k - a_k} \quad (5)$$

Nada que no supiéramos ya, el concepto de derivada parcial de un campo escalar coincide con el de derivada de una función real de variable real, pero ahora la relación entre ambas funciones se comprende mejor. Mientras f depende de N variables x_1, x_2, \dots, x_N , vemos que la función ψ de una sola variable, describe “parcialmente” el comportamiento de f . Sólo refleja la dependencia de f respecto de la variable x_k , cuando por así decirlo, las demás variables se mantienen constantes, pues para obtener $\psi(x_k)$ a partir de $f(x_1, \dots, x_{k-1}, x_k, x_{k+1}, \dots, x_N)$, tomamos $x_j = a_j$ para todo $j \in \Delta_N \setminus \{k\}$. Por eso decimos que la derivada de ψ en a_k es sólo una “derivada parcial” de f en a .

Consideremos por ejemplo el caso $N = 2$. Si Ω es un abierto de \mathbb{R}^2 y $f : \Omega \rightarrow \mathbb{R}$ es parcialmente derivable en un punto $(a, b) \in \Omega$, sus dos derivadas parciales vienen dadas por

$$\frac{\partial f}{\partial x}(a, b) = \lim_{x \rightarrow a} \frac{f(x, b) - f(a, b)}{x - a} \quad \text{y} \quad \frac{\partial f}{\partial y}(a, b) = \lim_{y \rightarrow b} \frac{f(a, y) - f(a, b)}{y - b}$$

Volviendo al caso general, es hora de mover el punto $a \in \Omega$, hasta ahora fijo. Dado $k \in \Delta_N$, decimos que f es **derivable con respecto a la k -ésima variable**, cuando lo es en todo $x \in \Omega$, en cuyo caso tenemos la **función derivada parcial** $\frac{\partial f}{\partial x_k} : \Omega \rightarrow \mathbb{R}$ dada por $x \mapsto \frac{\partial f}{\partial x_k}(x)$. Si esto ocurre para todo $k \in \Delta_N$ decimos simplemente que f es **parcialmente derivable**. Las funciones derivadas parciales, $\frac{\partial f}{\partial x_k} : \Omega \rightarrow \mathbb{R}$ con $k \in \Delta_N$, son entonces campos escalares análogos a f , están definidos en el mismo abierto Ω que f .

Si ahora usamos las coordenadas (x_1, x_2, \dots, x_N) de un punto genérico $x \in \Omega$, la igualdad

$$\frac{\partial f}{\partial x_k}(x) = \frac{\partial f}{\partial x_k}(x_1, \dots, x_{k-1}, x_k, x_{k+1}, \dots, x_N)$$

plantea una cuestión que conviene aclarar. En el segundo miembro, x_k es la k -ésima coordenada del punto $x \in \Omega$ en el que estamos calculando una derivada parcial, pero también aparece en el símbolo $\partial f / \partial x_k$ para indicar cuál de las N derivadas parciales de f en x estamos calculando.

En vez de representar un problema, este doble uso de la variable x_k nos debe ayudar a recordar la regla práctica para calcular derivadas parciales que resuelve la mayoría de los casos, sin necesidad de recurrir a la igualdad (5), como vamos a explicar.

El símbolo $\partial f / \partial x_k$ nos indica la derivada parcial que estamos calculando, así que, en la expresión $f(x_1, \dots, x_{k-1}, x_k, x_{k+1}, \dots, x_N)$, típicamente una fórmula en la que aparecen las N variables x_1, x_2, \dots, x_N , debemos pensar que x_k es la única variable, tratar a las demás como constantes, y calcular la derivada de la función de una variable que de esta forma tenemos en mente. Ni que decir tiene, para ello podemos usar todas las reglas de derivación para funciones reales de variable real que podamos necesitar.

Este procedimiento se entenderá muy fácilmente con un ejemplo concreto, que exponemos en el caso más sencillo, $N = 2$. Entonces f es una función de dos variables que habitualmente se denotan por x e y . Por tanto, sus derivadas parciales en un punto $a = (x_0, y_0) \in \Omega$, caso de que existan, vendrán dadas por

$$\frac{\partial f}{\partial x}(x_0, y_0) = \lim_{x \rightarrow x_0} \frac{f(x, y_0) - f(x_0, y_0)}{x - x_0} \quad \text{y} \quad \frac{\partial f}{\partial y}(x_0, y_0) = \lim_{y \rightarrow y_0} \frac{f(x_0, y) - f(x_0, y_0)}{y - y_0}$$

Cuando f es parcialmente derivable, sus derivadas parciales en un punto genérico $(x, y) \in \Omega$ se denotan entonces por $\frac{\partial f}{\partial x}(x, y)$ y $\frac{\partial f}{\partial y}(x, y)$, obteniendo dos funciones de dos variables, $\frac{\partial f}{\partial x}$ y $\frac{\partial f}{\partial y}$, de Ω en \mathbb{R} , que son las dos funciones derivadas parciales de f . Concretamente, para todo $(x, y) \in \Omega$, podemos escribir:

$$\frac{\partial f}{\partial x}(x, y) = \lim_{w \rightarrow x} \frac{f(w, y) - f(x, y)}{w - x} \quad \text{y} \quad \frac{\partial f}{\partial y}(x, y) = \lim_{w \rightarrow y} \frac{f(x, w) - f(x, y)}{w - y}$$

Sea por ejemplo $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ la función dada por $f(x, y) = e^x \sin y$ para todo $(x, y) \in \mathbb{R}^2$. Si en la expresión $e^x \sin y$, vemos y como constante, pensamos en la función $x \mapsto e^x \sin y$, producto de la exponencial por una constante, cuya función derivada es ella misma, y esto es válido para todo $y \in \mathbb{R}$. Análogamente, para cualquier constante $x \in \mathbb{R}$, la función $y \mapsto e^x \sin y$ es el producto de una constante por el seno, cuya función derivada es el producto de la misma constante por el coseno. Por tanto, f es parcialmente derivable en todo punto de \mathbb{R}^2 con

$$\frac{\partial f}{\partial x}(x, y) = e^x \sin y \quad \text{y} \quad \frac{\partial f}{\partial y}(x, y) = e^x \cos y \quad \forall (x, y) \in \mathbb{R}^2$$

Nótese que $\frac{\partial f}{\partial x} = f$. En general, la definición de f puede ser mucho más complicada, pero el mecanismo de razonamiento es el mismo.

Naturalmente, las dos variables de las que depende nuestra función no siempre se denotan por x e y . Incluso es posible que x e y sean dos funciones que estemos estudiando. Por ejemplo, cuando trabajamos con las coordenadas polares en el plano, usamos determinadas restricciones de las funciones $x, y : \mathbb{R}^2 \rightarrow \mathbb{R}$ definidas, para todo $(\rho, \theta) \in \mathbb{R}^2$, por

$$x(\rho, \theta) = \rho \cos \theta \quad \text{e} \quad y(\rho, \theta) = \rho \sin \theta$$

Es claro que x e y son parcialmente derivables en todo punto $(\rho, \theta) \in \mathbb{R}^2$ con

$$\frac{\partial x}{\partial \rho}(\rho, \theta) = \cos \theta, \quad \frac{\partial y}{\partial \rho}(\rho, \theta) = \sin \theta, \quad \frac{\partial x}{\partial \theta} = -y \quad y \quad \frac{\partial y}{\partial \theta} = x$$

Nótese que en la tercera y cuarta igualdad hemos podido abreviar, escribiendo una igualdad entre funciones. En este ejemplo y otros similares, para definir las funciones x e y , se suele escribir simplemente $x = \rho \cos \theta$ e $y = \rho \sin \theta$, y se entiende que x, y son funciones de dos variables ρ, θ . Dependiendo del uso que queramos hacer de x e y , deberemos precisar el conjunto en el que definimos tales funciones, es decir, indicar los valores que pueden tomar las variables ρ y θ . Por ejemplo, si escribimos $x = \rho \cos \theta$ e $y = \rho \sin \theta$ para $\rho > 0$ y $-\pi < \theta < \pi$, está claro que x e y se definen en $\mathbb{R}^+ \times]-\pi, \pi[$.

8.3. Vector gradiente

Calculemos ahora la diferencial de un campo escalar, a partir de las derivadas parciales. Mantenemos la notación: Ω es un abierto de \mathbb{R}^N , $f: \Omega \rightarrow \mathbb{R}$ un campo escalar y $a \in \Omega$.

Si f es diferenciable en a , usando que $Df(a)$ es lineal, junto con la igualdad (4), para todo $x = (x_1, x_2, \dots, x_N) \in \mathbb{R}^N$ tenemos

$$Df(a)(x) = Df(a) \left(\sum_{k=1}^N x_k e_k \right) = \sum_{k=1}^N x_k Df(a)(e_k) = \sum_{k=1}^N x_k \frac{\partial f}{\partial x_k}(a)$$

La última suma es el producto escalar de x por el vector cuyas coordenadas en la base usual de \mathbb{R}^N son las N derivadas parciales de f en a , que recibe el nombre de vector gradiente, o simplemente gradiente, del campo escalar f en el punto a .

Dicho más formalmente, cuando el campo f es parcialmente derivable en a , el **gradiente** de f en a es, por definición, el vector $\nabla f(a) \in \mathbb{R}^N$ dado por

$$\nabla f(a) = \left(\frac{\partial f}{\partial x_1}(a), \frac{\partial f}{\partial x_2}(a), \dots, \frac{\partial f}{\partial x_N}(a) \right) = \sum_{k=1}^N \frac{\partial f}{\partial x_k}(a) e_k$$

de modo que, para cada $k \in \Delta_N$, la k -ésima componente del vector gradiente de f en a es la derivada parcial de f en a , con respecto a la k -ésima variable.

Acabamos de ver que, cuando f es diferenciable en a , su diferencial se obtiene mediante el producto escalar por el vector gradiente de f en a .

Recíprocamente, si sólo suponemos que f es parcialmente derivable en a , disponemos del vector gradiente $\nabla f(a)$ y, como el producto escalar de \mathbb{R}^N es una forma bilineal simétrica, al hacer el producto escalar por $\nabla f(a)$, obtenemos una aplicación lineal de \mathbb{R}^N en \mathbb{R} que es la única posible diferencial de f en a . Así pues, la aplicación lineal $T: \mathbb{R}^N \rightarrow \mathbb{R}$ dada por

$$T(x) = (\nabla f(a) | x)$$

es la única posible diferencial de f en a , luego f será diferenciable en a si, y sólo si, T cumple la condición que caracteriza a la diferencial. En resumen, tenemos el siguiente resultado.

- Para un campo escalar $f : \Omega \rightarrow \mathbb{R}$, donde Ω es un abierto de \mathbb{R}^N , y un punto $a \in \Omega$, las siguientes afirmaciones son equivalentes:

- (i) f es diferenciable en a .
(ii) f es parcialmente derivable en a y se verifica que:

$$\lim_{x \rightarrow a} \frac{f(x) - f(a) - (\nabla f(a) | x - a)}{\|x - a\|} = 0 \quad (6)$$

En caso de que se cumplan (i) y (ii) se tiene:

$$Df(a)(x) = (\nabla f(a) | x) \quad \forall x \in \mathbb{R}^N \quad (7)$$

Conviene analizar en general, para cualquier aplicación lineal de \mathbb{R}^N en \mathbb{R} , la relación entre diferencial y gradiente que acaba de aparecer. Llegamos así a identificar totalmente los espacios normados $L(\mathbb{R}^N, \mathbb{R})$ y \mathbb{R}^N . Para ello es ahora esencial que estemos usando en \mathbb{R}^N la norma euclídea, a todos los efectos. En particular, notando como siempre $S = \{x \in \mathbb{R}^N : \|x\| = 1\}$, tenemos

$$\|T\| = \max \{ |T(x)| : x \in S \} \quad \forall T \in L(\mathbb{R}^N, \mathbb{R})$$

- Sea $\Phi : \mathbb{R}^N \rightarrow L(\mathbb{R}^N, \mathbb{R})$ la aplicación definida, para todo $y \in \mathbb{R}^N$, por $\Phi(y) = T_y$, donde $T_y(x) = (y | x)$ para todo $x \in \mathbb{R}^N$. Se tiene que Φ es lineal, biyectiva y preserva la norma, luego permite identificar totalmente los espacios normados $L(\mathbb{R}^N, \mathbb{R})$ y \mathbb{R}^N .

La linealidad del producto escalar en la segunda variable nos dice que $T_y \in L(\mathbb{R}^N, \mathbb{R})$ mientras que, de la linealidad en primera variable deducimos fácilmente que Φ es lineal.

Para ver que Φ preserva la norma, fijamos $y \in \mathbb{R}^N \setminus \{0\}$, pues si $y = 0$ no hay nada que comprobar. La desigualdad de Cauchy-Schwartz nos dice que $|T_y(x)| \leq \|y\|$ para todo $x \in S$, pero tomando $x = y / \|y\| \in S$ obtenemos $T_y(x) = \|y\|$, luego

$$\|\Phi(y)\| = \|T_y\| = \max \{ |T_y(x)| : x \in S \} = \|y\|$$

En particular, Φ es inyectiva. Finalmente, para ver que es sobreyectiva, fijada $T \in L(\mathbb{R}^N, \mathbb{R})$, tomamos $y = \sum_{k=1}^N T(e_k) e_k$ y tenemos claramente $T_y(e_k) = T(e_k)$ para todo $k \in \Delta_N$, de donde por ser T y T_y lineales, deducimos que $T = T_y = \Phi(y)$. ■

Recuérdese que \mathbb{R}^N también se identificó en su momento con $L(\mathbb{R}, \mathbb{R}^N)$, pero de otra forma. Un vector $y \in \mathbb{R}^N$ se identificó con la aplicación lineal $U_y \in L(\mathbb{R}, \mathbb{R}^N)$ dada por $U_y(t) = ty$ para todo $t \in \mathbb{R}$, usando el producto por escalares de \mathbb{R}^N . En cambio ahora, usamos el producto escalar de \mathbb{R}^N para identificar y con la aplicación lineal $T_y \in L(\mathbb{R}^N, \mathbb{R})$ dada por $T_y(x) = (y | x)$ para todo $x \in \mathbb{R}^N$.

Volviendo a un campo escalar $f : \Omega \rightarrow \mathbb{R}$, donde Ω es un abierto de \mathbb{R}^N , diferenciable en $a \in \Omega$, el resultado recién obtenido permite identificar la diferencial $Df(a)$ con el vector gradiente $\nabla f(a)$. Concretamente, la relación entre ambos es:

$$\nabla f(a) = \sum_{k=1}^N Df(a)(e_k) e_k \quad y \quad Df(a)(x) = (\nabla f(a) | x) \quad \forall x \in \mathbb{R}^N$$

8.4. Condición suficiente de diferenciabilidad

Frecuentemente, se puede probar la diferenciabilidad de un campo escalar, trabajando sólo con sus derivadas parciales. Ello nos permitirá obtener una caracterización muy cómoda de los campos escalares de clase C^1 . Ambos resultados son también útiles para campos vectoriales, sin más que aplicarlos a sus componentes, que son campos escalares.

Sea pues Ω un abierto de \mathbb{R}^N y supongamos que queremos probar la diferenciabilidad de un campo escalar $f : \Omega \rightarrow \mathbb{R}$ en un punto $a \in \Omega$. Sabemos que ello equivale a probar que f es parcialmente derivable en el punto a y que

$$\lim_{x \rightarrow a} \frac{f(x) - f(a) - (\nabla f(a) | x - a)}{\|x - a\|} = 0 \quad (8)$$

La idea es evitar el cálculo de este límite, a cambio de trabajar un poco más con las derivadas parciales.

En la práctica f suele ser parcialmente derivable, no sólo en a , sino en un entorno abierto de a . En tal caso, por el carácter local de la diferenciabilidad, podemos sustituir f por su restricción a dicho entorno, así que no perdemos generalidad suponiendo que f es parcialmente derivable en Ω . Entonces, en vez de comprobar la igualdad (8), puede ser más fácil comprobar la continuidad en el punto a de las derivadas parciales, es decir, comprobar que

$$\lim_{x \rightarrow a} \frac{\partial f}{\partial x_k}(x) = \frac{\partial f}{\partial x_k}(a) \quad \forall k \in \Delta_N \quad (9)$$

Pues bien, vamos a probar que (8) es consecuencia de (9), y de hecho, que una condición ligeramente más débil que (9) ya es suficiente para deducir (8). Concretamente, fijado $k \in \Delta_N$, podemos suponer solamente la existencia de la derivada parcial de f con respecto a la k -ésima variable en el punto a , mientras que las demás derivadas parciales sí se suponen definidas en todo punto de Ω y continuas en a . Esta hipótesis, obviamente más débil que (9), es ya suficiente para la diferenciabilidad de f en a .

- Sea Ω un abierto de \mathbb{R}^N , $f : \Omega \rightarrow \mathbb{R}$ un campo escalar, $a \in \Omega$ y $k \in \Delta_N$. Supongamos que se verifican las dos condiciones siguientes:

- (i) f es parcialmente derivable con respecto a la k -ésima variable en el punto a
- (ii) Para $j \in \Delta_N \setminus \{k\}$, f es parcialmente derivable con respecto a la j -ésima variable en todo punto $x \in \Omega$ y la función derivada parcial $\frac{\partial f}{\partial x_j} : \Omega \rightarrow \mathbb{R}$ es continua en a .

Entonces f es diferenciable en el punto a .

Salvo un obvio intercambio de variables, que no afecta a la diferenciabilidad de f , podemos suponer que $k = N$, es decir, que la derivada parcial para la que sólo se supone su existencia en el punto a , es la última. Tomamos $r \in \mathbb{R}^+$ tal que $B(a, r) \subset \Omega$, y se trata de probar (8). Para que se comprenda mejor el razonamiento, conviene destacar una igualdad evidente.

Si escribimos $a = (a_1, a_2, \dots, a_N)$, para $x = (x_1, x_2, \dots, x_N) \in B(a, r)$ se tiene:

$$\begin{aligned} f(x) - f(a) &= (f(x) - f(a_1, x_2, \dots, x_N)) \\ &\quad + (f(a_1, x_2, \dots, x_N) - f(a_1, a_2, x_3, \dots, x_N)) \\ &\quad + \dots + (f(a_1, \dots, a_{N-1}, x_N) - f(a)) \\ &= \sum_{j=1}^N (f(a_1, \dots, a_{j-1}, x_j, \dots, x_N) - f(a_1, \dots, a_j, x_{j+1}, \dots, x_N)) \end{aligned}$$

En la anterior igualdad se debe entender que $(a_1, \dots, a_{j-1}, x_j, \dots, x_N) = x$ para $j = 1$, así como que $(a_1, \dots, a_j, x_{j+1}, \dots, x_N) = a$ para $j = N$. Deducimos claramente que

$$f(x) - f(a) - (\nabla f(a) | x - a) = \sum_{j=1}^N R_j(x) \quad \forall x \in B(a, r) \quad (10)$$

donde, para cada $j \in \Delta_N$ hemos escrito:

$$R_j(x) = f(a_1, \dots, a_{j-1}, x_j, \dots, x_N) - f(a_1, \dots, a_j, x_{j+1}, \dots, x_N) - (x_j - a_j) \frac{\partial f}{\partial x_j}(a) \quad (11)$$

Trabajaremos ahora por separado con cada uno de los sumandos que han aparecido.

Dado $\varepsilon > 0$, de (i) obtenemos $\delta \in]0, r[$ tal que, para $x_N \in \mathbb{R}$ con $|x_N - a_N| < \delta$, se tiene

$$\left| f(a_1, \dots, a_{N-1}, x_N) - f(a) - (x_N - a_N) \frac{\partial f}{\partial x_N}(a) \right| \leq \frac{\varepsilon}{N} |x_N - a_N| \quad (12)$$

Por otra parte, la hipótesis (ii) nos permite conseguir que el mismo δ verifique además que, para todo $w \in B(x, \delta)$ se tenga

$$\left| \frac{\partial f}{\partial x_j}(w) - \frac{\partial f}{\partial x_j}(a) \right| \leq \frac{\varepsilon}{N} \quad \forall j \in \Delta_{N-1} \quad (13)$$

Para comprobar (8) y concluir así la demostración, fijamos $x = (x_1, x_2, \dots, x_N) \in B(x, \delta)$, y es importante tener en cuenta que, a partir de este momento, el vector x , y por tanto todas sus componentes, estarán fijos.

Por una parte tenemos que $|x_N - a_N| \leq \|x - a\| < \delta$ luego (12) nos dice que

$$|R_N(x)| \leq (\varepsilon/N) |x_N - a_N| \leq (\varepsilon/N) \|x - a\| \quad (14)$$

Por otra parte, fijamos $j \in \Delta_{N-1}$ para trabajar en el intervalo $J =]a_j - \delta, a_j + \delta[$ con la función $\psi : J \rightarrow \mathbb{R}$ definida por

$$\psi(t) = f(a_1, \dots, a_{j-1}, t, x_{j+1}, \dots, x_N) \quad \forall t \in J$$

Para cada $t \in J$, que f sea parcialmente derivable con respecto a la j -ésima variable en el punto $(a_1, \dots, a_{j-1}, t, x_{j+1}, \dots, x_N) \in \Omega$, significa que ψ es derivable en el punto t con

$$\psi'(t) = \frac{\partial f}{\partial x_j}(a_1, \dots, a_{j-1}, t, x_{j+1}, \dots, x_N)$$

Usando el teorema del valor medio, obtenemos $u \in \mathbb{R}$, con $|u - a_j| \leq |x_j - a_j|$, tal que

$$\Psi(x_j) - \Psi(a_j) = (x_j - a_j) \Psi'(u) = (x_j - a_j) \frac{\partial f}{\partial x_j}(a_1, \dots, a_{j-1}, u, x_{j+1}, \dots, x_N)$$

Tomando $w = (a_1, \dots, a_{j-1}, u, x_{j+1}, \dots, x_N)$, de la igualdad anterior deducimos que

$$R_j(x) = \Psi(x_j) - \Psi(a_j) - (x_j - a_j) \frac{\partial f}{\partial x_j}(a) = (x_j - a_j) \left(\frac{\partial f}{\partial x_j}(w) - \frac{\partial f}{\partial x_j}(a) \right)$$

Es claro que $\|w - a\| \leq \|x - a\| < \delta$, luego podemos usar (13) para obtener que

$$|R_j(x)| \leq (\varepsilon/N) |x_j - a_j| \leq (\varepsilon/N) \|x - a\| \quad \forall j \in \Delta_{N-1} \quad (15)$$

Teniendo en cuenta (10), de (14) y (15), deducimos finalmente que

$$|f(x) - f(a) - (\nabla f(a) | (x - a))| \leq \sum_{j=1}^N |R_j(x_j, x_{j+1}, \dots, x_N)| \leq \varepsilon \|x - a\|$$

Esto prueba que se verifica (8), así que f es diferenciable en el punto a . ■

El resultado anterior nos permitirá obtener una útil caracterización de los campos escalares de clase C^1 en términos de sus derivadas parciales.

Si Ω es de nuevo un abierto de \mathbb{R}^N y $f \in D(\Omega)$, con vistas a estudiar la continuidad de la función diferencial $Df : \Omega \rightarrow L(\mathbb{R}^N, \mathbb{R})$, para cada $x \in \Omega$, podemos identificar $Df(x)$ con $\nabla f(x)$, y considerar la **función gradiente** $\nabla f : \mathbb{R}^N \rightarrow \mathbb{R}^N$, que a cada punto $x \in \Omega$ hace corresponder el gradiente de f en x , que es un campo vectorial. Entonces Df será continua en un punto $a \in \Omega$ si, y sólo si, lo es ∇f , pues sabemos que

$$\|Df(x) - Df(a)\| = \|\nabla f(x) - \nabla f(a)\| \quad \forall x, a \in \Omega$$

Enunciamos ya la caracterización que buscamos:

- Si Ω es un abierto de \mathbb{R}^N y $f : \Omega \rightarrow \mathbb{R}$ un campo escalar, las siguientes afirmaciones son equivalentes:

- (i) $f \in C^1(\Omega)$
- (ii) f es parcialmente derivable en todo punto de Ω y su función gradiente es continua, es decir, $\nabla f \in \mathcal{C}(\Omega, \mathbb{R}^N)$.
- (iii) f es parcialmente derivable en todo punto de Ω y todas sus derivadas parciales son funciones continuas, es decir, $\partial f / \partial x_k \in \mathcal{C}(\Omega)$ para todo $k \in \Delta_N$.

La equivalencia entre (ii) y (iii) es evidente, pues las derivadas parciales son las componentes de la función gradiente. La afirmación (i) implica obviamente que f es diferenciable y del resultado anterior se deduce que (ii) también lo implica. Una vez que $f \in D(\Omega)$, la equivalencia entre (i) y (ii) ya se ha comentado. ■

8.5. Interpretación física del gradiente

El uso en Física del gradiente de un campo escalar se explica fácilmente a partir de una interpretación de las derivadas direccionales, que pasamos a explicar. Trabajamos por tanto con un abierto $\Omega \subset \mathbb{R}^N$ y un campo escalar $f : \Omega \rightarrow \mathbb{R}$, direccionalmente derivable en $a \in \Omega$.

Fijada una dirección $u \in S$, al desplazarnos una distancia $t > 0$ desde el punto a , en la dirección y el sentido del vector u , el campo f experimenta una variación de $f(a+tu) - f(a)$ unidades. Podemos decir por tanto que la derivada direccional $f'_u(a)$ es la *tasa de variación* del campo en el punto a y en la dirección u . En un “pequeño” desplazamiento desde a , en la dirección y sentido del vector u , el campo varía “aproximadamente” a razón de $f'_u(a)$ unidades de campo por unidad de longitud.

Si f es diferenciable en $a \in \Omega$, se tiene $f'_u(a) = (\nabla f(a) | u)$ para todo $u \in S$. Supongamos que $\nabla f(a) \neq 0$ y consideremos la dirección $v = \nabla f(a) / \|\nabla f(a)\|$. Para toda dirección $u \in S$, la desigualdad de Cauchy-Schwartz nos dice que

$$f'_u(a) = (\nabla f(a) | u) \leq \|\nabla f(a)\| = (\nabla f(a) | v) = f'_v(a)$$

de donde deducimos que

$$f'_v(a) = \max \{ f'_u(a) : u \in S \} > 0$$

Además, sabemos que la desigualdad de Cauchy-Schwartz sólo es una igualdad cuando en ella aparece el producto escalar de dos vectores linealmente dependientes. Por tanto, si $u \in S$ verifica que $f'_u(a) = f'_v(a)$, se tendrá $u = \lambda v$ con $\lambda \in \mathbb{R}$, pero entonces $f'_u(a) = \lambda f'_v(a)$ y deducimos que $\lambda = 1$, es decir, $u = v$.

Tenemos así una caracterización del vector gradiente normalizado: $v = \nabla f(a) / \|\nabla f(a)\|$ es la única dirección $v \in S$ que hace que la derivada direccional $f'_v(a)$ sea máxima. Obsérvese que esta caracterización del vector gradiente es independiente de la base que estemos usando en \mathbb{R}^N . Si cambiamos de base en \mathbb{R}^N , las derivadas direccionales no cambian, luego el vector gradiente seguirá siendo el mismo, aunque obviamente cambien sus coordenadas.

El significado físico de esta caracterización es claro, y se facilita si tenemos en cuenta que el vector $v = \nabla f(a) / \|\nabla f(a)\|$ tiene la misma dirección y sentido que $\nabla f(a)$. Por tanto, al desplazarnos desde el punto a en la dirección y el sentido del vector gradiente, conseguimos la máxima tasa de aumento del campo por unidad de longitud, o si se quiere, hacemos que el campo aumente lo más rápidamente posible. Son además la única dirección y el único sentido que producen dicha máxima tasa de aumento, que viene dada por la norma euclídea del vector gradiente, $f'_v(a) = \|\nabla f(a)\|$. Dicho más intuitivamente, un “pequeño” desplazamiento en la dirección y sentido del vector gradiente, hace que el campo aumente “aproximadamente” a razón de $\|\nabla f(a)\|$ unidades, de las que usamos para medirlo, por unidad de longitud.

Por otra parte, como $f'_{-v}(a) = -f'_v(a)$, al desplazarnos por la misma recta, pero en el sentido opuesto, el del vector $-\nabla f(a)$, conseguimos que el valor del campo disminuya lo más rápidamente posible, “aproximadamente” a razón de $\|\nabla f(a)\|$ unidades de campo por unidad de longitud.

Resaltamos que lo dicho es válido cuando f es diferenciable en a , no basta con que f sea parcialmente derivable en a , suponiendo además que $\nabla f(a) \neq 0$. Cuando $\nabla f(a) = 0$, todas las derivadas direccionales de f en a se anulan, luego el campo varía muy lentamente cerca del punto a , su tasa de variación es “prácticamente” nula en todas las direcciones. Se dice entonces que a es un *punto crítico* o un *punto estacionario* del campo f .

8.6. Plano tangente a una superficie explícita

Recordemos que una curva explícita en \mathbb{R}^2 es la gráfica de una función continua, definida en un intervalo y con valores reales, el tipo más sencillo de curva paramétrica. Análogamente, consideramos ahora el tipo más sencillo de superficie en \mathbb{R}^3 .

Llamamos **superficie explícita** en \mathbb{R}^3 a la gráfica de una función continua $f : \Omega \rightarrow \mathbb{R}$, donde Ω es un subconjunto no vacío, abierto y conexo, de \mathbb{R}^2 . Se trata por tanto del conjunto

$$\Sigma = \text{Gr } f = \{ (x, y, f(x, y)) : (x, y) \in \Omega \} \subset \mathbb{R}^3$$

Nótese que Σ determina de manera única a f , pues para $(x, y) \in \mathbb{R}^2$, es claro que $(x, y) \in \Omega$ si, y sólo si, existe $z \in \mathbb{R}$ tal que $(x, y, z) \in \Sigma$, en cuyo caso $f(x, y)$ es el único $z \in \mathbb{R}$ que verifica dicha condición. Se dice que la igualdad

$$z = f(x, y) \quad ((x, y) \in \Omega)$$

es la **ecuación explícita** de la superficie Σ .

Geoméricamente, es también natural considerar superficies explícitas de otros dos tipos, que se obtienen intercambiando los ejes de coordenadas. Concretamente, la misma función f también da lugar a las superficies explícitas

$$\Sigma_1 = \{ (f(y, z), y, z) : (y, z) \in \Omega \} \quad \text{y} \quad \Sigma_2 = \{ (x, f(x, z), z) : (x, z) \in \Omega \}$$

cuyas ecuaciones explícitas serán $x = f(y, z)$ con $(y, z) \in \Omega$ e $y = f(x, z)$ con $(x, z) \in \Omega$, respectivamente. Trabajaremos solamente con una superficie Σ del primer tipo, la gráfica de f , pues toda la discusión que haremos se traduce fácilmente para aplicarla a los otros dos tipos.

Cuando f es diferenciable en un punto $(x_0, y_0) \in \Omega$, veamos la relación entre el vector gradiente $\nabla f(x_0, y_0)$ y la superficie Σ . Para abreviar, escribimos

$$z_0 = f(x_0, y_0), \quad \alpha_0 = \frac{\partial f}{\partial x}(x_0, y_0) \quad \text{y} \quad \beta_0 = \frac{\partial f}{\partial y}(x_0, y_0)$$

y consideramos el plano afín Π definido por la ecuación

$$z - z_0 = \alpha_0(x - x_0) + \beta_0(y - y_0)$$

que también es una superficie explícita, concretamente $\Pi = \text{Gr } g$ donde $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ es la función definida por

$$g(x, y) = z_0 + \alpha_0(x - x_0) + \beta_0(y - y_0) \quad \forall (x, y) \in \mathbb{R}^2 \quad (16)$$

Usando la relación entre la diferencial de f y su gradiente, vemos que g es una función afín que ya hemos manejado anteriormente:

$$g(x, y) = f(x_0, y_0) + Df(x_0, y_0)((x, y) - (x_0, y_0)) \quad \forall (x, y) \in \mathbb{R}^2$$

Al comentar el significado analítico de la diferencial, vimos que g es la función afín que nos da una buena aproximación de f cerca del punto (x_0, y_0) , es decir,

$$\lim_{(x, y) \rightarrow (x_0, y_0)} \frac{|f(x, y) - g(x, y)|}{\|(x, y) - (x_0, y_0)\|} = 0 \quad (17)$$

Geométricamente, esto significa que la distancia (vertical) entre el punto $(x, y, f(x, y))$ de la superficie Σ y el correspondiente punto $(x, y, g(x, y))$ del plano Π , tiende a cero cuando ambos puntos se acercan a (x_0, y_0, z_0) , “mucho más rápidamente” que $\|(x, y) - (x_0, y_0)\|$, así que la siguiente definición está plenamente justificada.

Sea $\Sigma = \text{Gr } f$ una superficie explícita en \mathbb{R}^3 , donde $f : \Omega \rightarrow \mathbb{R}$ es una función continua en un abierto conexo $\Omega \subset \mathbb{R}^2$. Si f es diferenciable en un punto (x_0, y_0) y $z_0 = f(x_0, y_0)$, se dice que el plano Π de ecuación explícita

$$z - z_0 = (x - x_0) \frac{\partial f}{\partial x}(x_0, y_0) + (y - y_0) \frac{\partial f}{\partial y}(x_0, y_0) \quad (18)$$

es el **plano tangente** a la superficie Σ en el punto (x_0, y_0, z_0) .

Decimos también que el vector

$$\left(\frac{\partial f}{\partial x}(x_0, y_0), \frac{\partial f}{\partial y}(x_0, y_0), -1 \right) \in \mathbb{R}^3 \quad (19)$$

es un **vector normal** a la superficie Σ en el punto (x_0, y_0, z_0) . Ello se explica porque, en vista de (18), se trata de un vector normal al plano tangente Π , es claramente ortogonal a todos los vectores de la forma $v - u$ con $u, v \in \Pi$ es decir, a todas las rectas contenidas en Π . Tenemos así la interpretación geométrica de la diferencial $Df(x_0, y_0)$ o del vector gradiente $\nabla f(x_0, y_0)$.

Obsérvese que, para definir el vector que aparece en (19), normal al plano dado por (18), o lo que es lo mismo, considerar la función afín g dada por (16), bastaría disponer del gradiente de f en el punto (x_0, y_0) , es decir, que f fuese sólo parcialmente derivable en (x_0, y_0) . Sin embargo, el plano así obtenido sólo es una buena aproximación de la superficie Σ cerca del punto (x_0, y_0, z_0) cuando se cumple (17), que equivale a la diferenciabilidad de f en (x_0, y_0) .

Esta es la razón por la que sólo hemos definido el plano tangente y el vector normal a una superficie $\Sigma = \text{Gr } f$ en el punto $(x_0, y_0, z_0) \in \Sigma$, cuando f es diferenciable en (x_0, y_0) . Así pues, la noción geoméricamente relevante, la que justifica la definición de plano tangente y vector normal, es la diferenciabilidad, no basta la derivabilidad parcial.

Más adelante, la regla de la cadena para las derivadas parciales, nos permitirá reafirmar aún más la interpretación geométrica que acabamos de hacer.

8.7. Extremos absolutos y relativos

Recordemos ciertas nociones elementales, que conocemos bien en el caso de una función real de variable real, pero tienen perfecto sentido en contextos mucho más generales.

Si A es un conjunto no vacío arbitrario y $f : A \rightarrow \mathbb{R}$ una función, decimos que f tiene o alcanza un **máximo absoluto** en un punto $a \in A$, cuando $f(a) = \max f(A)$, es decir,

$$f(a) \geq f(x) \quad \forall x \in A$$

Obviamente, no debemos confundir el valor máximo de f , que es $f(a)$, con el punto a donde se alcanza dicho valor. Análogamente, decimos que f tiene un **mínimo absoluto** en $a \in A$ cuando $f(a) = \min f(A)$ es decir,

$$f(a) \leq f(x) \quad \forall x \in A$$

Obviamente, f tiene un mínimo absoluto en a si, y sólo si, la función $-f$ tiene un máximo absoluto en a , de modo que ambas nociones se reducen una a la otra. Decimos que f tiene un **extremo absoluto** en un punto $a \in A$ cuando tiene un máximo absoluto en a o un mínimo absoluto en a .

El problema de averiguar si f tiene un máximo o un mínimo absoluto en algún punto de A , y en su caso encontrarlos, es lo que, en términos muy generales, se conoce como un *problema de optimización* o un problema de extremos. Claramente, sin hipótesis sobre el conjunto A y la función f , este problema es intratable.

Conocemos una respuesta a este problema en un contexto muy general. Concretamente, sabemos que si A es un espacio métrico compacto y f es continua, entonces f tiene un máximo absoluto y un mínimo absoluto en sendos puntos de A , aunque nada sabemos sobre cómo encontrarlos. En casos más concretos, suele ser útil la siguiente noción de extremo relativo.

Supongamos que A es un subconjunto de un espacio métrico E . Decimos que $f : A \rightarrow \mathbb{R}$ tiene un **máximo relativo** en un punto $a \in A$, cuando existe $r \in \mathbb{R}^+$ verificando que

$$B(a, r) \subset A \quad \text{y} \quad f(a) \leq f(x) \quad \forall x \in B(a, r)$$

Obsérvese que en particular se ha de tener $a \in A^\circ$.

Análogamente, f tiene un **mínimo relativo** en $a \in A$, cuando existe $r \in \mathbb{R}^+$ tal que

$$B(a, r) \subset A \quad \text{y} \quad f(x) \leq f(a) \quad \forall x \in B(a, r)$$

De nuevo es claro que f tiene un mínimo relativo en a si, y sólo si, $-f$ tiene un máximo relativo en a . Finalmente, decimos que f tiene un **extremo relativo** en un punto $a \in A$, cuando tiene un máximo relativo o un mínimo relativo en a .

Observemos que las nociones de extremo absoluto y relativo no están relacionadas. Ciertamente, cuando f tiene un extremo relativo en a , la restricción de f a una bola $B(a, r) \subset A$ tiene un máximo absoluto en a , pero fuera de dicha bola, f puede tomar valores estrictamente mayores y menores que $f(a)$, en cuyo caso f no tendrá un extremo absoluto en a .

En sentido opuesto, es claro que si f tiene en a un extremo absoluto, entonces tendrá un extremo relativo si, y sólo si, $a \in A^\circ$. Por tanto, salvo que A sea abierto, f puede perfectamente tener un extremo absoluto en un punto $a \in A \setminus A^\circ$, que nunca puede ser un extremo relativo.

Pues bien, a la hora de encontrar los puntos en los que un campo escalar puede tener un extremo relativo, entra en juego el cálculo diferencial, igual que ocurría con funciones reales de variable real. Supongamos pues que A es un subconjunto no vacío de \mathbb{R}^N y que $f : A \rightarrow \mathbb{R}$ es un campo escalar diferenciable en un punto $a \in A^\circ$. Tanto el punto de vista físico como la intuición geométrica sugieren que, si f tiene un extremo relativo en a , se debe tener $\nabla f(a) = 0$. En efecto, si fuese $\nabla f(a) \neq 0$, el campo aumentaría en la dirección y el sentido del vector $\nabla f(a)$ y disminuiría en sentido opuesto, luego f no puede tener un extremo relativo en a . Igualmente, en el caso $N = 2$, cabe esperar que, si f tiene un extremo relativo en el punto $a = (a_1, a_2)$, el plano tangente a la superficie $\text{Gr } f$ en el punto $(a_1, a_2, f(a_1, a_2))$ tenga que ser horizontal, es decir, $\nabla f(a_1, a_2) = 0$.

Comprobamos fácilmente que las ideas anteriores son correctas, e incluso, esta vez no es necesario que f sea diferenciable en el punto a , basta la derivabilidad parcial:

- **Condición necesaria de extremo relativo.** Sea A un subconjunto no vacío de \mathbb{R}^N , y sea $f : A \rightarrow \mathbb{R}$ un campo escalar. Si f tiene un extremo relativo en un punto $a \in A^\circ$ y es parcialmente derivable en dicho punto, entonces $\nabla f(a) = 0$.

Supongamos que f tiene en a un máximo relativo, pues en el otro caso, sustituimos f por $-f$. Sea pues $r \in \mathbb{R}^+$ tal que $B(a, r) \subset A$ y $f(x) \leq f(a)$ para todo $x \in B(a, r)$. Fijado $k \in \Delta_N$, para todo $t \in]0, r[$ se tiene entonces

$$\frac{f(a + te_k) - f(a)}{t} \leq 0 \quad \text{luego} \quad \frac{\partial f}{\partial x_k}(a) = \lim_{t \rightarrow 0^+} \frac{f(a + te_k) - f(a)}{t} \leq 0$$

Pero análogamente, para $-r < t < 0$ se tiene también

$$\frac{f(a + te_k) - f(a)}{t} \geq 0 \quad \text{luego} \quad \frac{\partial f}{\partial x_k}(a) = \lim_{t \rightarrow 0^-} \frac{f(a + te_k) - f(a)}{t} \geq 0$$

Esto prueba que $\frac{\partial f}{\partial x_k}(a) = 0$ para todo $k \in \Delta_N$, es decir, $\nabla f(a) = 0$. ■

Como se hizo en la interpretación física del gradiente, cuando un campo escalar $f : A \rightarrow \mathbb{R}$ es parcialmente derivable en un punto $a \in A^\circ$, y verifica que $\nabla f(a) = 0$, se dice que a es un **punto crítico** de f . Así pues, los puntos en los que f es parcialmente derivable y tiene un extremo relativo, han de ser puntos críticos. Esto ayuda con frecuencia a encontrarlos.

8.8. Ejercicios

1. Calcular todas las derivadas direccionales en el punto $(-1, 0, 0)$ de la función $f: \mathbb{R}^3 \rightarrow \mathbb{R}$ definida por

$$f(x, y, z) = x^3 - 3xy + z^3 \quad \forall (x, y, z) \in \mathbb{R}^3$$

2. Sea J un intervalo abierto en \mathbb{R} y Ω un subconjunto abierto de \mathbb{R}^N . Si $f: J \rightarrow \Omega$ es una función derivable en un punto $a \in J$ y $g: \Omega \rightarrow \mathbb{R}$ es diferenciable en el punto $b = f(a)$, probar que la función $h = g \circ f: J \rightarrow \mathbb{R}$ es derivable en el punto a con

$$h'(a) = (\nabla g(b) | f'(a))$$

3. Sea Ω un subconjunto abierto de \mathbb{R}^N y $f, g: \Omega \rightarrow \mathbb{R}$ funciones parcialmente derivables en un punto $a \in \Omega$. Probar que las funciones $f + g$ y fg son parcialmente derivables en a con

$$\nabla(f + g)(a) = \nabla f(a) + \nabla g(a) \quad y$$

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

Suponiendo que $g(x) \neq 0$ para todo $x \in \Omega$, probar también que f/g es parcialmente derivable en a con

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

4. Fijado $p \in \mathbb{R}^*$, se considera la función $f: \mathbb{R}^N \setminus \{0\} \rightarrow \mathbb{R}$ definida por $f(x) = \|x\|^p$ para todo $x \in \mathbb{R}^N \setminus \{0\}$, donde $\|\cdot\|$ es la norma euclídea. Probar que $f \in C^1(\mathbb{R}^N \setminus \{0\})$ con

$$\nabla f(x) = p\|x\|^{p-2}x \quad \forall x \in \mathbb{R}^N \setminus \{0\}$$

Como consecuencia, encontrar una función $g \in C^1(\mathbb{R}^N, \mathbb{R}^N)$ que verifique

$$\nabla g(x) = x \quad \forall x \in \mathbb{R}^N$$

5. Calcular la ecuación del plano tangente a la superficie explícita de ecuación $z = x + y^3$ con $(x, y) \in \mathbb{R}^2$, en el punto $(1, 1, 2)$.
6. Sea Ω un subconjunto abierto y conexo de \mathbb{R}^2 y $f: \Omega \rightarrow \mathbb{R}$ una función diferenciable. Se considera la superficie explícita $S \subset \mathbb{R}^3$ dada por

$$S = \{(x, f(x, z), z) : (x, z) \in \Omega\}$$

Calcular la ecuación del plano tangente a S en un punto arbitrario $(x_0, y_0, z_0) \in S$.

7. Sea Ω un subconjunto abierto de \mathbb{R}^2 y $f: \Omega \rightarrow \mathbb{R}^2$ una función parcialmente derivable en todo punto de Ω . Probar que, si la función $\nabla f: \Omega \rightarrow \mathbb{R}^2$ está acotada, entonces f es continua. Usando la función $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ definida por

$$f(x, y) = \frac{xy}{\sqrt{x^2 + y^2}} \quad \forall (x, y) \in \mathbb{R}^2 \setminus \{(0, 0)\} \quad y \quad f(0, 0) = 0$$

comprobar que, con las mismas hipótesis, no se puede asegurar que f sea diferenciable.

Matriz jacobiana

Como último caso particular de la noción de diferenciabilidad, suponemos ahora que el espacio normado de partida es \mathbb{R}^N con $N > 1$, y el de llegada es \mathbb{R}^M , también con $M > 1$. Estudiamos por tanto la diferenciabilidad de una función definida en un abierto de \mathbb{R}^N y con valores en \mathbb{R}^M , es decir, de un campo vectorial. Dependiendo de los valores de N y M tenemos campos vectoriales muy variopintos, siendo lógicamente $N = M = 2$ y $N = M = 3$, los casos más interesantes. Usando las bases usuales de \mathbb{R}^N y \mathbb{R}^M , la diferencial de nuestra función en un punto, cuando existe, viene representada por una matriz $M \times N$ con coeficientes reales, que será la *matriz jacobiana*. Sus filas son los gradientes de M campos escalares en \mathbb{R}^N , las M componentes de nuestro campo vectorial.

A la composición de aplicaciones lineales corresponde entonces el producto de matrices, con lo que obtenemos una *regla de la cadena para las derivadas parciales*. Como aplicación geométrica que merece destacarse, cuando $N = 2$ y $M = 3$ estudiamos el plano tangente a una *superficie paramétrica* en \mathbb{R}^3 , generalizando lo que ya sabemos para superficies explícitas.

9.1. Matriz de una aplicación lineal

Toda aplicación lineal $T \in L(\mathbb{R}^N, \mathbb{R}^M)$ admite una expresión matricial que vamos a recordar. Denotamos por $\mathcal{M}_{M \times N}$ al conjunto de todas las matrices $M \times N$ (M filas y N columnas) con coeficientes reales, cuya estructura algebraica suponemos bien conocida. Para cada $x \in \mathbb{R}^N$, sus coordenadas en la base usual de \mathbb{R}^N forman una matriz columna que también denotamos por x , por lo que escribimos $x \in \mathcal{M}_{N \times 1}$. Análogamente, el vector $y = T(x) \in \mathbb{R}^M$ tiene sus coordenadas en la base usual de \mathbb{R}^M , que dan la matriz columna $y \in \mathcal{M}_{M \times 1}$. Existe entonces una única matriz $A \in \mathcal{M}_{M \times N}$ tal que, para todo $x \in \mathbb{R}^N$, obtenemos $y = T(x)$ como producto de matrices: $y = A \cdot x$.

De hecho, si escribimos $A = (\alpha_{jk})$ para indicar los coeficientes de la matriz A , se tiene

$$\alpha_{jk} = (\pi_j \circ T)(e_k) \quad \forall j \in \Delta_M, \quad \forall k \in \Delta_N \quad (1)$$

siendo $\{e_1, \dots, e_N\}$ la base usual de \mathbb{R}^N y $\{\pi_1, \dots, \pi_M\}$ las proyecciones coordenadas en \mathbb{R}^M .

Diremos que A es la **matriz de la aplicación lineal** T , pero debemos entender su unicidad: es la única matriz que representa a T en la forma $y = A \cdot x$ cuando, tanto para $x \in \mathbb{R}^N$ como para $y = T(x) \in \mathbb{R}^M$, usamos sus coordenadas en las bases usuales, escritas en forma de matrices columna. Tenemos de hecho un isomorfismo entre los espacios vectoriales $L(\mathbb{R}^N, \mathbb{R}^M)$ y $\mathcal{M}_{M \times N}$, que identifica cada $T \in L(\mathbb{R}^N, \mathbb{R}^M)$ con su matriz $A \in \mathcal{M}_{M \times N}$, recién definida.

9.2. Matriz jacobiana

En lo que sigue fijamos un abierto Ω de \mathbb{R}^N y una función $f : \Omega \rightarrow \mathbb{R}^M$. Cuando $M = 1$ tenemos un campo escalar en \mathbb{R}^N , mientras que cuando $N = 1$ se trata de una función de variable real con valores en \mathbb{R}^M , los casos ya estudiados. Lo que hagamos es válido en ambos casos, pero nos interesa lo que ocurre para $N > 1$ y $M > 1$. Escribimos $f = (f_1, f_2, \dots, f_M)$ para indicar las M componentes de f , campos escalares definidos en Ω . Recordamos que, para todo $j \in \Delta_M$ se tiene $f_j = \pi_j \circ f$.

Si f es diferenciable en un punto $a \in \Omega$, la matriz de la aplicación lineal $Df(a) \in L(\mathbb{R}^N, \mathbb{R}^M)$ recibe el nombre de **matriz jacobiana** de f en a y se denota por $Jf(a)$. Por tanto, la matriz jacobiana representa a la diferencial $Df(a)$ en el siguiente sentido: para todo $x \in \mathbb{R}^N = \mathcal{M}_{N \times 1}$, el vector $y = Df(a)(x) \in \mathbb{R}^M = \mathcal{M}_{M \times 1}$ se obtiene como producto de matrices: $y = Jf(a) \cdot x$.

De (1) deducimos que los coeficientes de la matriz jacobiana, $Jf(a) = (\alpha_{jk}) \in \mathcal{M}_{M \times N}$, vienen dados por $\alpha_{jk} = (\pi_j \circ Df(a))(e_k)$, para cualesquiera $j \in \Delta_M$ y $k \in \Delta_N$. Fijado $j \in \Delta_M$, sabemos que $f_j = \pi_j \circ f$ es diferenciable en a con $Df_j(a) = \pi_j \circ Df(a)$. Pero además, para cada $k \in \Delta_N$, $Df_j(a)(e_k)$ es la derivada parcial de f_j con respecto a la k -ésima variable en el punto a , luego

$$\alpha_{jk} = Df_j(a)(e_k) = \frac{\partial f_j}{\partial x_k}(a) \quad \forall j \in \Delta_M, \quad \forall k \in \Delta_N$$

Así pues, la matriz jacobiana de f en a viene dada por:

$$Jf(a) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \frac{\partial f_1}{\partial x_2}(a) & \dots & \dots & \frac{\partial f_1}{\partial x_N}(a) \\ \frac{\partial f_2}{\partial x_1}(a) & \frac{\partial f_2}{\partial x_2}(a) & \dots & \dots & \frac{\partial f_2}{\partial x_N}(a) \\ \vdots & \vdots & & & \vdots \\ \frac{\partial f_M}{\partial x_1}(a) & \frac{\partial f_M}{\partial x_2}(a) & \dots & \dots & \frac{\partial f_M}{\partial x_N}(a) \end{pmatrix}$$

Se recuerda fácilmente, pues para cada $j \in \Delta_M$, su j -ésima fila es el gradiente de f_j en a , es decir, las M filas de la matriz jacobiana son los gradientes de las M componentes de f . Por otra parte, para cada $k \in \Delta_N$, la k -ésima columna de la matriz jacobiana es el vector derivada parcial de f con respecto a la k -ésima variable en el punto a , luego las N columnas de la matriz jacobiana son las N derivadas parciales de f en a .

Cuando $M = 1$, tenemos un campo escalar, cuya matriz jacobiana es su gradiente, escrito como una matriz fila, $Jf(a) = \nabla f(a) \in \mathcal{M}_{1 \times N}$, de forma que, para todo $x \in \mathbb{R}^N$, el producto escalar $(\nabla f(a) | x)$ coincide con el producto de matrices $\nabla f(a) \cdot x = Jf(a) \cdot x$. Por otra parte, cuando $N = 1$, tenemos una función de una sola variable real, cuya matriz jacobiana es su vector derivada, escrito como una matriz columna, $Jf(a) = f'(a) \in \mathcal{M}_{M \times 1}$, de forma que, para todo $x \in \mathbb{R}$, el producto del vector $f'(a)$ por el escalar x coincide con el producto de matrices $f'(a) \cdot x = Jf(a) \cdot x$.

9.3. Regla de la cadena para las derivadas parciales

La noción de matriz jacobiana, junto con la regla general de la cadena, nos van a permitir calcular las derivadas parciales de las componentes de una composición de campos vectoriales diferenciables.

Mantenemos pues nuestra función $f : \Omega \rightarrow \mathbb{R}^M$, donde Ω es un abierto de \mathbb{R}^N , pero ahora consideramos un abierto U de \mathbb{R}^M tal que $f(\Omega) \subset U$ y otra función $g : U \rightarrow \mathbb{R}^P$ donde $P \in \mathbb{N}$ es arbitrario. Esto permite definir la composición $h = g \circ f : \Omega \rightarrow \mathbb{R}^P$. Suponemos que f es diferenciable en un punto $a \in \Omega$ y que g es diferenciable en el punto $b = f(a) \in U$, con lo que la regla de la cadena nos dice que h es diferenciable en a con

$$Dh(a) = Dg(b) \circ Df(a)$$

Veamos cómo se traduce esta igualdad en términos de las matrices jacobianas. Para $x \in \mathbb{R}^N$, sean $y = Df(a)(x) \in \mathbb{R}^M$ y $z = Dg(b)(y) \in \mathbb{R}^P$, con lo que $z = Dh(a)(x)$. Escribiendo los tres vectores x, y, z como matrices columna, tenemos

$$Jh(a) \cdot x = z = Jg(b) \cdot y = Jg(b) \cdot (Jf(a) \cdot x) = (Jg(b) \cdot Jf(a)) \cdot x \quad \forall x \in \mathbb{R}^N$$

donde hemos usado la asociatividad del producto de matrices. Deducimos claramente que

$$Jh(a) = Jg(b) \cdot Jf(a) \quad (2)$$

Esta igualdad matricial nos dará una regla práctica para calcular derivadas parciales, tan pronto como especifiquemos los coeficientes de las tres matrices que en ella aparecen. Para $Jf(a)$ ya lo hemos hecho antes:

$$Jf(a) = (\alpha_{jk}) = \left(\frac{\partial f_j}{\partial x_k}(a) \right) \in \mathcal{M}_{M \times N}$$

La función g tiene P componentes g_i con $i \in \Delta_P$, que son funciones de M variables reales. A efectos de escribir las derivadas parciales de cada componente, la j -ésima de esas variables debe denotarse lógicamente por y_j para todo $j \in \Delta_M$. De esta forma, para $i \in \Delta_P$ y $j \in \Delta_M$, la derivada parcial de g_i con respecto a la j -ésima variable, en el punto b , es $\frac{\partial g_i}{\partial y_j}(b)$, y ya podemos escribir explícitamente la matriz jacobiana:

$$Jg(b) = (\beta_{ij}) = \left(\frac{\partial g_i}{\partial y_j}(b) \right) \in \mathcal{M}_{P \times M}$$

Finalmente h tiene también sus P componentes $h_i = g_i \circ f$ con $i \in \Delta_P$, que son funciones de N variables reales, las mismas de las que depende f , esto es, x_k con $k \in \Delta_N$. Por tanto:

$$Jh(a) = (\lambda_{ik}) = \left(\frac{\partial h_i}{\partial x_k}(a) \right) \in \mathcal{M}_{P \times N}$$

En vista de (2) la definición del producto de matrices nos dice que, para cualesquiera $i \in \Delta_P$ y $k \in \Delta_N$, se tiene $\lambda_{ik} = \sum_{j=1}^M \beta_{ij} \alpha_{jk}$, es decir,

$$\frac{\partial h_i}{\partial x_k}(a) = \sum_{j=1}^M \frac{\partial g_i}{\partial y_j}(b) \frac{\partial f_j}{\partial x_k}(a) \quad \forall k \in \Delta_N, \quad \forall i \in \Delta_P \quad (3)$$

Podemos así calcular las derivadas parciales de todas las componentes de h , una vez que hayamos hecho lo mismo con f y g . También podemos usar (3) como sistema de ecuaciones, del que obtener ciertas derivadas parciales a partir de otras que conozcamos. Enseguida veremos ejemplos de aplicación de esta regla práctica para el cálculo de derivadas parciales.

9.4. Cambio de variables

La igualdad (3) será más fácil de entender, y de recordar, con una notación más intuitiva, la que se usa en la práctica para manejar diversos ejemplos de cambio de variables. Mantenemos las hipótesis que nos han permitido obtener (3), es decir, f es diferenciable en $a \in \Omega$ y g es diferenciable en $b = f(a) \in U$.

Pensemos que f , g y h describen la relación entre tres variables, $x \in \Omega$, $y \in U$ y $z \in \mathbb{R}^P$, mediante las tres igualdades

$$y = f(x), \quad z = g(y), \quad z = h(x)$$

siendo la tercera consecuencia de las dos primeras. Nótese el doble papel de la variable y , que depende de x mediante la función f , pero es también la variable de la que depende z mediante la función g . Como consecuencia z también depende de x mediante la función h . Nada de particular, es lo que ocurre siempre que consideramos una composición de funciones, entendidas como dependencia entre variables. La función $z = h(x)$ para $x \in \Omega$, se ha obtenido a partir de la función $z = g(y)$ para $y \in U$, mediante el cambio de variable $y = f(x)$ para $x \in \Omega$.

Si ahora usamos las coordenadas en las bases usuales de los tres espacios involucrados, las tres variables vectoriales x, y, z se convierten en tres bloques de variables que ya toman valores reales: $x = (x_1, x_2, \dots, x_N)$, $y = (y_1, y_2, \dots, y_M)$ y $z = (z_1, z_2, \dots, z_P)$. Estas $N + M + P$ variables están relacionadas mediante las componentes de f , g y h , por las igualdades

$$y_j = f_j(x_1, x_2, \dots, x_N), \quad z_i = g_i(y_1, y_2, \dots, y_M), \quad z_i = h_i(x_1, x_2, \dots, x_N) \quad (4)$$

válidas para cualesquiera $j \in \Delta_M$, $i \in \Delta_P$. El doble papel de las variables vectoriales y, z se ha transmitido a las variables reales y_1, y_2, \dots, y_M y z_1, z_2, \dots, z_P .

Las igualdades (4) sugieren sustituir en (3) cada función por la variable cuyos valores determina, teniendo en cuenta que g_i y h_i determinan a la misma variable z_i , la primera como función de y_1, y_2, \dots, y_M y la segunda como función de x_1, x_2, \dots, x_N . Así pues, escribimos

$$\frac{\partial y_j}{\partial x_k}(a) = \frac{\partial f_j}{\partial x_k}(a), \quad \frac{\partial z_i}{\partial y_j}(b) = \frac{\partial g_i}{\partial y_j}(b), \quad \frac{\partial z_i}{\partial x_k}(a) = \frac{\partial h_i}{\partial x_k}(a) \quad (5)$$

para cualesquiera $k \in \Delta_N$, $j \in \Delta_M$, $i \in \Delta_P$.

El doble uso de la variable y_j no tiene problema: en la primera igualdad está claro que y_j es una función de (x_1, x_2, \dots, x_N) , la función f_j que teníamos en (4), mientras que en la segunda igualdad y_j es una de las variables de las que depende z_i . El doble uso de la variable z_i es más delicado, pero también se comprende fácilmente. La dependencia de z_i con respecto a y_j viene determinada por la función g_i , como puede verse en (4), luego la derivada parcial de z_i con respecto a y_j no puede ser otra que la de g_i . Por la misma razón, en vista de la tercera igualdad de (4), la derivada parcial de z_i con respecto a x_k sólo puede ser la de h_i . En resumidas cuentas, lo que estamos haciendo es entender las derivadas parciales de unas variables respecto de otras como las derivadas parciales de las funciones que determinan la dependencia de las unas respecto de las otras. Esta notación parece complicada, e incluso formalmente discutible, pero es la más intuitiva y cómoda en la práctica, como enseguida veremos.

Así pues, al sustituir en (3) las igualdades (5), la regla de la cadena para el cálculo de derivadas parciales queda de la siguiente forma:

$$\frac{\partial z_i}{\partial x_k}(a) = \sum_{j=1}^M \frac{\partial z_i}{\partial y_j}(b) \frac{\partial y_j}{\partial x_k}(a) \quad \forall k \in \Delta_N \quad \forall i \in \Delta_P \quad (6)$$

De entrada esta igualdad es mucho más fácil de recordar que (3), sólo incluye los tres bloques inevitables de variables, $x \in \Omega$, $y \in U$ y $z \in \mathbb{R}^P$, formalmente han desaparecido las componentes de f, g y h , aunque obviamente siguen estando ahí, son las funciones con cuyas derivadas parciales estamos operando.

Además, (6) tiene una interpretación intuitiva que conviene comentar. Para $i \in \Delta_P$ y $k \in \Delta_N$, la derivada parcial $\frac{\partial z_i}{\partial x_k}(a)$ se obtiene pensando que z_i sólo depende de x_k , cuando el resto de variables x_h con $h \in \Delta_N \setminus \{k\}$ se mantienen constantes, dependencia que se produce, por así decirlo, a través de todas las variables intermedias y_j con $j \in \Delta_M$, pues todas ellas dependen de x_k . Esto permite interpretar la suma del segundo miembro de (6). Si fuese $M = 1$ y sólo hubiese una variable intermedia y_j , la regla de la cadena para dos funciones de una variable nos daría la derivada parcial como el producto $\frac{\partial z_i}{\partial y_j}(b) \frac{\partial y_j}{\partial x_k}(a)$, que es el j -ésimo sumando del segundo miembro de (6), así que podríamos decir que dicho sumando es la contribución de la variable y_j a la derivada parcial que estamos calculando. Entonces (6) nos dice que dicha derivada parcial se obtiene sumando las contribuciones de todas las variables intermedias. Por supuesto, esta interpretación no sirve como demostración de (6), pero ayuda a entender su significado. No obstante, la mejor forma de familiarizarse con la regla obtenida para el cálculo de derivadas parciales es ver cómo se usa en ejemplos concretos.

9.5. Coordenadas polares en el plano

Consideremos el caso particular en que $M = N = 2$ y f es la función que se usa para definir las coordenadas polares, pongamos por caso, en el semiplano superior. Concretamente estamos tomando $\Omega = \mathbb{R}^+ \times]0, \pi[$, y $f : \Omega \rightarrow \mathbb{R}^2$ es la función cuyas componentes, que vamos a denotar por x e y , vienen dadas por

$$x(\rho, \theta) = \rho \cos \theta \quad \text{e} \quad y(\rho, \theta) = \rho \sin \theta \quad \forall (\rho, \theta) \in \Omega \quad (7)$$

Es fácil ver que x e y , luego también f , son diferenciables en todo punto $(\rho, \theta) \in \Omega$, con

$$\frac{\partial x}{\partial \rho}(\rho, \theta) = \cos \theta, \quad \frac{\partial x}{\partial \theta}(\rho, \theta) = -\rho \sin \theta, \quad \frac{\partial y}{\partial \rho}(\rho, \theta) = \sin \theta, \quad \frac{\partial y}{\partial \theta}(\rho, \theta) = \rho \cos \theta \quad (8)$$

Obsérvese lo que ha ocurrido en nuestro caso con la primera igualdad de (5). Las variables de partida x_1 y x_2 son ρ y θ , mientras que las variables intermedias y_1 e y_2 serán x e y . Entonces la primera igualdad de (5) es automática, pues empezamos tomando $f_1 = x$ y $f_2 = y$. Ahora x e y pasarán a ser variables de las que dependerá otra función g , definida en $f(\Omega)$. Se comprueba sin dificultad que $f(\Omega)$ es el semiplano superior.

Consideremos un campo escalar $g : U \rightarrow \mathbb{R}$, donde $U = \{(x, y) \in \mathbb{R}^2 : y > 0\} = f(\Omega)$, y supongamos que g es diferenciable en todo punto de U . Por ejemplo, g puede describir la temperatura en el semiplano superior. Ahora tenemos $P = 1$, luego una sola variable $z \in \mathbb{R}$, que depende de x e y mediante la igualdad $z = g(x, y)$ y nos da la temperatura en el punto cuyas coordenadas cartesianas son (x, y) . Por tanto, para todo $(x, y) \in U$ escribimos

$$\frac{\partial z}{\partial x}(x, y) = \frac{\partial g}{\partial x}(x, y) \quad \text{y} \quad \frac{\partial z}{\partial y}(x, y) = \frac{\partial g}{\partial y}(x, y)$$

Tenemos aquí, en nuestro caso, la segunda igualdad de (5). La notación sugiere claramente que las funciones $\frac{\partial z}{\partial x}$ y $\frac{\partial z}{\partial y}$, las derivadas parciales de g , nos informan sobre cómo varía la temperatura al desplazarnos horizontal o verticalmente desde un punto arbitrario de U .

Para estudiar la temperatura en coordenadas polares, consideramos la función $h = g \circ f$, que es diferenciable en todo punto de Ω , y escribimos

$$z = g(x, y) = g(\rho \cos \theta, \rho \sin \theta) = h(\rho, \theta) \quad \forall (\rho, \theta) \in \Omega$$

Así pues, en nuestro caso, la tercera igualdad de (5) es la siguiente:

$$\frac{\partial z}{\partial \rho}(\rho, \theta) = \frac{\partial h}{\partial \rho}(\rho, \theta) \quad \text{y} \quad \frac{\partial z}{\partial \theta}(\rho, \theta) = \frac{\partial h}{\partial \theta}(\rho, \theta) \quad \forall (\rho, \theta) \in \Omega$$

Obsérvese que h es la función que describe la temperatura en coordenadas polares, es decir, para cada $(\rho, \theta) \in \Omega$, $z = h(\rho, \theta)$ es la temperatura en el punto $a = (x, y) = (\rho \cos \theta, \rho \sin \theta)$, el punto del semiplano superior cuyas coordenadas polares son ρ y θ . El doble uso de z tiene perfecto sentido: cada valor de z es la temperatura en un punto a del semiplano superior, que podemos calcular usando las coordenadas cartesianas de a , mediante la igualdad $z = g(x, y)$, o usando sus coordenadas polares, mediante $z = h(\rho, \theta)$.

Las derivadas parciales de h , es decir, las funciones $\frac{\partial z}{\partial \rho}$ y $\frac{\partial z}{\partial \theta}$, nos informan sobre como varía la temperatura al desplazarnos desde un punto arbitrario de U , por la recta que le une con el origen, o por una circunferencia centrada en el origen.

Pues bien, para $(\rho, \theta) \in \Omega$ arbitrario, la regla de la cadena para las derivadas parciales, escrita como en (6), con $(x, y) = (\rho \cos \theta, \rho \sin \theta)$, nos da:

$$\begin{aligned}\frac{\partial z}{\partial \rho}(\rho, \theta) &= \cos \theta \frac{\partial z}{\partial x}(x, y) + \sin \theta \frac{\partial z}{\partial y}(x, y) & y \\ \frac{\partial z}{\partial \theta}(\rho, \theta) &= -\rho \sin \theta \frac{\partial z}{\partial x}(x, y) + \rho \cos \theta \frac{\partial z}{\partial y}(x, y)\end{aligned}\quad (9)$$

Hemos hecho una serie de comentarios para poner de manifiesto el aspecto que toman en este caso concreto todas las igualdades que habíamos manejado en general, pero en la práctica el razonamiento es mucho más directo y sencillo: partiendo de (7) que no es más que la definición de las coordenadas polares, deducimos claramente (8), y entonces, viendo z como función de las variables x, y , definida en U , y también como función de ρ, θ , definida en Ω , tenemos (9).

Naturalmente, cuando conocemos explícitamente g , también tenemos explícitamente h , y es probable que podamos calcular sus derivadas parciales sin usar la regla de la cadena. Lo interesante de (9) es que permite calcular las derivadas parciales de h a partir de las de g , sin necesidad de conocer explícitamente g . Pero recíprocamente, si conocemos las derivadas parciales de h , podemos calcular las de g , pues de (9) se deduce claramente que

$$\begin{aligned}\cos \theta \frac{\partial z}{\partial \rho}(\rho, \theta) - \frac{\sin \theta}{\rho} \frac{\partial z}{\partial \theta}(\rho, \theta) &= \frac{\partial z}{\partial x}(x, y) & y \\ \sin \theta \frac{\partial z}{\partial \rho}(\rho, \theta) + \frac{\cos \theta}{\rho} \frac{\partial z}{\partial \theta}(\rho, \theta) &= \frac{\partial z}{\partial y}(x, y)\end{aligned}$$

9.6. Relación entre plano tangente y rectas tangentes

La regla de la cadena para las derivadas parciales nos va a dar una propiedad del plano tangente a una superficie explícita, que luego usaremos para motivar la definición del plano tangente a superficies más generales. Consideremos una superficie explícita $\Sigma = \text{Gr } f \subset \mathbb{R}^3$, donde $f: \Omega \rightarrow \mathbb{R}$ es una función continua en un abierto conexo $\Omega \subset \mathbb{R}^2$. Sabemos que si f es diferenciable en un punto $(x_0, y_0) \in \Omega$, tomando $z_0 = f(x_0, y_0)$ y $P_0 = (x_0, y_0, z_0) \in \Sigma$, el plano tangente Π , a la superficie Σ en el punto P_0 viene dado por la ecuación

$$z - z_0 = (x - x_0) \frac{\partial f}{\partial x}(x_0, y_0) + (y - y_0) \frac{\partial f}{\partial y}(x_0, y_0) \quad (10)$$

Sea ahora $C = \gamma(J)$ una curva paramétrica en \mathbb{R}^3 , tal que $P_0 \in C \subset \Sigma$. Se tiene por tanto que $\gamma: J \rightarrow \mathbb{R}^3$ es continua en un intervalo abierto $J \subset \mathbb{R}$, existe $t_0 \in J$ tal que $\gamma(t_0) = P_0$, y además, $\gamma(t) \in \Sigma$ para todo $t \in J$.

Si x, y, z son las tres componentes de γ , vemos que $(x(t), y(t)) \in \Omega$ y $z(t) = f(x(t), y(t))$ para todo $t \in J$. Así pues, las ecuaciones paramétricas de C son

$$x = x(t), \quad y = y(t), \quad z = f(x(t), y(t)) \quad \forall t \in J$$

Supongamos que $P_0 = \gamma(t_0)$ es punto regular de $C = \gamma(J)$, es decir, que γ es derivable en t_0 con $\gamma'(t_0) \neq 0$, para ver la relación entre la recta R , tangente a la curva C en el punto P_0 , y el plano tangente Π . Podemos entender la recta R como una *recta tangente* a la superficie Σ en el punto P_0 , y cabe esperar que se tenga $R \subset \Pi$, como efectivamente vamos a comprobar.

La regla de la cadena nos dice que

$$z'(t_0) = x'(t_0) \frac{\partial f}{\partial x}(x_0, y_0) + y'(t_0) \frac{\partial f}{\partial y}(x_0, y_0)$$

luego las ecuaciones paramétricas de R son:

$$\begin{aligned} x &= x_0 + t x'(t_0), & y &= y_0 + t y'(t_0) & y \\ z &= z_0 + t \left(x'(t_0) \frac{\partial f}{\partial x}(x_0, y_0) + y'(t_0) \frac{\partial f}{\partial y}(x_0, y_0) \right) & \forall t \in \mathbb{R} \end{aligned} \quad (11)$$

Comparando (11) y (10) vemos claramente que para todo $(x, y, z) \in R$ se tiene $(x, y, z) \in \Pi$, luego $R \subset \Pi$ como queríamos.

En resumen, podemos decir que *el plano tangente a la superficie Σ en el punto P_0 contiene a todas las rectas tangentes en dicho punto a curvas paramétricas que estén contenidas en la superficie Σ* , o si se quiere, a todas las rectas tangentes a la superficie Σ en P_0 . La definición del plano tangente se justificó por ser una buena aproximación de la superficie Σ cerca del punto P_0 , pero ahora hemos encontrado una segunda justificación, más intuitiva si cabe: el plano tangente a Σ en P_0 es el único plano que contiene a todas las rectas tangentes a Σ en P_0 .

9.7. Superficies en forma paramétrica

Generalizando la noción de superficie explícita, definimos las superficies paramétricas, de forma bastante análoga a lo que hicimos para las curvas paramétricas, sólo que intuitivamente ahora se trata de un objeto “bidimensional”, luego necesitamos dos parámetros.

Llamamos **superficie paramétrica** en \mathbb{R}^3 a la imagen de toda función continua $\Gamma : \Omega \rightarrow \mathbb{R}^3$ donde Ω es un subconjunto no vacío, abierto y conexo de \mathbb{R}^2 . Se trata por tanto del conjunto

$$\Sigma = \Gamma(\Omega) = \{\Gamma(t, s) : (t, s) \in \Omega\}$$

Las variables reales t y s son ahora los *parámetros*, y a cada valor $(t, s) \in \Omega$ de los mismos, corresponde un único punto $\Gamma(t, s)$ de la superficie Σ , luego podemos decir que la función Γ *parametriza* la superficie Σ . Obviamente, podemos tener otras parametrizaciones muy distintas de la misma superficie. En toda la discusión que sigue, usaremos Γ , y debe quedar claro que todo lo que hagamos depende de dicha parametrización.

Como hicimos con las curvas paramétricas, llamamos x, y, z a las tres componentes de la función Γ y decimos que

$$x = x(t, s), \quad y = y(t, s), \quad z = z(t, s), \quad ((t, s) \in \Omega)$$

son las *ecuaciones paramétricas* de la superficie $\Sigma = \Gamma(\Omega)$. Por ejemplo, el conjunto

$$\Sigma = \{ (x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = 1 \}$$

un cilindro vertical de radio 1, es una superficie paramétrica cuyas ecuaciones pueden ser

$$x = \cos t, \quad y = \sin t, \quad z = s, \quad (t, s \in \mathbb{R}) \quad (12)$$

Es claro que las superficies explícitas son un tipo muy particular de superficies paramétricas. Si $\Sigma = \text{Gr } f$ donde $f : \Omega \rightarrow \mathbb{R}$ es una función continua, definiendo $\Gamma(x, y) = (x, y, f(x, y))$ para todo $(x, y) \in \Omega$ tenemos una función continua $\Gamma : \Omega \rightarrow \mathbb{R}^3$ tal que $\Gamma(\Omega) = \text{Gr } f$. Digamos que una superficie explícita se puede siempre parametrizar usando como parámetros las dos primeras coordenadas de los puntos de la superficie. Podríamos usar otro par de coordenadas y considerar las superficies explícitas de la forma

$$\Sigma_1 = \{ (f(y, z), y, z) : (y, z) \in \Omega \}, \quad \text{o bien,} \quad \Sigma_2 = \{ (x, f(x, z), z) : (x, z) \in \Omega \}$$

cuyas ecuaciones explícitas serían $x = f(y, z)$ con $(y, z) \in \Omega$, o bien $y = f(x, z)$ con $(x, z) \in \Omega$.

Está claro que $\Sigma_1 = \Gamma_1(\Omega)$ y $\Sigma_2 = \Gamma_2(\Omega)$ donde

$$\Gamma_1(y, z) = (f(y, z), y, z) \quad \forall (y, z) \in \Omega \quad \text{y} \quad \Gamma_2(x, z) = (x, f(x, z), z) \quad \forall (x, z) \in \Omega$$

luego Σ_1 y Σ_2 también son superficies paramétricas. El cilindro definido en (12) es un ejemplo sencillo de superficie paramétrica que no es explícita en ninguno de los tres sentidos.

9.8. Plano tangente a una superficie paramétrica

Para extender la noción de plano tangente, fijamos una superficie paramétrica $\Sigma = \Gamma(\Omega)$ donde $\Gamma : \Omega \rightarrow \mathbb{R}^3$ es continua en un abierto conexo $\Omega \subset \mathbb{R}^2$, cuyas componentes serán x, y, z . Supongamos que Γ es diferenciable en un punto $(t_0, s_0) \in \Omega$, sea $P_0 = \Gamma(t_0, s_0) = (x_0, y_0, z_0)$ y observemos la matriz jacobiana, cuyas columnas son las derivadas parciales de Γ :

$$J\Gamma(t_0, s_0) = \begin{pmatrix} \frac{\partial x}{\partial t}(t_0, s_0) & \frac{\partial x}{\partial s}(t_0, s_0) \\ \frac{\partial y}{\partial t}(t_0, s_0) & \frac{\partial y}{\partial s}(t_0, s_0) \\ \frac{\partial z}{\partial t}(t_0, s_0) & \frac{\partial z}{\partial s}(t_0, s_0) \end{pmatrix} = \left(\frac{\partial \Gamma}{\partial t}(t_0, s_0), \frac{\partial \Gamma}{\partial s}(t_0, s_0) \right)$$

Por razones que se comprenderán enseguida, *supondremos que $J\Gamma(t_0, s_0)$ tiene rango 2*, es decir, que sus columnas son vectores linealmente independientes de \mathbb{R}^3 . La idea clave para definir el plano tangente a Σ en el punto P_0 se puede fácilmente adivinar: las rectas tangentes en P_0 a curvas paramétricas contenidas en la superficie Σ deben estar contenidas en el plano tangente que buscamos.

Usando que Ω es abierto, podemos encontrar $\delta > 0$ de forma que se tenga $J_1 \times J_2 \subset \Omega$ donde $J_1 =]t_0 - \delta, t_0 + \delta[$ y $J_2 =]s_0 - \delta, s_0 + \delta[$. Sean entonces $\gamma_1 : J_1 \rightarrow \mathbb{R}^3$ y $\gamma_2 : J_2 \rightarrow \mathbb{R}^3$ las funciones definidas por

$$\gamma_1(t) = \Gamma(t, s_0) \quad \forall t \in J_1 \quad \text{y} \quad \gamma_2(s) = \Gamma(t_0, s) \quad \forall s \in J_2$$

Tenemos dos curvas paramétricas $C_1 = \gamma_1(J_1)$ y $C_2 = \gamma_2(J_2)$, contenidas en la superficie Σ y verificando que $P_0 = \gamma_1(t_0) = \gamma_2(s_0)$. Además γ_1 es derivable en t_0 y γ_2 en s_0 , con

$$\gamma_1'(t_0) = \frac{\partial \Gamma}{\partial t}(t_0, s_0) \quad \text{y} \quad \gamma_2'(s_0) = \frac{\partial \Gamma}{\partial s}(t_0, s_0)$$

vectores que, gracias a la hipótesis sobre la matriz $J\Gamma(t_0, s_0)$, son linealmente independientes y, en particular, no nulos. Por tanto, $P_0 = \gamma_1(t_0) = \gamma_2(s_0)$ es un punto regular de las curvas C_1 y C_2 , cuyas rectas tangentes en P_0 son distintas. Ello nos permite considerar el único plano Π que contiene a dichas rectas tangentes, es decir, el único plano que pasa por el punto P_0 y tiene como vectores de dirección a las dos derivadas parciales de Γ en (t_0, s_0) .

Por tanto las ecuaciones paramétricas del plano Π pueden escribirse en la forma:

$$\begin{aligned} x &= x_0 + (t - t_0) \frac{\partial x}{\partial t}(t_0, s_0) + (s - s_0) \frac{\partial x}{\partial s}(t_0, s_0) \\ y &= y_0 + (t - t_0) \frac{\partial y}{\partial t}(t_0, s_0) + (s - s_0) \frac{\partial y}{\partial s}(t_0, s_0) \\ z &= z_0 + (t - t_0) \frac{\partial z}{\partial t}(t_0, s_0) + (s - s_0) \frac{\partial z}{\partial s}(t_0, s_0) \end{aligned} \quad (13)$$

Nótese que obtendríamos el mismo plano escribiendo t y s en lugar de $t - t_0$ y $s - s_0$. Escribir las ecuaciones de esta forma tiene la ventaja de que el punto P_0 se obtiene tomando en ellas $t = t_0$ y $s = s_0$, igual que ocurre con las ecuaciones de la superficie Σ . Más concretamente, si para cada $(t, s) \in \Omega$ consideramos el punto $P(t, s) = (x, y, z)$ dado por estas ecuaciones, tenemos $P(t_0, s_0) = P_0$. Por supuesto, la función $P : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ parametriza el plano $\Pi = P(\Omega)$, como superficie paramétrica que es. Las ecuaciones (13) son más fáciles de recordar y de manejar si las escribimos en forma matricial:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} + J\Gamma(t_0, s_0) \cdot \begin{pmatrix} t - t_0 \\ s - s_0 \end{pmatrix} \quad (14)$$

Decimos que Π es el **plano tangente** a la superficie Σ en el punto P_0 , definición que, como en el caso de una superficie explícita, se justifica de dos maneras. En primer lugar, el plano Π es una buena aproximación de la superficie Σ cerca del punto P_0 , en el sentido que sigue.

Para $(t, s) \in \Omega$, veamos la distancia del punto $\Gamma(t, s)$ de la superficie Σ al correspondiente punto $P(t, s)$ del plano Π . En vista de (14) tenemos

$$P(t, s) = \Gamma(t_0, s_0) + D\Gamma(t_0, s_0) \left((t, s) - (t_0, s_0) \right) \quad \forall t, s \in \mathbb{R}$$

y de la definición de diferenciabilidad deducimos que

$$\lim_{(t,s) \rightarrow (t_0,s_0)} \frac{\|\Gamma(t, s) - P(t, s)\|}{\|(t, s) - (t_0, s_0)\|}$$

Por tanto $\|\Gamma(t, s) - P(t, s)\|$ tiende a cero cuando $(t, s) \rightarrow (t_0, s_0)$ “mucho más rápidamente” que $\|(t, s) - (t_0, s_0)\|$. Este es el tipo de buena aproximación de la superficie Σ mediante el plano Π cerca del punto P_0 que justifica analíticamente la definición de plano tangente.

Para tener una justificación más geométrica, podemos usar curvas paramétricas contenidas en la superficie Σ , más generales que las curvas C_1 y C_2 usadas para encontrar el plano Π , y comprobar que sus rectas tangentes están contenidas en Π . Concretamente, sea $\phi: J \rightarrow \Omega$ una función continua en un intervalo abierto $J \subset \mathbb{R}$, tal que $\phi(\alpha_0) = (t_0, s_0)$ para algún $\alpha_0 \in J$. La función continua $\gamma = \Gamma \circ \phi$ nos da una curva paramétrica $C = \gamma(J)$ que claramente está contenida en la superficie Σ y verifica que $\gamma(\alpha_0) = P_0$. Nótese que esta es una forma sencilla de obtener multitud de curvas paramétricas contenidas en la superficie Σ y que contienen al punto P_0 , cada función del mismo tipo que ϕ da lugar a una curva de las requeridas.

Cuando ϕ es derivable en α_0 con $\phi'(\alpha_0) \neq 0$, la regla de la cadena nos dice que γ es derivable en α_0 y, escribiendo ambos vectores derivada como matrices columna, tenemos

$$\gamma'(\alpha_0) = J\Gamma(t_0, s_0) \cdot \phi'(\alpha_0)$$

Como $J\Gamma(t_0, s_0)$ tiene rango 2, deducimos que $\gamma'(\alpha_0) \neq 0$, luego $P_0 = \gamma(\alpha_0)$ es un punto regular de la curva $C = \gamma(J)$. Escribimos las ecuaciones paramétricas de la recta tangente R , como hicimos para el plano tangente Π , es decir, las expresamos en forma matricial, haciendo que al valor $\alpha = \alpha_0$ del parámetro corresponda el punto P_0 , obteniendo

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} + J\Gamma(t_0, s_0) \cdot \phi'(\alpha_0) \cdot (\alpha - \alpha_0) \quad (\alpha \in \mathbb{R}) \quad (15)$$

Está claro ahora que la recta R está contenida en el plano Π : si $(x, y, z) \in \mathbb{R}^3$ verifica (15) para algún $\alpha \in \mathbb{R}$, basta tomar $(t, s) \in \mathbb{R}^2$ dados por

$$\begin{pmatrix} t - t_0 \\ s - s_0 \end{pmatrix} = \phi'(\alpha_0) \cdot (\alpha - \alpha_0)$$

para obtener (14), luego $(x, y, z) \in \Pi$.

En resumen, para la definición de plano tangente a una superficie paramétrica, tenemos justificaciones análogas a las obtenidas para superficies explícitas: el plano tangente en cada punto P_0 , cuando existe, es una buena aproximación de la superficie cerca de P_0 , y las rectas tangentes a la superficie en dicho punto están contenidas en el plano tangente.

Resaltamos que, para definir el plano tangente a la superficie paramétrica $\Sigma = \Gamma(\Omega)$ en el punto $P_0 = \Gamma(t_0, s_0)$, y sobre todo, para poder justificarla como lo hemos hecho, es necesario suponer que Γ es diferenciable en (t_0, s_0) y que la matriz $J\Gamma(t_0, s_0)$ tiene rango 2. Cuando se cumplen estas condiciones, se dice que $P_0 = \Gamma(t_0, s_0)$ es *punto regular* de la superficie Σ , y si todos sus puntos son regulares, decimos que Σ es regular. Por tanto, $\Sigma = \Gamma(\Omega)$ es una *superficie regular* cuando Γ es diferenciable y $J\Gamma(t, s)$ tiene rango 2 para todo $(t, s) \in \Omega$. Debemos tener claro que estas nociones dependen de la parametrización Γ que usemos, para otra parametrización de la misma superficie, los puntos regulares podrían ser otros. De hecho, cuando $\Sigma = \Gamma(\Omega)$ es regular, sería más adecuado decir que Γ es una *parametrización regular* de la superficie Σ .

Nótese también que estas nociones de regularidad son análogas a las que usamos para una curva paramétrica $C = \gamma(J) \subset \mathbb{R}^M$ donde $\gamma: J \rightarrow \mathbb{R}^M$ es continua en un intervalo abierto $J \subset \mathbb{R}$. Dado $t_0 \in J$, necesitamos que γ sea derivable en t_0 con $\gamma'(t_0) \neq 0$, para tener recta tangente a la curva C en el punto $P_0 = \gamma(t_0)$. Pensemos que el vector derivada $\gamma'(t_0)$ es una matriz $M \times 1$ que tiene rango 1 cuando no se anula. Para superficies paramétricas tenemos una matriz 3×2 a la que exigimos tener rango 2.

Comprobemos que el plano tangente a una superficie explícita, considerada como superficie paramétrica, es el mismo que habíamos definido previamente. Si $f: \Omega \rightarrow \mathbb{R}$ es una función continua en un abierto $\Omega \subset \mathbb{R}^2$, la superficie explícita $\Sigma = \text{Gr } f$ se parametriza como sabemos mediante la función $\Gamma: \Omega \rightarrow \mathbb{R}^3$ dada por

$$\Gamma(x, y) = (x, y, f(x, y)) \quad \forall (x, y) \in \Omega$$

Cuando f es diferenciable en un punto $(x_0, y_0) \in \Omega$, es claro que Γ también lo es, con

$$J\Gamma(x_0, y_0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \frac{\partial f}{\partial x}(x_0, y_0) & \frac{\partial f}{\partial y}(x_0, y_0) \end{pmatrix}$$

luego $J\Gamma(x_0, y_0)$ tiene rango 2, y $P_0 = \Gamma(t_0, s_0) = (x_0, y_0, z_0)$ es un punto regular de Σ . Además, las ecuaciones paramétricas del plano tangente a Σ en P_0 pueden escribirse en la forma

$$x = x_0 + t, \quad y = y_0 + s, \quad z = z_0 + t \frac{\partial f}{\partial x}(x_0, y_0) + s \frac{\partial f}{\partial y}(x_0, y_0) \quad \forall t, s \in \mathbb{R}$$

luego equivalen claramente a (10). Volvemos pues a obtener el plano tangente, tal y como se definió para una superficie explícita.

Nótese que, si f es diferenciable en todo punto de Ω , entonces Γ es una parametrización regular de la superficie explícita $\Sigma = \text{Gr } f$.

Consideremos finalmente el cilindro cuyas ecuaciones paramétricas vimos en (12). En este caso, la función $\Gamma: \mathbb{R}^2 \rightarrow \mathbb{R}^3$ es diferenciable en todo punto $(t, s) \in \mathbb{R}^2$ con

$$J\Gamma(t, s) = \begin{pmatrix} -\sin t & 0 \\ \cos t & 0 \\ 0 & 1 \end{pmatrix}$$

Claramente, $J\Gamma(t, s)$ tiene rango 2, para todo $(t, s) \in \mathbb{R}^2$, luego el cilindro $\Sigma = \Gamma(\mathbb{R}^2)$ es una superficie regular. Su plano tangente en cualquier punto $(x_0, y_0, z_0) \in \Sigma$ tiene ecuaciones paramétricas

$$x = x_0 - t y_0, \quad y = y_0 + t x_0, \quad z = z_0 + s, \quad (t, s \in \mathbb{R})$$

9.9. Curvas y superficies de nivel

Usando otra vez la regla de la cadena para las derivadas parciales, vamos a probar una propiedad del gradiente de un campo escalar que tiene interés en Física. Fijamos por tanto un campo escalar $f : \Omega \rightarrow \mathbb{R}$, donde Ω es un abierto de \mathbb{R}^N . Llamamos conjuntos de nivel de f a los subconjuntos de Ω en los que toma cada uno de sus valores. Más concretamente, para cada valor $\lambda \in f(\Omega) \subset \mathbb{R}$, el **conjunto de nivel** λ del campo f viene dado por

$$N_\lambda = \{x \in \Omega : f(x) = \lambda\}$$

Obviamente, sin hipótesis sobre f , sus conjuntos de nivel pueden ser totalmente arbitrarios. Concentrándonos en los casos más interesantes, $N = 2$ o $N = 3$, consideramos respectivamente curvas o superficies paramétricas, que estén contenidas en un conjunto de nivel.

En el caso $N = 2$, decimos que una curva paramétrica $C = \gamma(J)$, donde $\gamma : J \rightarrow \Omega$ es continua en un intervalo abierto $J \subset \mathbb{R}$, es una **curva de nivel** del campo f , cuando C está contenida en un conjunto de nivel de f , es decir, cuando $f \circ \gamma$ es constante.

Si γ es derivable en un punto $t_0 \in J$ y f es diferenciable en el punto $P_0 = \gamma(t_0) \in \Omega$, la regla de la cadena nos dice que los vectores $\nabla f(P_0)$ y $\gamma'(t_0)$ son ortogonales, ya que

$$(\nabla f(P_0) | \gamma'(t_0)) = (f \circ \gamma)'(t_0) = 0$$

Claramente, esta afirmación tiene interés cuando ambos vectores son no nulos, pues entonces nos dice que *el vector gradiente $\nabla f(P_0)$ es ortogonal a la recta tangente a toda curva de nivel del campo f , para la que P_0 sea un punto regular*. Hablando con poca precisión, suele decirse que el gradiente de un campo escalar en \mathbb{R}^2 es ortogonal a las curvas de nivel. Si queremos desplazarnos en Ω de forma que el campo f se mantenga constante, es decir, siguiendo una curva de nivel, debemos hacerlo siempre en dirección ortogonal al vector gradiente.

En el caso $N = 3$, decimos que una superficie paramétrica $\Sigma = \Gamma(U)$, donde $\Gamma : U \rightarrow \Omega$ es continua en un abierto conexo $U \subset \mathbb{R}^2$, es una **superficie de nivel** del campo f , cuando Σ está contenida en un conjunto de nivel de f , es decir, cuando $f \circ \Gamma$ es constante. Suponiendo que Γ es diferenciable en $(t_0, s_0) \in U$ y que f es diferenciable en $P_0 = \Gamma(t_0, s_0) \in \Omega$, la regla de la cadena nos dice que el vector $\nabla f(P_0)$ es ortogonal a los vectores $\frac{\partial \Gamma}{\partial t}(t_0, s_0)$ y $\frac{\partial \Gamma}{\partial s}(t_0, s_0)$. En efecto, basta observar que

$$0 = \frac{\partial(f \circ \Gamma)}{\partial t}(t_0, s_0) = \left(\nabla f(P_0) \left| \frac{\partial \Gamma}{\partial t}(t_0, s_0) \right. \right)$$

e igual ocurre con la otra derivada parcial.

De nuevo, este resultado tiene interés cuando $\nabla f(P_0) \neq 0$ y $P_0 = \Gamma(t_0, s_0)$ es punto regular de Σ , pues entonces las dos derivadas parciales de Γ en (t_0, s_0) son vectores de dirección del plano tangente a Σ en P_0 , luego $\nabla f(P_0)$ es un vector normal a dicho plano tangente. Así pues, *el vector gradiente $\nabla f(P_0)$ es ortogonal al plano tangente a cualquier superficie de nivel del campo f , para la que P_0 sea un punto regular.* Dicho de forma imprecisa, el gradiente de un campo escalar en \mathbb{R}^3 es ortogonal a las superficies de nivel.

9.10. Un ejemplo ilustrativo

Hemos insistido en que todas las aplicaciones anteriores de la regla de la cadena para las derivadas parciales, sólo adquieren su verdadero sentido cuando las funciones involucradas son diferenciables. Sin embargo, cabe preguntarse si la propia regla de la cadena es válida para la composición de dos funciones parcialmente derivables, pues ello es obviamente cierto en el caso $N = M = 1$. El siguiente ejemplo, en el caso más sencillo, $N = 1$ y $M = 2$, prueba que la respuesta es negativa.

Sea $f: \mathbb{R} \rightarrow \mathbb{R}^2$ dada por $f(t) = (t, t)$ para todo $t \in \mathbb{R}$, y $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ definida por

$$g(x, y) = \frac{x^2 y}{x^2 + y^4} \quad \forall (x, y) \in \mathbb{R}^2 \setminus \{(0, 0)\}, \quad g(0, 0) = 0$$

Obviamente f es derivable con $f'(t) = (1, 1)$ para todo $t \in \mathbb{R}$, y $g(x, 0) = g(0, y) = 0$ para cualesquiera $x, y \in \mathbb{R}$, luego g es parcialmente derivable en el origen con $\nabla g(0, 0) = (0, 0)$. Puesto que $f(0) = (0, 0)$, si la regla de la cadena para las derivadas parciales pudiera aplicarse en este caso, para la función $h = g \circ f$ se tendría $h'(0) = 0$. Vemos sin embargo que

$$h(t) = \frac{t}{1+t^2} \quad \forall t \in \mathbb{R}, \quad \text{luego} \quad h'(0) = \lim_{t \rightarrow 0} \frac{1}{1+t^2} = 1$$

Observemos que f sí es diferenciable, pero la regla de la cadena para las derivadas parciales no puede aplicarse, lo que sólo puede deberse a una razón: g no es diferenciable en el origen.

Es fácil ver que g es continua, luego $\Sigma = \text{Gr } g$ es una superficie explícita, e ingenuamente podríamos pensar que el plano de ecuación $z = 0$ es tangente a Σ en el origen. Consideremos sin embargo la curva $C = \gamma(\mathbb{R})$ donde

$$\gamma(t) = (t, t, h(t)) \quad \forall t \in \mathbb{R}$$

Tenemos claramente $C \subset \Sigma$ y el origen es un punto regular de C pero la recta tangente a C en el origen tiene vector de dirección $\gamma'(0) = (1, 1, 1)$ luego no está contenida en el supuesto plano tangente, que por tanto no merece ese nombre.

Queda claro una vez más, que la noción relevante del cálculo diferencial, desde el punto de vista teórico y por sus aplicaciones, es la diferenciabilidad, no la derivabilidad parcial.

Teorema del valor medio

Podría decirse que hasta ahora sólo hemos sentado las bases para el estudio del cálculo diferencial en varias variables. Hemos introducido el concepto general o abstracto de función diferenciable y comprobado algunas propiedades básicas, como su carácter local o su relación con la continuidad. También hemos visto cómo se concreta dicho concepto en los tres casos particulares que más nos interesan, vector derivada, vector gradiente y matriz jacobiana, viendo al mismo tiempo diversas interpretaciones geométricas y físicas. Por último disponemos de algunos métodos para estudiar la diferenciabilidad de una función y calcular su diferencial, entre los que destaca la regla de la cadena.

Ha llegado ya el momento de estudiar los teoremas fundamentales del cálculo diferencial, empezando como es lógico por intentar generalizar el *teorema del valor medio*, que es sin duda el resultado más importante del cálculo diferencial en una variable. Obtendremos fácilmente una versión del teorema para funciones que parten de un espacio normado arbitrario y toman valores reales, que en particular se aplica a los campos escalares en \mathbb{R}^N . Para funciones con valores vectoriales las cosas se complican, la versión literal del teorema es falsa, incluso para funciones de una variable real con valores en \mathbb{R}^2 . Sin embargo, probaremos una versión más débil, que sí es cierta a plena generalidad y permite deducir algunas consecuencias interesantes.

10.1. Funciones con valores escalares

Recordemos el teorema del valor medio para funciones reales de variable real:

- Sean $a, b \in \mathbb{R}$ con $a < b$ y $f : [a, b] \rightarrow \mathbb{R}$ una función continua, que suponemos derivable en el intervalo abierto $]a, b[$.
 - (i) **Teorema del valor medio:** existe $c \in]a, b[$ tal que $f(b) - f(a) = f'(c)(b - a)$
 - (ii) **Desigualdad del valor medio:** si existe $M \in \mathbb{R}_0^+$ verificando que $|f'(x)| \leq M$ para todo $x \in]a, b[$, entonces $|f(b) - f(a)| \leq M(b - a)$.

Cuando se trabaja con funciones reales de variable real, la desigualdad del valor medio no suele destacarse, siendo como es, consecuencia obvia del teorema. Sin embargo, veremos que, cuando se trabaja con espacios normados, el teorema no siempre es cierto, mientras que la desigualdad es válida a plena generalidad. Por otra parte, la desigualdad es suficiente para conseguir varias consecuencias importantes del teorema. Por ejemplo, permite probar que toda función derivable en un intervalo, con derivada acotada, es lipschitziana y, en el caso de que la derivada sea idénticamente nula, la función es constante.

Con vistas a la generalización del teorema anterior, conviene hacer algunas observaciones. Por una parte, el papel del intervalo $[a, b]$ puede hacerlo cualquier segmento entre dos puntos de un espacio normado, mientras que $]a, b[$ debe ser el mismo segmento, excluidos sus extremos. Concretamente, si X es un espacio normado y $a, b \in X$, escribiremos

$$[a, b] = \{a + t(b - a) : t \in [0, 1]\} \quad \text{y} \quad]a, b[= \{a + t(b - a) : t \in]0, 1[\}$$

Obsérvese que, cuando $a \neq b$, único caso que nos interesa, se tiene $]a, b[= [a, b] \setminus \{a, b\}$. Tiene sentido decir que $[a, b]$ es el *segmento cerrado* de extremos a y b , pues se trata de un conjunto cerrado, e incluso compacto. Pero conviene notar que, salvo que X tenga dimensión 1, el conjunto $]a, b[$ no es abierto, de hecho $[a, b]$ tiene interior vacío. Por esta razón, no podemos trabajar con una función f definida solamente en el segmento $[a, b]$, pues no tendría sentido hablar de la diferenciabilidad de f . Una solución aceptable consiste en suponer que f está definida en un abierto que contenga al segmento $[a, b]$. Eso sí, supondremos solamente que f es continua en $[a, b]$ y diferenciable en todo punto de $]a, b[$.

Por otra parte, está claro que la tesis del teorema del valor medio tiene perfecto sentido, otra cosa es que sea cierta, para una función con valores en un espacio normado arbitrario Y . Piénsese que, en el caso $X = Y = \mathbb{R}$, multiplicar $b - a$ por la derivada $f'(c)$ es tanto como aplicar al número real $b - a$ la diferencial $Df(c)$. Por tanto, en general, podemos preguntarnos si existe un punto $c \in]a, b[$ que verifique $f(b) - f(a) = Df(c)(b - a)$, pues ambos miembros de esta igualdad son vectores de Y .

Pues bien, expresado de esta forma, el teorema del valor medio, y la desigualdad que de él se deduce, son ciertos, siempre que el espacio normado de llegada sea \mathbb{R} , es decir, para funciones con valores escalares:

Teorema del valor medio escalar. *Sea Ω un abierto de un espacio normado X , $a, b \in \Omega$ tales que $[a, b] \subset \Omega$ y $f : \Omega \rightarrow \mathbb{R}$ una función continua en $[a, b]$ y diferenciable en $]a, b[$. Entonces, existe $c \in]a, b[$ tal que*

$$f(b) - f(a) = Df(c)(b - a) \quad (1)$$

Como consecuencia, si $M \in \mathbb{R}_0^+$ verifica que $\|Df(x)\| \leq M$ para todo $x \in]a, b[$, se tendrá:

$$|f(b) - f(a)| \leq M \|b - a\| \quad (2)$$

Demostración. Definimos una función $\varphi : [0, 1] \rightarrow \mathbb{R}$, escribiendo $\varphi(t) = f(a + t(b - a))$ para todo $t \in [0, 1]$. Entonces φ es continua y, por la regla de la cadena, es derivable en $]0, 1[$ con derivada dada por

$$\varphi'(t) = Df(a + t(b - a))(b - a) \quad \forall t \in]0, 1[$$

El teorema del valor medio para funciones reales de variable real nos da un $t_0 \in]0, 1[$ tal que

$$f(b) - f(a) = \varphi(1) - \varphi(0) = \varphi'(t_0) = Df(a + t_0(b-a))(b-a)$$

Tomando $c = a + t_0(b-a) \in]a, b[$, obtenemos (1). Para deducir (2) basta usar la definición de la norma en $L(X, \mathbb{R})$:

$$|f(b) - f(a)| = |Df(c)(b-a)| \leq \|Df(c)\| \|b-a\| \leq M \|b-a\| \quad \blacksquare$$

En el caso particular de un campo escalar en \mathbb{R}^N , usando la desigualdad de Cauchy-Schwartz obtenemos:

- Sea Ω un abierto de \mathbb{R}^N , $a, b \in \Omega$ tales que $[a, b] \subset \Omega$ y $f : \Omega \rightarrow \mathbb{R}$ un campo escalar continuo en $[a, b]$ y diferenciable en $]a, b[$. Entonces, existe $c \in]a, b[$ tal que

$$f(b) - f(a) = (\nabla f(c) | b - a)$$

Por tanto, usando en \mathbb{R}^N la norma euclídea, si existe $M \in \mathbb{R}_0^+$ tal que $\|\nabla f(x)\| \leq M$ para todo $x \in]a, b[$, se tendrá:

$$|f(b) - f(a)| \leq M \|b - a\|$$

Como se habrá podido observar, la demostración del teorema del valor medio escalar ha consistido en una aplicación muy inmediata del ya conocido para funciones de una variable. De hecho, como sólo se trabaja con la función f en un segmento $[a, b]$, en esencia tenemos una función de una variable. Los resultados anteriores, generalizan formalmente el teorema del valor medio que ya conocíamos, pero en realidad sólo consisten en darse cuenta de que dicho teorema puede aplicarse fácilmente en situaciones más generales. Tiene mayor interés ver lo que ocurre en el caso de un campo vectorial, o incluso para una función entre espacios normados arbitrarios, como enseguida vamos a intentar.

10.2. Caso general

Observamos inmediatamente que la primera afirmación del teorema del valor medio escalar no es cierta para funciones con valores en \mathbb{R}^2 , incluso aunque sean funciones de una variable real. Concretamente, basta considerar la función $f : \mathbb{R} \rightarrow \mathbb{R}^2$ dada por $f(t) = (\cos t, \sin t)$ para todo $t \in \mathbb{R}$, que claramente es derivable en todo punto de \mathbb{R} con

$$f'(t) = (-\sin t, \cos t) \neq (0, 0) \quad \forall t \in \mathbb{R}$$

Tomando $a \in \mathbb{R}$ arbitrario y $b = a + 2\pi$ se tiene $f(b) = f(a)$. Si existiese $c \in [a, b]$ verificando que $f(b) - f(a) = Df(c)(b-a) = 2\pi f'(c)$, se tendría $f'(c) = (0, 0)$, lo cual es imposible.

Sin embargo, probaremos a plena generalidad la segunda de las afirmaciones del teorema del valor medio escalar, es decir, la desigualdad del valor medio. Ciertamente es más débil que el teorema, pero será suficiente para deducir interesantes consecuencias.

La dificultad para probarlo se concentra en el caso de funciones de una variable real, que es el que resolvemos previamente.

Lema. Sea Y un espacio normado y sean $g : [0, 1] \rightarrow Y$ y $\alpha : [0, 1] \rightarrow \mathbb{R}$ dos funciones continuas en $[0, 1]$ y derivables en $]0, 1[$, verificando que

$$\|g'(t)\| \leq \alpha'(t) \quad \forall t \in]0, 1[\quad (3)$$

Se tiene entonces la siguiente desigualdad:

$$\|g(1) - g(0)\| \leq \alpha(1) - \alpha(0) \quad (4)$$

Demostración. Fijado $\varepsilon > 0$, consideramos el conjunto

$$\Lambda = \{t \in [0, 1] : \|g(t) - g(0)\| \leq \alpha(t) - \alpha(0) + \varepsilon t + \varepsilon\}$$

y a poco que se piense, la demostración estará casi concluida si probamos que $1 \in \Lambda$.

Por ser g y α continuas, la función $\varphi : [0, 1] \rightarrow \mathbb{R}$ definida por

$$\varphi(t) = \|g(t) - g(0)\| - (\alpha(t) - \alpha(0)) - \varepsilon t \quad \forall t \in [0, 1]$$

también es continua, de lo que deduciremos dos consecuencias:

En primer lugar, como $\varphi(0) = 0$, la continuidad de φ en 0 nos permite encontrar $\eta \in]0, 1[$ tal que, para $t \in [0, \eta]$ se tenga $\varphi(t) < \varepsilon$, con lo que $[0, \eta] \subset \Lambda$.

Por otra parte, como $\Lambda = \{t \in [0, 1] : \varphi(t) \leq \varepsilon\}$, la continuidad de φ nos dice que Λ es un subconjunto cerrado de $[0, 1]$, luego es compacto, y en particular tiene máximo. Sea pues $t_0 = \max \Lambda$ y anotemos que $t_0 \geq \eta > 0$. Nuestro objetivo es probar que $t_0 = 1$, así que supondremos que $t_0 < 1$ para llegar a una contradicción.

Al ser $0 < t_0 < 1$, tenemos que g y α son derivables en t_0 , luego existe $\delta > 0$ tal que, para todo $t \in [0, 1]$ con $|t - t_0| \leq \delta$, se tiene

$$\begin{aligned} \|g(t) - g(t_0) - g'(t_0)(t - t_0)\| &\leq \frac{\varepsilon}{2}|t - t_0| & \text{y también} \\ |\alpha(t) - \alpha(t_0) - \alpha'(t_0)(t - t_0)| &\leq \frac{\varepsilon}{2}|t - t_0| \end{aligned}$$

Obviamente podemos suponer que $t_0 + \delta < 1$, para tomar $t = t_0 + \delta$ y obtener

$$\begin{aligned} \|g(t_0 + \delta) - g(t_0) - \delta g'(t_0)\| &\leq \frac{\varepsilon}{2}\delta & \text{así como} \\ |\alpha(t_0 + \delta) - \alpha(t_0) - \delta \alpha'(t_0)| &\leq \frac{\varepsilon}{2}\delta \end{aligned} \quad (5)$$

Llegaremos a contradicción viendo que $t_0 + \delta \in \Lambda$. Para ello, usando la primera desigualdad de (5), la hipótesis (3) con $t = t_0$, y el hecho de que $t_0 \in \Lambda$, tenemos:

$$\begin{aligned} \|g(t_0 + \delta) - g(0)\| &\leq \|g(t_0 + \delta) - g(t_0) - \delta g'(t_0)\| + \delta \|g'(t_0)\| + \|g(t_0) - g(0)\| \\ &\leq \frac{\varepsilon}{2}\delta + \delta \alpha'(t_0) + \alpha(t_0) - \alpha(0) + \varepsilon t_0 + \varepsilon \end{aligned} \quad (6)$$

Por otra parte, usando la segunda desigualdad de (5) tenemos también

$$\begin{aligned}\delta\alpha'(t_0) + \alpha(t_0) &= \alpha(t_0 + \delta) - (\alpha(t_0 + \delta) - \alpha(t_0) - \delta\alpha'(t_0)) \\ &\leq \alpha(t_0 + \delta) + \frac{\varepsilon}{2}\delta\end{aligned}\quad (7)$$

De (6) y (7) deducimos claramente que

$$\begin{aligned}\|g(t_0 + \delta) - g(0)\| &\leq \frac{\varepsilon}{2}\delta + \alpha(t_0 + \delta) + \frac{\varepsilon}{2}\delta - \alpha(0) + \varepsilon t_0 + \varepsilon \\ &= \alpha(t_0 + \delta) - \alpha(0) + \varepsilon(t_0 + \delta) + \varepsilon\end{aligned}$$

es decir, que $t_0 + \delta \in \Lambda$. Esto es una clara contradicción, ya que $t_0 + \delta > t_0 = \max \Lambda$.

Así pues hemos comprobado que $t_0 = 1$ y en particular $1 \in \Lambda$, es decir

$$\|g(1) - g(0)\| \leq \alpha(1) - \alpha(0) + 2\varepsilon$$

Como esto es válido para todo $\varepsilon \in \mathbb{R}^+$, tenemos (4), como queríamos. ■

En el caso $Y = \mathbb{R}^2$, aunque también se podría hacer en general, el resultado anterior tiene una interpretación física que lo hace muy plausible. Pensemos que g describe un movimiento curvilíneo en el plano y α un movimiento rectilíneo, durante el mismo intervalo de tiempo. La hipótesis (3) significa que la celeridad del primer móvil nunca supera a la del segundo. Entonces $\|g(1) - g(0)\|$ es la distancia entre las posiciones inicial y final del móvil más lento, que será menor o igual que la longitud de la trayectoria recorrida, pero esta longitud ha de ser a su vez menor o igual que $\alpha(1) - \alpha(0)$, que es la distancia recorrida por el móvil más rápido.

Desigualdad del valor medio. Sean X e Y espacios normados, Ω abierto de X , $a, b \in X$ tales que $[a, b] \subset \Omega$ y $f : \Omega \rightarrow Y$ una función. Supongamos que f es continua en $[a, b]$ y diferenciable en $]a, b[$, y que existe $M \in \mathbb{R}_0^+$ tal que $\|Df(x)\| \leq M$ para todo $x \in]a, b[$. Se tiene entonces:

$$\|f(b) - f(a)\| \leq M\|b - a\|$$

Demostración. Basta aplicar el lema anterior a las funciones $g : [0, 1] \rightarrow Y$ y $\alpha : [0, 1] \rightarrow \mathbb{R}$ definidas, para todo $t \in [0, 1]$, por

$$g(t) = f((1-t)a + tb) \quad \text{y} \quad \alpha(t) = M\|b - a\|t$$

Es claro que g y α son continuas en $[0, 1]$ y derivables en $]0, 1[$ con

$$\begin{aligned}\|g'(t)\| &= \|Df((1-t)a + tb)(b-a)\| \\ &\leq \|Df((1-t)a + tb)\| \|b-a\| \leq M\|b-a\| = \alpha'(t) \quad \forall t \in]0, 1[\end{aligned}$$

Aplicando pues el lema anterior, obtenemos la desigualdad buscada:

$$\|f(b) - f(a)\| = \|g(1) - g(0)\| \leq \alpha(1) - \alpha(0) = M\|b - a\| \quad \blacksquare$$

Como primera aplicación de este resultado, es natural ponerse en una situación que permita usarlo con cualquier par de puntos $a, b \in \Omega$, lo cual es bien fácil:

Corolario 1. Sean X e Y espacios normados, Ω un abierto convexo de X y $f : \Omega \rightarrow Y$ una función diferenciable. Supongamos que existe $M \in \mathbb{R}_0^+$ verificando que $\|Df(x)\| \leq M$ para todo $x \in \Omega$. Entonces f es lipschitziana con constante M , es decir:

$$\|f(b) - f(a)\| \leq M \|b - a\| \quad \forall a, b \in \Omega$$

Obsérvese que, en caso de que podamos tomar $M = 0$, esto es, cuando la diferencial Df es idénticamente nula, obtenemos que $f(b) = f(a)$ para cualesquiera $a, b \in \Omega$, es decir, que f es constante. Pero podemos conseguir el mismo resultado con una hipótesis más débil que la convexidad de Ω :

Corolario 2. Sean X e Y espacios normados, Ω un subconjunto abierto y conexo de X y $f : \Omega \rightarrow Y$ una función diferenciable tal que $Df(x) = 0$ para todo $x \in \Omega$. Entonces f es constante.

Demostración. Fijado $a \in \Omega$, consideramos el conjunto $A = \{x \in \Omega : f(x) = f(a)\}$ y se trata de probar que $A = \Omega$. Por ser f continua, A es un subconjunto cerrado de Ω , pero vamos a comprobar que también es abierto. Dado $x \in A$, como Ω es abierto, existe $r \in \mathbb{R}^+$ tal que $B(x, r) \subset \Omega$. Podemos entonces aplicar el corolario anterior, a la restricción de f al abierto convexo $B(x, r)$, cuya función diferencial es idénticamente nula, obteniendo que f es constante en $B(x, r)$. Por tanto tenemos que $f(y) = f(x) = f(a)$ para todo $y \in B(x, r)$, es decir, $B(x, r) \subset A$. Como $x \in A$ era arbitrario, hemos probado que A es abierto. Puesto que Ω es conexo y $A \neq \emptyset$, porque $a \in A$, concluimos que $A = \Omega$ como se quería. ■

Destacamos una consecuencia obvia: con las mismas hipótesis sobre Ω , si f y g son dos funciones diferenciables en Ω tales que $Df = Dg$, entonces $f - g$ es constante. Dicho de otra forma, una función diferenciable en un subconjunto abierto y conexo de un espacio normado, queda determinada por su función diferencial, salvo adición de una constante.

10.3. Aplicaciones a campos escalares o vectoriales

En el caso de un campo escalar o vectorial, es decir, cuando $X = \mathbb{R}^N$ e $Y = \mathbb{R}^M$, para usar el último corolario no necesitamos comprobar la diferenciabilidad de la función f , sino que basta trabajar con sus derivadas parciales:

- Sea Ω un abierto conexo de \mathbb{R}^N y $f : \Omega \rightarrow \mathbb{R}^M$ una función. Supongamos que todas las derivadas parciales de f son idénticamente nulas, es decir,

$$\frac{\partial f}{\partial x_k}(x) = 0 \quad \forall x \in \Omega, \quad \forall k \in \Delta_N$$

Entonces f es constante.

Trabajando con cada una de las componentes de f , basta considerar el caso $M = 1$. Es obvio que todas las derivadas parciales de f son funciones continuas, luego podemos usar la condición suficiente de diferenciabilidad, obteniendo que f es diferenciable en Ω . Como $Df(x) = 0$ para todo $x \in \Omega$, basta usar el Corolario 2. ■

Terminamos este tema con una aplicación del teorema del valor medio en su versión para funciones reales de variable real. Intuitivamente, si Ω es un abierto de \mathbb{R}^N con $N \geq 2$, y un campo escalar o vectorial $f : \Omega \rightarrow \mathbb{R}^M$ es derivable con respecto a una variable, siendo dicha derivada parcial idénticamente nula en Ω , parece natural pensar que f sólo dependa de las demás variables. Esto es obviamente falso cuando Ω no es conexo, como se puede fácilmente comprobar, y veremos más adelante que puede ser falso aunque Ω sea conexo. Previamente probaremos un resultado positivo en el caso de que Ω sea *convexo*, cuya demostración nos mostrará la posibilidad de considerar una hipótesis más general y nos hará comprender lo que puede fallar cuando no se cumple esa hipótesis.

- Sea Ω un abierto convexo de \mathbb{R}^N con $N \geq 2$, y fijado $k \in \Delta_N$, sea $f : \Omega \rightarrow \mathbb{R}^M$ una función derivable con respecto a la k -ésima variable en Ω , con

$$\frac{\partial f}{\partial x_k}(x) = 0 \quad \forall x \in \Omega$$

Si para $x = (x_1, x_2, \dots, x_N) \in \Omega$ escribimos $\hat{x} = (x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_N) \in \mathbb{R}^{N-1}$, y consideramos el conjunto $U = \{\hat{x} : x \in \Omega\}$, entonces existe una función $g : U \rightarrow \mathbb{R}^M$ tal que $f(x) = g(\hat{x})$ para todo $x \in \Omega$. Dicho de manera más intuitiva, f no depende de la k -ésima variable.

De nuevo, usando las componentes de f , basta considerar el caso $M = 1$. A poco que se piense, bastará comprobar lo siguiente:

$$x, u \in \Omega, \hat{x} = \hat{u} \implies f(x) = f(u) \quad (8)$$

Comprobada esta implicación, para $\hat{x} \in U$ con $x \in \Omega$, podremos definir $g(\hat{x}) = f(x)$. Esta definición es correcta, pues aunque x no es único, para otro $u \in \Omega$ que verifique $\hat{u} = \hat{x}$, tendremos $f(u) = f(x)$, luego el valor de $g(\hat{x})$ dependerá de \hat{x} , pero no de x . Obtenemos una función $g : U \rightarrow \mathbb{R}^M$, que obviamente cumple la condición requerida.

Para probar (8) escribimos $x = (x_1, x_2, \dots, x_N)$ y consideramos el conjunto

$$J = \{t \in \mathbb{R} : (x_1, \dots, x_{k-1}, t, x_{k+1}, \dots, x_N) \in \Omega\} \quad (9)$$

Usando que Ω es convexo, se comprueba rutinariamente que J también lo es, luego J es un intervalo. Consideramos ahora la función $\varphi : J \rightarrow \mathbb{R}$ definida por

$$\varphi(t) = f(x_1, \dots, x_{k-1}, t, x_{k+1}, \dots, x_N) \quad \forall t \in J$$

Como f es derivable con respecto a la k -ésima variable en todo punto de Ω , vemos claramente que φ es derivable en J con

$$\varphi'(t) = \frac{\partial f}{\partial x_k}(x_1, \dots, x_{k-1}, t, x_{k+1}, \dots, x_N) = 0 \quad \forall t \in J$$

Del teorema del valor medio para funciones reales de variable real, deducimos que φ es constante. Es obvio que $x_k \in J$ y, como $\hat{u} = \hat{x}$, escribiendo $u = (x_1, \dots, x_{k-1}, u_k, x_{k+1}, \dots, x_N)$, de $u \in \Omega$ deducimos que $u_k \in J$. Concluimos entonces que $f(x) = \varphi(x_k) = \varphi(u_k) = f(u)$, y hemos probado (8), concluyendo así la demostración. ■

Es fácil adivinar una hipótesis más general sobre Ω , para conseguir el mismo resultado. Basta observar que la convexidad de Ω sólo se ha usado para deducir que el conjunto J definido en (9) es un intervalo. Bastará por tanto suponer que, para todo $x \in \Omega$, el conjunto J es un intervalo. Geométricamente, esto significa que, si una recta R tiene la dirección del k -ésimo eje de coordenadas, el conjunto $R \cap \Omega$ ha de ser convexo, pudiendo ser vacío. Por ejemplo, si $N = 2$ y $k = 1$, la intersección con Ω de cualquier recta horizontal debe ser un conjunto convexo. Es claro que esta condición es más débil que la convexidad de Ω . De hecho, si Ω es convexo, entonces $R \cap \Omega$ es un conjunto convexo para toda recta R , cualquiera que sea su dirección. Veamos finalmente que, aunque Ω sea conexo, el resultado anterior puede no verificarse.

Consideremos el conjunto $\Omega = \mathbb{R}^2 \setminus \{(0, y) : y \geq 0\}$, que también podemos escribir en la forma $\Omega = A \cup B \cup C$ donde

$$A = \{(x, y) \in \mathbb{R}^2 : x < 0\}, \quad B = \{(x, y) \in \mathbb{R}^2 : y < 0\}, \quad C = \{(x, y) \in \mathbb{R}^2 : x > 0\}$$

tres semiplanos, que obviamente son conjuntos abiertos y convexos, luego conexos. Por tanto, vemos que Ω es abierto, y también conexo. De hecho, como $A \cap B \neq \emptyset$, vemos que $A \cup B$ es conexo, pero también $(A \cup B) \cap C \neq \emptyset$ luego $A \cup B \cup C = \Omega$ es conexo. Observamos que la intersección de Ω con la recta de ecuación $y = 1$ es el conjunto $\{(x, 1) : x \in \mathbb{R}\} \setminus \{(0, 1)\}$, que no es convexo. Esto hace que una función $f : \Omega \rightarrow \mathbb{R}$ pueda depender de la primera variable, aunque su primera derivada parcial sea idénticamente nula en Ω , como vamos a ver.

Consideramos la función $f : \Omega \rightarrow \mathbb{R}^2$ dada por

$$f(x, y) = y^2 \quad \forall (x, y) \in \mathbb{R}^+ \times \mathbb{R}^+, \quad f(x, y) = 0 \quad \forall (x, y) \in \Omega \setminus (\mathbb{R}^+ \times \mathbb{R}^+)$$

Como f es idénticamente nula en el conjunto abierto $A \cup B$, es derivable con respecto a la primera variable, con derivada parcial idénticamente nula en dicho conjunto. Por otra parte, en el primer cuadrante $\mathbb{R}^+ \times \mathbb{R}^+$, que también es abierto, tenemos claramente que

$$\frac{\partial f}{\partial x}(x, y) = 0 \quad \forall x, y \in \mathbb{R}^+$$

Para un punto de la forma $(x_0, 0)$ con $x_0 > 0$, observamos que $f(x, 0) = 0$ para todo $x \in \mathbb{R}^+$, de donde deducimos que

$$\frac{\partial f}{\partial x}(x_0, 0) = 0 \quad \forall x_0 \in \mathbb{R}^+$$

En resumen, f es derivable con respecto a la primera variable, y su primera derivada parcial es idénticamente nula en Ω . Sin embargo, vemos que f sí depende de la primera variable, puesto que $f(-1, 1) = 0 \neq 1 = f(1, 1)$.

Es fácil comprobar que f también es derivable con respecto a la segunda variable y su segunda derivada parcial es continua, luego de hecho se tiene que $f \in C^1(\Omega)$.

Función inversa

Como segundo resultado fundamental del cálculo diferencial, estudiaremos ahora la posible existencia, así como la diferenciabilidad, de la función inversa de una función diferenciable. Empezamos probando una *regla de diferenciación de la función inversa*, para funciones entre espacios normados arbitrarios, que generaliza casi al pie de la letra la que conocemos para funciones reales de una variable real.

Pasamos entonces al problema mucho más interesante que empieza por discutir la propia existencia de la función inversa, esto es, la inyectividad de una función diferenciable. Claro está que, una vez asegurada la inyectividad, también nos preguntamos si la función inversa es diferenciable. En el caso conocido de las funciones reales de una variable real, este problema se resolvía mediante el llamado “teorema de la función inversa” para tales funciones, que tiene una versión “local” y otra “global”. A la hora de su generalización a funciones de varias variables, estos dos resultados se comportan de manera muy diferente. La versión local se generaliza de forma casi literal, obteniendo el *teorema de la función inversa* en \mathbb{R}^N . Como consecuencia de este teorema fundamental, que como se ha dicho tiene carácter “local”, deducimos fácilmente un resultado que puede entenderse como una versión “global” del teorema, pero no generaliza lo que conocemos para funciones de una variable, es un resultado mucho más débil.

11.1. Diferencial de la función inversa

Recordemos la regla de derivación de la inversa, para funciones de una variable:

- Sea A un subconjunto no vacío de \mathbb{R} y $f : A \rightarrow \mathbb{R}$ una función inyectiva. Sea $B = f(A)$ y consideremos la función inversa $f^{-1} : B \rightarrow \mathbb{R}$. Si f es derivable en un punto $a \in A \cap A'$, entonces $b = f(a) \in B'$ y las siguientes afirmaciones son equivalentes:

- (i) f^{-1} es derivable en el punto b
- (ii) f^{-1} es continua en b y $f'(a) \neq 0$

Caso de que se cumplan (i) y (ii), se tiene: $(f^{-1})'(b) = \frac{1}{f'(a)}$.

Con vistas a generalizar este resultado, de forma que pueda aplicarse a una función entre dos espacios normados arbitrarios, conviene aclarar dos cuestiones.

Tendremos un abierto Ω de un espacio normado X y una función inyectiva $f : \Omega \rightarrow Y$, donde Y es otro espacio normado, que será diferenciable en un punto $a \in \Omega$. Surge aquí la única pequeña diferencia con el caso particular que acabamos de recordar: para poder hablar de la diferenciabilidad de f^{-1} en el punto $b = f(a)$, necesitamos que b sea, no sólo punto de acumulación del conjunto $B = f(\Omega)$, sino punto interior del mismo. Las hipótesis sobre f no permiten asegurar que esto ocurra, luego deberemos suponerlo. Al fin y al cabo, cuando $b \notin B^\circ$, no tiene sentido plantearse el problema que estamos estudiando.

La otra cuestión, más relevante, se refiere a la forma de generalizar la afirmación (ii), más concretamente, la condición $f'(a) \neq 0$ que en ella aparece. Es una ingenuidad pensar que la condición vaya a ser $Df(a) \neq 0$. Recordemos que, en el caso particular, la diferencial $Df(a)$ consiste en multiplicar por $f'(a)$. Por tanto, decir que $f'(a) \neq 0$, equivale a decir que $Df(a)$ es biyectiva, y entonces, la expresión de $(f^{-1})'(b)$ toma la forma $Df^{-1}(b) = Df(a)^{-1}$. En general, si queremos tener la misma expresión para la diferencial de la función inversa, no sólo necesitamos que $Df(a)$ sea biyectiva, sino también que su inversa $Df(a)^{-1}$ sea continua. Como se trata de una aplicación lineal de Y en X , su continuidad será automática cuando Y tenga dimensión finita, pero no en general. Esto motiva la siguiente definición.

Si X e Y son espacios normados, llamamos **homeomorfismo lineal** de X sobre Y a toda aplicación lineal y continua $T : X \rightarrow Y$, que sea biyectiva y tal que T^{-1} también sea continua. Más brevemente, estamos exigiendo que $T \in L(X, Y)$ sea biyectiva con $T^{-1} \in L(Y, X)$. La nomenclatura es coherente con la noción de homeomorfismo que se usa en Topología: biyección continua entre dos espacios topológicos, cuya función inversa también es continua. En nuestro caso, un homeomorfismo lineal entre dos espacios normados X e Y no es más que lo que su nombre indica: un homeomorfismo de X sobre Y que además es lineal, lo que obviamente hace que la función inversa también sea lineal. Notemos que, cuando dos espacios normados son *linealmente homeomorfos*, es decir, existe un homeomorfismo lineal de uno en otro, son idénticos como espacios vectoriales (pues tenemos una biyección lineal entre ellos) y también como espacios topológicos (pues son homeomorfos). Podemos por tanto pensar que se trata de un mismo espacio vectorial, en el que tenemos dos normas que pueden ser distintas, pero son equivalentes, pues dan lugar a la misma topología.

Los comentarios hechos anteriormente hacen ver que el siguiente resultado generaliza, de manera casi literal, la regla de derivación de la función inversa que acabamos de recordar.

Teorema (Regla de diferenciación de la inversa). Sean X e Y espacios normados, Ω un abierto de X , $f : \Omega \rightarrow Y$ una función inyectiva y $f^{-1} : B \rightarrow X$ su inversa, donde $B = f(\Omega)$. Supongamos que f es diferenciable en un punto $a \in \Omega$ y que $b = f(a) \in B^\circ$. Entonces, las siguientes afirmaciones son equivalentes:

- (i) f^{-1} es diferenciable en el punto b .
- (ii) f^{-1} es continua en b y $Df(a)$ es un homeomorfismo lineal de X sobre Y .

En caso de que se cumplan (i) y (ii), se tiene: $Df^{-1}(b) = Df(a)^{-1}$.

Demostración. (i) \Rightarrow (ii). Como f^{-1} es diferenciable en b , también es continua en b . Si ahora $\text{Id}_X \in L(X, X)$ e $\text{Id}_Y \in L(Y, Y)$ son las funciones identidad en X e Y respectivamente, vemos que $f^{-1} \circ f$ es la restricción a Ω de Id_X . Sabemos entonces que $f^{-1} \circ f$ es diferenciable en el punto a y su diferencial es la propia Id_X . Análogamente, la diferencial de $f \circ f^{-1}$ en el punto b es Id_Y . Aplicando entonces dos veces la regla de la cadena, obtenemos

$$Df^{-1}(b) \circ Df(a) = \text{Id}_X \quad \text{y} \quad Df(a) \circ Df^{-1}(b) = \text{Id}_Y$$

Esto prueba que $Df(a)$ es biyectiva con $Df(a)^{-1} = Df^{-1}(b) \in L(Y, X)$, luego $Df(a)$ es un homeomorfismo lineal de X sobre Y , como queríamos. De paso, hemos probado ya la última afirmación del enunciado: $Df^{-1}(b) = Df(a)^{-1}$.

(ii) \Rightarrow (i). Para abreviar escribimos $T = Df(a) \in L(X, Y)$ y, por hipótesis, sabemos que T es biyectiva con $T^{-1} \in L(Y, X)$. Como ya hemos usado otras veces, la diferenciabilidad de f en el punto a equivale a la continuidad en a de la función $\Phi : \Omega \rightarrow \mathbb{R}$ definida por

$$\Phi(x) = \frac{\|f(x) - f(a) - T(x-a)\|}{\|x-a\|} \quad \forall x \in \Omega \setminus \{a\} \quad \text{y} \quad \Phi(a) = 0$$

Por tanto Φ es continua en el punto a y evidentemente verifica que

$$\|f(x) - f(a) - T(x-a)\| = \Phi(x) \|x-a\| \quad \forall x \in \Omega \quad (1)$$

Definimos entonces $\Lambda : B \rightarrow \mathbb{R}$ por

$$\Lambda(y) = \|f^{-1}(y) - f^{-1}(b) - T^{-1}(y-b)\| \quad \forall y \in B$$

de modo que bastará probar la igualdad

$$\lim_{y \rightarrow b} \frac{\Lambda(y)}{\|y-b\|} = 0 \quad (2)$$

Fijamos $y \in B$ y escribimos $x = f^{-1}(y) \in \Omega$, con lo cual tenemos

$$\begin{aligned} \Lambda(y) &= \|x - a - T^{-1}(y-b)\| = \|T^{-1}(T(x-a) - (y-b))\| \\ &\leq \|T^{-1}\| \|T(x-a) - (f(x) - f(a))\| \\ &= \|T^{-1}\| \Phi(x) \|x-a\| \end{aligned} \quad (3)$$

donde hemos usado la definición de la norma en $L(Y, X)$ y la igualdad (1). Por otra parte, tenemos también

$$\begin{aligned} \|x-a\| &= \|T^{-1}(T(x-a))\| \leq \|T^{-1}\| \|T(x-a)\| \\ &\leq \|T^{-1}\| (\|T(x-a) - (f(x) - f(a))\| + \|f(x) - f(a)\|) \\ &= \|T^{-1}\| (\Phi(x) \|x-a\| + \|y-b\|) \end{aligned}$$

o lo que es lo mismo,

$$(1 - \|T^{-1}\| \Phi(x)) \|x-a\| \leq \|T^{-1}\| \|y-b\| \quad (4)$$

Queremos deducir de (4) una estimación de $\|x - a\|$ que podamos usar en (3). Para ello necesitamos que $\|T^{-1}\| \Phi(x) < 1$, lo cual es cierto para y suficientemente próximo a b , como vamos a ver. En primer lugar, por ser Φ continua en el punto a con $\Phi(a) = 0$, existe $\rho > 0$ tal que, para $z \in \Omega$ con $\|z - a\| < \rho$ se tiene $\|T^{-1}\| \Phi(z) < 1$. Como ahora f^{-1} es continua en b , existe $\delta > 0$ tal que, para $w \in B$ con $\|w - b\| < \delta$ se tiene $\|f^{-1}(w) - f^{-1}(b)\| < \rho$. Volviendo al punto $y \in B$ que habíamos fijado, pero suponiendo además que $\|y - b\| < \delta$, usamos lo anterior con $w = y$ para obtener $\|f^{-1}(y) - f^{-1}(b)\| < \rho$, es decir, $\|x - a\| < \rho$. Esto permite tomar $z = x$ para concluir que $\|T^{-1}\| \Phi(x) < 1$ como queríamos.

Por tanto, si $y \in B$ verifica que $\|y - b\| < \delta$, para $x = f^{-1}(y)$ tenemos, usando (4), que

$$\|x - a\| \leq \frac{\|T^{-1}\| \|y - b\|}{1 - \|T^{-1}\| \Phi(x)}$$

y de (3) deducimos entonces que

$$\Lambda(y) \leq \frac{\|T^{-1}\|^2 \Phi(x) \|y - b\|}{1 - \|T^{-1}\| \Phi(x)}$$

En resumen, hemos encontrado $\delta > 0$ tal que

$$y \in B, \quad 0 < \|y - b\| < \delta \implies 0 \leq \frac{\Lambda(y)}{\|y - b\|} \leq \frac{\|T^{-1}\|^2 \Phi(f^{-1}(y))}{1 - \|T^{-1}\| \Phi(f^{-1}(y))} \quad (5)$$

Finalmente, como f^{-1} es continua en b y Φ es continua en $f^{-1}(b) = a$, tenemos que $\Phi \circ f^{-1}$ es continua en b , es decir,

$$\lim_{y \rightarrow b} \Phi(f^{-1}(y)) = \Phi(f^{-1}(b)) = \Phi(a) = 0$$

Esta igualdad, junto con (5), nos da directamente (2), que era la igualdad buscada. ■

La implicación fácil del teorema anterior tiene una consecuencia que merece la pena resaltar:

- Sean U y V abiertos de sendos espacios normados X e Y . Si existe una aplicación biyectiva $f : U \rightarrow V$ que es diferenciable en un punto $a \in U$ y f^{-1} es diferenciable en el punto $f(a)$, entonces X e Y son linealmente homeomorfos. En particular, si X tiene dimensión finita $N \in \mathbb{N}$, también Y tendrá dimensión N .

En efecto, por el teorema anterior, $Df(a)$ es un homeomorfismo lineal de X sobre Y . ■

A partir de ahora trabajamos en espacios de dimensión finita, y en vista de la observación anterior, de la misma dimensión, es decir, $X = Y = \mathbb{R}^N$ aunque las normas de X e Y pueden ser dos normas cualesquiera en \mathbb{R}^N . Al aplicar el teorema anterior, $Df(a)^{-1}$ será continua siempre que $Df(a)$ sea biyectiva, pues $Df(a)^{-1}$ es lineal y está definida en un espacio normado de dimensión finita. Por tanto, para asegurar que f^{-1} es diferenciable en b , bastará comprobar que es continua en b y que $Df(a)$ es biyectiva.

Para cada aplicación lineal $T \in L(\mathbb{R}^N, \mathbb{R}^N)$ recordemos la matriz cuadrada $A_T \in \mathcal{M}_{N \times N}$ que la representa en la base usual de \mathbb{R}^N , cuyo determinante denotamos por $\det A_T$. Sabemos que T es biyectiva si, y sólo si, $\det A_T \neq 0$. Si ahora Ω es un abierto de \mathbb{R}^N y $f : \Omega \rightarrow \mathbb{R}^N$ una función diferenciable en un punto $a \in \Omega$, para $T = Df(a)$ sabemos que $A_T = Jf(a)$ es la matriz jacobiana de f en a y se dice que $\det Jf(a)$ es el **determinante jacobiano** de f en a . Según hemos dicho, $Df(a)$ es biyectiva si, y sólo si, $\det Jf(a) \neq 0$. Así pues, en nuestro caso particular, el teorema anterior toma la siguiente forma:

- Sea Ω un abierto de \mathbb{R}^N , $f : \Omega \rightarrow \mathbb{R}^N$ una función inyectiva y $f^{-1} : B \rightarrow \mathbb{R}^N$ su inversa, donde $B = f(\Omega)$. Supongamos que f es diferenciable en $a \in \Omega$ y que $b = f(a) \in B^\circ$. Entonces, las siguientes afirmaciones son equivalentes:

- (i) f^{-1} es diferenciable en el punto b .
- (ii) f^{-1} es continua en b y $\det Jf(a) \neq 0$.

En caso de que se cumplan (i) y (ii), se tiene que $Df^{-1}(b) = Df(a)^{-1}$, o lo que es lo mismo, $Jf^{-1}(b) = Jf(a)^{-1}$.

Acerca del determinante jacobiano, resaltamos un hecho elemental que será útil enseguida. Es fácil ver que la aplicación $T \mapsto \det A_T$, definida en el espacio normado $L(\mathbb{R}^N, \mathbb{R}^N)$ y con valores en \mathbb{R} , es continua. En efecto, para cada $j, k \in \Delta_N$ sabemos que el coeficiente que aparece en la j -ésima fila y la k -ésima columna de la matriz A_T viene dado por

$$\alpha_{jk}(T) = (T e_k | e_j) \quad \forall T \in L(\mathbb{R}^N, \mathbb{R}^N)$$

donde $\{e_1, e_2, \dots, e_N\}$ es la base usual de \mathbb{R}^N . Entonces la aplicación $\alpha_{jk} : L(\mathbb{R}^N, \mathbb{R}^N) \rightarrow \mathbb{R}$ es continua, puesto que es lineal y está definida en un espacio normado de dimensión finita. Por tanto, la función $T \mapsto \det A_T$ se puede expresar como combinación lineal de productos de funciones continuas, luego es una función continua. Apliquemos ahora esta observación al caso particular del determinante jacobiano de una función diferenciable:

- Sea Ω un abierto de \mathbb{R}^N y $f : \Omega \rightarrow \mathbb{R}^N$ una función diferenciable. Si Df es continua en un punto $a \in \Omega$, entonces la función $x \mapsto \det Jf(x)$, definida en Ω y con valores en \mathbb{R} , también es continua en a .

En efecto, se trata de la composición de la función Df , de Ω en $L(\mathbb{R}^N, \mathbb{R}^N)$, que es continua en a , con la función $T \mapsto \det A_T$ de $L(\mathbb{R}^N, \mathbb{R}^N)$ en \mathbb{R} , que es continua, como hemos visto. ■

11.2. Teorema de la función inversa

En la regla de diferenciación de la función inversa que acabamos de estudiar, trabajamos con una función que de entrada se supone inyectiva. Sin embargo, igual que se hace con las funciones reales de una variable real, podemos intentar averiguar si una función diferenciable es inyectiva, usando su diferencial. El significado analítico de la diferencial invita a pensar que podemos tener éxito a nivel “local”, como pasamos a explicar.

Hablando informalmente, una función f que sea diferenciable en un punto a , se aproxima bien, “cerca” de a , por una función afín y continua g , que se obtiene componiendo $Df(a)$ con una traslación. Cuando $Df(a)$ es inyectiva, está claro que g también lo es, y es lógico esperar que f sea inyectiva donde sabemos que debe comportarse de forma parecida a como lo hace g , es decir, en un entorno suficientemente pequeño de a . Con las hipótesis adecuadas, probaremos rigurosamente que esta idea intuitiva es correcta, obteniendo así el que puede considerarse como el resultado más importante del cálculo diferencial en varias variables: el teorema de la función inversa en \mathbb{R}^N . Para que se comprenda mejor el enunciado de dicho teorema, recordemos lo que ocurría en el caso $N = 1$, es decir, la versión local del teorema de la función inversa para funciones reales de variable real:

- Sea $I \subset \mathbb{R}$ un intervalo no trivial y $f : I \rightarrow \mathbb{R}$ una función derivable en I . Supongamos que f' es continua en un punto $a \in I$, con $f'(a) \neq 0$. Entonces existe un $\delta > 0$ tal que f es inyectiva en el intervalo $I_\delta = I \cap]a - \delta, a + \delta[$. Además, si llamamos φ a la restricción de f a I_δ y consideramos el intervalo $J_\delta = f(I_\delta)$, se tiene que φ^{-1} es derivable en J_δ con $(\varphi^{-1})'(f(x)) = 1/f'(x)$ para todo $x \in I_\delta$.

Vamos a conseguir una generalización casi literal del teorema anterior, válida para funciones de N variables. El único cambio se debe a que sólo podemos usar la diferenciabilidad de una función en puntos interiores de su conjunto de definición. Por tanto, el papel de los intervalos que aparecen en el enunciado anterior, lo harán convenientes abiertos de \mathbb{R}^N . La demostración es muy diferente de la que se hace en \mathbb{R} , basada en la noción de monotonía. En \mathbb{R}^N las cosas son más complicadas, como se podrá comprobar.

Teorema de la función inversa (local). Sea Ω un abierto de \mathbb{R}^N y $f : \Omega \rightarrow \mathbb{R}^N$ una función diferenciable, es decir, $f \in D(\Omega, \mathbb{R}^N)$. Supongamos que la función diferencial Df es continua en un punto $a \in \Omega$ y que $\det Jf(a) \neq 0$. Entonces existe un abierto U de \mathbb{R}^N , con $a \in U \subset \Omega$, para el que se verifican las siguientes afirmaciones:

- (i) f es inyectiva en U
- (ii) $V = f(U)$ es un abierto de \mathbb{R}^N
- (iii) $\det Jf(x) \neq 0$ para todo $x \in U$
- (iv) Si $\varphi = f|_U$, entonces $\varphi^{-1} \in D(V, \mathbb{R}^N)$, con $D\varphi^{-1}(f(x)) = Df(x)^{-1}$ para todo $x \in U$.

Demostración. Usaremos cualquier norma en \mathbb{R}^N , y empezamos trabajando en un caso particular más cómodo, del que al final se deducirá fácilmente el caso general. Concretamente suponemos por ahora que $a = f(a) = 0$ y que $Df(0) = \text{Id}$ es la identidad en \mathbb{R}^N .

En este caso, la idea intuitiva en la que se basa el teorema está muy clara: en un entorno del origen, f debe comportarse de forma similar a como lo hace la identidad, una función que cumple obviamente todas las afirmaciones del teorema. Así pues, procede trabajar con la diferencia entre ambas, es decir, la función $g : \Omega \rightarrow \mathbb{R}^N$ dada por

$$g(x) = x - f(x) \quad \forall x \in \Omega$$

Es claro que $g(0) = 0$, así como que g es diferenciable con

$$Dg(x) = \text{Id} - Df(x) \quad \forall x \in \Omega$$

Como, por hipótesis, Df es continua en el origen, vemos claramente que Dg también lo es, con $Dg(0) = 0$, luego existe $\delta_1 > 0$ tal que $B(0, \delta_1) \subset \Omega$ y

$$\|Dg(x)\| \leq 1/2 \quad \forall x \in B(0, \delta_1)$$

Por otra parte, recordemos que la continuidad de Df en el origen también implica la de la función $x \mapsto \det Jf(x)$, siendo $\det Jf(0) = 1$, luego existe $\delta_2 > 0$ tal que $B(0, \delta_2) \subset \Omega$ y

$$\det Jf(x) \neq 0 \quad \forall x \in B(0, \delta_2)$$

Tomamos ahora $r = (1/3) \min\{\delta_1, \delta_2\}$ con lo cual tenemos:

$$x \in \mathbb{R}^N, \|x\| < 3r \implies x \in \Omega, \|Dg(x)\| \leq 1/2 \text{ y } \det Jf(x) \neq 0 \quad (6)$$

Como $B(0, 3r)$ es un subconjunto abierto y convexo de \mathbb{R}^N , un corolario de la desigualdad del valor medio nos asegura que g es lipschitziana en dicho abierto, más concretamente:

$$x, z \in B(x, 3r) \implies \|g(x) - g(z)\| \leq (1/2) \|x - z\| \quad (7)$$

En particular, tomando $z = 0$ tenemos

$$x \in B(x, 3r) \implies \|g(x)\| \leq (1/2) \|x\| \quad (8)$$

Llegamos al paso clave de la demostración, que consiste en usar el teorema del punto fijo de Banach, para probar lo siguiente:

(*) Para cada $y_0 \in \overline{B}(0, r)$ existe un único $x_0 \in \overline{B}(0, 2r)$ tal que $f(x_0) = y_0$. Además, si de hecho $\|y_0\| < r$, se tiene también $\|x_0\| < 2r$.

Fijado $y_0 \in \overline{B}(0, r)$, para $x \in \overline{B}(0, 2r)$ se tiene

$$f(x) = y_0 \iff g(x) = x - y_0 \iff g(x) + y_0 = x \quad (9)$$

luego buscamos un punto fijo de la función $x \mapsto g(x) + y_0$. Así pues, tomamos $E = \overline{B}(0, 2r)$, que es un espacio métrico completo, subconjunto cerrado de \mathbb{R}^N , y definimos $h(x) = g(x) + y_0$ para todo $x \in E$. Usando (8) tenemos

$$\|h(x)\| \leq \|g(x)\| + \|y_0\| \leq (1/2)\|x\| + \|y_0\| \leq 2r \quad \forall x \in E$$

luego $h(E) \subset E$. Además, h es contractiva, pues de (7) deducimos claramente que,

$$\|h(x) - h(z)\| = \|g(x) - g(z)\| \leq (1/2) \|x - z\| \quad \forall x, z \in E$$

Por el teorema del punto fijo, existe un único $x_0 \in E$ tal que $h(x_0) = x_0$ y, en vista de (9), x_0 es el único punto de $\overline{B}(0, 2r)$ tal que $f(x_0) = y_0$. Por último, si $\|y_0\| < 2r$, razonando como antes tenemos $\|x_0\| = \|h(x_0)\| \leq (1/2)\|x_0\| + \|y_0\| \leq r + \|y_0\| < 2r$. Queda así comprobada la afirmación (*), y el resto de la demostración se obtendrá ya sin dificultad.

Concretamente tomamos

$$U = B(0, 2r) \cap f^{-1}(B(0, r))$$

Como f es continua, el conjunto $f^{-1}(B(0, r))$ es abierto, luego U también es abierto, y es claro que $0 \in U \subset \Omega$. Comprobamos ahora todas las afirmaciones del teorema.

(i). Si $x, z \in U$ verifican que $f(x) = f(z)$, tomando $y_0 = f(x) \in B(0, r)$, de (*) deducimos que existe un único $x_0 \in B(0, 2r)$ tal que $f(x_0) = y_0$, pero $x, z \in B(0, 2r)$, luego $x = z = x_0$ y hemos probado que f es inyectiva en U .

(ii). Tomando $V = f(U)$, es claro que $V \subset B(0, r)$ y comprobamos enseguida que se da la igualdad, con lo que V es una bola abierta. Para cada $y_0 \in B(0, r)$, usando de nuevo (*) tenemos un $x_0 \in B(0, 2r)$ tal que $f(x_0) = y_0$, pero entonces $x_0 \in U$, luego $y_0 \in f(U)$ como queríamos.

(iii). Para $x \in U$ tenemos $\|x\| < 3r$ y (6) nos dice que $\det Jf(x) \neq 0$.

(iv). Usando la regla de diferenciación de la función inversa, y teniendo en cuenta (iii), basta probar que φ^{-1} es continua, y de hecho veremos que es lipschitziana.

Para $y_1, y_2 \in V$, sean $x_1 = \varphi^{-1}(y_1)$ y $x_2 = \varphi^{-1}(y_2)$, con lo que $x_1, x_2 \in U$, $f(x_1) = y_1$ y también $f(x_2) = y_2$. Por definición de g , tenemos $x_1 = y_1 + g(x_1)$ y $x_2 = y_2 + g(x_2)$. Por tanto, usando (7) obtenemos

$$\|x_1 - x_2\| \leq \|y_1 - y_2\| + \|g(x_1) - g(x_2)\| \leq \|y_1 - y_2\| + (1/2)\|x_1 - x_2\|$$

Deducimos claramente que

$$\|\varphi^{-1}(y_1) - \varphi^{-1}(y_2)\| = \|x_1 - x_2\| \leq 2\|y_1 - y_2\| \quad \forall y_1, y_2 \in V$$

Queda así probado el teorema cuando $a = f(a) = 0$ y $Df(a) = \text{Id}$. Completamos ahora la demostración, viendo que el caso general se deduce del que ya tenemos resuelto.

Sea $\Omega_0 = \{z \in \mathbb{R}^N : z + a \in \Omega\}$ que es un abierto de \mathbb{R}^N , como imagen inversa de Ω por una traslación, y verifica que $0 \in \Omega_0$. Como, por hipótesis, $\det Jf(a) \neq 0$, tenemos que $T = Df(a)$ es biyectiva y usaremos su inversa. Consideramos entonces la función $f_0 : \Omega_0 \rightarrow \mathbb{R}^N$ dada por

$$f_0(z) = T^{-1}(f(z+a) - f(a)) \quad \forall z \in \Omega_0 \quad (10)$$

que verifica $f_0(0) = 0$. La regla de la cadena nos dice que f_0 es diferenciable con

$$Df_0(z) = T^{-1} \circ Df(z+a) \quad \forall z \in \Omega_0 \quad (11)$$

y en particular $Df_0(0) = T^{-1} \circ T = \text{Id}$. También deducimos que Df_0 es continua en el origen, pues para $z \in U_0$, se tiene

$$\|Df_0(z) - Df_0(0)\| = \|T^{-1} \circ (Df(z+a) - T)\| \leq \|T^{-1}\| \|Df(z+a) - Df(a)\|$$

y basta tener en cuenta que Df es continua en a . En resumen, f_0 verifica las hipótesis del teorema, en el caso particular que ya hemos resuelto.

Por tanto, tenemos $0 \in U_0 = U_0^\circ \subset \Omega_0$, con f_0 inyectiva en U_0 , el conjunto $V_0 = f_0(U_0)$ es abierto, $\det Jf_0(z) \neq 0$ para todo $z \in U_0$ y $\varphi_0^{-1} \in D(V_0, \mathbb{R}^N)$ donde $\varphi_0 = f_0|_{U_0}$. Sólo queda hacer el camino de vuelta, traduciendo toda esta información en términos de f .

Para ello, sea $U = \{z + a : z \in U_0\}$ que claramente es un abierto de \mathbb{R}^N con $a \in U \subset \Omega$. A partir de (10) deducimos claramente que

$$f(x) = T(f_0(x-a)) + f(a) \quad \forall x \in \Omega \quad (12)$$

mientras que (11) se traduce en

$$Df(x) = T \circ Df_0(x-a) \quad \forall x \in \Omega \quad (13)$$

Comprobamos ya, de forma bastante rutinaria, que f tiene en U las propiedades requeridas.

(i). Si $f(x_1) = f(x_2)$ con $x_1, x_2 \in U$, usando (12) tenemos $T(f_0(x_1-a)) = T(f_0(x_2-a))$, pero T es biyectiva, luego $f_0(x_1-a) = f_0(x_2-a)$. Como $x_1-a, x_2-a \in U_0$ y f_0 es inyectiva en U_0 , concluimos que $x_1 = x_2$, luego f es inyectiva en U .

(ii). Veamos que $V = f(U)$ es abierto. Como V_0 es abierto y T es un homeomorfismo, $T(V_0)$ es abierto, y de (12) se deduce que $V = \{v \in \mathbb{R}^N : v - f(a) \in T(V_0)\}$, luego V es abierto.

(iii). Para $x \in U$ tenemos $x-a \in U_0$, luego $\det Jf_0(x-a) \neq 0$, así que $Df_0(x-a)$ es biyectiva. Como T también lo es, deducimos de (13) que $Df(x)$ es biyectiva, es decir, $\det Jf(x) \neq 0$.

(iv). Si $\varphi = f|_U$, veamos la relación entre φ^{-1} y φ_0^{-1} . Para $y \in V$ se tiene $y - f(a) \in T(V_0)$, luego $T^{-1}(y - f(a)) \in V_0$, lo que permite tomar $x = (\varphi_0^{-1} \circ T^{-1})(y - f(a)) + a \in U$, y usando (12) vemos claramente que $f(x) = y$. Esto prueba que

$$\varphi^{-1}(y) = (\varphi_0^{-1} \circ T^{-1})(y - f(a)) + a \quad \forall y \in V$$

donde vemos claramente que φ^{-1} es continua. Como φ es diferenciable y $D\varphi(x) = Df(x)$ es biyectiva para todo $x \in U$, la regla de diferenciación de la función inversa nos dice que φ^{-1} es diferenciable con $D\varphi^{-1}(f(x)) = Df(x)^{-1}$ para todo $x \in U$. ■

El teorema anterior se resume en la existencia y diferenciabilidad de la función φ^{-1} , a la que conviene dar un nombre. No es inversa de f , pues f puede no ser inyectiva, es la inversa de la restricción de f a un entorno abierto de a . Por ello se dice que φ^{-1} es una **inversa local** de f en el punto a . Nótese que φ^{-1} está definida en un entorno abierto de $f(a)$.

11.3. Coordenadas polares, cilíndricas y esféricas

Veamos ejemplos sencillos que muestran la utilidad del teorema anterior. Consideremos el abierto $\Omega = \mathbb{R}^+ \times \mathbb{R} \subset \mathbb{R}^2$ y la función $f : \Omega \rightarrow \mathbb{R}^2$ definida por

$$f(\rho, \theta) = (\rho \cos \theta, \rho \sin \theta) \quad \forall (\rho, \theta) \in \Omega$$

Es fácil ver que $f(\Omega) = \mathbb{R}^2 \setminus \{(0,0)\}$, pero f no es inyectiva. Para cada $(\rho, \theta) \in \Omega$, se dice que el punto $(x, y) = f(\rho, \theta)$ tiene **coordenadas polares** (ρ, θ) , pero no son únicas. De hecho, el *radio polar* ρ es único, pero el *ángulo polar* θ no lo es.

Se tiene $f \in C^1(\Omega, \mathbb{R}^2)$ con $\det Jf(\rho, \theta) = \rho \neq 0$ para todo $(\rho, \theta) \in \Omega$, luego existe una inversa local de f en cada punto de Ω . Esto significa que, en un entorno de cada punto del plano, excluido el origen, podemos definir sin ambigüedad coordenadas polares, y obtenerlas a partir de las cartesianas, mediante una función diferenciable.

Esto puede comprobarse directamente, pues f es una función sencilla. Dado $(\rho_0, \theta_0) \in \Omega$, es fácil encontrar un abierto concreto U , con $(\rho_0, \theta_0) \in U \subset \Omega$, tal que f es inyectiva en U , ver que $V = f(U)$ es abierto, e incluso, si $\varphi = f|_U$, calcular explícitamente φ^{-1} , ver que es diferenciable y calcular su diferencial en cada punto de V . Sin embargo, debemos pensar que el teorema seguiría siendo útil, aunque f fuese más complicada, y no pudiésemos encontrar explícitamente el abierto U y la función φ^{-1} , debiendo contentarnos con saber que existen. Por tanto, para mostrar esa utilidad mediante f , debemos extraer alguna consecuencia del teorema que no precise calcular explícitamente U y φ^{-1} , como puede ser la siguiente:

- Sea $g: \mathbb{R}^2 \setminus \{(0,0)\} \rightarrow \mathbb{R}$ un campo escalar y consideremos la función $h: \mathbb{R}^+ \times \mathbb{R} \rightarrow \mathbb{R}$ definida por:

$$h(\rho, \theta) = g(\rho \cos \theta, \rho \sin \theta) \quad \forall (\rho, \theta) \in \mathbb{R}^+ \times \mathbb{R}$$

Entonces g es diferenciable si, y sólo si, lo es h .

Nótese que $h = g \circ f$ donde f es la función con la que veníamos trabajando. La regla de la cadena nos da una implicación: h será diferenciable siempre que g lo sea. El recíproco sería igual de fácil si f fuese biyectiva y f^{-1} fuese diferenciable, pues escribiríamos $g = h \circ f^{-1}$ y volveríamos a usar la regla de la cadena. Pero f no es inyectiva, no tiene “inversa global”. Sin embargo, queremos comprobar una propiedad local, la diferenciabilidad de g , y para ello podemos usar una inversa local de f .

Fijado $(x_0, y_0) \in \mathbb{R}^2 \setminus \{(0,0)\}$, tomamos $(\rho_0, \theta_0) \in \Omega = \mathbb{R}^+ \times \mathbb{R}$ tal que $f(\rho_0, \theta_0) = (x_0, y_0)$. Entonces existe un abierto U de \mathbb{R}^2 , con $(\rho_0, \theta_0) \in U \subset \Omega$, tal que f es inyectiva en U , el conjunto $V = f(U)$ es abierto y φ^{-1} es diferenciable, donde $\varphi = f|_U$. Para $(x, y) \in V$ se tiene que $\varphi^{-1}(x, y) \in U$ luego $f(\varphi^{-1}(x, y)) = \varphi(\varphi^{-1}(x, y)) = (x, y)$, de donde deducimos que

$$g(x, y) = g(f(\varphi^{-1}(x, y))) = h(\varphi^{-1}(x, y)) \quad \forall (x, y) \in V$$

Así pues, tenemos $g|_V = h \circ \varphi^{-1}$. Como h y φ^{-1} son diferenciables, la regla de la cadena nos dice que $g|_V$ es diferenciable. Por el carácter local de la diferenciabilidad, g es diferenciable en (x_0, y_0) como queríamos demostrar. ■

En \mathbb{R}^3 podemos definir coordenadas muy similares a las polares del plano, pero ahora hay que excluir, no sólo el origen, sino uno de los ejes de coordenadas. Para concretar excluimos el tercero, y trabajamos con el abierto $G = \mathbb{R}^3 \setminus \{(0,0,z) : z \in \mathbb{R}\}$. Se tiene entonces $G = f(\Omega)$, donde $\Omega = \mathbb{R}^+ \times \mathbb{R}^2$ y $f: \Omega \rightarrow \mathbb{R}^3$ viene dada por

$$f(\rho, \theta, z) = (\rho \cos \theta, \rho \sin \theta, z) \quad \forall (\rho, \theta, z) \in \Omega$$

con lo que f no es inyectiva. Para cada $(\rho, \theta, z) \in \Omega$, decimos que el punto $f(\rho, \theta, z) \in G$ tiene **coordenadas cilíndricas** (ρ, θ, z) , que no son únicas.

Es claro que $f \in C^1(\Omega, \mathbb{R}^3)$ con $\det Jf(\rho, \theta) = \rho \neq 0$ para todo $(\rho, \theta) \in \Omega$. En un entorno de cada punto de G , podemos definir coordenadas cilíndricas, como función diferenciable de las cartesianas. Razonando como en el plano, vemos que un campo escalar definido en G es diferenciable en coordenadas cartesianas si, y sólo si, lo es en coordenadas cilíndricas.

Veamos otras coordenadas en G , más útiles que las cilíndricas. Tenemos ahora $G = f(\Omega)$, donde tomamos $\Omega = \{(r, \theta, \varphi) \in \mathbb{R}^3 : r > 0, |\varphi| < \pi/2\}$ y $f : \Omega \rightarrow \mathbb{R}^3$ viene dada por

$$f(r, \theta, \varphi) = (r \cos \theta \cos \varphi, r \sin \theta \cos \varphi, r \sin \varphi) \quad \forall (r, \theta, \varphi) \in \Omega$$

Para cada $(r, \theta, \varphi) \in \Omega$, decimos que $f(r, \theta, \varphi)$ tiene **coordenadas esféricas** (r, θ, φ) , que no son únicas.

Es fácil ver que $f \in C^1(\Omega, \mathbb{R}^3)$, con $\det Jf(r, \theta, \varphi) = r^2 \cos \varphi \neq 0$ para todo $(r, \theta, \varphi) \in \Omega$. Todo lo dicho sobre las coordenadas cilíndricas es también válido para las esféricas.

11.4. Función inversa global

Como hicimos con la versión local, recordemos el teorema de la función inversa global para funciones reales de variable real:

- Sea $I \subset \mathbb{R}$ un intervalo no trivial y $f : I \rightarrow \mathbb{R}$ una función derivable en I con $f'(x) \neq 0$ para todo $x \in I$. Entonces f es inyectiva, $J = f(I)$ es un intervalo y f^{-1} es derivable en J con $(f^{-1})'(f(x)) = 1/f'(x)$ para todo $x \in I$.

A diferencia de lo que ocurrió con el teorema local, la generalización del resultado anterior que cabría esperar, no es cierta. Concretamente, sustituyendo como es natural el intervalo I por un abierto $\Omega \subset \mathbb{R}^N$, que se puede suponer conexo o incluso convexo, podríamos pensar que una función diferenciable $f : \Omega \rightarrow \mathbb{R}^N$ verificando que $\det Jf(x) \neq 0$ para todo $x \in \Omega$, tenga que ser inyectiva, pero esto es falso: basta considerar cualquiera de las funciones que acabamos de usar para estudiar distintos tipos de coordenadas. Todas ellas son inyectivas en un entorno de cada punto de un abierto convexo Ω , podríamos decir que son localmente inyectivas, pero ninguna es inyectiva en Ω . Debemos pues aceptar que, para funciones de varias variables, la versión global del teorema de la función inversa va a ser mucho más débil que en el caso de una variable, pues la inyectividad de la función no será parte de la tesis, sino que habrá de suponerse como hipótesis. Enunciamos ya el resultado global que podemos conseguir.

Teorema de la función inversa (global). Sea Ω un abierto de \mathbb{R}^N y $f \in C^1(\Omega, \mathbb{R}^N)$. Supongamos que f es inyectiva y que $\det Jf(x) \neq 0$ para todo $x \in \Omega$. Entonces $W = f(\Omega)$ es abierto y f^{-1} es diferenciable con $Df^{-1}(f(x)) = Df(x)^{-1}$ para todo $x \in \Omega$. De hecho se tiene que $f^{-1} \in C^1(W, \mathbb{R}^N)$.

Demostración. Sea $y \in W$ y $x = f^{-1}(y) \in \Omega$. Como Df es continua en x y $\det Jf(x) \neq 0$, podemos aplicar el teorema local, obteniendo un abierto U de \mathbb{R}^N , con $x \in U \subset \Omega$, verificando las propiedades que vamos a comentar.

En primer lugar, que f sea inyectiva en U no es nada nuevo, pues f es inyectiva en Ω . Pero también tenemos que $V = f(U)$ es abierto, con $y = f(x) \in f(U) = V \subset W$, luego $y \in W^\circ$. Antes de seguir con el razonamiento, como $y \in W$ era arbitrario, tenemos ya probado que W es abierto. De vuelta al punto $y \in W$ antes fijado, si $\varphi = f|_U$, está claro que $\varphi^{-1} = f^{-1}|_V$. Como el teorema local nos dice que φ^{-1} es diferenciable, por el carácter local de la diferenciabilidad, deducimos que f^{-1} es diferenciable en el punto y . Otra vez $y \in W$ era arbitrario, luego f^{-1} es diferenciable. La regla de derivación de la función inversa nos da la diferencial de f^{-1} en cada punto de W . Para ver que Df^{-1} es continua, comprobaremos previamente el siguiente resultado.

- Sea \mathcal{G} el subconjunto de $L(\mathbb{R}^N, \mathbb{R}^N)$ formado por las aplicaciones lineales biyectivas. Entonces, la aplicación $T \mapsto T^{-1}$, de \mathcal{G} en sí mismo, es continua.

Empezamos comprobando la continuidad de la función $\alpha : \mathcal{G} \rightarrow \mathbb{R}^+$ definida por

$$\alpha(T) = \frac{1}{\|T^{-1}\|} \quad \forall T \in \mathcal{G}$$

De hecho, para $S, T \in \mathcal{G}$ tenemos

$$\begin{aligned} |\alpha(T) - \alpha(S)| &= \left| \frac{\|S^{-1}\| - \|T^{-1}\|}{\|T^{-1}\| \|S^{-1}\|} \right| \leq \frac{\|S^{-1} - T^{-1}\|}{\|T^{-1}\| \|S^{-1}\|} = \frac{\|T^{-1} \circ (T - S) \circ S^{-1}\|}{\|T^{-1}\| \|S^{-1}\|} \\ &\leq \frac{\|T^{-1}\| \|T - S\| \|S^{-1}\|}{\|T^{-1}\| \|S^{-1}\|} = \|T - S\| \end{aligned}$$

Así pues, vemos que de hecho α es no expansiva, luego continua. Deducimos obviamente que la función $1/\alpha$, es decir, la función $T \mapsto \|T^{-1}\|$, de \mathcal{G} en \mathbb{R}^+ , también es continua.

Finalmente, dada $T \in \mathcal{G}$ y una sucesión $\{T_n\}$ de elementos de \mathcal{G} con $\{T_n\} \rightarrow T$, bastará comprobar que $\{T_n^{-1}\} \rightarrow T^{-1}$. En efecto, para todo $n \in \mathbb{N}$ se tiene

$$\|T_n^{-1} - T^{-1}\| = \|T^{-1} \circ (T - T_n) \circ T_n^{-1}\| \leq \|T^{-1}\| \|T - T_n\| \|T_n^{-1}\|$$

Como ya sabemos que $\{\|T_n^{-1}\|\} \rightarrow \|T^{-1}\|$, vemos que $\{\|T_n^{-1} - T^{-1}\|\} \rightarrow 0$. □

Para terminar la demostración del teorema de la función inversa global, teníamos ya

$$Df^{-1}(y) = [Df(f^{-1}(y))]^{-1} \quad \forall y \in W$$

y vemos que $Df^{-1} = J \circ Df \circ f^{-1}$, donde J es la función $T \mapsto T^{-1}$, de \mathcal{G} en sí mismo, cuya continuidad acabamos de comprobar. Como f^{-1} es continua, porque es diferenciable, y Df es continua por hipótesis, concluimos que Df^{-1} es continua, como composición de tres funciones continuas. Así pues, $f^{-1} \in C^1(W, \mathbb{R}^N)$ como queríamos demostrar. ■

Tema 12

Función implícita

Estudiamos ahora un tercer resultado fundamental del cálculo diferencial en varias variables, el *teorema de la función implícita*. Como motivación, explicaremos la forma en que el teorema de la función inversa resuelve “localmente” ciertos sistemas de ecuaciones, para después hacer un planteamiento más ambicioso, consistente en intentar resolver sistemas de ecuaciones mucho más generales. Conseguiremos este objetivo, también desde un punto de vista “local”. Por tanto el teorema de la función implícita es formalmente más general que el de la inversa. No obstante, veremos que la demostración del primero se consigue fácilmente a partir del segundo. Por tanto, en esencia ambos teoremas son equivalentes, pero el de la función implícita permite obtener aplicaciones relevantes de forma más directa. Está mejor preparado para usarlo en la práctica.

12.1. Planteamiento del problema

Puede decirse que el teorema de la función inversa resuelve localmente algunos sistemas de ecuaciones, en el sentido que vamos a explicar. Dada una función $f : A \rightarrow \mathbb{R}^N$, donde A es un subconjunto de \mathbb{R}^N , veamos la igualdad

$$f(x) = y \tag{1.a}$$

como una ecuación que involucra dos variables vectoriales $x \in A$ e $y \in \mathbb{R}^N$. Sus soluciones son todos los pares $(x, y) \in A \times \mathbb{R}^N$ que la verifican.

Si consideramos las componentes de x e y , así como las de la función f , entonces (1.a) se convierte en el sistema de ecuaciones

$$\begin{aligned} f_1(x_1, x_2, \dots, x_N) &= y_1 \\ f_2(x_1, x_2, \dots, x_N) &= y_2 \\ \dots \dots \dots \dots \dots & \\ f_N(x_1, x_2, \dots, x_N) &= y_N \end{aligned} \tag{1.b}$$

en el que intervienen $2N$ variables reales $x_1, x_2, \dots, x_N, y_1, \dots, y_N$, que sólo están sometidas a la restricción $(x_1, x_2, \dots, x_N) \in A$.

Volvamos a la ecuación (1.a), sin olvidar que equivale al sistema (1.b). Resolverla, con x como dato e y como incógnita, es obvio: las soluciones son todos los pares $(x, f(x))$ con $x \in A$. Lo podemos resumir en una igualdad entre conjuntos, que también es obvia:

$$\{(x, y) \in A \times \mathbb{R}^N : f(x) = y\} = \{(x, f(x)) : x \in A\}$$

Pero el problema no trivial va en sentido opuesto: resolver la ecuación (1.a), con y como dato y x como incógnita, es decir, expresar las soluciones en la forma $(g(y), y)$ para conveniente función g , o si se quiere, “despejar” x como función de y . Claro está que esto equivale a invertir la función f .

Cuando f es inyectiva, la respuesta es la función $g = f^{-1}$, definida en el conjunto $f(A)$. Lo podemos expresar como una igualdad de conjuntos análoga a la anterior:

$$\{(x, y) \in \mathbb{R}^N \times A : f(x) = y\} = \{(g(y), y) : y \in f(A)\} \quad (1.c)$$

Esta igualdad resuelve el problema *globalmente*, pues en ambos miembros aparecen *todas* las soluciones de (1.a), o si se quiere, g está definida en el conjunto más grande posible, es la inversa *global* de f .

En general f no tiene por qué ser inyectiva, pero el teorema de la función inversa permite resolver *localmente* el problema. Partiendo de una solución (a, b) de la ecuación (1.a), es decir, tomando $a \in A$ y $b = f(a)$, y con hipótesis adecuadas, el teorema nos da dos abiertos U y V , con $a \in U \subset A$ y $b \in V \subset f(A)$, tales que $f|_U$ es una biyección de U sobre V . Escribamos este resultado como una igualdad entre conjuntos, para compararlo con (1.c). Tomando $W = U \times V$ tenemos un abierto de $\mathbb{R}^N \times \mathbb{R}^N$ con $(a, b) \in W \subset A \times \mathbb{R}^N$ y, si $g = (f|_U)^{-1}$, se tiene:

$$\{(x, y) \in W : f(x) = y\} = \{(g(y), y) : y \in V\} \quad (1.d)$$

Obsérvese por qué esta igualdad sólo resuelve localmente nuestro problema. A diferencia de lo que ocurría en (1.c), g no está definida en todo el conjunto $f(A)$, sino sólo en V : un entorno de b , contenido en $f(A)$. Este carácter local también se refleja en el primer miembro de (1.d), donde no aparecen todas las soluciones de (1.a), sino sólo las que pertenecen a $W = U \times V$, un entorno de la solución de partida (a, b) .

Pues bien, hacemos ahora un planteamiento más ambicioso, sustituyendo la ecuación (1.a) por otra más general, que será de la forma

$$F(x, y) = 0 \quad (2.a)$$

donde x e y son variables vectoriales, en principio independientes, y queremos saber hasta qué punto (2.a) equivale a expresar una de ellas como función de la otra, pues ahora la asimetría del problema ha desaparecido. Aunque intercambiamos los papeles de las variables, parece más natural expresar y como función de x . Dicho intuitivamente, en la ecuación $F(x, y) = 0$, queremos “despejar” y como función de x . En el caso particular de la función inversa, era esencial que las variables x e y se movieran en espacios de la misma dimensión, $x, y \in \mathbb{R}^N$, pero ahora esta limitación está fuera de lugar, pues ya no se trata de invertir ninguna función. Para $x \in \mathbb{R}^N$ e $y \in \mathbb{R}^M$, tiene perfecto sentido que y se pueda expresar como función de x .

Por tanto, F estará definida en un abierto de $\mathbb{R}^N \times \mathbb{R}^M$, con valores, en principio, en \mathbb{R}^k con $k \in \mathbb{N}$, pero enseguida vemos que debe ser $k = M$. En efecto, no olvidemos que seguimos trabajando con sistemas de ecuaciones, así que (2.a) será la ecuación vectorial que resume un sistema de k ecuaciones con $N + M$ incógnitas y queremos despejar M de ellas en función de las otras N . Si queremos despejar M incógnitas, y no más de M , debemos tener M ecuaciones.

Tendremos por tanto $F : \Omega \rightarrow \mathbb{R}^M$ donde Ω es un abierto de $\mathbb{R}^N \times \mathbb{R}^M$. Escribiendo de nuevo $x = (x_1, x_2, \dots, x_N)$, $y = (y_1, y_2, \dots, y_M)$ y $F = (F_1, F_2, \dots, F_M)$, visualizamos el sistema con el que vamos a trabajar. Consta de M ecuaciones que involucran $N + M$ variables reales.

$$\begin{aligned} F_1(x_1, \dots, x_N, y_1, \dots, y_M) &= 0 \\ F_2(x_1, \dots, x_N, y_1, \dots, y_M) &= 0 \\ \dots \dots \dots \dots \dots \dots \dots \\ F_M(x_1, \dots, x_N, y_1, \dots, y_M) &= 0 \end{aligned} \tag{2.b}$$

Pretendemos ahora que la igualdad $F(x, y) = 0$ sea equivalente a una de la forma $y = \psi(x)$ para conveniente función ψ . Digamos que *resolver* la ecuación (2.a), o lo que es lo mismo, resolver el sistema (2.b), es tanto como encontrar la función ψ .

Claramente, este problema es mucho más general que el de la función inversa, luego no debemos esperar nada mejor que lo obtenido en ese caso particular. Por tanto, no aspiramos a resolver globalmente la ecuación (2.a), sino tan sólo localmente. Comparando con lo dicho para la ecuación (1.a), no buscamos una igualdad análoga a (1.c) sino a (1.d). Para ello debemos disponer de una solución de partida, es decir, de un par $(a, b) \in \Omega$ tal que $F(a, b) = 0$, cuya existencia no está ahora garantizada. Supondremos que tal solución existe, pues en otro caso, el conjunto de soluciones de la ecuación (2.a) es vacío y no hay nada que estudiar.

Nuestro objetivo es, por tanto, encontrar un abierto W de $\mathbb{R}^N \times \mathbb{R}^M$, con $(a, b) \in W \subset \Omega$, otro abierto U de \mathbb{R}^N y una función $\psi : U \rightarrow \mathbb{R}^M$ tales que

$$\{(x, y) \in W : F(x, y) = 0\} = \{(x, \psi(x)) : x \in U\} \tag{2.d}$$

Este es precisamente el contenido del teorema de la función implícita. Resuelve localmente la ecuación (2.a) mediante la igualdad (2.d) igual que el teorema de la función inversa resolvía localmente (1.a) mediante la igualdad (1.d).

Tendremos así multitud de soluciones de la ecuación (2.a), todas las de la forma $(x, \psi(x))$ con $x \in U$, que de hecho son todas las soluciones en el conjunto W , es decir, suficientemente próximas a la solución inicial (a, b) . Para $(x, y) \in W$ se tiene $F(x, y) = 0$ si, y sólo si, $y = \psi(x)$, luego podemos decir que la función ψ está “implícita” en la ecuación $F(x, y) = 0$, de ahí el nombre del teorema. Lo probaremos con ciertas hipótesis acerca de la diferenciabilidad de F , y obtendremos que ψ también es diferenciable. Las hipótesis son análogas a las del teorema de la función inversa, y la demostración consiste en aplicar dicho teorema a una función construida a partir de F , así que el nuevo teorema es consecuencia fácil del que ya conocemos, pero a su vez es más general, luego en realidad ambos teoremas son equivalentes.

Nótese finalmente que el problema de la existencia de una función implícita tiene interés incluso en el caso $N = M = 1$ y, aunque buscamos una función real de variable real ψ , el resultado se escapa del ámbito del cálculo en una variable, puesto que la función F está definida en un abierto de \mathbb{R}^2 .

12.2. Teorema de la función implícita

Vamos a probar exactamente el resultado que hemos anunciado:

Teorema. Sea Ω un abierto de $\mathbb{R}^N \times \mathbb{R}^M$, $F \in D(\Omega, \mathbb{R}^M)$ y $(a, b) \in \Omega$ tal que $F(a, b) = 0$. Consideremos el abierto $\Omega_a \subset \mathbb{R}^M$ y la función $F_a \in D(\Omega_a, \mathbb{R}^M)$ dados por

$$\Omega_a = \{y \in \mathbb{R}^M : (a, y) \in \Omega\} \quad \text{y} \quad F_a(y) = F(a, y) \quad \forall y \in \Omega_a$$

Supongamos que DF es continua en (a, b) y que $DF_a(b)$ es biyectiva. Entonces existen, un abierto W , con $(a, b) \in W \subset \Omega$, un abierto $U \subset \mathbb{R}^N$ y una función $\psi \in D(U, \mathbb{R}^M)$, tales que:

$$\{(x, y) \in W : F(x, y) = 0\} = \{(x, \psi(x)) : x \in U\} \quad (3)$$

Demostración. Consistirá simplemente en aplicar el teorema de la función inversa local a la función $H : \Omega \rightarrow \mathbb{R}^N \times \mathbb{R}^M$ definida por

$$H(x, y) = (x, F(x, y)) \quad \forall (x, y) \in \Omega$$

que claramente verifica $H(a, b) = (a, 0)$.

Empezamos observando que H es diferenciable, pues sus dos componentes lo son. Pero conviene calcular explícitamente la diferencial de H en cada punto de Ω .

Para ello, consideramos las proyecciones lineales naturales de $\mathbb{R}^N \times \mathbb{R}^M$ sobre \mathbb{R}^N y \mathbb{R}^M , que denotamos por π_1 y π_2 respectivamente, es decir, escribimos:

$$\pi_1(x, y) = x \quad \text{y} \quad \pi_2(x, y) = y \quad \forall (x, y) \in \mathbb{R}^N \times \mathbb{R}^M$$

Por otra parte J_1 y J_2 serán las inyecciones lineales de \mathbb{R}^N y \mathbb{R}^M en $\mathbb{R}^N \times \mathbb{R}^M$, dadas por

$$J_1(x) = (x, 0) \quad \forall x \in \mathbb{R}^N \quad \text{y} \quad J_2(y) = (0, y) \quad \forall y \in \mathbb{R}^M$$

De esta forma, para todo $(x, y) \in \Omega$, tenemos claramente

$$H(x, y) = (x, 0) + (0, F(x, y)) = J_1(x) + J_2(F(x, y)) = J_1(\pi_1(x, y)) + J_2(F(x, y))$$

lo que se resume escribiendo: $H = J_1 \circ (\pi_1|_{\Omega}) + J_2 \circ F$. Deducimos claramente que

$$DH(x, y) = J_1 \circ \pi_1 + J_2 \circ DF(x, y) \quad \forall (x, y) \in \Omega \quad (4)$$

Entonces, también para todo $(x, y) \in \Omega$, tenemos

$$\|DH(x, y) - DH(a, b)\| = \|J_2 \circ (DF(x, y) - DF(a, b))\| \leq \|J_2\| \|DF(x, y) - DF(a, b)\|$$

y como DF es continua en (a, b) , vemos que DH también lo es. Para aplicar el teorema de la función inversa local, sólo queda comprobar que $DH(a, b)$ es biyectiva, para lo cual usaremos la hipótesis sobre la función F_a .

Observamos que para todo $y \in \Omega_a$ se tiene $F_a(y) = F((a, 0) + (0, y)) = F(J_1(a) + J_2(y))$, y la regla de la cadena nos da

$$DF_a(y) = DF(a, y) \circ J_2 \quad \forall y \in \Omega_a, \quad \text{luego} \quad DF_a(b) = DF(a, b) \circ J_2 \quad (5)$$

Para comprobar que $DH(a, b)$ es biyectiva, como se trata de una aplicación lineal entre dos espacios vectoriales de la misma dimensión, bastará ver que es inyectiva. Suponemos por tanto que $DH(a, b)(u, v) = (0, 0)$ con $(u, v) \in \mathbb{R}^N \times \mathbb{R}^M$, para probar que $u = v = 0$. En efecto, usando (4) tenemos

$$(0, 0) = DH(a, b)(u, v) = (u, 0) + (0, DF(a, b)(u, v))$$

de donde deducimos, primero que $u = 0$, y entonces que

$$0 = DF(a, b)(0, v) = (DF(a, b) \circ J_2)(v) = (DF_a(b))(v)$$

donde, para la última igualdad, hemos usado (5). Como por hipótesis, $DF_a(b)$ es biyectiva, obtenemos $v = 0$, como queríamos.

El teorema de la función inversa nos da un abierto W de $\mathbb{R}^N \times \mathbb{R}^M$, con $(a, b) \in W \subset \Omega$, y un abierto $G = H(W)$ de \mathbb{R}^M , con $(a, 0) \in G$, tales que $H|_W$ es una biyección de W sobre G , cuya inversa es diferenciable en G . Dicha inversa es por tanto una biyección $K : G \rightarrow W$ que es diferenciable y verifica que $H(K(x, z)) = (x, z)$ para todo $(x, z) \in G$.

Tomando $U = J_1^{-1}(G) = \{x \in \mathbb{R}^N : (x, 0) \in G\}$, tenemos un abierto de \mathbb{R}^N tal que $a \in U$, y la última igualdad nos dice que

$$H(K(x, 0)) = (x, 0) \quad \forall x \in U \quad (6)$$

Consideremos ahora las dos componentes de la función $x \mapsto K(x, 0)$, de U en W , pues la segunda es la función $\psi : U \rightarrow \mathbb{R}^M$ que buscamos. Más concretamente, definimos

$$\varphi(x) = \pi_1(K(x, 0)) \quad \text{y} \quad \psi(x) = \pi_2(K(x, 0)) \quad \forall x \in U$$

Pero φ es fácil de calcular. Para $x \in U$, tenemos $(\varphi(x), \psi(x)) = K(x, 0) \in W$ y (6) nos da

$$(x, 0) = H(\varphi(x), \psi(x)) = (\varphi(x), F(\varphi(x), \psi(x)))$$

luego $\varphi(x) = x$ para todo $x \in U$. Deducimos que

$$(x, \psi(x)) \in W \quad \text{y} \quad F(x, \psi(x)) = 0 \quad \forall x \in U \quad (7)$$

Claramente ψ es diferenciable, pues basta observar que $\psi = \pi_2 \circ K \circ (J_1|_U)$. Sólo nos queda comprobar que W , U y ψ verifican la igualdad (3).

Una inclusión la tenemos en (7), pues si $x \in U$ e $y = \psi(x)$, vemos en (7) que $(x, y) \in W$ y $F(x, y) = 0$.

Recíprocamente, si $(x, y) \in W$ y $F(x, y) = 0$, tenemos que $(x, 0) = H(x, y) \in G$, luego $x \in U$. Además, también sabemos que $K(x, 0) \in W$ y $H(K(x, 0)) = (x, 0)$, pero H es inyectiva en W , luego $(x, y) = K(x, 0) = (x, \psi(x))$, de donde $y = \psi(x)$ como queríamos demostrar. ■

Merece la pena repasar brevemente la demostración anterior para entender mejor la idea clave que en ella hemos usado. Al considerar la función H , lo que hemos hecho es modificar la ecuación (2.a) y añadirle otra cuya solución es obvia, para conseguir una ecuación del mismo tipo que (1.a), que involucra dos variables, $(x, y) \in \Omega$ y $(u, v) \in \mathbb{R}^N \times \mathbb{R}^M$. Concretamente, en vez de la ecuación $F(x, y) = 0$, hemos considerado la ecuación más general $F(x, y) = v$ junto con la igualdad $x = u$, que se engloban en la ecuación

$$H(x, y) = (u, v) \quad (8.a)$$

que es del mismo tipo que (1.a), sólo que sus dos variables se mueven en \mathbb{R}^{N+M} en lugar de hacerlo en \mathbb{R}^N .

Equivalentemente, hemos generalizado y agrandado el sistema de ecuaciones (2.a), para considerar el sistema

$$\begin{aligned} x_1 &= u_1 \\ x_2 &= u_2 \\ &\dots \\ x_N &= u_N \\ F_1(x_1, \dots, x_N, y_1, \dots, y_M) &= v_1 \\ F_2(x_1, \dots, x_N, y_1, \dots, y_M) &= v_2 \\ &\dots \dots \dots \dots \dots \dots \\ F_M(x_1, \dots, x_N, y_1, \dots, y_M) &= v_M \end{aligned} \quad (8.b)$$

que es del mismo tipo que (1.b), sólo que con $N + M$ ecuaciones, que involucran $2(N + M)$ variables reales.

Como ya se ha dicho, la demostración ha consistido en usar el teorema de la función inversa para resolver localmente la ecuación (8.a) o el sistema (8.b), es decir, para invertir localmente la función H . Las hipótesis sobre F nos han permitido ver que H verifica las condiciones que requiere dicho teorema, entre las que resaltaremos una. Si observamos las derivadas parciales de las componentes de H vemos claramente la razón por la que, de ser $DF_a(b)$ biyectiva, hemos podido comprobar que $DH(a, b)$ también lo es. De hecho, es fácil ver que los determinantes de las matrices jacobianas $JF_a(b)$ y $JH(a, b)$ coinciden.

Pues bien, como $H(a, b) = (a, 0)$, el teorema de la función inversa nos ha dado los dos abiertos W y $G = H(W)$ en $\mathbb{R}^N \times \mathbb{R}^M$, con $(a, b) \in W$ y $(a, 0) \in G$, y la función $K : G \rightarrow W$ que es una inversa local de H . Por tanto, para $(x, y) \in W$ y $(u, v) \in \mathbb{R}^N \times \mathbb{R}^M$ se tiene

$$x = u, \quad F(x, y) = v \iff H(x, y) = (u, v) \iff (u, v) \in G, \quad (x, y) = K(u, v)$$

Por una parte, esto sugiere olvidarnos de la variable u , sustituyéndola por x . Por otra, recordemos que la igualdad $F(x, y) = v$ sólo nos interesa para $v = 0$. Esto explica que hayamos usado el abierto $U = \{x \in \mathbb{R}^N : (x, 0) \in G\}$. Siempre para $(x, y) \in W$, obtenemos que

$$F(x, y) = 0 \iff H(x, y) = (x, 0) \iff x \in U, \quad (x, y) = K(x, 0)$$

Finalmente, esto indica que la primera componente de la función $x \mapsto K(x, 0)$ debe ser la identidad en U , mientras que la segunda es la función $\psi : U \rightarrow \mathbb{R}^M$ que buscábamos.

12.3. Algunas observaciones adicionales

Manteniendo la notación del teorema anterior, conviene hacer algunos comentarios sobre la forma de aplicarlo en la práctica y la nomenclatura que suele usarse.

Como $F(a, b) = 0$, de (3) deducimos que $a \in U$ y que $\psi(a) = b$. Veamos además que, fijados los abiertos W y U , la función ψ que verifica (3) es *única*. En efecto si $\psi_1 : U \rightarrow \mathbb{R}^N$ también verifica (3), para cada $x \in U$ tomamos $y = \psi_1(x)$ y de (3) deducimos que $(x, y) \in W$ con $F(x, y) = 0$, pero como ψ también verifica (3), tenemos $\psi(x) = y = \psi_1(x)$.

Así pues, la función ψ está definida en un entorno de a , verifica que $\psi(a) = b$ y está determinada por la igualdad (3). Para $(x, y) \in W$ se tiene que $F(x, y) = 0$ si, y sólo si, $y = \psi(x)$, luego podemos entender que la función ψ está “implícita” en la ecuación $F(x, y) = 0$. Por eso, se dice que *la ecuación (2.a) define a la variable y como función implícita de x , en un entorno del punto a , con $y = b$ para $x = a$* . Naturalmente se está hablando de la función ψ , pero entendida como relación entre dos variables, igual que hemos hecho otras veces.

Es importante especificar que $y = b$ para $x = a$, porque puede existir $c \in \mathbb{R}^M$ con $c \neq b$ tal que $F(a, c) = 0$. Entonces, si podemos aplicar el teorema con la solución de partida (a, c) en vez de (a, b) , expresaremos y como función implícita de x , también en un entorno de a , pero con $y = c$ para $x = a$, una función implícita distinta de la que teníamos antes.

En la práctica, se suele aludir a las variables reales x_1, \dots, x_N e y_1, \dots, y_N que aparecen en el sistema (2.b) que son las componentes de x e y . Si $b = (b_1, \dots, b_M)$, se dice entonces que *el sistema (2.b) define a y_1, y_2, \dots, y_M como funciones implícitas de x_1, x_2, \dots, x_N en un entorno del punto (a_1, a_2, \dots, a_N) , con $(y_1, \dots, y_M) = (b_1, \dots, b_M)$ para $(x_1, \dots, x_N) = (a_1, \dots, a_N)$* . Incluso, como hemos hecho otras veces, dichas funciones implícitas se denotan con el mismo nombre de las variables, escribiendo $y_j = y_j(x_1, \dots, x_N)$ para todo $j \in \Delta_M$.

Conviene comentar brevemente las hipótesis del teorema anterior. Que F sea diferenciable y que DF sea continua, no ya en el punto (a, b) , sino incluso en Ω , son hipótesis muy poco restrictivas. Se trata simplemente de que la ecuación $F(x, y) = 0$ sea manejable con técnicas de cálculo diferencial. En la práctica estas dos hipótesis se suelen comprobar con un simple vistazo a la función F .

Nótese por ejemplo que, si las componentes de F son funciones polinómicas, estas hipótesis se cumplen obviamente y, aún en este caso tan particular, el sistema de ecuaciones (2.b) puede ser extraordinariamente complicado, por no hablar de lo que ocurre si las componentes de F involucran funciones trascendentes como la exponencial o las trigonométricas.

La última hipótesis, que $DF_a(b)$ sea biyectiva, parece más rebuscada, pero es muy natural y también muy fácil de comprobar. Como $DF_a(b)$ es biyectiva si, y sólo si, su determinante jacobiano no se anula, veamos cual es ese determinante.

Se comprueba fácilmente que

$$\frac{\partial F_a}{\partial y_j}(b) = \frac{\partial F}{\partial y_j}(a, b) \quad \forall j \in \Delta_M$$

sin más que escribir la definición de las derivadas parciales de ambos campos vectoriales, pues ambas definiciones son idénticas.

Así pues, las columnas de la matriz $JF_a(b)$ son las M últimas columnas de $JF(a, b)$. Por tanto, basta considerar la matriz $JF(a, b) \in \mathcal{M}_{M \times (N+M)}$ y comprobar que el determinante de la submatriz cuadrada formada por sus últimas M columnas no se anula.

Esto sugiere usar la siguiente notación, para la matriz jacobiana $JF_a(b)$ y su determinante, que es muy cómoda en la práctica. Si $F = (F_1, \dots, F_M)$ son las componentes de F , escribimos:

$$JF_a(b) = \frac{\partial(F_1, \dots, F_M)}{\partial(y_1, \dots, y_M)}(a, b) \quad \text{y} \quad \det JF_a(b) = \det \left(\frac{\partial(F_1, \dots, F_M)}{\partial(y_1, \dots, y_M)}(a, b) \right)$$

Esta notación indica claramente que nos referimos a la matriz cuadrada y su determinante, que tienen como j -ésima columna al vector derivada parcial de F con respecto a la variable y_j en el punto (a, b) , para todo $j \in \Delta_M$.

Pensemos además que en la práctica, lo que pretendemos es expresar M variables, que no tienen por qué ser las M últimas, ni llamarse y_1, \dots, y_M , como funciones implícitas de las restantes. La notación anterior indica claramente las variables que pretendemos despejar.

Observemos también que, la auténtica hipótesis para poder aplicar el teorema de la función implícita es que la matriz $JF(a, b)$ tenga rango M , pues entonces habrá M columnas que forman una submatriz cuadrada con determinante no nulo y esas columnas nos indican las variables que podemos expresar como funciones implícitas de las restantes.

Se comprende ahora el papel que juega esta hipótesis, por analogía con el caso en que F es lineal y $DF(a, b) = F$. Dicho intuitivamente, que $DF(a, b)$ tenga rango menor que M , significa que alguna de las ecuaciones de nuestro sistema es consecuencia de las restantes, al menos en un entorno del punto (a, b) . Digamos que el rango de la matriz $DF(a, b)$ es el número de ecuaciones verdaderamente independientes, y esto explica que sea también el máximo número de variables que el teorema anterior nos permite expresar como funciones implícitas de las restantes. Por poner un ejemplo muy obvio, si añadimos una ecuación al sistema repitiendo una de las que ya teníamos, el sistema pasará a tener $M + 1$ ecuaciones, pero eso no nos va a permitir despejar más variables.

Finalmente, merece la pena destacar el caso particular $N = M = 1$ del teorema anterior:

- Sean Ω un abierto de \mathbb{R}^2 , $F \in D(\Omega)$ y $(a, b) \in \Omega$ tal que $F(a, b) = 0$. Supongamos que las dos derivadas parciales de F son continuas en el punto (a, b) y que $\frac{\partial F}{\partial y}(a, b) \neq 0$. Entonces existen un abierto W de \mathbb{R}^2 , con $(a, b) \in W \subset \Omega$, un abierto $U \subset \mathbb{R}$ y una función derivable $\psi : U \rightarrow \mathbb{R}$, tales que:

$$\{(x, y) \in W : F(x, y) = 0\} = \{(x, \psi(x)) : x \in U\}$$