Cálculo Diferencial multivariable (resumen)

Luis Navas

Nociones básicas de Cálculo Diferencial en $\mathbb{R}^n$. Como requisito previo para leer estos apuntes conviene conocer los conceptos básicos del Álgebra Lineal y la Geometría.

1 Geometría y Álgebra

En el Cálculo Diferencial multivariable se estudian las funciones de un número finito pero arbitrario de variables reales. Es decir, se establece un número n de variables iniciales o «de entrada» $x_1, x_2, \ldots, x_n$ y otro número m de variables «de salida» que dependen de éstas, $y_1, y_2, \ldots, y_m$. Esta dependencia se simboliza por

$$(x_1, x_2, \ldots, x_n) \mapsto (y_1, y_2, \ldots, y_m).$$

Equivale a dar m funciones $f_1, f_2, \ldots, f_m$ de las variables $x_1, x_2, \ldots, x_n$, que describen la relación entre las variables:

$$y_k = f_k(x_1, x_2, \ldots, x_n), \quad k = 1, 2, \ldots, m. \tag{1.1}$$

Se emplea entonces la notación de **n-tuplas** para escribir, de manera abreviada,

$$x = (x_1, x_2, \ldots, x_n), \quad y = (y_1, y_2, \ldots, y_m).$$

Si escribimos todas las relaciones entre las variables como funciones, queda

$$(y_1, y_2, \ldots, y_m) = \big(f_1(x_1, x_2, \ldots, x_n), f_2(x_1, x_2, \ldots, x_n), \ldots, f_m(x_1, x_2, \ldots, x_n)\big). \tag{1.2}$$

Como esto es largo de escribir, se abrevia

$$y = F(x), \quad F = (f_1, f_2, \ldots, f_m), \quad y_k = f_k(x), \quad k = 1, 2, \ldots, m. \tag{1.3}$$

De este modo, F es una función que transforma n variables reales en m variables reales:

$$F : \mathbb{R}^n \to \mathbb{R}^m.$$

A las funciones f_k las llamamos **componentes** de F.

Cuando hay una sola componente, es decir, cuando $m = 1$ y por tanto $F : \mathbb{R}^n \to \mathbb{R}$, solemos escribir la función en minúsculas, f en vez de F, y decimos para distinguir este caso que es un **campo escalar**. En el caso general llamamos a una $F : \mathbb{R}^n \to \mathbb{R}^m$ un **campo vectorial**.

Nota 1.1. ▷ En general, las funciones de n variables no estarán definidas para cualquier n-tupla $x = (x_1, x_2, \dots, x_n)$ sino sólo para x en algún subconjunto de $\mathbb{R}^n$, su **dominio**. ◁

Ejemplo 1.2. ▷ Si (r, θ) son las variables iniciales y (x, y) las finales,

$$(x, y) = (r\cos\theta, r\,\mathrm{sen}\,\theta)$$

describe una función $F : \mathbb{R}^2 \to \mathbb{R}^2$ con componentes dadas por

$$x = f_1(r, \theta) = r\cos\theta, \qquad y = f_2(r, \theta) = r\,\mathrm{sen}\,\theta.$$

Otros ejemplos son

$$(u, v) \mapsto \big(\cos u\,\mathrm{sen}\,v, \mathrm{sen}\,u\,\mathrm{sen}\,v, \cos v\big) : \mathbb{R}^2 \to \mathbb{R}^3,$$

$$t \mapsto \big(\cos t, \mathrm{sen}\,t, t\big) : \mathbb{R} \to \mathbb{R}^3.$$

◁

Las **funciones coordenadas** sobre $\mathbb{R}^n$ son los campos escalares

$$\pi_k : \mathbb{R}^n \to \mathbb{R}, \quad \pi_k(x_1, x_2, \dots, x_n) = x_k, \quad k = 1, 2, \dots, n.$$

La **función identidad** o **vector de posición** ó **radio vector** es el campo vectorial cuyas componentes son las coordenadas, es decir,

$$\vec{\mathbf{r}}(x) = x$$

para todo $x \in \mathbb{R}^n$.

En $\mathbb{R}^n$ tenemos el **producto interior** ó **métrica euclídea**

$$x \cdot y \stackrel{\text{def}}{=} x_1 y_1 + x_2 y_2 + \cdots + x_n y_n \tag{1.4}$$

que tiene las propiedades siguientes:

1. **(positividad)** $x \cdot x \geqslant 0$, con igualdad si y sólo si $x = 0$.
2. **(simetría)** $a \cdot b = b \cdot a$.
3. **(bilinealidad)** Se cumplen las leyes distributivas

$$(a + b) \cdot c = a \cdot c + b \cdot c, \qquad a \cdot (b + c) = a \cdot b + a \cdot c$$

y para una constante $k \in \mathbb{R}$,

$$k(a \cdot b) = (ka) \cdot b = a \cdot (kb).$$

En particular de esto se deduce que $a \cdot 0 = 0 \cdot a = 0$ (en el producto 0 es el **vector** nulo y en el resultado es el **escalar** cero).

Estas propiedades implican que la métrica «no tiene núcleo», es decir,

$$a \cdot b = a \cdot c \quad \forall\, a \in \mathbb{R}^n \iff b = c. \tag{1.5}$$

La métrica sirve para definir la **longitud** de una n-tupla mediante

$$\|x\| \stackrel{\text{def}}{=} \sqrt{x \cdot x} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}. \tag{1.6}$$

La longitud, también llamada **norma euclídea**, satisface

1. $\|x\| \geqslant 0$, con igualdad si y sólo si $x = 0$ (la n-tupla $(0, 0, \ldots, 0)$) **(positividad)**.
2. $\|kx\| = |k|\|x\|$ si $k \in \mathbb{R}$ **(homogeneidad)**.
3. $\|x + y\| \leqslant \|x\| + \|y\|$ (**desigualdad triangular**).

La **distancia (euclídea)** entre dos n-tuplas está dada por

$$\Delta(x, y) = \|x - y\| = \sqrt{\sum_{k=1}^{n} (x_k - y_k)^2}$$

y satisface

1. $\Delta(x, y) \geqslant 0$, con igualdad si y sólo si $x = y$ **(positividad)**.
2. $\Delta(y, x) = \Delta(x, y)$ **(simetría)**.
3. $\Delta(x, y) \leqslant \Delta(x, z) + \Delta(z, y)$ **(desigualdad triangular)**.

Se tiene la **desigualdad CBS** (Cauchy-Bunyakovski-Schwarz)

$$(x \cdot y)^2 \leqslant \|x\|^2 \|y\|^2 \tag{1.7}$$

con igualdad si y sólo si x, y son linealmente dependientes, o sea, si $y = \lambda x$ para algún $\lambda \in \mathbb{R}$. La desigualdad CBS admite las formulaciones equivalentes

$$|x \cdot y| \leqslant \|x\| \|y\| \tag{1.8}$$

con la misma condición de igualdad, y también

$$x \cdot y \leqslant \|x\| \|y\| \tag{1.9}$$

con igualdad si y sólo si alguno de x, y es nulo ó $y = \lambda x$ con $\lambda > 0$.

La desigualdad CBS permite definir el **ángulo** entre n-tuplas mediante

$$\measuredangle(x, y) = \arccos \frac{x \cdot y}{\|x\| \|y\|} \in [0, \pi]. \tag{1.10}$$

Por definición se mide el ángulo θ en **radianes**, eligiendo el que es menor que la mitad de una circunferencia. El arco mayor es el ángulo complementario, es decir $2\pi - \theta$.

La **ortogonalidad** de dos n-tuplas corresponde a la relación algebraica

$$x \perp y \iff x \cdot y = 0. \tag{1.11}$$

La **base estándar** de $\mathbb{R}^n$ es la formada por los vectores

$$\mathbb{e}_1 = (1, 0, \ldots, 0), \quad \mathbb{e}_2 = (0, 1, 0, \ldots, 0), \quad \ldots \quad \mathbb{e}_n = (0, 0, \ldots, 1).$$

Es una **base ortonormal**, es decir,

$$\mathbb{e}_i \cdot \mathbb{e}_j = \delta_{ij} \stackrel{\text{def}}{=} \begin{cases} 1 & i = j, \\ 0 & i \neq j. \end{cases}$$

2 Límites y continuidad

La definición de límite para funciones $F : \mathbb{R}^n \to \mathbb{R}^m$ es una extensión de la definición de límite en el caso $n = m = 1$ de funciones reales de una variable. Basta con cambiar el valor absoluto por la longitud. Por ejemplo,

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

significa que dado $\epsilon > 0$ existe $\delta > 0$ tal que

$$0 < \|x - a\| < \delta \implies \|F(x) - b\| < \epsilon.$$

Dicho esto, no deberían extrañar las siguientes **propiedades básicas de los límites**.

Proposición 2.1 (Propiedades de los límites). *Suponiendo que existen los límites de los campos vectoriales* $F, G : \mathbb{R}^n \to \mathbb{R}^m$ *y de los campos escalares* $f, g : \mathbb{R}^n \to \mathbb{R}$ *en el punto* $a \in \mathbb{R}^n$*, entonces al operar sobre ellos tenemos:*

1. *Si* $f \leqslant g$*, entonces* $\lim_{x \to a} f(x) \leqslant \lim_{x \to a} g(x)$*.*
2. $\lim_{x \to a} kF(x) = k \lim_{x \to a} F(x)$ *para* $k \in \mathbb{R}$*.*
3. $\lim_{x \to a}(F(x) + G(x)) = \lim_{x \to a} F(x) + \lim_{x \to a} G(x)$*.*
4. $\lim_{x \to a} f(x)G(x) = \left(\lim_{x \to a} f(x)\right)\left(\lim_{x \to a} G(x)\right)$*.*
5. $\lim_{x \to a} F(x) \cdot G(x) = \lim_{x \to a} F(x) \cdot \lim_{x \to a} G(x)$*.*
6. $\lim_{x \to a} \dfrac{F(x)}{g(x)} = \dfrac{\lim_{x \to a} F(x)}{\lim_{x \to a} g(x)}$ *si* $\lim_{x \to a} g(x) \neq 0$*.*

Estas propiedades se resumen diciendo que los límites **conmutan** *con las operaciones aritméticas y con la relación de orden.*

Las demostraciones de todo lo que estamos comentando aquí son completamente análogas a las demostraciones en el caso $n = m = 1$. Se pueden buscar en un libro de Cálculo en 1 variable y «cambiando lo que haya que cambiar» (en general basta con cambiar el valor absoluto por la norma) se obtiene la demostración en el caso multivariable.

En general no están definidos ni el **producto** ni el **cociente** de dos vectores en $\mathbb{R}^m$, pero sí el producto de un vector por un escalar. En los casos donde sí tenga sentido «multiplicar vectores» ó «dividir vectores» tendremos la propiedad análoga de que el límite conmuta con el producto y el cociente. Esto ocurre, por ejemplo, en el conjunto $\mathbb{C}$ de números complejos visto como el plano $\mathbb{R}^2$. También en $\mathbb{R}^3$ con el producto vectorial $\vec{v} \times \vec{w}$.

Análogamente, $F : \mathbb{R}^n \to \mathbb{R}^m$ es **continua** en $a \in \mathbb{R}^n$ si

$$\lim_{x \to a} F(x) = F(a).$$

Se tiene, con las demostraciones análogas al caso de funciones de una variable, el teorema de **«conservación de la continuidad»**

Teorema 2.2 (Conservación de la continuidad). *Las sumas, múltiplos, productos, cocientes por campos escalares que no se anulan y composiciones de funciones continuas son funciones continuas. Aquí «producto» se refiere al producto de un campo escalar por uno vectorial o al producto interior de dos campos vectoriales.*

Es fácil comprobar directamente con la definición, que las funciones constantes y las funciones coordenadas son continuas. Por tanto los **polinomios multivariables** (combinaciones lineales de productos de coordenadas) como

$$2x_1^2 - 6x_2x_3^6 + x_1x_2x_7$$

son continuos en todo $\mathbb{R}^n$ y las **funciones racionales multivariable** (cocientes de polinomios multivariables), como

$$\frac{xy^2}{1+x+y}, \quad \frac{x+4y+8z}{1-x^2-y^4}, \quad \frac{2x_1^2 - 6x_2x_3^6 + x_1x_2x_7}{x_1 - x_2 + 5x_3x_4}$$

son continuas en los puntos donde no se anula el denominador.

En general, la conservación de la continuidad nos permite ver inmediatamente a partir de una fórmula si la función que describe es continua. Por ejemplo

$$\frac{\operatorname{sen}(\log(1+x^2y^2))e^{xyz} - x^3y^4z^2 + \cos(\cos(xyz))}{(1-x)(1-y)(1-z)}$$

$$\operatorname{tg}(xyz) + \frac{\sin\left(\sqrt{1-x^2-y^2-z^2}\right)}{\sqrt{1-x^2-y^2-z^2}} + xy \operatorname{arc\,tg}(1+e^y)$$

son continuas en sus dominios de definición (hallar este dominio puede ser lo más complicado).

Sólo para funciones que no reconocemos como combinaciones de otras cuya continuidad ya conocemos, tendremos que recurrir a la definición.

Respecto a composiciones, en general, pero no siempre

$$\lim_{x \to a} F(x) = b, \qquad \lim_{y \to b} G(y) = G(b) \implies \lim_{x \to a} G(F(x)) = G(b). \tag{2.1}$$

Basta con que G sea continua en b, ó que $F(x) \neq b$, por ejemplo.

Una propiedad que a menudo simplifica el entender los límites y la continuidad es que «los límites se pueden hacer por componentes»

Proposición 2.3 (Límites y continuidad y componentes). *Sea*

$$F = (f_1, f_2, \ldots, f_m) : \mathbb{R}^n \to \mathbb{R}^m.$$

Entonces

$$\lim_{x \to a} F(x) = b \iff \lim_{x \to a} f_k(x) = b_k, \quad k = 1, 2, \ldots, m. \tag{2.2}$$

En particular, $F : \mathbb{R}^n \to \mathbb{R}^m$ **es continua si y sólo si lo es cada componente** f_k**.**

▪ *Demostración:* Para la continuidad, por ejemplo, si F es continua entonces $f_k = \pi_k \circ F$ es continua al ser composición de continuas. En sentido inverso, si cada f_k es continua, la combinación $F = \sum_{k=1}^n f_k \mathbb{e}_k$ es continua. ■

Las nociones de límite y continuidad se pueden expresar en el lenguaje de **entornos**, que lleva a estudiar la **Topología General**.

Definición 2.1 *(Bolas y esferas).*

- *La* **bola abierta** *de centro* $c \in \mathbb{R}^n$ *y radio* $r > 0$ *es el conjunto*

$$\mathbb{B}(c, r) \stackrel{\text{def}}{=} \left(x \in \mathbb{R}^n : \|x - c\| < r\right).$$

- *La* **bola cerrada** *de centro* $c \in \mathbb{R}^n$ *y radio* $r > 0$ *es el conjunto*

$$\overline{\mathbb{B}}(c, r) \stackrel{\text{def}}{=} \left(x \in \mathbb{R}^n : \|x - c\| \leqslant r\right).$$

- *La* **esfera** *de centro* $c \in \mathbb{R}^n$ *y radio* $r > 0$ *es el conjunto*

$$\mathbb{S}(c, r) \stackrel{\text{def}}{=} \left(x \in \mathbb{R}^n : \|x - c\| = r\right).$$

- *Un subconjunto* $\mathcal{A} \subseteq \mathbb{R}^n$ *es* **abierto** *si cada punto* $c \in \mathcal{A}$ *es el centro de alguna bola contenida en* $\mathcal{A}$*, y por tanto de toda bola de radio suficientemente pequeño; es decir, si* $\mathbb{B}(c, r) \subseteq \mathcal{A}$ *para* $r \approx 0^+$.
- *Un* **entorno** *de un punto es cualquier conjunto abierto que lo contiene.*
- *Un subconjunto* A *es* **cerrado** *si su* **complementario** A^c *es abierto.*

Nota 2.1. ▷ Puede verificarse como ejercicio que una bola abierta es un abierto. Primero, debe entenderse por qué esto **no** es una tautología. Por tanto, como entorno de un punto siempre puede suponerse sin pérdida de generalidad una bola abierta centrada en él. ◁

Si usamos la notación

$$x \approx a$$

como sinónimo de la frase «para x en algún entorno de a», y como sentido riguroso de las frases «para x cerca de a», «para x aproximadamente igual á a», etc. y la notación

$$x \approx a^*$$

como lo mismo pero con la condición añadida que $x \neq a$, entonces

$$x \approx a \implies F(x) \approx b$$

es otra manera de expresar que $\lim_{x \to a} F(x) = b$, pues si lo detallamos, dice que dado un entorno prefijado de b, por ejemplo, la bola $\mathbb{B}(b, \epsilon)$, existe un entorno de a, por ejemplo, la bola $\mathbb{B}(a, \delta)$, tal que

$$x \in \mathbb{B}(a, \delta), x \neq a \implies F(x) \in \mathbb{B}(b, \epsilon)$$

y sustituyendo la definición de bola, queda

$$0 < \|x - a\| < \delta \implies \|F(x) - b\| < \epsilon,$$

la definición habitual de límite. De manera similar

$$x \approx a \implies F(x) \approx F(a)$$

es otra manera de expresar que F es continua en a. Desde este punto de vista queda clara la intuición de que la continuidad es la propiedad de que «pequeños cambios en las causas (las variables de entrada) producen pequeños cambios en los efectos (las variables de salida)».

Definición 2.2 *(Interior, exterior y frontera). Sea $X \subseteq \mathbb{R}^n$ un subconjunto y $a \in \mathbb{R}^n$ un punto. Denotamos por X^c al* **complementario** *de X, es decir, los puntos que no están en X.*

- *El* **interior** *de X, denotado por $\operatorname{int} X$ ó X^o, consta de los puntos $a \in \mathbb{R}^n$ tales que $\mathbb{B}(a,r) \subseteq X$ para algún $r > 0$. Es decir, puntos que son centros de una bola contenida en X.*
- *El* **exterior** *de X, denotado por $\operatorname{ext} X$, consta de los puntos $a \in \mathbb{R}^n$ tales que $\mathbb{B}(a,r) \subseteq X^c$ para algún $r > 0$. Es decir, puntos que son centros de una bola disjunta de X.*
- *La* **frontera** *de X, denotada por ∂X, consta de los puntos $a \in \mathbb{R}^n$ que no están ni en el interior ni el exterior de X. Es decir, tales que* **cualquier** *bola $\mathbb{B}(a,r)$ centrada en a ni está contenida en X ni en X^c, y por tanto $\mathbb{B}(a,r)$ contiene tanto algún punto de X como también algún punto de X^c.*

Las tres opciones interior/exterior/frontera son mutuamente excluyentes, por tanto

$$\mathbb{R}^n = \operatorname{int} X \sqcup \operatorname{ext} X \sqcup \partial X$$

es una subdivisión de $\mathbb{R}^n$ en tres partes disjuntas, aunque alguna posiblemente sea vacía.

Definición 2.3. *El* **cierre** *de un subconjunto $X \subseteq \mathbb{R}^n$, denotado $\overline{X}$, es la unión de su interior con su frontera.*

Nota 2.2. ▷ Un subconjunto X es abierto si y sólo si es igual a su interior, o sea, todos sus puntos son interiores. Un subconjunto X es cerrado si y sólo si es igual a su cierre. Equivalentemente, si su complementario es igual a su exterior. El interior y exterior de cualquier subconjunto X siempre son abiertos. La frontera y el cierre de cualquier subconjunto X siempre son cerrados. ◁

Ejemplo 2.3. ▷ La frontera tanto de la bola abierta $\mathbb{B}(a,r)$ como de la bola cerrada $\overline{\mathbb{B}}(a,r)$ es la esfera $\mathbb{S}(a,r)$. El exterior tanto de la bola abierta $\mathbb{B}(a,r)$ como de la bola cerrada $\overline{\mathbb{B}}(a,r)$ consta de los puntos x con $\|x-a\| > r$. ◁

Ejemplo 2.4. ▷ El interior de un rectángulo n-dimensional $R = \bigtimes_{k=1}^n [a_k, b_k]$ es $\bigtimes_{k=1}^n (a_k, b_k)$. Su frontera consta de los puntos $x = (x_1, x_2, \dots, x_n)$ donde $x_k = a_k$ ó $x_k = b_k$ para algún k con $1 \leqslant k \leqslant n$. ◁

Ejemplo 2.5. ▷ El subconjunto $X = [0,1] \cup \{2\}$ de $\mathbb{R}$ tiene interior $(0,1)$, frontera $\{0,1,2\}$ y exterior $(-\infty, 0) \sqcup (1,2) \sqcup (2,\infty)$. ◁

Ejemplo 2.6. ▷ El subconjunto $X = \mathbb{Z}$ de números enteros en $\mathbb{R}$ es cerrado, tiene interior vacío, frontera igual a sí mismo, y exterior igual a su complementario. ◁

Ejemplo 2.7. ▷ El subconjunto $X = \mathbb{Q}$ de números racionales en $\mathbb{R}$ tiene interior y exterior vacíos: su frontera es todo $\mathbb{R}$: $\partial\mathbb{Q} = \mathbb{R}$, porque dado $a \in \mathbb{R}$, cualquier intervalo abierto $(a-r, a+r)$ contiene tanto números racionales como números irracionales. ◁

3 Derivadas

La herramienta fundamental en el Cálculo Diferencial multivariable es la **derivación parcial**. Para un campo escalar $f : \mathbb{R}^n \to \mathbb{R}$, se define

$$\frac{\partial f}{\partial x_k}(a) \overset{\text{def}}{=} \lim_{t\to 0} \frac{f(a_1, \dots, a_k + t, \dots, a_n) - f(a_1, \dots, a_k, \dots, , a_n)}{t}. \tag{3.1}$$

En términos prácticos, esto significa que se deriva respecto a la variable x_k manteniendo constantes las demás, de este modo derivando igual que en una variable. Por ejemplo

$$\frac{\partial}{\partial x}(x^2 \operatorname{sen}(y)) = 2x \operatorname{sen}(y), \qquad \frac{\partial}{\partial y}(x^2 \operatorname{sen}(y)) = x^2 \cos(y).$$

$$\frac{\partial}{\partial r}(r\cos\theta) = \cos\theta, \qquad \frac{\partial}{\partial \theta}(r\cos\theta) = -r\operatorname{sen}\theta.$$

$$\frac{\partial}{\partial \theta}(\rho\cos\theta\operatorname{sen}\phi) = -\rho\operatorname{sen}\theta\operatorname{sen}\phi, \quad \frac{\partial}{\partial \rho}(\rho\cos\theta\operatorname{sen}\phi) = \cos\theta\operatorname{sen}\phi,$$

Para un campo vectorial, la derivada parcial se aplica a cada componente,

$$\frac{\partial}{\partial x_k}(f_1, f_2, \dots, f_m) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_m}\right). \tag{3.2}$$

Por ejemplo,

$$\frac{\partial}{\partial r}(r\cos\theta, r\operatorname{sen}\theta) = (\cos\theta, \operatorname{sen}\theta), \qquad \frac{\partial}{\partial \theta}(r\cos\theta, r\operatorname{sen}\theta) = (-r\operatorname{sen}\theta, r\cos\theta).$$

Otras notaciones para la derivada parcial respecto á x_k son

$$D_k, \quad \partial_k, \quad \partial_{x_k}.$$

La derivada parcial es un caso particular de **derivada direccional**. Dado un vector $v \in \mathbb{R}^n$, la derivada de un campo escalar $f : \mathbb{R}^n \to \mathbb{R}^m$ en a en dirección v es

$$D_v f(a) \overset{\text{def}}{=} \lim_{t\to 0} \frac{f(a + tv) - f(a)}{t}.$$

Si $F = (f_1, f_2, \dots, f_m)$, la derivada direccional se aplica componente a componente:

$$D_v F(a) = \big(D_v f_1(a), D_v f_2(a), \dots, D_v f_m(a)\big). \tag{3.3}$$

Las derivadas parciales son las derivadas en las direcciones de la base estándar $\mathbb{e}_1, \mathbb{e}_2, \dots, \mathbb{e}_n$.

Nota 3.1. ▷ Existen funciones $f : \mathbb{R}^n \to \mathbb{R}$ que en un punto a tienen todas las derivadas direccionales $D_v f(a)$, en cualquier dirección, pero f no es continua en a. Por este motivo, la mera existencia de todas las derivadas direccionales, o simplemente las parciales, no es el concepto más adecuado de diferenciabilidad para una función multivariable. ◁

Definición 3.1 *(Matriz jacobiana). Sea* $F : \mathbb{R}^n \to \mathbb{R}^m$ *un campo vectorial con componentes* $F = (f_1, f_2, \ldots, f_m)$. *Si existen las derivadas parciales* $D_i f_j(a)$ *para todos los índices* $1 \leqslant i \leqslant n$ *y* $1 \leqslant j \leqslant m$, *la* ***matriz jacobiana*** *ó* ***matriz diferencial*** *de* F *en* a *es la matriz de* m *filas y* n *columnas*

$$[dF_a] \stackrel{\text{def}}{=} \big(D_i f_j(a)\big)_{\substack{1\leqslant i\leqslant n\\ 1\leqslant j\leqslant m}} \in \mathbb{R}^{m\times n} \tag{3.4}$$

indexando las filas con las componentes (salidas) y las columnas con las variables (entradas).

Si las m variables $(y_1, y_2, \ldots, y_m)$ dependen de las n variables $(x_1, x_2, \ldots, x_n)$, también se denota a la matriz jacobiana de la función $F : \mathbb{R}^n \to \mathbb{R}^m$ correspondiente por

$$\frac{\partial(y_1, y_2, \ldots, y_m)}{\partial(x_1, x_2, \ldots, x_n)} = \begin{array}{c} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_m \\ \\ \end{array} \begin{array}{c} \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_j} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \frac{\partial y_i}{\partial x_1} & \cdots & \frac{\partial y_i}{\partial x_j} & \cdots & \frac{\partial y_i}{\partial x_n} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_j} & \cdots & \frac{\partial y_m}{\partial x_n} \end{pmatrix} \\ \begin{array}{ccccc} x_1 & \cdots & x_j & \cdots & x_n \end{array} \end{array} \tag{3.5}$$

Si el número de variables n de entrada es el mismo que el número de variables m de salida, entonces la matriz diferencial es cuadrada $n \times n$. En este caso, el **determinante jacobiano** o simplemente **jacobiano** es el determinante de la matriz jacobiana,

$$J_F(a) = \det dF_a = \left|\frac{\partial(y_1, y_2, \ldots, y_n)}{\partial(x_1, x_2, \ldots, x_n)}\right| \tag{3.6}$$

Ejemplo 3.2. ▷ La transformación $P(r, \theta) = (r\cos\theta, r\,\mathrm{sen}\,\theta)$ tiene matriz jacobiana

$$[dP]_{(r,\theta)} = \begin{array}{c} x \\ y \\ \\ \end{array} \begin{array}{c} \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{pmatrix} \\ \begin{array}{cc} r & \theta \end{array} \end{array} = \begin{pmatrix} \cos\theta & -r\,\mathrm{sen}\,\theta \\ \mathrm{sen}\,\theta & r\cos\theta \end{pmatrix} \tag{3.7}$$

y el determinante jacobiano es $r\cos^2\theta + r\,\mathrm{sen}^2\,\theta = r$. ◁

Ejemplo 3.3. ▷ Sea $x = u^2 - v^2$, $y = 2uv$ y $z = u + v$. La matriz jacobiana es

$$\frac{\partial(x, y, z)}{\partial(u, v)} = \begin{array}{c} x \\ y \\ z \\ \\ \end{array} \begin{array}{c} \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} \end{pmatrix} \\ \begin{array}{cc} u & v \end{array} \end{array} = \begin{pmatrix} 2u & -2v \\ 2v & 2u \\ 1 & 1 \end{pmatrix} \in \mathbb{R}^{3\times 2}.$$

Las dimensiones corresponden a las 2 variables de entrada (u, v) y 3 de salida (x, y, z). ◁

Ejemplo 3.4. ▷ Sea $x = u, y = u^2, z = u^3, t = u^4$.

$$\frac{\partial(x,y,z,t)}{\partial(t)} = \begin{array}{cc} \begin{array}{c} x\\[2ex] y\\[2ex] z\\[2ex] t\end{array} & \begin{pmatrix} \frac{\partial x}{\partial u}\\ \frac{\partial y}{\partial u}\\ \frac{\partial z}{\partial u}\\ \frac{\partial t}{\partial u}\end{pmatrix} \\ & u \end{array} = \begin{pmatrix}1\\2u\\3u^2\\4u^3\end{pmatrix} \in \mathbb{R}^{4\times 1}$$

La jacobiana es un vector columna ya que se representa una función $\mathbb{R}^1 \to \mathbb{R}^4$. ◁

Ejemplo 3.5. ▷ Sea $x = u^2 v + \cos(v) + uvw + te^t$.

$$\frac{\partial(x)}{\partial(u,v,w,t)} = \begin{array}{cc} x & \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} & \frac{\partial x}{\partial t}\end{pmatrix} \\ & \begin{array}{cccc} u & v & w & t\end{array}\end{array} = \begin{pmatrix} 2uv+vw & u^2-\operatorname{sen}(v)+uw & uv & e^t+te^t\end{pmatrix}$$

La jacobiana es un vector fila ya que representa una función $\mathbb{R}^4 \to \mathbb{R}$. ◁

Ejemplo 3.6 (La derivada en una variable). ▷ Se comprueba inmediatamente a partir de las definiciones que para funciones reales de una variable real, $f(x) : \mathbb{R} \to \mathbb{R}$, la única derivada «parcial» es la derivada ordinaria $df/dx = f'(x)$, y la matriz jacobiana es simplemente la matriz $[df_x] \in \mathbb{R}^{1\times 1}$ cuya única entrada es $f'(x)$. Por ejemplo, si $y = e^x$

$$\frac{\partial(y)}{\partial(x)} = \begin{array}{cc} y & \begin{pmatrix} \frac{\partial y}{\partial x}\end{pmatrix} \\ & x \end{array} = (e^x)$$

entendiendo la jacobiana como matriz 1×1. ◁

Ejemplo 3.7 (El vector tangente). ▷ Consideremos una **curva**, es decir, una función vectorial de una variable, $\gamma(t) : \mathbb{R} \to \mathbb{R}^n$. Dicho tipo de función tiene una representación geométrica como **camino** ó **trayectoria** en $\mathbb{R}^n$. Su diferencial en un punto será una matriz $[d\gamma]_t \in \mathbb{R}^{n\times 1}$, es decir, un vector columna. La única derivada parcial es la derivada d/dt respecto a la única variable t. Si escribimos $\gamma(t)$ en componentes, entonces tenemos

$$\gamma(t) = \sum_{k=1}^{n} \gamma_k(t)\mathbb{e}_k = \big(\gamma_1(t), \ldots, \gamma_n(t)\big) \implies [d\gamma_t] = \begin{pmatrix} \gamma_1'(t)\\ \vdots \\ \gamma_n'(t)\end{pmatrix} \tag{3.8}$$

donde $\gamma_k'(t)$ es la derivada habitual en una variable de la componente $\gamma_k : \mathbb{R} \to \mathbb{R}$. Llamamos al vector correspondiente a esta matriz el **vector tangente** de γ en t y lo denotamos por $\gamma'(t)$ al igual que una derivada escalar en una variable. Por ejemplo,

$$\gamma(t) = (t\cos(t), t\operatorname{sen}(t), t^2), \qquad \gamma'(t) = (\cos(t) - t\operatorname{sen}(t), \operatorname{sen}(t) + t\cos(t), 2t)$$

De hecho, esta curva describe una **hélice** sobre un **paraboloide**. ◁

Ejemplo 3.8 (El vector gradiente). ▷ Si $f : \mathbb{R}^n \to \mathbb{R}$ es un campo escalar, su matriz jacobiana en un punto a es una matriz $[df_a] \in \mathbb{R}^{1\times n}$, es decir, un vector fila, compuesto por las n derivadas parciales del campo. Este vector se denota por

$$\nabla f(a) \stackrel{\text{def}}{=} \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n}\right) \tag{3.9}$$

y se llama **gradiente** del campo en a. Por ejemplo,

$$f(x,y,z,t) = x^2 + y^2 + z^2 - t^2 \qquad \Longrightarrow \qquad \nabla f(x,y,z,t) = 2(x,y,z,-t)$$

y para un vector $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$,

$$f(x) = f(x_1, x_2, \dots, x_n) = x_1^2 + x_2^2 + \cdots + x_n^2 \qquad \Longrightarrow \qquad \nabla f(x) = 2x.$$

El gradiente tiene varias propiedades que lo hacen útil en la descripción de fenómenos físicos. Las comentaremos más adelante. ◁

En general, teniendo en cuenta las definiciones anteriores, vemos que:

- La **fila** i-ésima de la matriz diferencial es el **gradiente** de la **componente** i-ésima f_i:

$$\text{fila } i \text{ de } [dF_a] = \nabla f_i(a) = \left(\frac{\partial f_i}{\partial x_1}, \frac{\partial f_i}{\partial x_2}, \dots, \frac{\partial f_i}{\partial x_n}\right)$$

$$\begin{array}{c} \nabla f_1 \\[2ex] \vdots \\[2ex] \nabla f_i \\[2ex] \vdots \\[2ex] \nabla f_m \end{array} \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_j} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \frac{\partial f_i}{\partial x_1} & \cdots & \frac{\partial f_i}{\partial x_j} & \cdots & \frac{\partial f_i}{\partial x_n} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_j} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix} \tag{3.10}$$

- La **columna** j-ésima de la matriz diferencial es el **vector derivada parcial** respecto a x_j:

$$\text{columna } j \text{ de } [dF_a] = D_jF(a) = \left(\frac{\partial f_1}{\partial x_j}, \frac{\partial f_2}{\partial x_j}, \dots, \frac{\partial f_m}{\partial x_j}\right)$$

$$\begin{array}{c} \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_j} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \frac{\partial f_i}{\partial x_1} & \cdots & \frac{\partial f_i}{\partial x_j} & \cdots & \frac{\partial f_i}{\partial x_n} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_j} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix} \\ \begin{matrix} D_1F & \cdots & D_jF & \cdots & D_nF \end{matrix} \end{array} \tag{3.11}$$

Como la columna k-ésima de una matriz es la imagen del vector $\mathbb{e}_k = (0, \ldots, \underset{k}{1}, \ldots, 0)$, la matriz jacobiana está determinada por las relaciones

$$[dF_a](\mathbb{e}_k) = D_k F(a), \quad k = 1, 2, \ldots, n, \tag{3.12}$$

donde $\mathbb{e}_1, \ldots, \mathbb{e}_n$ es la base estándar de $\mathbb{R}^n$. Por tanto, sobre un vector v cualquiera, con coordenadas $v = (v_1, v_2, \ldots, v_n)$, se tendrá

$$[dF_a](v) = \sum_{k=1}^{n} v_k D_k F(a). \tag{3.13}$$

Cabe preguntarse qué representa $[dF_a](v)$ para un vector cualquiera v. Como sobre $v = \mathbb{e}_k$ es la derivada parcial, es decir, en dirección de $\mathbb{e}_k$, se podría aventurar que será la derivada direccional $D_v F(a)$ para cualquier v. Sin embargo esto **no** es verdad sin más hipótesis.

Como sabemos del Álgebra Lineal, para dar un enfoque general a este tipo de consideraciones, es preciso pensar en **operadores lineales** en vez de las matrices que los representan.

Definición 3.2 *(La diferencial como operador lineal). Si $F : \mathbb{R}^n \to \mathbb{R}^m$ y existen todas las derivadas parciales de F en a, llamamos* **diferencial** *de F en a al operador lineal dF_a cuya matriz es la matriz jacobiana. Es decir, sobre la base estándar $\mathbb{e}_1, \mathbb{e}_2, \ldots, \mathbb{e}_n$ es*

$$dF_a(\mathbb{e}_k) = D_k F(a) \tag{3.14}$$

y sobre un vector $v = (v_1, v_2, \ldots, v_n)$ en general

$$dF_a(v) = \sum_{k=1}^{n} v_k D_k F(a). \tag{3.15}$$

Definición 3.3 *(Diferenciabilidad). Un campo vectorial $F : \mathbb{R}^n \to \mathbb{R}^m$ es* **diferenciable** *en a si existen todas sus derivadas parciales en a y además se cumple la* **aproximación lineal**

$$F(a + h) = F(a) + dF_a(h) + \|h\|\epsilon(h), \quad \lim_{h \to 0} \epsilon(h) = 0. \tag{3.16}$$

Equivalentemente, si

$$\lim_{h \to 0} \frac{\|F(a + h) - F(a) - dF_a(h)\|}{\|h\|} = 0 \tag{3.17}$$

Usando la **O-notación de Landau** la fórmula de aproximación (3.16) se expresa de manera más concisa como

$$F(a + h) = F(a) + dF_a(h) + o(\|h\|) \quad (h \to 0) \tag{3.18}$$

resaltando que el **error** en la aproximación, es decir, la diferencia entre $F(a + h)$ y la aproximación de primer orden $F(a) + dF_a(h)$, es infinitesimal relativo a la norma de la «perturbación» h. Por otra parte, una aplicación lineal tiene orden $O(\|h\|)$, luego la fórmula expresa que el error es «mucho más pequeño» que la aproximación.

Nota 3.9. ▷ La definición que hemos dado, requiriendo la existencia de las derivadas parciales, depende de haber elegido la base estándar $\mathbb{e}_1, \mathbb{e}_2, \dots, \mathbb{e}_n$ de $\mathbb{R}^n$. Hemos optado por este enfoque en vez de uno más abstracto, que no depende de la elección de una base. En el siguiente resultado, vemos que de hecho no es necesaria la elección previa de una base para hablar de la diferenciabilidad. ◁

Proposición 3.1 (Caracterización de la diferenciabilidad como aproximación). *Un campo vectorial* $F : \mathbb{R}^n \to \mathbb{R}^m$ *es diferenciable en* a *si y sólo si existe un operador lineal* $T : \mathbb{R}^n \to \mathbb{R}^m$ *que satisface la fórmula de aproximación*

$$F(a+h) = F(a) + T(h) + o(\|h\|) \qquad (h \to 0). \tag{3.19}$$

En este caso, dicho operador es **único**, *igual a la diferencial:* $T = dF_a$, *y además*

$$dF_a(v) = D_v F(a) \quad \forall v \in \mathbb{R}^n, \tag{3.20}$$

o sea, el valor de la diferencial dF_a *en* v *es la derivada de* F *en* a *en dirección* v.

▪ *Demostración:* A partir de la fórmula de aproximación lineal (3.19), poniendo $h = tv$,

$$F(a+tv) - F(a) = T(tv) + \|tv\|\epsilon(tv) = tT(v) + |t|\|v\|\epsilon(tv)$$
$$\frac{F(a+tv) - F(a)}{t} = T(v) + \operatorname{sgn}(t)\|v\|\epsilon(tv)$$

donde $\operatorname{sgn}(t) = \pm 1$ es el signo de t, y como $\epsilon \to 0$ cuando $t \to 0$, queda $T(v) = D_v F(a)$. En particular, existen todas las derivadas parciales de F en a y satisfacen $T(\mathbb{e}_k) = D_k F(a)$. Como por definición también $dF_a(\mathbb{e}_k) = D_k F(a)$, al coincidir T y dF_a sobre una base, son iguales, y entonces también $dF_a(v) = T(v) = D_v F(a)$ para **todo** vector v. ■

Dicho en palabras, la importante fórmula (3.20) dice que para funciones diferenciables, el cálculo analítico de la derivada en un punto en una dirección v equivale al cálculo algebraico de aplicar la diferencial en ese punto al vector de dirección.

Nota 3.10. ▷ Resaltamos que es posible que existan todas las derivadas parciales, es decir, que esté definida la diferencial, y por tanto (3.20) se cumple para los vectores $v = \mathbb{e}_k$, ya que así lo hemos definido, pero si F no cumple la fórmula de aproximación, es decir, no es diferenciable, entonces puede ser que no se cumpla (3.20) para **otros** vectores v. ◁

Ejemplo 3.11 (La diferencial de una aplicación lineal). ▷ Una aplicación **lineal** $T : \mathbb{R}^n \to \mathbb{R}^m$ es diferenciable en cualquier punto $a \in \mathbb{R}^n$, con diferencial en cada punto igual a ella misma:

$$[dT_a] = [T] \tag{3.21}$$

pues para una aplicación lineal la aproximación (3.16) es **exacta**, es decir, de **error nulo**:

$$T(a+h) = T(a) + T(h) + 0$$

y por la Proposición 3.1 se concluye que T es diferenciable en a con $dT_a = T$. ◁

Nota 3.12. ▷ La relación $dT_a = T$ para una aplicación lineal, que se resume diciendo que «una aplicación lineal es su propia diferencial» no tiene nada que ver con la ecuación diferencial de la función exponencial, que se resume diciendo que «la función exponencial es su propia derivada». Es un buen ejercicio para verificar el correcto entendimiento de estos conceptos, explicar por qué no tiene nada que ver. ◁

Otro «defecto» que se remedia con el concepto de diferenciabilidad es el fenómeno que una función puede tener todas las derivadas direccionales en un punto pero ni siquiera ser continua en ese punto si no cumple la fórmula de aproximación.

Proposición 3.2 (La diferenciabilidad implica la continuidad). *Si una función $F : \mathbb{R}^n \to \mathbb{R}^m$ es diferenciable en a, es continua en a.*

▪ *Demostración:* Usando la O-notación, si F es diferenciable en a, entonces

$$F(a+h) = F(a) + O(\|h\|) \qquad (h \to 0)$$

porque $dF_a = O(\|h\|)$ y el error es $o(\|h\|)$. Haciendo $h \to 0$ queda $F(a+h) \to F(a)$, o sea, F es continua en a. ■

Los dos resultados anteriores explican que el concepto de diferenciabilidad, es decir, existencia de las derivadas direccionales junto con la fórmula de aproximación lineal (3.16), es mejor que sólo la existencia de las derivadas, ya que entonces, al igual que con las funciones de 1 variable, la diferenciabilidad es una condición más fuerte que la continuidad. Aún así, verificar la fórmula de aproximación (3.16) ó el límite equivalente (3.17) no es demasiado práctico. El siguiente teorema suele bastar para la mayoría de aplicaciones.

Definición 3.4 *(Clases de diferenciabilidad). Si existen las derivadas parciales de f hasta orden p en un dominio, y son continuas, decimos que f es* **de clase** *C^p en ese dominio. Si f es infinitamente diferenciable, automáticamente todas sus derivadas son continuas, y decimos que f es de clase C^∞.*

Teorema 3.3 (Criterio C^1 de diferenciabilidad). *Sea $F : \mathbb{R}^n \to \mathbb{R}^m$ y $a \in \mathbb{R}^n$. Si F es de clase C^1, entonces F es diferenciable.*

Como criterio es más práctico que la Definición 3.3, pero **no es una equivalencia:** existen funciones diferenciables que no son C^1.

Nota 3.13 (La propiedad maximizante del gradiente). ▷ La fórmula (3.20) aplicada al caso de un campo escalar $f : \mathbb{R}^n \to \mathbb{R}$ diferenciable en a, da

$$D_v f(a) = [df_a](v) = \nabla f(a) \cdot v \tag{3.22}$$

pensando en el gradiente como vector fila $1 \times n$ que actúa sobre un vector columna $n \times 1$. Esta importante fórmula dice que la derivada direccional de f en a en dirección v es el producto interior del gradiente en a con v. La desigualdad CBS (1.8) implica entonces que

$$|D_v f(a)| \leqslant \|\nabla f(a)\| \|v\|$$

y **normalizando**, es decir, restringiendo a direcciones con $\|v\| = 1$, **unitarias**, queda que

$$|D_v f(a)| \leqslant \|\nabla f(a)\|$$

y además, la igualdad se alcanza si y sólo si v es paralelo á $\nabla f(a)$. Esta es la **propiedad maximizante del gradiente**: la dirección (normalizada) de máxima variación de un campo escalar es la dirección del gradiente, y la magnitud de la variación máxima es la longitud del gradiente. Si v apunta en el mismo sentido que $\nabla f(a)$, es el máximo crecimiento, y si apunta en sentido opuesto, es decir, de $-\nabla f(a)$, es el máximo decrecimiento.

Por ejemplo, en un mapa topográfico se muestra en dos dimensiones el relieve de un paisaje. Sobre cada punto (x, y) de mapa hay una función $z = h(x, y)$ que mide la altura sobre el nivel del mar de ese punto. El gradiente ∇h representa la dirección en la que uno debe caminar para ascender o descender lo más rápidamente posible a lo largo del relieve. ◁

Acabamos esta sección con otro resultado importante.

Teorema 3.4 (Simetría de las parciales segundas). *Si* $F : \mathbb{R}^n \to \mathbb{R}^m$ *es de clase* C^2 *(existen todas sus derivadas parciales de segundo orden y son continuas) entonces*

$$\frac{\partial^2 F}{\partial x_i \partial x_j} = \frac{\partial^2 F}{\partial x_j \partial x_i} \tag{3.23}$$

▪ *Demostración:* Queda fuera del alcance de estos apuntes. ■

El teorema anterior es falso en general si no se supone la continuidad de las derivadas parciales segundas. Para un campo escalar $f : \mathbb{R}^n \to \mathbb{R}$ con derivadas parciales segundas en un dominio, la matriz formada por las derivadas parciales segundas se conoce como la **matriz Hessiana** de f,

$$Hf_a \stackrel{\text{def}}{=} \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)_{1 \leqslant i,j \leqslant n} \in \mathbb{R}^{n \times n}. \tag{3.24}$$

El teorema dice entonces que la matriz Hessiana de un campo escalar C^2 es **simétrica**.

Nota 3.14. ▷ La **diagonalización** de la matriz Hessiana es de suma importancia en el estudio de los problemas de **optimización** (máximos y mínimos) de funciones multivariables. Forma parte de la teoría de los **multiplicadores de Lagrange**. ◁

4 Reglas de derivación multivariable

Para campos escalares, las reglas de derivación para derivadas direccionales son idénticas a las de funciones de una variable.

Proposición 4.1 (Reglas de derivación direccionales). *Supongamos que* $f, g : \mathbb{R}^n \to \mathbb{R}$ *son campos escalares y que existen* $D_v f(a)$ *y* $D_v g(a)$*. Entonces*

1. $D_v(kf)(a) = kD_v f(a)$ *si* k *es una constante.*
2. $D_v(f+g)(a) = D_v f(a) + D_v g(a)$.
3. $D_v(fg)(a) = f(a)D_v g(a) + g(a)D_v f(a)$.
4. $D_v(f/g)(a) = \big(g(a)D_v f(a) - f(a)D_v g(a)\big)/g(a)^2$ *si* g *no es nula en un entorno de* a.
5. $D_v(\phi \circ f)(a) = \phi'(f(a))D_v f(a)$ *si* $\phi : \mathbb{R} \to \mathbb{R}$ *es derivable en* $f(a)$.

▪ *Demostración:* Para un campo escalar f, poniendo $\sigma(t) = a + tv$ (la ecuación de la recta que pasa por a en dirección v) la función

$$f_\sigma(t) \stackrel{\text{def}}{=} f(a+tv)$$

es una función real de una variable real. Su derivada ordinaria en t_0 es la derivada direccional:

$$f'_\sigma(t_0) = \lim_{t\to 0} \frac{f_\sigma(t_0+t) - f_\sigma(t_0)}{t} = \lim_{t\to 0} \frac{f(a+t_0v+tv) - f(a+t_0v)}{t} = D_v f(a+t_0v)$$

por tanto las reglas enunciadas son simplemente las reglas de derivación de funciones reales de una variable real, traducidas a derivadas direccionales mediante esta observación. ■

La regla de la cadena es distinta, pues no tiene sentido componer dos campos escalares salvo que sea $n = 1$. Hemos dado una posible versión, para la composición de un campo escalar a la izquierda con una función real de una variable.

Proposición 4.2 (Reglas de derivación para gradientes). *Supongamos que* $f, g : \mathbb{R}^n \to \mathbb{R}$ *son campos escalares y que existen* $\nabla f(a)$ *y* $\nabla g(a)$*. Entonces*

1. $\nabla(kf)(a) = k\nabla f(a)$ *si* k *es una constante.*
2. $\nabla(f+g)(a) = \nabla f(a) + \nabla g(a)$.
3. $\nabla(fg)(a) = f(a)\nabla g(a) + g(a)\nabla f(a)$.
4. $\nabla(f/g)(a) = \big(g(a)\nabla f(a) - f(a)\nabla g(a)\big)/g(a)^2$ *si* g *no es nula en un entorno de* a.
5. $\nabla(\phi \circ f)(a) = \phi'(f(a))\nabla f(a)$ *si* $\phi : \mathbb{R} \to \mathbb{R}$ *es derivable en* $f(a)$.

▪ *Demostración:* Las reglas de derivación para gradientes se deducen de las reglas para derivadas direccionales derivando componente a componente, teniendo en cuenta que el gradiente tiene como componentes a las derivadas parciales, que son un caso particular de derivadas direccionales. ■

Una aplicación importante de estas reglas es facilitar los cálculos sin tener que recurrir constantemente a la definición y al cálculo en coordenadas.

Ejemplo 4.1. ▷ Sea $r(p) = \|p\|$, el campo escalar «radio» (distancia al origen). En coordenadas tenemos

$$r = \sqrt{x_1^2 + \cdots + x_n^2}.$$

Si elevamos al cuadrado evitamos la raíz cuadrada y tenemos de manera fácil

$$r(x)^2 = x_1^2 + \cdots x_n^2 \quad \therefore \quad \frac{\partial(r^2)}{\partial x_k} = 2x_k \quad \therefore \quad \nabla(r^2) = 2\vec{r}$$

donde $\vec{r}(p) = p$ es el campo posición ó «radio vector». Por otra parte,

$$\nabla(r^2) = 2r\nabla(r)$$

por tanto, si $r \neq 0$, es decir, fuera del origen,

$$\nabla(r) = \frac{\vec{r}}{r}. \tag{4.1}$$

Por supuesto se puede calcular $\nabla(r)$ directamente en coordenadas, pero así evitamos tratar la raíz en la expresión $r = \sqrt{x_1^2 + \cdots + x_n^2}$.

Entonces tenemos, para cualquier potencia $p \in \mathbb{R}$, si $r \neq 0$,

$$\nabla(r^p) = pr^{p-1}\nabla(r) = pr^{p-1}\frac{\vec{r}}{r} = pr^{p-2}\,\vec{r}$$

en particular, para $p = -1$

$$\nabla\left(-\frac{1}{r}\right) = \frac{\vec{r}}{r^3}$$

es un campo inverso cuadrado. Para el logaritmo (neperiano) tenemos

$$\nabla(\log(r)) = \frac{1}{r}\nabla r = \frac{1}{r}\frac{\vec{r}}{r} = \frac{\vec{r}}{r^2}.$$

◁

Otro caso útil es el de las derivadas de curvas. Tenemos en cuenta que en general no tiene sentido componer dos curvas y tampoco dividirlas.

Proposición 4.3 (Reglas de derivación para curvas). *Supongamos que $\alpha, \beta : \mathbb{R} \to \mathbb{R}^n$ son curvas y que existen los vectores tangentes $\alpha(t)'$ y $\beta'(t)$. Entonces*

1. $(k\alpha)'(t) = k\alpha'(t)$ *si* k *es una constante.*

2. $(\alpha + \beta)'(t) = \alpha'(t) + \beta'(t)$.

3. $(\alpha \cdot \beta)'(t) = \alpha(t) \cdot \beta'(t) + \alpha'(t) \cdot \beta((t)$ *para el producto interior.*

4. $(\alpha \times \beta)'(t) = \alpha(t) \times \beta'(t) + \alpha'(t) \times \beta((t)$ *para curvas en* 3 *dimensiones.*

5. $(\alpha\beta)'(t) = \alpha'(t)\beta(t) + \alpha(t)\beta'(t)$ *para curvas en un espacio de matrices* $\mathbb{R}^{m\times n}$.

6. $(\alpha/\phi)'(t) = \big(\phi(t)\alpha'(t) - \phi'(t)\alpha(t)\big)/\phi(t)^2$ *si* $\phi : \mathbb{R} \to \mathbb{R}$ *es derivable en* t.

7. $(\alpha \circ \phi)'(t) = \phi'(t)\alpha'(\phi(t))$ *si* $\phi : \mathbb{R} \to \mathbb{R}$ *es derivable en* t *y* α *es derivable en* $\phi(t)$.

▪ *Demostración:* Derivar componente a componente y aplicar las reglas en una variable. ■

Ejemplo 4.2. ▷ La regla de la cadena que hemos dado equivale a cambiar la velocidad con la cual la curva recorre su imagen. Por ejemplo, sea

$$\gamma(t) = (\cos t, \operatorname{sen} t) \quad \therefore \quad \gamma'(t) = (-\operatorname{sen} t, \cos t)$$

(una circunferencia) y sea $\phi(t) = t^2$. Entonces

$$\gamma(\phi(t)) = \gamma(t^2) = (\cos(t^2), \operatorname{sen}(t^2)),$$

por tanto

$$(\gamma \circ \phi)'(t) = (-2t \operatorname{sen}(t^2), 2t \cos(t^2)) = 2t(-\operatorname{sen}(t^2), \cos(t^2)) = \phi'(t)\gamma'(\phi(t))$$

verificándose la regla de la cadena para la composición con una función de una variable. ◁

Ejemplo 4.3. ▷ Veamos un ejemplo de la regla del producto. Sean

$$\alpha(t) = (\cos t, \operatorname{sen} t, t), \qquad \beta(t) = (-\operatorname{sen} t, \cos t, t^2).$$

Entonces

$$\alpha'(t) = (-\operatorname{sen} t, \cos t, 1), \qquad \beta'(t) = (-\cos t, -\operatorname{sen} t, 2t)$$

Ahora,

$$(\alpha \cdot \beta)(t) = t^3, \quad \therefore \quad (\alpha \cdot \beta)'(t) = 3t^2$$

y por otra parte

$$\begin{aligned} &\alpha'(t) \cdot \beta(t) + \alpha(t) \cdot \beta'(t) \\ &= (-\operatorname{sen} t, \cos t, 1) \cdot (-\operatorname{sen} t, \cos t, t^2) + (\cos t, \operatorname{sen} t, t) \cdot (-\cos t, -\operatorname{sen} t, 2t) \\ &= \cancel{\operatorname{sen}^2 t} + \cancel{\cos^2 t} + t^2 \cancel{- \cos^2 t} \cancel{- \operatorname{sen}^2 t} + 2t^2 = 3t^2 \end{aligned}$$

verificándose la regla del producto para el producto interior. ◁

Debemos prestar atención al orden de los factores si el producto no es conmutativo, como es el caso con el producto vectorial $\times$ en $\mathbb{R}^3$ ó el producto de matrices.

Ejemplo 4.4. ▷ Sean

$$\alpha(t) = \begin{pmatrix} \cos(t) & -\operatorname{sen}(t) \\ \operatorname{sen}(t) & \cos(t) \end{pmatrix}, \qquad \beta(t) = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$$

Las derivadas se calculan componente a componente:

$$\alpha'(t) = \begin{pmatrix} -\operatorname{sen}(t) & -\cos(t) \\ \cos(t) & -\operatorname{sen}(t) \end{pmatrix} \qquad \beta'(t) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

El producto de las matrices es

$$\alpha(t)\beta(t) = \begin{pmatrix} \cos(t) & t\cos(t) - \operatorname{sen}(t) \\ \operatorname{sen}(t) & t\operatorname{sen}(t) + \cos(t) \end{pmatrix}$$

con lo cual

$$(\alpha\beta)'(t) = \begin{pmatrix} -\operatorname{sen}(t) & \cancel{\cos(t)} - t\operatorname{sen}(t) - \cancel{\cos(t)} \\ \cos(t) & \cancel{\operatorname{sen}(t)} + t\cos(t) - \cancel{\operatorname{sen}(t)} \end{pmatrix} = \begin{pmatrix} -\operatorname{sen}(t) & -t\operatorname{sen}(t) \\ \cos(t) & t\cos(t) \end{pmatrix}$$

Por otra parte

$$\alpha'(t)\beta(t) = \begin{pmatrix} -\operatorname{sen}(t) & -\cos(t) \\ \cos(t) & -\operatorname{sen}(t) \end{pmatrix} \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -\operatorname{sen}(t) & -t\operatorname{sen}(t) - \cos(t) \\ \cos(t) & t\cos(t) - \operatorname{sen}(t) \end{pmatrix}$$

$$\alpha(t)\beta'(t) = \begin{pmatrix} \cos(t) & -\operatorname{sen}(t) \\ \operatorname{sen}(t) & \cos(t) \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & \cos(t) \\ 0 & \operatorname{sen}(t) \end{pmatrix}$$

y sumando se comprueba que efectivamente $(\alpha\beta)'(t) = \alpha'(t)\beta(t) + \alpha(t)\beta'(t)$. ◁

Finalmente, las reglas de derivación más generales son para las matrices jacobianas. Engloban todas las anteriores, sobre todo si se generalizan las operaciones de producto y cociente debidamente. Aparte de las fórmulas, se trata de la «conservación de la diferenciabilidad» que permite, igual que ocurre con la continuidad, ver inmediatamente que una función, combinación de funciones más sencillas, es diferenciable.

Teorema 4.4 (Reglas de derivación con diferenciales). *Sean* $F, G : \mathbb{R}^n \to \mathbb{R}^m$ *diferenciables en* a. *Sus combinaciones lineales y su producto interior son diferenciables en* a*, con*

1. $d(kF)_a = k\,dF_a$ *si* k *es una constante.*

2. $d(F+G)_a = dF_a + dG_a$.

3. $d(F \cdot G)_a = F(a) \cdot dG_a + G(a) \cdot dF_a$.

Equivalentemente, en términos de las matrices jacobianas,

1. $[d(kF)_a] = k[dF_a]$ *si* k *es una constante.*

2. $[d(F+G)_a] = [dF_a] + [dG_a]$.

3. $[d(F \cdot G)_a] = [F(a)] \cdot [dG_a] + [G(a)] \cdot [dF_a]$.

Las dimensiones son $[dF_a], [dG_a] \in \mathbb{R}^{m\times n}$ *y para el producto* $F \cdot G : \mathbb{R}^n \to \mathbb{R}$*, luego* $[d(F\cdot G)_a] \in \mathbb{R}^{1\times n}$ *y se interpretan los vectores como filas:* $[F(a)], [G(a)] \in \mathbb{R}^{1\times m}$.

▪ *Demostración:* La estrategia de demostración de cada apartado es la misma: se escriben las fórmulas de aproximación para F y para G, se realiza la operación deseada sobre las fórmulas de aproximación, y en el resultado se identifica una nueva fórmula de aproximación con sus distintas partes: la lineal, que por la Proposición 3.1 dará la fórmula correspondiente, y el error, que habrá de verificarse es del orden de $o(\|h\|)$ cuando $h \to 0$. Por ejemplo,

$$F(a+h) = F(a) + dF_a(h) + o(\|h\|)$$
$$G(a+h) = G(a) + dG_a(h) + o(\|h\|)$$

luego, sumando ambas fórmulas, queda

$$(F+G)(a+h) = (F+G)(a) + (dF_a(h) + dG_a(h)) + o(\|h\|)$$

y como $dF_a + dG_a$ es un operador lineal, se concluye por la Proposición 3.1 que $F+G$ es diferenciable en a con $d(F+G)_a = dF_a + dG_a$. ■

Teorema 4.5 (Regla de la cadena). *Sean* $F : \mathbb{R}^n \to \mathbb{R}^m$ *y* $G : \mathbb{R}^m \to \mathbb{R}^p$ *campos vectoriales. Si* F *es diferenciable en* $a \in \mathbb{R}^n$ *y* G *es diferenciable en* $b = F(a) \in \mathbb{R}^m$*, entonces la composición* $G \circ F$ *es diferenciable en* a*, con*

$$d(G \circ F)_a = dG_{F(a)} dF_a. \tag{4.2}$$

Equivalentemente, la matriz jacobiana de la composición es el producto de las matrices jacobianas de cada eslabón de la cadena,

$$[d(G \circ F)_a] = [dG_{F(a)}]\big([dF_a]\big). \tag{4.3}$$

▪ *Demostración:* Se sigue la misma estrategia que en la demostración del Teorema 4.4. ■

Corolario 4.6 (Diferencial de la inversa). *Si* $F : \mathbb{R}^n \to \mathbb{R}^n$ *es invertible y diferenciable en* a*, y* F^{-1} *es diferenciable en* $F(a)$*, entonces*

$$[d(F^{-1})_{F(a)}] = [dF_a]^{-1}. \tag{4.4}$$

▪ *Demostración:* La regla de la cadena aplicada a la composición $F^{-1} \circ F = I = F \circ F^{-1}$ (la identidad) implica que $[d(F^{-1})_{F(a)}][dF_a] = [dF_a][d(F^{-1})_{F(a)}] = [dI_a] = I$. ■

Corolario 4.7. *El determinante jacobiano es multiplicativo sobre cadenas. Es decir,*

$$\det d(G \circ F)_a = \det dG_{F(a)} \det dF_a. \tag{4.5}$$

En particular, si F *es invertible con inversa* F^{-1}*, entonces*

$$\det d(F^{-1})_{F(a)} = \frac{1}{\det dF_a}. \tag{4.6}$$

▪ *Demostración:* $[d(G \circ F)_a] = [dG_{F(a)}][dF_a]$ y el determinante de un producto de matrices es el producto de los determinantes de cada factor. ■

Corolario 4.8 (La regla de la cadena en coordenadas). *Sean* $z_1, z_2, \ldots, z_m$ *variables que dependen de manera diferenciable de variables* $y_1, y_2, \ldots, y_l$ *que a su vez dependen de manera diferenciable de* $x_1, x_2, \ldots, x_n$*. Entonces*

$$\frac{\partial z_i}{\partial x_j} = \sum_{k=1}^{l} \frac{\partial z_i}{\partial y_k} \frac{\partial y_k}{\partial x_j}, \tag{4.7}$$

para cualquier par de índices (i, j) *con* $1 \leqslant j \leqslant n$ *y* $1 \leqslant i \leqslant m$.

▪ *Demostración:* Es la fórmula para el producto de matrices, aplicada a las jacobianas. ■

La ecuación (4.7) es fácil de recordar: para derivar una variable z_i que depende de otra x_j, sumamos sobre las variables «intermedias» y_k que se «cancelan» simbólicamente:

$$\sum_{k=1}^{l} \frac{\partial z_i}{\partial y_k} \frac{\partial y_k}{\partial x_j} = \frac{\partial z_i}{\partial x_j}.$$

En la **notación de Einstein** se suprime incluso el símbolo $\sum$ del sumatorio, conviniendo en que la expresión del sumando general ya indica que se suma sobre todos los índices k.

Ejemplo 4.5. ▷ Sean x, y, w, r, θ variables relacionadas por

$$x = r\cos\theta, \quad y = r\operatorname{sen}\theta, \quad w = x^2 + xy + y^2.$$

Si primero sustituimos en w para expresarla en términos de r, θ, obtenemos

$$w = r^2\cos^2\theta + r^2\cos\theta\operatorname{sen}\theta + r^2\operatorname{sen}^2\theta = r^2 + r^2\cos\theta\operatorname{sen}\theta,$$

por tanto, derivando la composición, o sea la «cadena completa»,

$$\frac{\partial w}{\partial r} = 2r + 2r\operatorname{sen}\theta\cos\theta.$$

Si empleamos la regla de la cadena, derivando cada «eslabón», queda

$$\frac{\partial w}{\partial r} = \frac{\partial w}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial r} = (2x + y)\cos\theta + (x + 2y)\operatorname{sen}\theta,$$

y si ahora sustituimos para expresar x, y en términos de r, θ, queda

$$\begin{aligned}\frac{\partial w}{\partial r} &= (2r\cos\theta + r\operatorname{sen}\theta)\cos\theta + (r\cos\theta + 2r\operatorname{sen}\theta)\operatorname{sen}\theta \\ &= 2r\cos^2\theta + 2r\operatorname{sen}\theta\cos\theta + r\operatorname{sen}^2\theta \\ &= 2r + 2r\operatorname{sen}\theta\cos\theta,\end{aligned}$$

que es la misma expresión para $\partial w/\partial r$ obtenida anteriormente. La forma de las ecuaciones determinará cuál de los dos procedimientos es más corto, si derivar la cadena completa o cada eslabón. Suele ser más práctico lo segundo. ◁

De las reglas de derivación generales se pueden deducir todas las particulares expuestas anteriormente. La regla de la cadena para la composición de un campo escalar $f : \mathbb{R}^n \to \mathbb{R}$ con una curva $\gamma : \mathbb{R} \to \mathbb{R}^n$ es un caso particular de especial importancia.

Proposición 4.9. *Si $\gamma : \mathbb{R} \to \mathbb{R}^n$ es una curva, diferenciable en t y $f : \mathbb{R}^n \to \mathbb{R}$ es un campo escalar diferenciable en $\gamma(t)$, entonces la derivada ordinaria de la composición $f \circ \gamma : \mathbb{R} \to \mathbb{R}$ en t es*

$$(f \circ \gamma)'(t) = \nabla f\big(\gamma(t)\big) \cdot \gamma'(t). \tag{4.8}$$

▪ *Demostración:* El gradiente $\nabla f(\gamma(t))$ corresponde a la diferencial $[df_{\gamma(t)}]$ de f en $\gamma(t)$, de dimensiones $1 \times n$ y el vector tangente a la diferencial $[d\gamma_t]$ de γ en t, de dimensiones $n \times 1$, y el producto matricial $[df_{\gamma(t)}][d\gamma_t]$ es precisamente el producto interior de los dos vectores, igual que en (3.22), dando como resultado la diferencial $[d(f \circ \gamma)_t]$ de dimensiones 1×1, es decir, una matriz con única componente la derivada ordinaria $(f \circ \gamma)'(t)$. ■

Nota 4.6 (Propiedad de ortogonalidad del gradiente). ▷ La Proposición 4.9 tiene una importante consecuencia geométrica. Los **conjuntos de nivel** de un campo escalar $f : \mathbb{R}^n \to \mathbb{R}$ son los conjuntos donde el campo es constante, es decir,

$$N_c(f) \overset{\text{def}}{=} \big\{x : f(x) = c\big\}.$$

Si $\gamma : \mathbb{R} \to \mathbb{R}^n$ es una curva que recorre el conjunto de nivel $N_c(f)$, es decir, si

$$f(\gamma(t)) = c,$$

para todo t en el dominio de la curva, entonces, al derivar, por (4.8) queda

$$0 = (f \circ \gamma)'(t) = \nabla f\big(\gamma(t)\big) \cdot \gamma'(t),$$

es decir, **el gradiente es ortogonal al vector tangente de cualquier trayectoria en un conjunto de nivel**:

$$\nabla f(p) \perp \vec{\mathbf{t}}_p$$

si $p \in N_c(f)$ y $\vec{\mathbf{t}}_p$ es el tangente a una trayectoria en $N_c(f)$ que pasa por el punto p. Geométricamente, esto significa que el gradiente es ortogonal al propio conjunto $N_c(f)$.

Si el campo es de dos variables, $f(x, y) : \mathbb{R}^2 \to \mathbb{R}$, los conjuntos de nivel son **curvas**. Ejemplos de éstas son las **isobaras** o **isotermas** correspondientes a las curvas de presión y temperatura constantes, medidas a una altura fija dada (por ejemplo, el nivel del mar). Si el campo es de tres variables, $f(x, y, z) : \mathbb{R}^3 \to \mathbb{R}$, los conjuntos de nivel son **superficies**. En general, si f representa un **potencial**, se llaman superficies **equipotenciales**.

En el ejemplo de la Nota 3.13, con un mapa topográfico con función de altura $z = h(x, y)$, las curvas de nivel son las «plataformas» de altura constante. Para ascender o descender lo más rápidamente debemos movernos en la dirección del gradiente, que ahora sabemos es moverse en dirección perpendicular a las curvas de nivel. $\triangleleft$

Ejemplo 4.7. $\triangleright$ El campo radial r, que en un punto p da la distancia al origen $\|p\|$, tiene como superficies equipotenciales las esferas centradas en el origen si $c > 0$, el origen si $c = 0$ y el vacío si $c < 0$. La ortogonalidad del gradiente $\nabla r = r^{-1}\vec{\mathbf{r}}$ significa que el radio $\vec{\mathbf{r}}$ es perpendicular a la esfera. $\triangleleft$

Ejemplo 4.8. $\triangleright$ Las curvas equipotenciales del campo $f(x, y) = y^2 - x^2$ son hipérbolas si $c \neq 0$ y el par de rectas asintóticas $y = \pm x$ si $c = 0$. El gradiente $\nabla f(x, y) = 2(-x, y)$ es ortogonal a estas hipérbolas. Por ejemplo, $(1, 2) \in N_3$ y el vector $(-1, 2)$ es perpendicular a la hipérbola $y^2 - x^2 = 3$ en el punto $(1, 2)$. $\triangleleft$