2 Diferenciabilidad

2.1 Diferencial en un punto

Notación. Dados espacios normados E, F, denotamos por $\mathcal{L}(E,F)$ el conjunto de las aplicaciones lineales *acotadas,* es decir $T : E \to F$ lineales con $\|Tv\| \leq C\,\|v\|$ para alguna constante C.

Definición 57. *Sean: E, F espacios normados, un punto $x_0 \in E$ y un abierto $U \subseteq E$ entorno de x_0. Decimos que $f : U \to F$ es **diferenciable en** x_0 si existe $T \in \mathcal{L}(E,F)$ tal que:*

$$\lim_{h \to \vec{0}_E} \frac{f(x_0+h) - f(x_0) - Th}{\|h\|} \;=\; \vec{0}_F\,. \tag{11}$$

El límite en esa fórmula significa lo siguiente: para todo $\varepsilon > 0$ existe un $\delta > 0$ tal que

$$0 < \|h\| < \delta \implies \frac{\|f(x_0+h) - f(x_0) - Th\|}{\|h\|} < \varepsilon\,, \tag{12}$$

de donde: $\|h\| < \delta \;\Rightarrow\; \|\, f(x_0+h) - f(x_0) - Th\,\| \leq \varepsilon\,\|h\|$.

Primeras propiedades:

a) T es única si existe, en cuyo caso se llama **diferencial** de f en x_0 y se denota $(df)_{x_0}$.

b) Si f es diferenciable en x_0 entonces es continua en x_0.

c) Toda $T \in \mathcal{L}(E,F)$ es diferenciable en todo punto y coincide con sus diferenciales.

d) Si f es constante entonces es diferenciable en todo punto con diferencial nula.

Demostración de a). Sean T_1, T_2 cumpliendo (11). Definimos $T = T_1 - T_2$, que es lineal y tal que $\lim_{h\to\vec{0}_E}(Th)/\|h\| = \vec{0}_F$. Dado $\varepsilon > 0$ tomamos $\delta > 0$ tal que $0 < \|h\| < \delta \;\Rightarrow\; \dfrac{\|Th\|}{\|h\|} < \varepsilon$. A cada vector unitario ω le asociamos $h = (\delta/2)\,\omega$, que cumple $\|h\| < \delta$, y entonces:

$$\varepsilon \;>\; \frac{\|T(h)\|}{\|h\|} \;=\; \frac{\|(\delta/2)\,T(\omega)\|}{\|(\delta/2)\,\omega\|} \;=\; \frac{(\delta/2)\,\|T(\omega)\|}{\delta/2} \;=\; \|T(\omega)\|\,,$$

es decir $\|T(\omega)| < \varepsilon$ para todo $\varepsilon > 0$, luego $\|T(\omega)\| = 0$ y, como esto último se cumple para todos los vectores unitarios ω, tenemos $T \equiv 0$ y $T_1 = T_2$.

Demostración de b). Sea $T = (df)_{x_0}$. Tomamos el valor δ_1 tal que se cumple (12) con $\varepsilon = 1$:

$$\|h\| < \delta_1 \implies \frac{\|f(x_0+h) - f(x_0) - (df)_{x_0}h\|}{\|h\|} \;<\; 1 \implies \|f(x_0+h) - f(x_0) - (df)_{x_0}h\| \;\leq\; \|h\|\,,$$

de donde:

$$\|h\| < \delta_1 \implies \|f(x_0+h) - f(x_0)\| \;=\; \big\|\,\big[f(x_0+h) - f(x_0) - (df)_{x_0}h\big] + (df)_{x_0}h\,\big\| \;\leq\; \big(1 + \|T\|\big)\,\|h\|\,.$$

Para cuaquier otro $\varepsilon > 0$ definimos $\delta = \min\{\,\delta_1\,,\, \varepsilon/(1+\|T\|)\} > 0$ y tenemos:

$$\|h\| < \delta \implies \|f(x_0+h) - f(x_0)\| < \varepsilon\,,$$

y queda visto que f es continua en x_0. $\square$

Importante. En el caso especial $E = \mathbb{R}^n$ y $F = \mathbb{R}^m$, sabemos que todas las normas en E son equivalentes y lo mismo para F (apartado 1.11). En este caso, pues, tanto la diferenciabilidad de f en x_0 como la diferencial $(df)_{x_0}$ son independientes de qué normas se elijan en E y en F.

Por supuesto, no toda aplicación continua en x_0 es diferenciable en x_0: la diferenciabilidad es una propiedad estrictamente más exigente que la continuidad.

Caso particular: $E = F = \mathbb{R}$. Cada $T \in \mathcal{L}(\mathbb{R},\mathbb{R})$ está dada por una constante real m, de manera que $T(h) = mh$ para todo $h \in \mathbb{R}$. Dado un entorno U de x_0 en $\mathbb{R}$, una función $f : U \to \mathbb{R}$ es diferenciable en x_0 si y sólo si existe una constante real m tal que:

$$0 \;=\; \lim_{h\to 0}\left|\frac{f(x_0+h)-f(x_0)-mh}{h}\right| \;=\; \lim_{h\to 0}\left|\frac{f(x_0+h)-f(x_0)}{h}-m\right| \;,$$

es decir si y sólo si $\lim_{h\to 0}\dfrac{f(x_0+h)-f(x_0)}{h}$ existe y es igual a m. Esto equivale a que f sea derivable en x_0 con derivada finita $f'(x_0) = m$; dicho de otra manera, el grafo de f tiene en $x = x_0$ tangente no vertical, con pendiente finita m.
Como hay muchas funciones continuas no derivables $\mathbb{R} \to \mathbb{R}$, queda claro que continuidad no implica diferenciabilidad.

Proposición 58. *Una función vectorial* $f \equiv \begin{bmatrix} f_1 \\ \vdots \\ f_m \end{bmatrix}$, *con valores en* $\mathbb{R}^m$, *es diferenciable en* x_0 *si y sólo si las funciones escalares* $f_1,\ldots,f_m$ *son todas diferenciables en* x_0, *en cuyo caso para todo* $h \in E$ *la imagen* $(df)_{x_0}h$ *es el vector de* $\mathbb{R}^m$ *cuyas entradas son los números* $(df_1)_{x_0}h\,,\,\ldots\,,\,(df_m)_{x_0}h$.

Proposición 59. (Linealidad). *Con* E, F, x_0, U *como antes, sean* $U \overset{f}{\underset{g}{\rightrightarrows}} F$ *ambas diferenciables en* x_0 *y* $c \in \mathbb{R}$. *Entonces* $f+g : U \to F$ *y* $cf : U \to F$ *son diferenciables en* x_0, *además:*

$$\boxed{d(f+g)_{x_0} = (df)_{x_0} + (dg)_{x_0} \quad \text{y} \quad \big(d(cf)\big)_{x_0} = c(df)_{x_0}}$$

Las demostraciones de las proposiciones 58 y 59 se dejan como ejercicio.

Proposición 60. (Regla del producto). *Sean* $U \overset{f}{\underset{g}{\rightrightarrows}} \mathbb{R}$ *diferenciables en* x_0. *La función producto* $fg : U \to \mathbb{R}$, *dada por* $(fg)(v) = f(v)g(v)$, *es diferenciable en* x_0 *y*

$$\boxed{d(fg)_{x_0} = (df)_{x_0}(\cdot)g(x_0) + f(x_0)(dg)_{x_0}}$$

Esto también funciona para $f : U \to \mathbb{R}$, $g : U \to F$ *y* $fg : U \to F$.

El caso de f y g escalares lo demostramos en el apartado 2.5.

Proposición 61. (Regla de la cadena). *Sean* E, F, G *tres espacios normados,* $x_0 \in E$, U *abierto de* E *y entorno de* x_0, $f : U \to F$ *diferenciable en* x_0. *Hacemos* $y_0 = f(x_0) \in F$, *tenemos un abierto* V *de* F *entorno de* y_0 *y* $g : V \to G$ *diferenciable en* y_0. *Como hemos visto que* f *es continua en* x_0, *existe un abierto* U' *de* E *con* $x_0 \in U'$ *y* $f(U') \subseteq V$. *La situación* $U' \xrightarrow{f} V \xrightarrow{g} G$ *nos permite definir la aplicación compuesta* $g \circ f : U' \to G$, *que resulta ser diferenciable en* x_0 *con diferencial igual a la compuesta de las diferenciales:*

$$\boxed{d(g\circ f)_{x_0} \;=\; (dg)_{y_0} \circ (df)_{x_0} \;=\; (dg)_{f(x_0)} \circ (df)_{x_0}}$$

Demostración. Definimos los restos $R_1 : E \to F$ y $R_2 : F \to G$, dados por:

$$R_1 h \;=\; f(x_0+h) - f(x_0) - (df)_{x_0}h \quad , \quad R_2 k \;=\; g(y_0+k) - g(y_0) - (dg)_{y_0}k \;.$$

Entonces:

$$\begin{aligned}
f(x_0+h) &= f(x_0)+(df)_{x_0}h+R_1h \;=\; y_0+k \;\;, \quad \text{donde } k \;=\; (df)_{x_0}h+R_1h\;, \\
g\circ f(x_0+h) &= g(y_0+k) \;=\; g(y_0)+(dg)_{y_0}k+R_2k \;=\; g\circ f(x_0)+(dg)_{y_0}k+R_2k\;, \\
g\circ f(x_0+h)-g\circ f(x_0) &= (dg)_{y_0}k+R_2k \;=\; (dg)_{y_0}\big(\,(df)_{x_0}h+R_1h\,\big)+R_2\big(\,(df)_{x_0}h+R_1h\,\big) \;= \\
&= (dg)_{y_0}(df)_{x_0}h+(dg)_{y_0}R_1h+R_2\big(\,(df)_{x_0}h+R_1h\,\big)\;,
\end{aligned}$$

así llegamos a $g\circ f(x_0+h)-g\circ f(x_0)-\big((dg)_{y_0}\circ(df)_{x_0}\big)h=Rh$, donde:

$$Rh \;=\; (dg)_{y_0}R_1h+R_2\big(\,(df)_{x_0}h+R_1h\,\big)\;, \tag{13}$$

y sólo queda ver que $\dfrac{Rh}{\|h\|}\to\vec{0}_G$ cuando $h\to\vec{0}_E$.
Sean C_1,C_2 las constantes tales que $\|(df)_{x_0}(h)\|\;\le\;C_1\,\|h\|$ y $\|(dg)_{y_0}(k)\|\;\le\;C_2\,\|k\|$. Dados $\varepsilon_1,\varepsilon_2>0$ cualesquiera, existen $\delta_1=\delta_1(\varepsilon_1)$ y $\delta_2=\delta_2(\varepsilon_2)$ ambos positivos y cumpliendo:

$$\|h\|<\delta_1 \implies \|R_1h\|\le\varepsilon_1\|h\| \quad , \quad \|k\|<\delta_2 \implies \|R_2k\|\le\varepsilon_2\|k\|\;.$$

Definimos $\delta=\delta(\varepsilon_1,\varepsilon_2)=\min\left\{\,\delta_1(\varepsilon_1)\,,\;\dfrac{\delta_2(\varepsilon_2)}{C_1+\varepsilon_1}\right\}>0$. Para $\|h\|<\delta$, por una parte tenemos:

$$\|(dg)_{y_0}R_1h\|\le C_2\,\|R_1h\|\le C_2\varepsilon_1\|h\|\;, \tag{14}$$

y por otra parte:

$$\|(df)_{x_0}h+R_1h\|\le C_1\,\|h\|+\varepsilon_1\|h\|=(C_1+\varepsilon_1)\|h\|<\delta_2\;,$$

de donde:

$$\|R_2\big(\,(df)_{x_0}h+R_1h\big)\|\le\varepsilon_2\big\|\,(df)_{x_0}h+R_1h\,\big\|\le\varepsilon_2\,(C_1+\varepsilon_1)\|h\|\;. \tag{15}$$

Juntando (13), (14) y (15), llegamos a:

$$0<\|h\|<\delta \implies \frac{\|Rh\|}{\|h\|}\le C_2\varepsilon_1+\varepsilon_2C_1+\varepsilon_2\varepsilon_1\;.$$

Dado cualquier $\varepsilon>0$, elegimos $\varepsilon_1,\varepsilon_2>0$ que cumplan $C_2\varepsilon_1+\varepsilon_2C_1+\varepsilon_2\varepsilon_1<\varepsilon$. El correspondiente número $\delta(\varepsilon_1,\varepsilon_2)$ es positivo y nos da:

$$0<\|h\|<\delta \implies \frac{\|Rh\|}{\|h\|}<\varepsilon\;,$$

y efectivamente $Rh/\|h\|\to\vec{0}_G$ cuando $h\to\vec{0}_E$. □

2.2 Derivada respecto de un vector

Definición 62. *Sean E,F espacios normados, $x_0\in E$ y U entorno de x_0 en E. Sean $f:U\to F$ y $v\in E$. La* ***derivada en*** x_0 ***de*** f ***respecto de*** v *es el siguiente límite, si es que existe:*

$$D_vf(x_0) \;=\; \frac{d}{dt}\bigg|_{t=0} f(x_0+tv) \;=\; \lim_{t\to0}\frac{1}{t}\,\big(\,f(x_0+tv)-f(x_0)\,\big)\;.$$

Si $\|\omega\|=1$ entonces (y sólo entonces) $D_\omega f(x_0)$ se llama ***derivada direccional.***

Propiedades:

a) $D_{\vec{0}_E}f(x_0)$ siempre existe y es $\vec{0}_F$.

b) Homogénea de grado 1 en v: $D_vf(x_0)$ existe $\Rightarrow$ $D_{cv}f(x_0)$ existe y es $c\cdot D_vf(x_0)$.

c) Si f es diferenciable en x_0 entonces en x_0 hay derivada respecto de cualquier vector y:

$$D_v f(x_0) \;=\; (df)_{x_0} v \;.$$

Demostración de b). Es trivial para $c = 0$. Supongamos $c \neq 0$. Entonces podemos escribir:

$$\lim_{t\to 0} \frac{1}{t} \big(f(x_0 + tcv) - f(x_0) \big) \;=\; c \cdot \lim_{t\to 0} \frac{1}{ct} \big(f(x_0 + ctv) - f(x_0) \big) \;,$$

y definiendo $t' = ct$ nos queda:

$$D_{cv} f(x_0) \;=\; c \cdot \lim_{t'\to 0} \frac{1}{t'} \big(f(x_0 + t'v) - f(x_0) \big) \;=\; c \cdot D_v f(x_0) \;.$$

Demostración de c). Es trivial para $v = \vec{0}_E$. Supongamos v no nulo. Definimos el resto:

$$R(h) \;=\; f(x_0 + h) - f(x_0) - (df)_{x_0} h \;,$$

y razonamos así:

$$\begin{aligned} D_v f(x_0) - (df)_{x_0} v \;&=\; \lim_{t\to 0} \frac{1}{t} \big(f(x_0 + tv) - f(x_0) \big) - (df)_{x_0} v \;= \\ &=\; \lim_{t\to 0} \frac{1}{t} \big[(df)_{x_0}(tv) + R(tv) \big] - (df)_{x_0} v \;=\; \lim_{t\to 0} \frac{1}{t} R(tv) \;. \end{aligned}$$

Introducimos la identidad $\dfrac{1}{t} = (\operatorname{sig} t) \dfrac{1}{|t|}$ y llegamos a:

$$D_v f(x_0) - (df)_{x_0} v \;=\; \|v\| \cdot \lim_{t\to 0} (\operatorname{sig} t) \, \frac{R(tv)}{\|tv\|} \;,$$

pero $\operatorname{sig} t$ es función acotada de t (sólo toma los valores 1 y -1) mientras que $R(tv)/\|tv\| \to \vec{0}_F$ cuando $t \to 0$, luego el límite es nulo:

$$D_v f(x_0) - (df)_{x_0} v \;=\; \|v\| \cdot \vec{0}_F \;=\; \vec{0}_F \;.$$

$\square$

2.3 Jacobianas y regla de la cadena

Ahora tenemos U abierto de $\mathbb{R}^n$, una función $f : U \to \mathbb{R}^m$ y un punto $x_0 \in U$. Tanto la diferenciabilidad de f en x_0 como la diferencial $(df)_{x_0}$ son independientes de qué normas se utilicen en $\mathbb{R}^n$ y en $\mathbb{R}^m$, gracias al teorema 48 del apartado 1.11.
Considerando la base estándar $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ de $\mathbb{R}^n$, cada derivada $D_{\mathbf{e}_i} f(x_0)$ es, en realidad, la **i-ésima derivada parcial** de f en x_0. En las diversas notaciones que se utilizan:

$$D_{\mathbf{e}_i} f(x_0) \;=\; \left. \frac{\partial}{\partial x_i} \right|_{x_0} f \;=\; f_{x_i}(x_0) \;=\; \partial_{x_i} f(x_0) \;=\; D_i f(x_0) \;,$$

y es un vector de $\mathbb{R}^m$ (un vector columna de altura m, aunque a veces lo escribamos "tumbado").

Repasemos las condiciones para que f sea diferenciable en x_0. Primera condición: las derivadas parciales $f_{x_i}(x_0)$ tienen que existir para $i = 1, \ldots, n$. Segunda condición: el único candidato posible a diferencial de f en x_0 es la siguiente aplicación lineal:

$$\mathbb{R}^n \longrightarrow \mathbb{R}^m \quad , \quad v = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \;\longmapsto\; v_1 \, f_{x_1}(x_0) + \cdots + v_n \, f_{x_n}(x_0) \;,$$

es decir la aplicación lineal $v \mapsto Av$, siendo A la matriz $m \times n$ siguiente:

$$A \;=\; Df_{x_0} \;\stackrel{\text{def}}{=}\; \big[\, f_{x_1}(x_0) \,|\, f_{x_2}(x_0) \,|\, \cdots \,|\, f_{x_n}(x_0) \,\big]_{m\times n} \,,$$

que conocemos con el nombre de **matriz jacobiana de f en x_0**.

Recuerda: las derivadas parciales de f se meten en la matriz jacobiana *como columnas.*

Si $f \equiv \begin{bmatrix} f_1 \\ \vdots \\ f_m \end{bmatrix}$ entonces las *filas de Df* son las jacobianas de las componentes $f_1, \ldots, f_m$ de f:

$$Df \;=\; \begin{bmatrix} Df_1 \\ \hline Df_2 \\ \hline \vdots \\ \hline Df_m \end{bmatrix}_{m\times n} .$$

Explicado esto, la tercera condición de diferenciabilidad dice: la matriz Df_{x_0}, además de existir, debe cumplir que $\dfrac{f(x_0+h) - f(x_0) - Df_{x_0}h}{\|h\|}$ tienda a $\mathbf{0}$ (en $\mathbb{R}^m$) cuando $h \to \mathbf{0}$ (en $\mathbb{R}^n$). Las **oes de Landau** proporcionan una manera conveniente de escribir esto último.

Definición 63. (Oes de Landau). *Sean f, g dos funciones definidas en un entorno de x_0; la g escalar y positiva en $x \neq x_0$; la f escalar o vectorial.*
*Decimos que f pertenece a la **clase o pequeña de** g (se escribe $f = \mathrm{o}(g)$) si $\lim_{x\to x_0} \dfrac{f}{g} = \mathbf{0}$.*
*Decimos que f pertenece a la **clase o grande de** g (se escribe $f = \mathrm{O}(g)$) si existe una constante C tal que $\|f\| \leq C\, g$ en un entorno de x_0.*

Notación: por tradición, $\mathrm{o}(g)$ y $\mathrm{O}(g)$ no denotan esas clases de funciones sino que denotan un elemento cualquiera dentro de la clase, por eso se escribe $f = \mathrm{o}(g)$ o bien $f = \mathrm{O}(g)$ para indicar que f pertenece a una u otra clase.

Con la notación de las oes de Landau, la tercera condición para que f sea diferenciable en x_0 podemos escribirla así:

$$f(x_0+h) - f(x_0) - Df_{x_0}h \;=\; \mathrm{o}(\|h\|) \quad \text{o bien} \quad f(x_0+h) \;=\; f(x_0) + Df_{x_0}h + \mathrm{o}(\|h\|) \,.$$

La fórmula $f = \mathrm{O}(1)$ significa que f es acotada en un entorno de x_0, mientras que $f = \mathrm{o}(1)$ significa que $\lim_{x\to x_0} f(x) = \mathbf{0}$. Si $\beta > \alpha$ entonces $\varphi = \mathrm{O}(\|x-x_0\|^\beta) \Longrightarrow \varphi = \mathrm{o}(\|x-x_0\|^\alpha)$. En particular, si $\beta > 1$ entonces $\varphi = \mathrm{O}(\|x-x_0\|^\beta) \Longrightarrow \varphi = \mathrm{o}(\|x-x_0\|)$, luego una condición suficiente (no necesaria) para que f sea diferenciable en x_0 es:

$$f(x_0+h) - f(x_0) - Df_{x_0}h \;=\; \mathrm{O}(\|h\|^\beta) \quad \text{para algún } \beta > 1 \,.$$

Pasando de las diferenciales a sus matrices, la **regla de la cadena** se escribe así:

$$\boxed{D(g\circ f)_{x_0} \;=\; (Dg)_{f(x_0)}\, Df_{x_0}}$$

Veamos ahora dos casos particulares de jacobianas y de regla de la cadena.

Camino en $\mathbb{R}^n$. Viene dado por un intervalo $I \subseteq \mathbb{R}$ y una aplicación $f : I \to \mathbb{R}^n$ diferenciable en cada $t \in I$. Tiene una descripción $f(t) \equiv \big(x_1(t)\,,\,\ldots\,,\,x_n(t)\big)$, siendo $x_1(t), \ldots x_n(t) : I \to \mathbb{R}$ funciones escalares diferenciables en todo $t \in I$. Este caso es excepcional por dos razones:

– Permitimos algunos dominios no abiertos, como por ejemplo $I = [a, b)$.

– Dado $t_0 \in I$, que existan las derivadas $x_1'(t_0), \dots, x_n'(t_0)$ es suficiente para que $f(t)$ sea diferenciable en t_0.

La matriz jacobiana es una *columna:* $Df_{t_0} = f'(t_0) = \begin{bmatrix} x_1'(t_0) \\ \vdots \\ x_n'(t_0) \end{bmatrix}$, es decir un vector de $\mathbb{R}^n$, que llamamos **vector derivada** o **vector velocidad.** Los **vectores tangentes** al camino en $t = t_0$ son los múltiplos $cf'(t_0)$ con $c \in \mathbb{R}$.

Sean ahora $U \subseteq \mathbb{R}^n$ un abierto, conteniendo a la imagen $f(I)$ del camino, y $g : U \to \mathbb{R}^m$ función diferenciable en cada punto $f(t) \in f(I)$. La compuesta $g \circ f : I \to \mathbb{R}^m$ es también un camino diferenciable y la regla de la cadena $d(g \circ f)_{t_0} = (dg)_{f(t_0)} \circ (df)_{t_0}$ se traduce en la igualdad matricial $D(g \circ f)_{t_0} = (Dg)_{f(t_0)} \cdot Df_{t_0}$, es decir columna = rectángulo · columna:

$$\mathbb{R}^m \ni (g \circ f)'(t_0) \;=\; (Dg)_{f(t_0)} \cdot f'(t_0) \;=\; \big[\, g_{x_1} \,|\, g_{x_2} \,|\, \cdots \,|\, g_{x_n} \,\big]_{m \times n} \begin{pmatrix} x_1'(t_0) \\ \vdots \\ x_n'(t_0) \end{pmatrix} ,$$

y recuperamos una de las expresiones habituales de la regla de la cadena:

$$\boxed{\frac{d}{dt}\, g\big(\, x_1(t) \,,\, \dots \,,\, x_n(t) \,\big) \;=\; x_1'(t)\, g_{x_1}\big(f(t)\big) + \cdots + x_n'(t)\, g_{x_n}\big(f(t)\big) \;\in \mathbb{R}^m}$$

Función escalar. Sea $V \subseteq \mathbb{R}^m$ un abierto y $g : V \to \mathbb{R}$ una función escalar diferenciable en todo $y^0 \in V$. Entonces la jacobiana Dg_{y^0} es $1 \times m$, es decir una *matriz fila:*

$$Dg_{y^0} \;=\; \big[\; g_{y_1}(y^0) \;\; g_{y_2}(y^0) \;\; \cdots \;\; g_{y_m}(y^0) \;\big]_{1 \times m} ,$$

que, al contrario del ejemplo anterior, no representa un vector de $\mathbb{R}^m$ sino una *función lineal:*

$$\mathbb{R}^m \longrightarrow \mathbb{R} \quad , \quad \mathbb{R}^m \ni \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix} \longmapsto \big[\, g_{y_1}(y^0) \;\; g_{y_2}(y^0) \;\; \cdots \;\; g_{y_m}(y^0) \,\big] \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix} ,$$

es decir la función $v \longmapsto g_{y_1}(y^0)\, v_1 + \cdots + g_{y_m}(y^0)\, v_m$, que es la diferencial $v \mapsto (dg)_{y^0} v$.
Existe un único vector $w \in \mathbb{R}^m$ que cumple la identidad $(dg)_{y^0}(v) \equiv w \cdot v$. Este vector se llama **gradiente** de g en y^0 y se denota ∇g_{y^0} o $\operatorname{grad} g_{y^0}$:

$$\nabla g_{y^0} \;=\; \text{grad } g_{y^0} \;=\; \big(\, g_{y_1}(y^0) \,,\, \dots \,,\, g_{y_m}(y^0) \,\big) ,$$

y es el único vector ∇g_{y^0} tal que $\boxed{(dg)_{y^0}(v) = \nabla g_{y^0} \cdot v}$ para todo $v \in \mathbb{R}^m$.

Sea ahora $U \subseteq \mathbb{R}^n$ un abierto y $f : U \to \mathbb{R}^m$ una función diferenciable en cada punto $x^0 \in U$. Supongamos que además $f(U) \subseteq V$, con lo cual existe la compuesta $g \circ f : U \to \mathbb{R}$ que es una función escalar diferenciable en todo punto $x^0 \in U$. En este caso la regla de la cadena se expresa en términos de jacobianas como fila = fila · rectángulo:

$$D(g \circ f)_{x^0} \;=\; Dg_{f(x^0)} \cdot Df_{x^0} ,$$

igualdad que expresa una identidad entre dos funciones lineales $\mathbb{R}^n \to \mathbb{R}$. Al ser una igualdad entre filas $1 \times n$, equivale a n igualdades escalares:

$$\boxed{\frac{\partial}{\partial x_{\mathbf{i}}}\Big|_{x^0} g\big(f_1(x) \,,\, \dots \,,\, f_m(x)\big) \;=\; g_{y_1}\big(f(x^0)\big)\, \frac{\partial f_1}{\partial x_{\mathbf{i}}}(x^0) + \cdots + g_{y_m}\big(f(x^0)\big)\, \frac{\partial f_m}{\partial x_{\mathbf{i}}}(x^0)} \quad \mathbf{i} = 1, \dots, n ,$$

de nuevo una manera habitual de escribir la regla de la cadena. La correpondiente fórmula para vectores gradientes es:

$$\text{grad } (g \circ f)_{x^0} \;=\; (\text{grad } g)_{f(x^0)} \cdot Df_{x^0} .$$

2.4 Ejemplos especiales

El propósito de este apartado es aportar evidencia de que la diferenciabilidad de f en un punto x_0 es una propiedad muy exigente. Las dos condiciones siguientes son necesarias, pero no suficientes, para dicha diferenciabilidad:
(1) que $D_v f(x_0)$ exista para todo v y dependa linealmente de v,
(2) que f sea continua en x_0.

Empecemos por $f(x,y) \equiv \sqrt[3]{x^3+y^3} : \mathbb{R}^2 \to \mathbb{R}$. Esta función es homogénea de grado 1, con lo cual existe la derivada en el origen respecto de cualquier vector y además:

$$D_v f(\mathbf{0}) \;=\; f(v) \quad \text{para todo } \; v \in \mathbb{R}^2 \,.$$

Pero $f(v)$ no es una función lineal. Una manera rápida de verlo es mirar el grafo de f y comprobar que no es un plano (es una superficie curvilínea). Otra manera de convencerse es demostrar que para ninguna pareja de constantes a, b se cumple $(ax+by)^3 \equiv x^3+y^3$, luego f no es de la forma $ax+by$.

Consideremos ahora $f(x,y) = \begin{cases} \dfrac{x^3 y}{x^6+y^2} & \text{si} \quad (x,y) \neq (0,0) \\[2ex] 0 & \text{si} \quad (x,y) = (0,0) \end{cases}$

Para esta función $D_v f(\mathbf{0})$ existe para todo $v \in \mathbb{R}^2$ y depende linealmente de v... por la sencilla razón de que es siempre nula: $D_v f(0,0) = 0$ para todo $v \in \mathbb{R}^2$.

Si $v_1 = 0$ o $v_2 = 0$, entonces $f(tv) \equiv 0$ luego $\dfrac{f(tv) - f(0,0)}{t} \to 0$ cuando $t \to 0$.

Si $a \neq 0 \neq b$, entonces:

$$\frac{f(ta,tb) - f(0,0)}{t} \;=\; \frac{a^3 b\, t^4}{t\,(a^6 t^6 + b^2 t^2)} \;=\; \frac{a^3 b}{a^6 t^4 + b^2} \cdot t \;\longrightarrow\; \frac{a^3}{b} \cdot 0 \;=\; 0 \quad \text{cuando} \;\; t \to 0 \,.$$

Sin embargo esta f es discontinua en $(0,0)$: resulta que es constante a lo largo de los caminos $\gamma(t) = (t, c \cdot t^3)$, $t > 0$. Por ejemplo $f(t,t^3) = 1/2$ mientras que $f(t, 2t^3) = 2/5$ para todo $t > 0$. Como estos caminos tienden al punto $(0,0)$ cuando $t \to 0$, ni siquiera existe el *límite de dos variables* $\lim_{(x,y)\to(0,0)} f(x,y)$.

Tercer ejemplo: $f(x,y) = \begin{cases} \dfrac{x^3 y}{x^4+y^2} & \text{si} \quad (x,y) \neq (0,0) \\[2ex] 0 & \text{si} \quad (x,y) = (0,0) \end{cases}$

Se demuestra, igual que hemos hecho en el segundo ejemplo, que $D_v f(0,0)$ existe y es nula para todo $v \in \mathbb{R}^2$.

Veamos que esta f sí es continua en $(0,0)$. Tenemos $|x^2 y| \leq \dfrac{x^4+y^2}{2}$ por la desigualdad aritmético-geométrica (es decir, Young para $p = 2$), luego $|f| \leq |x|/2 \leq \|(x,y)\|/2$ en todo el plano $\mathbb{R}^2$. Al ser $f = \mathrm{O}(\|(x,y)\|)$, es $f = \mathrm{o}(1)$ es decir $\lim_{(x,y)\to(0,0)} f(x,y) = 0 = f(0,0)$.

Veamos que esta f no es diferenciable en $(0,0)$. Consideramos el camino $\gamma(t) \equiv (t, t^2)$, que es diferenciable en todo $t \in \mathbb{R}$ y pasa por $(0,0)$ cuando $t = 0$. Si f fuera diferenciable en $(0,0)$ tendría que ser $(df)_{(0,0)} = 0$ y también sería $\left.\dfrac{d}{dt}\right|_{t=0} f \circ \gamma(t) = 0$, por la regla de la cadena. Pero $f \circ \gamma(t) \equiv t/2$, cuya derivada en $t = 0$ es $1/2$.

Lo que le ocurre a esta función es lo siguiente: cumple la regla de la cadena a lo largo de rectas, pero no a lo largo de otras curvas cuando pasan por $(0,0)$.

2.5 Derivadas continuas

Los tres ejemplos del apartado anterior nos avisan de que hay situaciones en las que es delicado decidir si una función de varias variables es diferenciable o no. El siguiente teorema es inmensamente útil porque describe una situación (bastante frecuente) en la que desaparecen esas dificultades.

Teorema 64. *Sea U abierto de $\mathbb{R}^n$ y $x_0 \in U$. Para que $f : U \to \mathbb{R}^m$ sea diferenciable en x_0 es suficiente (no necesario) que en un entorno de x_0 existan $f_{x_1}, \ldots, f_{x_n}$ y sean continuas en x_0.*

Demostración. Veamos primero que el caso $m = 1$ implica el caso general. Pongamos $f \equiv (f_1, \ldots, f_m)$. Si las funciones vectoriales $f_{x_1}, \ldots, f_{x_n}$ son continuas en x_0 entonces los gradientes de las funciones escalares f_j son continuos en x_0 y, por el caso $m = 1$ del teorema, cada f_j es diferenciable en x_0 y por lo tanto también f.
Nos quedamos, pues con el caso $m = 1$. Haremos la demostración cuando $n = 2$ y al final diremos brevemente cómo extenderla a n general. Sea, pues U abierto de $\mathbb{R}^2$ y sea $f : U \to \mathbb{R}$ tal que f_{x_1}, f_{x_2} existen en un entorno del punto $x_0 = (a, b) \in U$ y son continuas en x_0. Podemos suponer que dicho entorno es la bola $B(x_0, r)$ para algún $r > 0$.
Para cualquier punto $(c, d) \in B(x_0, r)$ vamos a estudiar la diferencia $f(c, d) - f(x_0)$. Para ello unimos x_0 con (c, d) mediante un *camino poligonal* formado por un segmento horizontal que une $x_0 = (a, b)$ con el punto intermedio $x' = (c, b)$, seguido de un segmento vertical que empieza en x' y termina en $x'' = (c, d)$. El esquema es:

$$x_0 = (a, b) \quad , \quad \text{segmento horizontal} \quad , \quad x' = (c, b) \quad , \quad \text{segmento vertical} \quad , \quad x'' = (c, d) \; .$$

En esta demostración utilizamos una norma $\|\cdot\|$ que cumpla lo siguiente:

$$|x| = \|(x, 0)\| \leq \|(x, y)\| \geq \|(0, y)\| = |y| \quad \text{para todo} \quad (x, y) \in \mathbb{R}^2 \; . \tag{16}$$

Esta propiedad la tienen muchas normas, entre otras las normas p. Una vez que se cumple (16), la bola $\overline{B}(x_0, \|x'' - x_0\|)$ contiene los tres vértices x_0, x', x'' del camino poligonal y, como es convexa, contiene el camino entero y así f está definida en todos los puntos de dicho camino.
La restricción $f|_{\text{segmento horizontal}}$ es la función de una variable $f(t, b)$ con t entre a y c. Le aplicamos el teorema de los incrementos finitos y resulta un punto $z_1 = (\theta_1, b)$, situado en el segmento horizontal, tal que $f(x') - f(x_0) = (c - a)\, f_{x_1}(\theta_1, b) = (c - a)\, f_{x_1}(z_1)$.
La restricción $f|_{\text{segmento vertical}}$ es la función de una variable $f(c, t)$ con t entre b y d. Le aplicamos el teorema de los incrementos finitos y resulta un punto $z_2 = (c, \theta_2)$, situado en el segmento vertical, tal que $f(x'') - f(x') = (d - b)\, f_{x_2}(c, \theta_2) = (d - b)\, f_{x_2}(z_2)$.
El incremento de f a lo largo del camino poligonal es la suma de los incrementos a lo largo de sus segmentos:

$$f(x'') - f(x_0) = \big(f(x') - f(x_0) \big) + \big(f(x'') - f(x') \big) = (c - a)\, f_{x_1}(z_1) + (d - b)\, f_{x_2}(z_2) \; . \tag{17}$$

Escribamos ahora:

$$f_{x_1}(z_1) = f_{x_1}(x_0) + \text{error}_1 \quad , \quad f_{x_2}(z_2) = f_{x_2}(x_0) + \text{error}_2 \; . \tag{18}$$

Como la bola $\overline{B}(x_0, \|x'' - x_0\|)$ contiene el camino poligonal, contiene los puntos intermedios z_1, z_2. A medida que x'' se acerca a x_0, el radio $\|x'' - x_0\|$ de esa bola tiende a cero y los puntos z_1, z_2 tienden ambos a x_0. Como las funciones f_{x_1}, f_{x_2} son continuas en x_0, tenemos:

$$\text{error}_1 \longrightarrow 0 \;\; \text{y} \;\; \text{error}_2 \longrightarrow 0 \quad \text{cuando} \;\; x'' \to x_0 \; .$$

Por otra parte, juntanto (17) con (18) y definiendo $\text{error} = \text{error}_1\,(c - a) + \text{error}_2\,(d - b)$ sale:

$$\begin{aligned} f(x'') - f(x_0) &= f_{x_1}(x_0)\,(c - a) + \text{error}_1\,(c - a) + f_{x_2}(x_0)\,(d - b) + \text{error}_2\,(d - b) = \\ &= \begin{bmatrix} f_{x_1}(x_0) & f_{x_2}(x_0) \end{bmatrix} \begin{pmatrix} c - a \\ d - b \end{pmatrix} + \text{error} = \\ &= Df_{x_0} \cdot (x'' - x_0) + \text{error} \; . \end{aligned}$$

Ya sólo nos falta ver que error $= \mathrm{o}(\|x'' - x_0\|)$. Ahora bien, por la condición (16) se tiene:

$$\frac{|c-a|}{\|\,(c-a,d-b)\,\|} \leq 1 \quad \text{y} \quad \frac{|d-b|}{\|\,(c-a,d-b)\,\|} \leq 1\;,$$

de donde:

$$\frac{|\text{error}|}{\|x''-x_0\|} \leq \frac{|\text{error}_1|\cdot|c-a| + |\text{error}_2|\cdot|d-b|}{\|x''-x_0\|} \leq |\text{error}_1|\cdot 1 + |\text{error}_2|\cdot 1 \longrightarrow 0 \quad \text{cuando } x'' \to x_0\;,$$

y efectivamente error $= \mathrm{o}(\|x'' - x_0\|)$, lo cual prueba que f es diferenciable en x_0 si $n = 2$.

En el caso $n = 3$, unimos x_0 con otro punto x''' mediante un camino poligonal formado con cuatro vértices x_0, x', x'', x''' y tres segmentos: uno paralelo al eje x_1, el segundo paralelo al eje x_2 y el tercero paralelo al eje x_3. Se obtendrán tres puntos intermedios z_1, z_2, z_3, cada uno situado en un segmento del camino poligonal, y el procedimiento es enteramente análogo a lo que hemos hecho para $n = 2$. Igual para n más grande. □

La función $f(x,y) = \begin{cases} (x^2+y^2)\,\mathrm{sen}\,\dfrac{1}{x^2+y^2} & \text{si} \quad (x,y) \neq (0,0) \\[2ex] 0 & \text{si} \quad (x,y) = (0,0) \end{cases}$ es diferenciable en $(0,0)$ pero tiene f_{x_1}, f_{x_2} discontinuas en ese punto, mostrando así que la condición suficiente proporcionada por el teorema anterior no es una condición necesaria.

Definición 65. *Decimos que $f : U \to \mathbb{R}^m$ es **de clase** $\mathcal{C}^1$ **en** U, y se indica por $f \in \mathcal{C}^1(U, \mathbb{R}^m)$ o simplemente $f \in \mathcal{C}^1(U)$, si $f_{x_1}, \ldots, f_{x_n}$ existen y son continuas en todo U.*

Las funciones de clase $\mathcal{C}^1$ son diferenciables en todo punto de su dominio.

Demostración de la proposición 60, caso escalar. La función producto prod : $\mathbb{R}^2 \to \mathbb{R}$, dada por $(x,y) \mapsto xy$, tiene jacobiana $D\,\mathrm{prod} = [y \;\; x]$, claramente continua, luego $\mathrm{prod} \in \mathcal{C}^1(\mathbb{R}^2)$ y es diferenciable en todo punto de $\mathbb{R}^2$. Dados un abierto $U \subseteq \mathbb{R}^n$ y $f, g : U \to \mathbb{R}$ diferenciables en $x_0 \in U$, el producto se describe como compuesta $fg \equiv \mathrm{prod} \circ \begin{bmatrix} f \\ g \end{bmatrix}$ y podemos aplicar la regla de la cadena:

$$D(fg)_{x_0} \;=\; D\,\mathrm{prod}_{(x,y)=\left(f(x_0),g(x_0)\right)} \cdot \begin{bmatrix} Df \\ Dg \end{bmatrix}_{x_0} \;=\; \begin{bmatrix} y & x \end{bmatrix}_{(x,y)=\left(f(x_0),g(x_0)\right)} \cdot \begin{bmatrix} Df_{x_0} \\ Dg_{x_0} \end{bmatrix}\;,$$

resultando la regla del producto: $D(fg)_{x_0} = (Df_{x_0})\,g(x_0) + f(x_0)\,Dg_{x_0}$. □

De las propiedades que hemos visto para la diferencial (suma de funciones, producto de funciones, etc.) y las que hemos visto para las funciones continuas, se deducen:

(1) $f \equiv (f_1, \ldots, f_m) : U \to \mathbb{R}^m$ es de clase $\mathcal{C}^1$ si y sólo si las f_j son todas de clase $\mathcal{C}^1$.
(2) Si $f, g : U \to \mathbb{R}^m$ son de clase $\mathcal{C}^1$ y $c \in \mathbb{R}$, entonces $f + g$ y cf son de clase $\mathcal{C}^1$.
(3) Si $f, g : U \to \mathbb{R}$ son de clase $\mathcal{C}^1$, entonces fg es de clase $\mathcal{C}^1$.
(4) La compuesta de aplicaciones $\mathcal{C}^1$ es $\mathcal{C}^1$.

En particular, toda aplicación polinómica es de clase $\mathcal{C}^1$. Más aún, combinando (1), (2), (3) y (4) tantas veces como sea necesario, es fácil deducir que si f viene dada (componente a componente) por una *fórmula elemental* que no plantee ningún problema en el abierto U (es decir, ningún denominador se hace cero, los radicandos y logaritmandos se mantienen estrictamente positivos, las cantidades dentro de un valor absoluto o de la función sig no se anulan) entonces $f \in \mathcal{C}^1(U)$. Como primer uso de estas ideas, las dos fracciones del apartado 2.4 son de clase $\mathcal{C}^1$ en $\mathbb{R}^2 \backslash \{(0,0)\}$, e igual el ejemplo que acabamos de dar $(x^2+y^2)\,\mathrm{sen}\left(1/(x^2+y^2)\right)$, luego son diferenciables en cada punto de $\mathbb{R}^2 \setminus \{(0,0)\}$ sin que haga falta analizarlos más.

La función $f = \sqrt[3]{x^3+y^3}$ tiene radicando nulo (solamente) a lo largo de la recta $L = \{x+y=0\}$ y por lo tanto es $\mathcal{C}^1$ en el abierto $\mathbb{R}^2 \setminus L$. Veamos que no es diferenciable en nigún punto de L. Para el punto $(0,0)$ ya lo hemos visto en el apartado 2.4. Para $x_0 = (a,-a)$, con $a \neq 0$, consideramos el camino $\gamma(t) \equiv x_0 + (t,t)$ y vemos que $f \circ \gamma(t) \equiv \sqrt[3]{2t\,(3a^2+t^2)}$ tiene derivada infinita en $t=0$, que es cuando $\gamma(t)$ pasa por x_0, luego f no es diferenciable en ese punto. Es, sin embargo, continua en todo $\mathbb{R}^2$ porque es compuesta de funciones continuas.

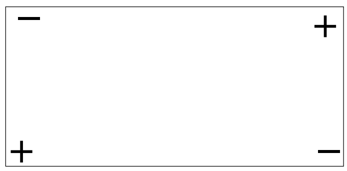
2.6 Derivadas cruzadas

Teorema 66. (Schwarz). *Sea $f(x_1,x_2)$ tal que:*
(1) f_{x_1}, f_{x_2} y $f_{x_1x_2} = \dfrac{\partial}{\partial x_2} f_{x_1}$ existen cerca de $x_0 = (a,b)$. (2) $f_{x_1x_2}$ es continua en x_0.
Entonces existe $f_{x_2x_1} = \dfrac{\partial}{\partial x_1} f_{x_2}$ en x_0 y $f_{x_2x_1}(x_0) = f_{x_1x_2}(x_0)$.

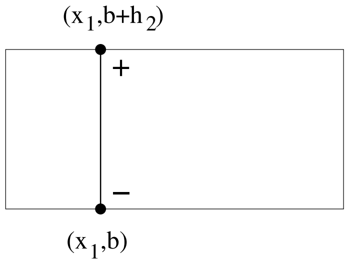
Demostración. Aquí utilizaremos la norma euclídea estándar, denotada $\|\cdot\|$.
Para cada $h = (h_1,h_2)$ consideramos el rectángulo de lados paralelos a los ejes cuyos vértices son $x_0 = (a,b), (a+h_1,b), (a,b+h_2), (a+h_1,b+h_2) = x_0+h$.

Para h pequeño todo el rectángulo está contenido en el entorno de x_0 donde existen $f_{x_1}, f_{x_2}, f_{x_1x_2}$. Definimos:

$$\Sigma(h) \overset{\text{def}}{=} f(a,b) - f(a+h_1,b) - f(a,b+h_2) + f(a+h_1,b+h_2)\ .$$



Fijado h, definimos la función $g(x_1) = f(x_1,b+h_2) - f(x_1,b)$,



que tiene derivada $g'(x_1) = f_{x_1}(x_1,b+h_2) - f_{x_1}(x_1,b)$. Además esta función permite escribir:

$$\Sigma(h) \ = \ g(a+h_1) - g(a)\ ,$$

luego existe $\xi = \xi(h)$, número intermedio entre a y $a+h_1$, tal que:

$$\Sigma(h) \ = \ h_1 \cdot g'(\xi) \ = \ h_1 \cdot \big(f_{x_1}(\xi,b+h_2) - f_{x_1}(\xi,b) \big)\ .$$

Como $f_{x_1x_2}$ existe en todo el rectángulo, hay un número $\eta = \eta(h)$ entre b y $b+h_2$, tal que

$$f_{x_1}(\xi,b+h_2) - f_{x_1}(\xi,b) \ = \ h_2 \cdot f_{x_1x_2}(\xi,\eta)\ ,$$

de donde:
$$\Sigma(h) \;=\; h_1\, h_2\, f_{x_1 x_2}(\xi,\eta)\,.$$
El punto (ξ,η) es interior al rectángulo, cuyo punto más alejado de x_0 es x_0+h (porque estamos utilizando la norma euclídea estándar), por lo tanto $(\xi,\eta)\to x_0$ cuando $h\to(0,0)$ y, como $f_{x_1x_2}$ es continua en x_0, tenemos:
$$\lim_{\substack{h\to(0,0)\\ h_1\neq 0\\ h_2>0}} \frac{\Sigma(h)}{h_1h_2} \;=\; f_{x_1x_2}(x_0)\,. \tag{19}$$
Sea $\varphi(h)=\Sigma(h)/(h_1h_2)$, definida en $\{h\,:\,h_1\neq 0, h_2>0\}$. Cuando un límite de dos variables $\lim_{h\to(0,0)}\varphi(h)$ existe y es finito, no siempre se puede calcular como un límite de límites $\lim_{h_1\to 0}\lim_{h_2\to 0}\varphi(h_1,h_2)$ porque, fijado $h_1\neq 0$ el límite $\lim_{h_2\to 0}\varphi(h_1,h_2)$ puede no existir por ser $(h_1,0)$ distinto del punto $(0,0)$ donde φ tiene límite. Pero en el caso que nos ocupa sí que existe, fijado un valor $h_1\neq 0$, el límite:
$$\begin{aligned}
\lim_{\substack{h_2\to 0\\ h_2>0}} \frac{\Sigma(h_1,h_2)}{h_1h_2} \;&=\; \frac{1}{h_1}\cdot\lim_{h_2\to 0}\left(\frac{f(a+h_1,b+h_2)-f(a+h_1,b)}{h_2}-\frac{f(a,b+h_2)-f(a,b)}{h_2}\right) =\\
&=\; \frac{1}{h_1}\cdot\big(\,f_{x_2}(a+h_1,b)-f_{x_2}(a,b)\,\big)\,.
\end{aligned}$$
Gracias a esto, (19) implica que existe $\displaystyle\lim_{h_1\to 0}\frac{f_{x_2}(a+h_1,b)-f_{x_2}(a,b)}{h_1}$ y es igual a $f_{x_1x_2}(x_0)$, lo que significa que existe $\displaystyle\frac{\partial}{\partial x_1}\Big|_{x_0} f_{x_2}$ y es igual a $f_{x_1x_2}(x_0)$, que es lo afirmado por el teorema. $\square$

Consideremos $f(x,y)=\begin{cases}\dfrac{x^3y}{x^2+y^2} & \text{si}\quad (x,y)\neq(0,0)\\[2ex] \quad 0 & \text{si}\quad (x,y)=(0,0)\end{cases}$ Esta función es diferenciable en $(0,0)$ con derivadas parciales nulas en dicho punto, porque $|f|\leq x^2/2=\mathrm{O}(\|(x,y)\|^2)$ y por lo tanto:
$$f(x,y)-f(0,0)-0\cdot x-0\cdot y \;=\; \mathrm{O}(\|(x,y)\|^2) \;=\; \mathrm{o}(\|(x,y)\|)\,.$$
Derivando en $(x,y)\neq(0,0)$ y añadiendo los valores $f_x(0,0)=f_y(0,0)=0$, calculamos:
$$f_x(0,y) \;=\; 0 \;\text{ para todo } y \quad,\quad f_y(x,0) \;=\; x \;\text{ para todo } x\,,$$
de donde $f_{xy}(0,0)=0$, mientras que $f_{yx}(0,0)=1$. Ahora sabemos que tanto f_{xy} como f_{yx} son discontinuas en $(0,0)$ (ya no lo vamos a comprobar), pues el teorema de Schwarz dice que tendrían el mismo valor en $(0,0)$ si una de ellas fuera continua en dicho punto.

2.7 Derivadas de orden mayor

Definición 67. *Sea $U\subseteq\mathbb{R}^n$ un abierto. Se dice que $f:U\to\mathbb{R}^m$ es* **de clase $\mathcal{C}^k$ en** *U si existen y son continuas en todo U las derivadas parciales de f de órdenes desde cero hasta k (entendiendo que la derivada de orden cero es la propia f).*
Decimos que f es $\mathcal{C}^\infty$, o que es **suave,** *si es $\mathcal{C}^k$ para todo k.*

Decir $f\in\mathcal{C}^0$ es lo mismo que decir que f es continua.

El teorema de Schwarz implica que si f es $\mathcal{C}^2$ entonces se tiene $f_{x_ix_j}\equiv f_{x_jx_i}$ para i,j cualesquiera, es decir que para tal función el orden de derivación no importa en las derivadas segundas.
Si f es $\mathcal{C}^3$, aplicando el teorema de Schwarz varias veces deducimos identidades como las siguientes:
$$f_{xxy}=f_{xyx}=f_{yxx}\quad,\quad f_{xyz}=\begin{cases} f_{xzy}=f_{zxy}\\ f_{yxz}=f_{yzx}=f_{zyx}\end{cases}$$

y para una tal f el orden de derivación tampoco importa en las derivadas terceras.
Ahora bien, la mayoría de las funciones $\mathcal{C}^3$ nos darán $f_{xxy} \neq f_{xyy}$, lo que significa que es importante cuántas veces se ha derivado respecto de cada variable independiente.
En general, si f es $\mathcal{C}^k$ entonces en las derivadas hasta orden k no importa el orden de derivación pero sí importa (y mucho) el número de veces que se ha derivado respecto de cada variable. Para codificar esto es a veces cómoda la notación de los **multíndices.** Para una función de n variables independientes, un multíndice es una n-upla $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ de enteros no negativos. Por ejemplo, a una función de cuatro variables $f(x_1, x_2, x_3, x_4)$ y al multíndice $\alpha = (0, 3, 0, 2)$ les corresponde la siguiente derivada parcial quinta:

$$D^\alpha f \;=\; \left(\frac{\partial}{\partial x}\right)^\alpha f \;=\; D^{(0,3,0,2)} f \;=\; f_{x_2 x_2 x_2 x_4 x_4} \;,$$

que resulta de derivar f ninguna vez respecto de x_1, tres veces respecto de x_2, ninguna vez respecto de x_3 y dos veces respecto de x_4.
En general $D^\alpha f$ es una derivada parcial de orden $|\alpha| \stackrel{\text{def}}{=} \alpha_1 + \cdots + \alpha_n$, que se define así:

$$D^\alpha f \;=\; \left(\frac{\partial}{\partial x}\right)^\alpha f \;=\; \frac{\partial^{|\alpha|}}{\partial_{x_1}^{\alpha_1} \partial_{x_2}^{\alpha_2} \cdots \partial_{x_n}^{\alpha_n}}\, f \;.$$

Esta notación puede ser desventajosa para derivadas de orden pequeño, pero es útil para las de orden alto.

Propiedades (se incluye el caso $k = \infty$):

(1) $f \equiv (f_1, \ldots, f_m)$ es $\mathcal{C}^k$ si y sólo si cada f_j es $\mathcal{C}^k$.
(2) Si $f, g : U \to \mathbb{R}^m$ son $\mathcal{C}^k$ y $c \in \mathbb{R}$, entonces $f + g$ y cf son $\mathcal{C}^k$.
(3) Si $f, g : U \to \mathbb{R}$ son $\mathcal{C}^k$, entonces fg es $\mathcal{C}^k$.
(4) La compuesta de aplicaciones $\mathcal{C}^k$ es $\mathcal{C}^k$.

Vale decir lo mismo que para la clase $\mathcal{C}^1$: una fórmula elemental define una aplicación $\mathcal{C}^\infty$ en el abierto en el que no se anule ningún denominador, los radicandos y logaritmandos permanezcan positivos y las cantidades dentro de un valor absoluto o de la función sig no se anulen.
En particular, las dos fracciones del apartado 2.4 y $(x^2+y^2)\,\mathrm{sen}\big(1/(x^2+y^2)\big)$ definen funciones $\mathcal{C}^\infty$ en $\mathbb{R}^2 \setminus \{(0,0)\}$. La función vista al final del apartado 2.6 es $\mathcal{C}^\infty$ en $\mathbb{R}^2 \setminus \{(0,0)\}$, luego el origen $(0,0)$ es el único punto donde presenta el fenómeno $f_{xy} \neq f_{yx}$.
Cualquier aplicación polinómica es $\mathcal{C}^\infty$ en todo $\mathbb{R}^n$.
El espacio $M_{n\times n}(\mathbb{R})$ de las matrices $n \times n$ puede identificarse con $\mathbb{R}^{n^2}$ y entonces la función determinante $\det : M_{n\times n}(\mathbb{R}) \to \mathbb{R}$ es un polinomio (de grado n) en n^2 variables y es por lo tanto $\mathcal{C}^\infty$.
El conjunto $\mathrm{GL}(n, \mathbb{R})$ de las matrices invertibles $n \times n$ es la preimagen del abierto $\mathbb{R} \setminus \{0\}$ por la función determinante, luego es un abierto de $\mathbb{R}^{n^2}$. En este abierto la función $A \mapsto A^{-1}$, que lleva cada matriz a su inversa, es $\mathcal{C}^\infty$ porque cada una de sus n^2 funciones componentes es un cociente de dos polinomios y el denominador no se anula en el abierto.

2.8 Desarrollo de Taylor

Definición 68. *Sea $U \subseteq \mathbb{R}^n$ un abierto y $x^0 \in U$. Sea $f \in \mathcal{C}^2(U)$ una función escalar. La **matriz hessiana** de f en x^0 es el cuadrado formado por la derivadas segundas de f en x^0:*

$$\mathit{Hess}(f)_{x^0} \;=\; \big[\, f_{x_i x_j}(x^0) \,\big]_{n\times n} \;,$$

*que, por el teorema de Schwarz, es una matriz simétrica. La **forma hessiana** de f en x^0 es la forma cuadrática $\mathit{Hess}(f)_{x^0}(\cdot) : \mathbb{R}^n \to \mathbb{R}$ correspondiente a esta matriz simétrica:*

$$\mathrm{Hess}(f)_{x^0}(v) \;\stackrel{\text{def}}{=}\; v^t\, \mathrm{Hess}(f)_{x^0}\, v \;=\; \sum_{1\le i,j\le n} v_i\, v_j\, f_{x_i x_j}(x^0) \;.$$

Teorema 69. *En las condiciones de la definición anterior, tenemos un desarrollo:*

$$f(x^0+h) \;=\; f(x^0) + \frac{1}{1!}\,(df)_{x^0}(h) + \frac{1}{2!}\,\mathrm{Hess}(f)_{x^0}(h) + R(h)\ ,$$

donde el resto $R(h)$ es un $\mathrm{o}(\|h\|^2)$ en cuanto f sea $\mathcal{C}^2$, y es un $\mathrm{O}(\|h\|^3)$ si f es $\mathcal{C}^3$ o mejor.

Demostración. Como todas las normas en $\mathbb{R}^n$ son equivalentes, la clase $\mathrm{o}(\|h\|^k)$ es la misma para todas ellas. Igual ocurre con $\mathrm{O}(\|h\|^k)$. Basta elegir una norma y demostrar el teorema para ella. En esta demostración $\|\cdot\|$ denota la norma euclídea estándar en $\mathbb{R}^n$.

Fijamos una bola $B(x^0,r)$ contenida en el dominio de f. Dado $x=x^0+h\in B(x^0,r)$, el segmento rectilíneo $[x^0,x]$ está contenido en $\overline{B}(x^0,\|h\|)$ y a fortiori en $B(x^0,r)$. Esto permite definir la siguiente función escalar de una variable:

$$g(t) \;=\; f(x^0+th) \quad , \quad 0\le t\le 1\ .$$

Como f es al menos $\mathcal{C}^2$, tenemos $g(t)\in\mathcal{C}^2[0,1]$. El teorema de Taylor para funciones de una variable nos dice que existe un valor intermedio $\theta\in(0,1)$ tal que:

$$g(1)-g(0) \;=\; (1-0)\,g'(0) + \frac{1}{2!}\,(1-0)^2\,g''(\theta)\ ,$$

es decir $f(x^0+h)-f(x^0) = g'(0)+(1/2)\,g''(\theta) = (df)_{x^0}h + (1/2)\,g''(\theta)$. Calculamos:

$$\begin{aligned} g''(t) \;&=\; \frac{d}{dt}\frac{d}{dt}\,f(x^0+th) \;=\; \frac{d}{dt}\sum_{1\le i\le n} h_i\, f_{x_i}(x^0+th) \;= \\ &=\; \sum_{1\le i\le n} h_i\,\frac{d}{dt}\,f_{x_i}(x^0+th) \;=\; \sum_{1\le i,j\le n} h_i\,h_j\,f_{x_ix_j}(x^0+th)\ , \end{aligned}$$

luego $g''(\theta)=\sum_{1\le i,j\le n} h_ih_j\,f_{x_ix_j}(z)$, con $z=x^0+\theta h\ \in\ \overline{B}(x^0,\|h\|)$. En definitiva:

$$\begin{aligned} f(x^0+h) \;&=\; f(x^0)+(df)_{x^0}h+\frac{1}{2}\,\mathrm{Hess}(f)_z(h) \;= \\ &=\; f(x^0)+(df)_{x^0}h+\frac{1}{2}\,\mathrm{Hess}(f)_{x^0}(h) + R\ , \end{aligned}$$

donde $R=(1/2)\,\mathrm{Hess}(f)_z(h)-(1/2)\,\mathrm{Hess}(f)_{x^0}(h)=(1/2)\,\sum_{1\le i,j\le n} h_ih_j\,\big(f_{x_ix_j}(z)-f_{x_ix_j}(x^0)\big)$. Por otra parte, como para todo i es $|h_i|\le\|h\|$, tenemos $|h_ih_j|/\|h\|^2\le 1$ para todo par i,j, luego:

$$\frac{|R|}{\|h\|^2} \;\le\; \frac{1}{2}\sum_{1\le i,j\le n} 1\cdot|\,f_{x_ix_j}(z)-f_{x_ix_j}(x^0)\,|\ .$$

Como $z\in\overline{B}(x^0,\|h\|)$, se tiene $z\to x^0$ cuando $h\to\mathbf{0}$ y, como cada $f_{x_ix_j}$ es continua:

$$f_{x_ix_j}(z)-f_{x_ix_j}(x^0)\;\to\;0 \quad \text{cuando}\ \ h\to\mathbf{0} \quad , \quad \text{para todo par}\ i,j\ ,$$

y deducimos que $R/\|h\|^2\to 0$ cuando $h\to\mathbf{0}$, es decir $R=\mathrm{o}(\|h\|^2)$.

Supongamos ahora que f es al menos $\mathcal{C}^3$. Entonces $g(t)$ es $\mathcal{C}^3$ y, de nuevo por el teorema de Taylor para funciones de una variable, existe un valor intermedio $\widetilde{\theta}\in(0,1)$ tal que:

$$g(1)-g(0) \;=\; (1-0)\,g'(0)+\frac{1}{2!}\,(1-0)^2\,g''(0)+\frac{1}{3!}\,(1-0)^3\,g'''(\widetilde{\theta})\ .$$

Ahora calculamos $g'''(t)=\sum_{1\le i,j,k\le n} h_i\,h_j\,h_k\,f_{x_ix_jx_k}(x^0+th)$, y así:

$$f(x^0+h)-f(x^0) \;=\; (df)_{x^0}h+\frac{1}{2}\,\mathrm{Hess}(f)_{x^0}(h)+\widetilde{R}\ ,$$

donde $\widetilde{R} = (1/6) \sum_{1\le i,j,k\le n} h_i\, h_j\, h_k\, f_{x_ix_jx_k}(\widetilde{z})$ y $\widetilde{z} = x^0 + \widetilde{\theta} h$ está en $\overline{B}(x^0, \|h\|)$. Para cada terna ijk tenemos $|h_ih_jh_k| \le \|h\|^3$, luego $|\widetilde{R}| \le (1/6)\, \|h\|^3 \sum_{1\le i,j,k\le n} |f_{x_ix_jx_k}(z)|$. Cada función $f_{x_ix_jx_k}$ es continua y por lo tanto acotada cerca de x^0 y, como es un conjunto finito de funciones, encontramos una cota común: $|f_{x_ix_jx_k}(x)| \le M$ para toda terna ijk y todo punto x cercano a x^0. Cuando h es pequeño el punto intermedio $\widetilde{z}$ es cercano a x^0 y se verifica $|f_{x_ix_jx_k}(\widetilde{z})| \le M$ para ijk cualesquiera, con lo cual:

$$|\widetilde{R}| \le \frac{1}{6}\, \|h\|^3 \,(\underbrace{M + \cdots + M}_{n^3 \text{ sumandos}}) \;=\; \frac{1}{6}\, n^3 M\, \|h\|^3 ,$$

y queda visto que $\widetilde{R} = \mathrm{O}(\|h\|^3)$ cuando f es $\mathcal{C}^3$. □

En general si f es de clase $\mathcal{C}^{k+1}$ entonces la función auxiliar $g(t) = f(x^0 + th)$ admite desarrollos de Taylor de orden k y en particular:

$$g(1)-g(0) \;=\; (1-0)\, g'(0) + \frac{1}{2!}\,(1-0)^2\, g''(0) + \cdots + \frac{1}{k!}\,(1-0)^k\, g^{(k)}(0) + \frac{1}{(k+1)!}\,(1-0)^{k+1}\, g^{(k+1)}(\theta_k) ,$$

con $\theta_k \in (0,1)$, con lo cual el punto $y = x^0 + \theta_k\, h$ está en la bola $\overline{B}(x^0, \|h\|)$ y es tal que:

$$\begin{aligned} f(x^0+h) - f(x^0) \;&=\; \sum_{s=1}^{k} \frac{1}{s!} \sum_{1\le i_1,\dots,i_s\le n} h_{i_1}\, h_{i_2}\, \cdots\, h_{i_s} \cdot f_{x_{i_1}x_{i_2}\cdots x_{i_s}}(x^0) \;+ \\ &\quad + \frac{1}{(k+1)!} \sum_{1\le i_1,\dots,i_{k+1}\le n} h_{i_1}\, h_{i_2}\, \cdots\, h_{i_{k+1}} \cdot f_{x_{i_1}x_{i_2}\cdots x_{i_{k+1}}}(y) , \end{aligned}$$

fórmula que se conoce como **desarrollo de Taylor de orden k de f en x^0**. La parte con $s = 1$ coincide con $(df)_{x^0} h$ y la parte con $s = 2$ coincide con $(1/2)\mathrm{Hess}(f)_{x^0}(h)$. La última suma, con las derivadas de orden $k+1$ evaluadas en y, es el **resto de Taylor de orden $k+1$** y es de clase $\mathrm{O}\big(\|h\|^{k+1}\big)$.

2.9 Extremos locales

Definición 70. *Sea $U \subseteq \mathbb{R}^n$ un abierto y $x^0 \in U$. Sea $f : U \to \mathbb{R}$ una función escalar diferenciable en x^0. Decimos que x^0 es un **punto crítico** de f si $(df)_{x^0} = 0$.*
*El punto x^0 es un **máximo local** si existe un entorno U de x^0 tal que $f(x) \le f(x^0)$ para todo $x \in U$; es, además, **estricto** si U puede elegirse tal que $f(x) < f(x^0)$ para todo $x \in U \setminus \{x^0\}$.*
*El punto x^0 es un **mínimo local** si existe un entorno U de x^0 tal que $f(x) \ge f(x^0)$ para todo $x \in U$; es, además, **estricto** si U puede elegirse tal que $f(x) > f(x^0)$ para todo $x \in U \setminus \{x^0\}$.*

Lema 71. *Si f es diferenciable en x^0 y $(df)_{x^0} \ne 0$, entonces x^0 no es máximo local ni mínimo local de f.*
Sea $f \in \mathcal{C}^2$ y x^0 punto crítico de f. Si existe un vector v con Hess$(f)_{x^0}(v) > 0$ entonces x^0 no es máximo local. Si existe un vector v tal que Hess$(f)_{x^0}(v) < 0$ entonces x^0 no es mínimo local.

Demostración. Supongamos f diferenciable en x^0 y que hay un vector v tal que el número $(df)_{x^0}(v_0)$ es no nulo. Consideramos la función escalar $g(t) = f(x^0 + tv)$, $t \in (-\varepsilon, \varepsilon)$. Se tiene $g(0) = f(x^0)$ y $g'(0) = (df)_{x^0}(v) \ne 0$, digamos por ejemplo $g'(0) > 0$. Para ε suficientemente pequeño, es $g(t) > g(0)$ en $t \in (0, \varepsilon)$ y $g(t) < g(0)$ en $t \in (-\varepsilon, 0)$. Encontramos así puntos $x = x^0 + tv$ arbitrariamente cercanos a x^0, algunos con $f(x) > f(x^0)$ y otros con $f(x) < f(x^0)$. Luego x^0 no es ni máximo local ni mínimo local. El caso $g'(0) < 0$ es análogo.

Supongamos ahora $f \in \mathcal{C}^2$ y $(df)_{x^0} = 0$. Ahora es $g(0) = f(x^0)$ y $g'(0) = 0$ para la función $g(t)$ construida a partir de cualquier vector v. Pero si $\mathrm{Hess}(f)_{x^0}(v) > 0$ entonces la correspondiente función g tiene $g''(0) > 0$, con lo cual $g(t) > g(0)$ para $t \ne 0$ pequeño (positivo o negativo). Los puntos $x = x^0 + tv$, con $t \ne 0$ pequeño, son arbitrariamente cercanos a x^0 y en ellos f vale más que en x^0, que no es, pues, máximo local. Del mismo modo, si un vector v cumple $\mathrm{Hess}(f)_{x^0}(v) < 0$ entonces hay puntos $x = x^0 + tv$ arbitrariamente cercanos a x^0 en los que f vale menos que en x^0, que por lo tanto no es mínimo local. □

Corolario 72. *Si f es diferenciable en x^0, para que x^0 sea máximo local o mínimo local es necesario (no suficiente) que sea punto crítico.*
Sea f de clase $\mathcal{C}^2$ y x^0 punto crítico de f. Para que x^0 sea máximo local es necesario (no suficiente) que $Hess(f)_{x^0}$ sea ***semidefinida negativa:*** *$Hess(f)_{x^0}(v) \leq 0$ para todo $v \in \mathbb{R}^n$.*
Para que x^0 sea mínimo local es necesario (no suficiente) que $Hess(f)_{x^0}$ sea ***semidefinida positiva:*** *$Hess(f)_{x^0}(v) \geq 0$ para todo $v \in \mathbb{R}^n$.*
Si $Hess(f)_{x^0}$ es ***indefinida*** *(degenerada o no) entonces x^0 no es ni máximo local ni mínimo local.*

La función $f(x,y) = x^2 + y^3$ proporciona un ejemplo en el que $\mathrm{Hess}(f)_{(0,0)}$ es semidefinida positiva pero $(0,0)$ no es mínimo local: el término cúbico y^3 no afecta a la diferencial ni a la hessiana, pero hace que para $t < 0$ sea $f(0,t) < f(0,0)$.

Teorema 73. *Sea f de clase $\mathcal{C}^2$ y x^0 un punto crítico suyo.*
Para que x^0 sea un máximo local estricto es suficiente (no necesario) que $Hess(f)_{x^0}$ sea definida negativa. Para que x^0 sea un mínimo local estricto es suficiente (no necesario) que $Hess(f)_{x^0}$ sea definida positiva.

Demostración. Supongamos la forma hessiana $Q(\cdot) = \mathrm{Hess}(f)_{x^0}(\cdot)$ definida positiva. La esfera unidad $S = \{v \in \mathbb{R}^n \,:\, \|v\| = 1\}$ es cerrada y acotada, por lo tanto compacta, y la función:

$$S \longrightarrow \mathbb{R} \quad , \quad v \longmapsto Q(v) \;=\; \frac{Q(v)}{\|v\|^2}\,,$$

es continua y positiva, luego alcanza un valor mínimo positivo $\lambda > 0$ en S. Resulta así la desigualdad:

$$Q(v) \;\geq\; \lambda\,\|v\|^2 \quad \text{para todo } \; v \in S\,,$$

cuyos miembros, el de la izquierda y el de la derecha, son ambos homogéneos de grado 2. Deducimos que esta desigualdad se cumple para todo vector $v \in \mathbb{R}^n$, no solo para $\|v\| = 1$.
Dado el desarrollo de Taylor $f(x^0 + h) = f(x^0) + 0 + (1/2)\,Q(h) + R(h)$ y dado $\varepsilon > 0$, existe un $\delta = \delta(\varepsilon) > 0$ tal que si $\|h\| < \delta$ entonces $|R(h)| \leq \varepsilon\,\|h\|^2 \leq \dfrac{\varepsilon}{\lambda}\,Q(h)$. Tomamos ε menor que $\lambda/2$, por ejemplo $\varepsilon = \lambda/10$, y el δ correspondiente. Entonces:

$$\|h\| \,<\, \delta \;\Longrightarrow\; |R(h)| \;\leq\; (0'1)\,Q(h) \;\Longrightarrow\; (0'4)\,Q(h) \;\leq\; (1/2)\,Q(h) + R(h) \;\leq\; (0'6)\,Q(h)\,.$$

Sumando $f(x^0)$ a los tres miembros de esta última desigualdad y poniendo $x = x^0 + h$, obtenemos:

$$f(x^0) + (0'4)\,Q(x - x^0) \;\leq\; f(x) \;\leq\; f(x^0) + (0'6)\,Q(x - x^0) \quad , \quad \text{para } \; \|x - x^0\| < \delta\,.$$

Para $x \in B(x^0, \delta) \setminus \{x^0\}$ se tiene

$$Q(x - x^0) \;>\; 0 \quad \text{y} \quad f(x) \;\geq\; f(x^0) + (0'4)\,Q(x - x^0) \;>\; f(x^0)\,,$$

luego x^0 es mínimo local estricto de f.

Si $Q(\cdot)$ es definida negativa se procede de manera análoga. □

Ejemplo de que la condición no es necesaria: la hessiana en $(0,0)$ de la función $f(x,y) = x^2 + y^4$ es degenerada, sin embargo $(0,0)$ sí es mínimo local estricto de f.