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 $\lim\limits_{h\to 0} \dfrac{f(x_0+h) - f(x_0)}{h}$ 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 $\|(df)_{x_0}(h)\|\;\leq\;C_1\,\|h\|$ y $\|(dg)_{y_0}(k)\|\;\leq\;C_2\,\|k\|$. 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:

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

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:

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

y por otra parte:

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

de donde:

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

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

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

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:*

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

*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:

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

□

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 $\lim\limits_{x\to x_0} \dfrac{f}{g} = \mathbf{0}$.*
*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 $\varphi = \mathrm{O}(\|x-x_0\|^\beta) \implies \varphi = \mathrm{o}(\|x-x_0\|^\alpha)$. En particular, si $\beta > 1$ entonces $\varphi = \mathrm{O}(\|x-x_0\|^\beta) \implies \varphi = \mathrm{o}(\|x-x_0\|)$, 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:

$$\boxed{D(g\circ f)_{x_0} \;=\; (Dg)_{f(x_0)}\, Df_{x_0}} \tag{16}$$

Veamos primero el caso particular más importante de esta regla

Definición 64. *Sea $U \subseteq \mathbb{R}^n$ un abierto. Un* **Camino diferenciable en** *U viene dado por un intervalo $I \subseteq \mathbb{R}$ y una aplicación $x(t) : I \to \mathbb{R}^n$ diferenciable en cada $t \in I$ y tal que $x(I) \subseteq U$.*

Todo camino tiene una descripción $x(t) \equiv \big(x_1(t)\,,\,\ldots\,,\,x_n(t)\big)$, 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 $x_1'(t_0),\ldots,x_n'(t_0)$ es suficiente para que $x(t)$ sea diferenciable en t_0.

La matriz jacobiana es una *columna:* $Dx_{t_0} = x'(t_0) = \begin{bmatrix} x_1'(t_0) \\ \vdots \\ x_n'(t_0) \end{bmatrix}$, 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 $c\,x'(t_0)$ con $c \in \mathbb{R}$.
Sean ahora $V \subseteq \mathbb{R}^m$ un abierto y sea $y(t) : I \to V$ un camino diferenciable en V. Sea también $g(y_1,\ldots,y_m) : V \to \mathbb{R}^k$ una función de m variables diferenciable en cada punto $y(t) \in y(I)$. La compuesta $g \circ y : I \to \mathbb{R}^k$ es a su vez un camino diferenciable en $\mathbb{R}^k$ y la regla de la cadena $d(g \circ y)_{t_0} = (dg)_{y(t_0)} \circ (dy)_{t_0}$ se traduce en la igualdad matricial $D(g \circ y)_{t_0} = (Dg)_{y(t_0)}\, Dy_{t_0}$, es decir columna = rectángulo · columna :

$$\mathbb{R}^k \ni (g \circ y)'(t_0) \;=\; (Dg)_{y(t_0)} \cdot y'(t_0) \;=\; \big[\, g_{y_1} \,|\, g_{y_2} \,|\, \cdots \,|\, g_{y_m} \,\big]_{k\times m} \begin{pmatrix} y_1'(t_0) \\ \vdots \\ y_m'(t_0) \end{pmatrix} ,$$

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

$$\boxed{\frac{d}{dt}\, g\big(\, y_1(t)\,,\,\ldots\,,\,y_m(t)\,\big) \;=\; y_1'(t)\, g_{y_1}\big(y(t)\big) + \cdots + y_m'(t)\, g_{y_m}\big(y(t)\big) \;\in \mathbb{R}^k} \tag{17}$$

Supongamos que el camino en (17) es rectilíneo de velocidad constante, es decir $y(t) = y_0 + tv$ con $v = (a_1,\ldots,a_m)$ un vector fijado. Observemos que en este caso:

$$y(0) \;=\; y_0 \quad , \quad y'(0) \;=\; v \quad , \quad \left.\frac{d}{dt}\right|_{t=0} g\big(y(t)\big) \;=\; D_v g(y_0)\,,$$

la tercera igualdad es por la misma definición de la derivada en y_0 respecto del vector v, luego en este caso particular la regla de la cadena dice que $D_v g(y_0) \;=\; (Dg)_{y_0} v$, es decir:

$$D_{a_1\mathbf{e}_1+\cdots+a_m\mathbf{e}_m}\, g\,(y_0) \;=\; a_1\, D_{\mathbf{e}_1} g(y_0) + \cdots + a_m\, D_{\mathbf{e}_m} g(y_0)\,.$$

Dicho de otra manera:

> *Pedir que $v \mapsto D_v g(y_0)$ sea una función lineal $\mathbb{R}^m \to \mathbb{R}^k$ no es otra cosa que pedir que g cumpla la regla de la cadena (17) para* **caminos rectilíneos.**

Recordemos que ésta era la segunda condición para que g sea diferenciable en y_0. Si además se satisface la tercera condición (la que involucra $\mathrm{o}(\|h\|)$) entonces la regla (17) también se cumple para caminos curvilíneos.

Ahora mostraremos que el caso general de la regla de la cadena consiste en realidad en varios casos particulares de (17). Además de la función $g : V \to \mathbb{R}^k$, ahora tenemos un abierto $U \subseteq \mathbb{R}^n$ y una función de n variables $f(x_1,\ldots,x_n) : U \to V$ diferenciable en todo punto de U y con sus valores en V. Fijado un punto $x_0 \in U$ y el correspondiente $y_0 = f(x_0) \in V$, para $\varepsilon > 0$ suficientemente pequeño están definidos los n caminos siguientes (el índice i va de 1 a n):

$$\alpha_i(t) : (-\varepsilon, \varepsilon) \longrightarrow V \quad , \quad \alpha_i(t) \;=\; f(x_0 + t\,\mathbf{e}_i) \;=\; f\big(\, x_{0,1}\,,\,\ldots\,,\,x_{0,i-1}\,,\,x_{0,i} + t\,,\,x_{0,i+1}\,,\,\ldots\,x_{0,n}\,\big)\,,$$

que satisfacen $\alpha_i(0) = y_0$, $\alpha_i{}'(0) = f_{x_i}(x_0) = (Df)_{x_0}\mathbf{e}_i$ y pueden ser curvilíneos (en cuanto f tenga un poco de complejidad). Pedir que g cumpla la regla (17) para estos n caminos:

$$\left.\frac{d}{dt}\right|_{t=0} g\big(\alpha_i(t)\big) \;=\; (Dg)_{y_0}\, \alpha_i'(0) \quad , \quad i = 1,\ldots,n\,,$$

es decir, pedir que sea $(g\circ f)_{x_i}(x_0) = (Dg)_{y_0}\, f_{x_i}(x_0)$ para $i=1,\ldots,n$, es exactamente lo mismo que pedir la siguiente igualdad matricial:

$$\big[\,(g\circ f)_{x_1}\,|\,\cdots\,|\,(g\circ f)_{x_n}\,\big]_{x_0} \;=\; \big(Dg\big)_{y_0}\,\big[\,f_{x_1}\,|\,\cdots\,|\,f_{x_n}\,\big]_{x_0}\;,$$

que es la fórmula (16), la regla general de la cadena para matrices jacobianas.
La manera práctica de expresar esto es con la siguiente fórmula, en la que el índice $\mathbf{i}$ va desde 1 hasta n:

$$\boxed{\frac{\partial}{\partial x_{\mathbf{i}}}\bigg|_{x_0} g\big(f_1(x)\,,\,\ldots\,,\,f_m(x)\big) \;=\; g_{y_1}\big(f(x_0)\big)\,\frac{\partial f_1}{\partial x_{\mathbf{i}}}(x_0)+\cdots+g_{y_m}\big(f(x_0)\big)\,\frac{\partial f_m}{\partial x_{\mathbf{i}}}(x_0)} \qquad (18)$$

Comentarios. (1) La regla (18) es la (17) con la operación $\dfrac{d}{dt}\bigg|_0$ reemplazada por la $\dfrac{\partial}{\partial x_{\mathbf{i}}}\bigg|_{x_0}$ en todas partes, resultado natural porque $\dfrac{\partial}{\partial x_{\mathbf{i}}}\bigg|_{x_0}$ es la derivada respecto de t a lo largo de $x_0+t\,\mathbf{e_i}$.
(2) Ahora sabemos que (16) equivale a lo siguiente: f es diferenciable en x_0 y g cumple la regla (17) a lo largo de caminos diferenciables *cualesquiera* (rectilíneos o curvilíneos).
(3) Es muy importante la siguiente interpretación de (17): fijada g, la derivada $(g\circ y)'(0)$ de g a lo largo del camino $y(t)$ sólo depende de y_0 y el vector velocidad $v=y'(0)$, de manera que si tenemos dos caminos $y(t),\tilde{y}(t)$ pasando por y_0 con velocidad v:

$$y(0)\;=\;y_0\;=\;\tilde{y}(0)\quad,\quad y'(0)\;=\;v\;=\;\tilde{y}'(0)\;,$$

entonces $(g\circ y)'(0)=(g\circ\tilde{y})'(0)$ y este vector es igual a $D_v g(y_0)$, es decir el resultado $(g\circ y_0)'(0)$ donde $y_0(t)$ es el camino rectilíneo $y_0(t)=x_0+t\,v$.

En definitiva, pues, la regla de la cadena afirma dos cosas:

– Fijados g,y_0,v, los infinitos caminos $y(t)$ que pasan por y_0 con velocidad v dan todos *el mismo resultado* en la derivada $(g\circ y)'(0)$. Esto incluye al camino rectilíneo.

– Fijada g y fijado $y_0=y(0)$, el vector $(g\circ y)'(0)$ depende *linealmente* del vector $y'(0)$.

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 (es decir, que se cumpla la regla de la cadena a lo largo de caminos rectilíneos),
(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 $g(x,y)=\begin{cases}\dfrac{x^3y}{x^6+y^2} & \text{si}\quad (x,y)\neq(0,0)\\[2ex] 0 & \text{si}\quad (x,y)=(0,0)\end{cases}$

Para esta función $D_v g(\mathbf{0})$ existe para todo $v \in \mathbb{R}^2$ y depende linealmente de v... por la sencilla razón de que es siempre nula: $D_v g(0,0) = 0$ para todo $v \in \mathbb{R}^2$.
Si $v_1 = 0$ o $v_2 = 0$, entonces $g(tv) \equiv 0$ luego $\dfrac{g(tv) - f(0,0)}{t} \to 0$ cuando $t \to 0$.
Si $a \neq 0 \neq b$, entonces:

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

Sin embargo esta g 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* $\lim_{(x,y)\to(0,0)} g(x,y)$.

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

Se demuestra, igual que hemos hecho en el segundo ejemplo, que $D_v h(0,0)$ existe y es nula para todo $v \in \mathbb{R}^2$, luego función lineal de v. Por lo explicado en el apartado 2.3, la función h cumple la regla de la cadena a lo largo de rectas al pasar por $(0,0)$.
Esta h 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 $|h| \leq |x|/2 \leq \|(x,y)\|/2$ en todo el plano $\mathbb{R}^2$. Al ser $h = \mathrm{O}(\|(x,y)\|)$, se tiene $\lim_{(x,y)\to(0,0)} h(x,y) = 0 = f(0,0)$.
Pero h no es diferenciable en $(0,0)$. Consideramos los caminos $\alpha(t) = (t,0)$ y $\gamma(t) = (t,t^2)$, que cuando $t = 0$ pasan por $(0,0)$ con la misma velocidad $(1,0)$. Como α es rectilíneo, se tiene $(h \circ \alpha)'(0) = D_{(1,0)} h(0,0) = 0$, en cambio $(h \circ \gamma)'(0) = \left.\dfrac{d}{dt}\right|_{t=0} \dfrac{t}{2} = \dfrac{1}{2} \neq 0$.
Lo que le ocurre a h 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 $(0,0)$, por ejemplo:

$$(h \circ \gamma)'(0) = \frac{1}{2} \neq 0 = [0 \;\; 0] \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \big(Dh_{(0,0)}\big)\, \gamma'(0) .$$

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 65. *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 para cada $j \in \{1, \ldots, m\}$ las derivadas escalares $f_{jx_1}, \ldots, f_{jx_n}$ 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:

$x_0 = (a,b)$, segmento horizontal , $x' = (c,b)$, segmento vertical , $x'' = (c,d)$.

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

$$|x| \;=\; \|(x,0)\| \;\leq\; \|(x,y)\| \;\geq\; \|(0,y)\| \;=\; |y| \quad \text{para todo} \quad (x,y)\in\mathbb{R}^2\,. \tag{19}$$

Esta propiedad la tienen muchas normas, entre otras las normas p. Una vez que se cumple (19), 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:

$$f(x'')-f(x_0) \;=\; \big(f(x')-f(x_0)\big)+\big(f(x'')-f(x')\big) \;=\; (c-a)\,f_{x_1}(z_1)+(d-b)\,f_{x_2}(z_2)\,. \tag{20}$$

Escribamos ahora:

$$f_{x_1}(z_1) \;=\; f_{x_1}(x_0)+\text{error}_1 \quad , \quad f_{x_2}(z_2) \;=\; f_{x_2}(x_0)+\text{error}_2\,. \tag{21}$$

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:

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

Por otra parte, juntanto (20) con (21) 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 $\text{error}=\text{o}(\|x''-x_0\|)$. Ahora bien, por la condición (19) 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 $\text{error}=\text{o}(\|x''-x_0\|)$, 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. $\square$

La función $f(x,y) = \begin{cases} (x^2+y^2)\,\mathrm{sen}\,\dfrac{1}{x^2+y^2} & \text{si} \quad (x,y) \neq (0,0) \\ 0 & \text{si} \quad (x,y) = (0,0) \end{cases}$ 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 66. *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 $\mathrm{prod} : \mathbb{R}^2 \to \mathbb{R}$, 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 $\mathcal{C}^1$ (ver apartado 2.7).

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 $\mathbb{R}^2 \setminus \{(0,0)\}$, 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.

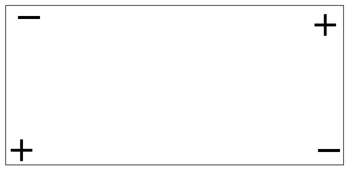
2.6 Derivadas cruzadas

Teorema 67. (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 $f_{x_2x_1}(x_0) = f_{x_1x_2}(x_0)$.

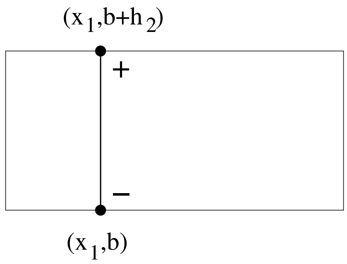
Demostración. Aquí utilizaremos la norma euclídea estándar, denotada $\|\cdot\|$.
Para cada $h = (h_1, h_2)$ consideramos el rectángulo de lados paralelos a los ejes cuyos vértices son $x_0 = (a,b), (a+h_1, b), (a, b+h_2), (a+h_1, b+h_2) = x_0 + h$.

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:

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



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:

$$\Sigma(h) \ = \ h_1 \, h_2 \, f_{x_1x_2}(\xi, \eta) \ .$$

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:

$$\lim_{\substack{h \to (0,0) \\ h_1 \neq 0 \\ h_2 > 0}} \frac{\Sigma(h)}{h_1 h_2} \ = \ f_{x_1x_2}(x_0) \, . \tag{22}$$

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:

$$\begin{aligned} \lim_{\substack{h_2 \to 0 \\ h_2 > 0}} \frac{\Sigma(h_1, h_2)}{h_1 h_2} \ &= \ \frac{1}{h_1} \cdot \lim_{h_2 \to 0} \left(\frac{f(a+h_1, b+h_2) - f(a+h_1, b)}{h_2} - \frac{f(a, b+h_2) - f(a,b)}{h_2} \right) \ = \\ &= \ \frac{1}{h_1} \cdot \big(f_{x_2}(a+h_1, b) - f_{x_2}(a,b) \big) \ . \end{aligned}$$

Gracias a esto, (22) implica que existe $\lim_{h_1\to 0} \dfrac{f_{x_2}(a+h_1,b)-f_{x_2}(a,b)}{h_1}$ y es igual a $f_{x_1x_2}(x_0)$, lo que significa que existe $\left.\dfrac{\partial}{\partial x_1}\right|_{x_0} f_{x_2}$ y es igual a $f_{x_1x_2}(x_0)$, que es lo afirmado por el teorema. $\square$

Consideremos $f(x,y)=\begin{cases} \dfrac{x^3y}{x^2+y^2} & \text{si} \quad (x,y)\neq(0,0) \\ 0 & \text{si} \quad (x,y)=(0,0) \end{cases}$ 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 68. *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:

$$D^\alpha f \;=\; \left(\frac{\partial}{\partial x}\right)^\alpha f \;=\; D^{(0,3,0,2)}f \;=\; f_{x_2x_2x_2x_4x_4}\;,$$

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 cuyo orden es la **longitud del multíndice** $|\alpha|\overset{\text{def}}{=}\alpha_1+\cdots+\alpha_n$ y 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) $f \equiv (f_1, \dots, f_m)$ 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) Todo polinomio es $\mathcal{C}^\infty$ (sus derivadas de todos los órdenes existen y son polinomios).
(4) Si $f, g : U \to \mathbb{R}$ son $\mathcal{C}^k$, entonces fg es $\mathcal{C}^k$.
(5) La compuesta de aplicaciones $\mathcal{C}^k$ es $\mathcal{C}^k$.

Demostramos (5) por inducción sobre k. Es cierto para $k = 0$, en cuyo caso dice que la compuesta de continuas es continua. Sea ahora $k \geq 1$ y supongámoslo cierto para $k - 1$. Dadas f, g de clase $\mathcal{C}^k$, en particular son $\mathcal{C}^1$ y diferenciables en todo punto. La regla de la cadena nos da entonces la siguiente identidad entre funciones matriz:

$$D(g \circ f) \;\equiv\; \big[(Dg) \circ f\big]\, Df \;.$$

El primer factor $(Dg)\circ f$ es la compuesta de una función $\mathcal{C}^k$ con una función matriz de clase $\mathcal{C}^{k-1}$, luego es $\mathcal{C}^{k-1}$ por la hipótesis de inducción. El segundo factor Df es otra función matriz de clase $\mathcal{C}^{k-1}$. Si podemos argumentar que el producto de funciones matriz de clase $\mathcal{C}^{k-1}$ es $\mathcal{C}^{k-1}$, tendremos $D(g \circ f) \in \mathcal{C}^{k-1}$, de donde $g \circ f \in \mathcal{C}^k$. La multiplicación de matrices es una función vectorial $\mathbb{R}^{sm} \times \mathbb{R}^{mn} \to \mathbb{R}^{sn}$ cuyas sn componentes son polinomios cuadráticos, luego es una función $\mathcal{C}^\infty$. Si $A(x), B(x)$ son funciones matriz $\mathcal{C}^{k-1}$, entonces el producto $A(x)B(x)$ es la compuesta de $\big(A(x), B(x)\big) \in \mathcal{C}^{k-1}$ con la multiplicación, y por lo tanto es $\mathcal{C}^{k-1}$ por la hipótesis de inducción. Esto completa la prueba de (5).
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 $f_{xy} \neq f_{yx}$.
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) y es $\mathcal{C}^\infty$.
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 cuadrático de Taylor

Definición 69. *Sea $U \subseteq \mathbb{R}^n$ un abierto y $a \in U$. Sea $f \in \mathcal{C}^2(U, \mathbb{R})$ una función escalar. La **matriz hessiana** de f en a es el cuadrado formado por la derivadas segundas de f en a:*

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

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

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

Teorema 70. *En las condiciones de la definición anterior, tenemos un desarrollo:*

$$f(a+h) \;=\; f(a) + \frac{1}{1!}\,(df)_a(h) + \frac{1}{2!}\,\mathrm{Hess}(f)_a(h) + R(h) \;,$$

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

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

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

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 $f(a+h) - f(a) = g'(0) + (1/2)\, g''(\theta) = (df)_a h + (1/2)\, g''(\theta)$. Calculamos:

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

luego $g''(\theta) = \sum_{1\le i,j\le n} h_i h_j\, f_{x_i x_j}(z)$, con $z = a + \theta h \;\in\; \overline{B}(a,\|h\|)$. En definitiva:

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

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

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

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

$$f_{x_i x_j}(z) - f_{x_i x_j}(a) \;\to\; 0 \quad \text{cuando} \;\; h \to \mathbf{0} \quad , \quad \text{para todo par } \; i,j \; ,$$

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

$$f(a+h) - f(a) \;=\; (df)_a h + \frac{1}{2}\,\mathrm{Hess}(f)_a(h) + \widetilde{R} \; ,$$

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

2.9 Extremos locales

Definición 71. *Sea $U \subseteq \mathbb{R}^n$ un abierto y $a \in U$. Sea $f : U \to \mathbb{R}$ una función escalar diferenciable en a. Decimos que a es un* ***punto crítico*** *de f si $(df)_a = 0$.*
El punto a es un ***máximo local*** *si existe un entorno V de a tal que $f(x) \le f(a)$ para todo $x \in V$; es, además,* ***estricto*** *si V puede elegirse tal que $f(x) < f(a)$ para todo $x \in V \setminus \{a\}$.*
El punto a es un ***mínimo local*** *si existe un entorno V de a tal que $f(x) \ge f(a)$ para todo $x \in V$; es, además,* ***estricto*** *si V puede elegirse tal que $f(x) > f(a)$ para todo $x \in V \setminus \{a\}$.*

Lema 72. *Si f es diferenciable en a y $(df)_a \neq 0$, entonces a no es máximo local ni mínimo local de f.*
Sean $f \in \mathcal{C}^2$ y $a \in U$ crítico. Si existe un vector v con Hess$(f)_a(v) > 0$ entonces a no es máximo local. Si existe un vector w tal que Hess$(f)_a(w) < 0$ entonces a no es mínimo local.

Demostración. Supongamos f diferenciable en a y que hay un vector v tal que el número $(df)_a(v)$ es no nulo. Consideramos la función escalar $g(t) = f(a + tv)$, $t \in (-\varepsilon, \varepsilon)$. Se tiene $g(0) = f(a)$ y $g'(0) = (df)_a(v) \neq 0$, 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 $x = a+tv$ arbitrariamente cercanos al punto a, algunos con $f(x) > f(a)$ y otros con $f(x) < f(a)$. Luego a no es ni máximo local ni mínimo local. El caso $g'(0) < 0$ es análogo.

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

Corolario 73. *Si f es diferenciable en a, para que a sea máximo local o mínimo local es necesario (no suficiente) que sea punto crítico.*
Sea f de clase $\mathcal{C}^2$ y $a \in U$ un punto crítico de f. Para que a sea máximo local es necesario (no suficiente) que Hess$(f)_a$ sea ***semidefinida negativa:*** *Hess$(f)_a(v) \le 0$ para todo $v \in \mathbb{R}^n$. Para que a sea mínimo local es necesario (no suficiente) que Hess$(f)_a$ sea* ***semidefinida positiva:*** *Hess$(f)_a(v) \ge 0$ para todo $v \in \mathbb{R}^n$.*
Si Hess$(f)_a$ es ***indefinida*** *(degenerada o no) entonces a 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 $y < 0$ sea $f(0, y) < f(0, 0)$, no importa lo pequeño que sea y.

Teorema 74. *Sea f de clase $\mathcal{C}^2$ y $a \in U$ un punto crítico suyo.*
Para que a sea un máximo local estricto es suficiente (no necesario) que Hess$(f)_a$ sea definida negativa. Para que a sea un mínimo local estricto es suficiente (no necesario) que Hess$(f)_a$ sea definida positiva.

Demostración. Escribamos $Q(\cdot) = \mathrm{Hess}(f)_a(\cdot)$ y supongamos $Q(\cdot)$ 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:

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

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 $f(a+h) = f(a) + 0 + (1/2)\, Q(h) + R(h)$ 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 $f(a)$ a los tres miembros de esta última desigualdad y poniendo $x = a + h$, obtenemos:

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

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

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

luego a es mínimo local estricto de f.

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

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.

2.10 Polinomios de Taylor

Definición 75. *Sean $U \subseteq \mathbb{R}^n$ un abierto y $f \in \mathcal{C}^k(U)$. Dado un punto $a \in U$, el* **polinomio de Taylor de orden k de f en** a *es el único polinomio de n variables $P_k(x) = P_k(x_1,\ldots,x_n)$ que tiene grado $\leq k$ y cumple lo siguiente:*

$$0 \;\leq\; |\alpha| \;\leq k \;\Longrightarrow\; D^\alpha P_k(a) \;=\; D^\alpha f(a)\;, \tag{23}$$

es decir, cada derivada en a de orden entre 0 y k da el mismo resultado para P_k que para f.

Se puede calcular el polinomio de Taylor por el **método de los coeficientes indeterminados,** o sea tratando a los coeficientes de P_k como incógnitas. Entonces (23) se convierte en un sistema con el mismo número de ecuaciones que de incógnitas (concretamente $\binom{n+k}{n}$, si f es una función escalar). La matriz de dicho sistema (cuadrada e invertible) es muy fea si como base del espacio de polinomios elegimos los productos de potencias de $x_1,\ldots,x_n$. Es, en cambio, muy sencilla si como base del espacio de polinomios elegimos los productos de potencias de $x_1 - a_1, \ldots, x_n - a_n$. Esto quiere decir que planteamos:

$$P_k(x_1,\ldots,x_n) \;=\; \sum_{0 \leq |\alpha| \leq k} c_{\alpha_1 \cdots \alpha_n}\, (x_1 - a_1)^{\alpha_1} \cdots (x_n - a_n)^{\alpha_n}\;,$$

y entonces la contribución a $D^\alpha P(a)$ de los términos $c_{\beta_1\cdots\beta_n}\, (x_1 - a_1)^{\beta_1} \cdots (x_n - a_n)^{\beta_n}$ con $\beta \neq \alpha$ es nula, con lo cual cada una de las condiciones (23) sólo contiene una incógnita y el sistema queda así:

$$0 \;\leq |\alpha| \;\leq k \;\Longrightarrow\; \alpha_1! \cdots \alpha_n!\, c_{\alpha_1\cdots\alpha_n} \;=\; D^\alpha f(a)\;.$$

Llegados a este punto se puede elegir entre *dos maneras equivalentes* de escribir el polinomio de Taylor. Una de ellas es con un sólo sumando por cada multíndice:

$$P_k(x) \;=\; \sum_{0 \leq |\alpha| \leq k} \frac{D^\alpha f(a)}{\alpha_1! \cdots \alpha_n!}\, (x_1 - a_1)^{\alpha_1} \cdots (x_n - a_n)^{\alpha_n}\;, \tag{24}$$

donde en cada monomio se escriben, de izquierda a derecha, primero la potencia de $x_1 - a_1$, luego la de $x_2 - a_2$ y así hasta la de $x_n - a_n$. La otra manera de escribir P_k es considerar las distintas expresiones que resultan de reordenar los factores de primer grado $x_i - a_i$ y repartir el coeficiente constante a partes iguales entre ellas. Ejemplos:

$$\begin{aligned}
&7\,(x_1-a_1)^2(x_5-a_5) \;= \\
&\quad = \frac{7}{3}\,(x_1-a_1)^2(x_5-a_5) + \frac{7}{3}\,(x_1-a_1)(x_5-a_5)(x_1-a_1) + \frac{7}{3}\,(x_5-a_5)(x_1-a_1)^2 \;, \\
&-11\,(x_1-a_1)(x_2-a_2)(x_3-a_3) \;=\; \sum_{\sigma\in S_3} \frac{-11}{3!}\,\big(x_{\sigma(1)}-a_{\sigma(1)}\big)\big(x_{\sigma(2)}-a_{\sigma(2)}\big)\big(x_{\sigma(3)}-a_{\sigma(3)}\big)\;.
\end{aligned}$$

Supuesto $|\alpha| = s$, al reordenar los factores de primer grado de $(x_1-a_1)^{\alpha_1}\cdots(x_n-a_n)^{\alpha_n}$ salen $\dfrac{s!}{\alpha_1!\cdots\alpha_n!}$ expresiones $\big(x_{i_1}-a_{i_1}\big)\cdots\big(x_{i_s}-a_{i_s}\big)$ diferentes. Por tanto:

$$c\,(x_1-a_1)^{\alpha_1}\cdots(x_n-a_n)^{\alpha_n} \;=\; \sum c\cdot\frac{\alpha_1!\cdots\alpha_n!}{s!}\,\big(x_{i_1}-a_{i_1}\big)\cdots\big(x_{i_s}-a_{i_s}\big)\;,$$

donde la suma se extiende a todas las posibles sucesiones $i_1,\ldots,i_s$ en las que el índice 1 aparece α_1 veces, el índice 2 aparece α_2 veces, ... y el índice n aparece α_n veces. Como en realidad el coeficiente c era igual a $D^\alpha f(a)/\alpha_1!\cdots\alpha_n!$, tenemos que:

$$c\cdot\frac{\alpha_1!\cdots\alpha_n!}{s!} \;=\; \frac{1}{s!}\,D^\alpha f(a)\;.$$

La segunda manera de escribir el polinomio de Taylor de orden k en a es, pues, la siguiente:

$$P_k(x) \;=\; \sum_{s=0}^{k}\frac{1}{s!}\sum_{1\le i_1,\ldots,i_s\le n} f_{x_{i_1}\cdots x_{i_s}}(a)\cdot(x_{i_1}-a_{i_1})\cdots(x_{i_s}-a_{i_s})\;. \tag{25}$$

Ambas fórmulas (24) y (25) son válidas tanto si f es escalar como si es vectorial $f : U \to \mathbb{R}^m$. Insistimos en que ambas definen el mismo polinomio (el único de grado k que cumple (23)) sólo que en (24) ningún monomio se repite mientras que en (25) cada monomio se repite tantas veces como distintas expresiones resultan de reordenar sus factores $x_i - a_i$.

Ejemplo. Si $k = 1$ las fórmulas (24) y (25) dan ambas el siguiente resultado:

$$P_1(x) \;=\; f(a) + (Df)_a\,(x-a)\;,$$

y el hecho de que el correspondiente resto $R(x) = f(x) - P_1(x)$ sea un $\mathrm{o}(\|x-a\|)$ equivale a la definición dada en el apartado 2.1 de que f sea diferenciable en el punto a.

Ejemplo. Sea $f(\mathbf{x}) = f(x,y)$ función de dos variables y sea $a = (0,5)$. Para cada multíndice $\alpha = (\alpha_1,\alpha_2)$, escribimos f_{α_1,α_2} para denotar $D^\alpha f(0,5)$. Entonces el polinomio de Taylor de orden 2 de f en a queda de la siguiente manera según la fórmula (24):

$$\begin{aligned}
&\frac{1}{0!0!}f_{0,0} + \frac{1}{1!0!}f_{1,0}x + \frac{1}{0!1!}f_{0,1}(y-5) + \frac{1}{2!0!}f_{2,0}x^2 + \frac{1}{1!1!}f_{1,1}x(y-5) + \frac{1}{0!2!}f_{0,2}(y-5)^2 \;= \\
&= \;f_{0,0} + f_{1,0}\cdot x + f_{0,1}\cdot(y-5) + \frac{1}{2}f_{2,0}\cdot x^2 + f_{1,1}\cdot x(y-5) + \frac{1}{2}f_{0,2}\cdot(y-5)^2\;,
\end{aligned} \tag{26}$$

mientras que según la fórmula (25) queda así:

$$\begin{aligned}
&\frac{1}{0!}f(0,5) \;+\; \frac{1}{1!}\Big(f_x(0,5)x + f_y(0,5)(y-5)\Big) + \\
&+\frac{1}{2!}\Big(f_{xx}(0,5)x^2 + f_{xy}(0,5)x(y-5) + f_{yx}(0,5)(y-5)x + f_{yy}(0,5)(y-5)^2\Big) \;= \\
&= \;f_{0,0} + f_{1,0}\cdot x + f_{0,1}\cdot(y-5) + \\
&+\frac{1}{2}f_{2,0}\cdot x^2 + \Big(\frac{1}{2}f_{1,1}\cdot x(y-5) + \frac{1}{2}f_{1,1}\cdot(y-5)x\Big) + \frac{1}{2}f_{0,2}\cdot(y-5)^2\;,
\end{aligned}$$

exactamente igual que la fórmula (26), sólo que con el término $f_{1,1}\cdot x(y-5)$ repartido como suma de dos mitades iguales. Estas mitades son las distintas expresiones que pueden salir como

resultado de reordenar los factores x e $y-5$ en $f_{1,1}\cdot x(y-5)$, divididas por el número 2 de tales resultados.
Vemos que, cuando f es escalar, la expresión de $P_2(\mathbf{x})$ según (25) es:

$$f(a)+Df_a\begin{bmatrix} x-0 \\ y-5 \end{bmatrix}+\frac{1}{2}[x-0 \quad y-5]\begin{bmatrix} f_{1,1} & f_{1,2} \\ f_{2,1} & f_{2,2} \end{bmatrix}\begin{bmatrix} x-0 \\ y-5 \end{bmatrix},$$

es decir el polinomio del teorema 70 del apartado 2.8:

$$f(a)+Df_a(\mathbf{x}-a)+(1/2)\,(\mathbf{x}-a)^t\,\mathrm{Hess}(f)_a\,(\mathbf{x}-a)\ .$$

Igual para funciones escalares de n variables y cualquier otro punto a. El teorema 70 afirma, pues, que si $P_2(x)$ es el polinomio de la definición 75 con $k=2$ entonces:

$$f(x)-P_2(x)\;=\;\begin{cases} \mathrm{o}(\|x-a\|^2) & \text{si} \quad f\in\mathcal{C}^2\ , \\ \\ \mathrm{O}(\|x-a\|^3) & \text{si} \quad f\in\mathcal{C}^3\ . \end{cases}$$

Ejemplo. Sean de nuevo $f(\mathbf{x})=f(x,y)$ y $a=(0,5)$. El polinomio de Taylor de orden 3 de f en a queda de la siguiente manera según la fórmula (24):

$$\begin{aligned} &f_{0,0}+f_{1,0}x+f_{0,1}(y-5)+\frac{1}{2}f_{2,0}x^2+f_{1,1}x(y-5)+\frac{1}{2}f_{0,2}(y-5)^2+ \\ &\quad+\frac{1}{3!0!}f_{3,0}x^3+\frac{1}{2!1!}f_{2,1}x^2(y-5)+\frac{1}{1!2!}f_{1,2}x(y-5)^2+\frac{1}{0!3!}f_{0,3}(y-5)^3\ , \end{aligned}$$

es decir:

$$\begin{aligned} &f_{0,0}+f_{1,0}\cdot x+f_{0,1}\cdot(y-5)+\frac{1}{2}f_{2,0}\cdot x^2+f_{1,1}\cdot x(y-5)+\frac{1}{2}f_{0,2}\cdot(y-5)^2+ \\ &\quad+\frac{1}{6}f_{3,0}\cdot x^3+\frac{1}{2}f_{2,1}\cdot x^2(y-5)+\frac{1}{2}f_{1,2}\cdot x(y-5)^2+\frac{1}{6}f_{0,3}\cdot(y-5)^3\ , \end{aligned} \tag{27}$$

mientras que según la fórmula (25) queda así:

$$\begin{aligned} &f_{0,0}+f_{1,0}\cdot x+f_{0,1}\cdot(y-5)+ \\ &+\frac{1}{2}\Big(f_{2,0}\cdot x^2+f_{1,1}\cdot x(y-5)+f_{1,1}\cdot(y-5)x+f_{0,2}\cdot(y-5)^2\Big)+ \\ &+\frac{1}{6}\Big(f_{xxx}(0,5)\cdot x^3+f_{yyy}(0,5)\cdot(y-5)^3\Big)+ \\ &+\frac{1}{6}\Big(f_{xxy}(0,5)\cdot x^2(y-5)+f_{xyx}(0,5)\cdot x(y-5)x+f_{yxx}(0,5)\cdot(y-5)x^2\Big)+ \\ &+\frac{1}{6}\Big(f_{xyy}(0,5)\cdot x(y-5)^2+f_{yxy}(0,5)\cdot(y-5)x(y-5)+f_{yyx}(0,5)\cdot(y-5)^2x\Big)\ , \end{aligned}$$

exactamente igual que la fórmula (27), sólo que con el término $f_{1,1}\cdot x(y-5)$ repartido en dos mitades iguales y cada uno de los términos $(1/2)\,f_{2,1}\cdot x^2(y-5)$ y $(1/2)\,f_{1,2}\cdot x(y-5)^2$ repartidos en tres partes iguales.

El siguiente teorema es a veces muy útil. Daremos un ejemplo de su uso después de demostrarlo.

Teorema 76. *Fijamos un entero positivo k. Sean un abierto $U\subseteq\mathbb{R}^n$, una función $f:U\to\mathbb{R}^m$ de clase $\mathcal{C}^k$ y un polinomio $P(x):\mathbb{R}^n\to\mathbb{R}^m$ de grado $\le k$. Dado $a=(a_1,\dots,a_n)\in U$, son equivalentes:*

1. *$P(x)$ es el polinomio de Taylor de orden k de f en a, según la definición 75.*
2. *La función $R(x)=f(x)-P(x)$ es un $o(\|x-a\|^k)$ cerca de a.*

Si f es también de clase $\mathcal{C}^{k+1}$, entonces la función $R(x)$ es un $O(\|x-a\|^{k+1})$ cerca de a.

El caso particular $k=2, m=1$ de la implicación 1.$\Longrightarrow$2. es el teorema 70 para $f \in \mathcal{C}^2$.

Demostración del teorema 76.
1.$\Longrightarrow$2. Sea $h = x - a$. Definimos la función auxiliar de una variable $g(t) = f(a+th)$, consideramos el siguiente desarrollo de Taylor:

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

donde $0<\theta<1$, y desarrollamos:

$$g^{(s)}(t) \;=\; \sum_{1\le i_1\,\ldots,i_s\le n} f_{x_{1_1}\cdots x_{i_s}}(a+th)\, h_{i_1}\cdots h_{i_s} \quad , \quad \text{para } s=1,\ldots,k\ .$$

Entonces, teniendo en cuenta la fórmula (25) para P_k, conseguimos llegar a la expresión:

$$f(x)-P_k(x) \;=\; \sum_{1\le i_1\,\ldots,i_k\le n} \big(f_{x_{1_1}\cdots x_{i_k}}(y) - f_{x_{1_1}\cdots x_{i_k}}(a) \big)\, h_{i_1}\cdots h_{i_k} \quad , \quad \text{con } y \;=\; a+\theta h\ ,$$

de donde:

$$\frac{|R_k|}{\|h\|^k} \;\le\; \sum_{1\le i_1\,\ldots,i_k\le n} \big|\, f_{x_{1_1}\cdots x_{i_k}}(y) - f_{x_{1_1}\cdots x_{i_k}}(a)\,\big| \;=\; \mathrm{o}(1)\ ,$$

luego $R_k(x) = \mathrm{o}(\|h\|^k)$.

2.$\Longrightarrow$1. Sean $P(x)$ un polinomio de grado $\le k$ que cumple 2. y $P_k(x)$ el polinomio de Taylor de orden k de f en a. Sabemos que $f(x) = P_k(x) + \mathrm{o}(\|x-a\|^k)$ por la primera parte de la demostración, de donde $P(x)-P_k(x) = \big(P(x)-f(x)\big) + \mathrm{o}(\|x-a\|^k)$. Luego $P(x)$ cumple 2. si y sólo si $P(x)-P_k(x)$ es un $\mathrm{o}(\|x-a\|^k)$.
Hacemos de nuevo $h = x-a$ y utilizamos como base del espacio de polinomios los productos de potencias de $h_1,\ldots,h_n$. Como $P(x)-P_k(x)$ tiene grado $\le k$, hay una identidad:

$$P(x)-P_k(x) \;=\; Q_0 + Q_1(h) + Q_2(h) + \cdots + Q_k(h)\ ,$$

donde Q_0 es una constante, $Q_1(h)$ es una forma lineal en h, $Q_2(h)$ es una forma cuadrática en h,... y $Q_k(h)$ es un polinomio homogéneo de grado k en h.
Haremos uso de lo siguiente, que es muy fácil de demostrar:

$$\left.\begin{aligned} \varphi(h)\ \text{ homogénea de grado } s \text{ en } h \\ \varphi(h) \;=\; \mathrm{o}(\|h\|^s) \end{aligned}\right\} \Longrightarrow\ \varphi(h)\equiv 0\ . \tag{28}$$

Supongamos que $P(x)$ cumple 2. Empezamos razonando así:

$$Q_0 \;=\; \big(P(x)-P_k(x)\big) - \big(\, Q_1(h)+\cdots+Q_k(h)\,\big) \;=\; \mathrm{o}(\|h\|^k) + \mathrm{o}(1) \;=\; \mathrm{o}(1)\ ,$$

y, haciendo $s=0$ en (28), deducimos $Q_0 = 0$. Pero entonces:

$$Q_1(h) \;=\; \big(P(x)-P_k(x)\big) - \big(\, Q_2(h)+\cdots+Q_k(h)\,\big) \;=\; \mathrm{o}(\|h\|^k) + \mathrm{o}(\|h\|) \;=\; \mathrm{o}(\|h\|)\ ,$$

y, haciendo $s=1$ en (28), deducimos $Q_1(h)\equiv 0$. Este proceso continúa sin problema hasta $s=k-1$ inclusive, luego en realidad $Q_0, Q_1(h),\ldots,Q_{k-1}(h)$ son todos nulos y se tiene:

$$Q_k(h) \;=\; P(x)-P_k(x) \;=\; \mathrm{o}(\|h\|^k)\ ,$$

luego $Q_k(h) = P(x)-P_k(x)\equiv 0$, es decir $P(x)\equiv P_k(x)$ que es lo que se afirma en 1.

El caso $f\in\mathcal{C}^{k+1}$. La función auxiliar $g(t) = f(a+th)$ es ahora de clase $\mathcal{C}^{k+1}$ y admite el siguiente desarrollo:

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

donde $0 < \widetilde{\theta} < 1$. Desarrollando cada $g^{(s)}(t) = \dfrac{d^s}{dt^s} f(a+th)$ y evaluando en $t = 0$ o en $t = \widetilde{\theta}$, según proceda, obtenemos:

$$f(x) - P_k(x) \;=\; \frac{1}{(k+1)!} \sum_{1 \leq 1_1, \ldots, i_{k+1} \leq n} f_{x_{i_1} \cdots x_{i_{k+1}}}(z)\, h_{1_1} \cdots h_{1_{k+1}} \quad , \quad \text{con } \; z = a + \widetilde{\theta} h \;,$$

y es muy fácil ver que $f(x) - P_k(x) = \mathrm{O}(\|h\|^{k+1}) = \mathrm{O}(\|x-a\|^{k+1})$, para x cercano al punto a. □

La utilidad de la parte 1.$\Longrightarrow$2. del teorema 76 es obvia: generaliza y refuerza el teorema 70 del apartado 2.8. Vamos a mostrar, con un ejemplo, que la parte 2.$\Longrightarrow$1. es también muy útil. Consideramos la función $f(x,y) = \dfrac{e^{xy}}{1-y^2}$, que es $\mathcal{C}^\infty$ cerca de $(0,0)$. El cálculo elemental de las derivadas parciales sucesivas de un cociente da lugar a expresiones sumamente largas (si no lo crees, prueba a hallar las derivadas sucesivas de $\tan t = \operatorname{sen} t / \cos t$). Sin embargo, con muy pocos cálculos vamos a ver que se tienen los siguiente valores:

$$f(0,0) = 1 \;,\; f_{xy}(0,0) = 1 \;,\; f_{yy}(0,0) = 2 \;,\; f_{xxyy}(0,0) = 2 \;,\; f_{xyyy}(0,0) = 6 \;,\; f_{yyyy}(0,0) = 24 \;,$$

y todas las demás derivadas de órdenes de 0 a 5 de f en $(x,y) = (0,0)$ son nulas.
Empezamos por recordar el desarrollo $e^t = 1 + t + (1/2)\, t^2 + \mathrm{O}(|t|^3)$, que nos da:

$$e^{xy} \;=\; 1 + xy + \frac{1}{2}\, x^2y^2 + \mathrm{O}(|xy|^3) \;=\; 1 + xy + \frac{1}{2}\, x^2y^2 + \mathrm{O}\big(\|(x,y)\|^6\big) \;.$$

Recordamos también que $\dfrac{1}{1-t} = 1 + t + t^2 + \mathrm{O}(|t|^3)$, de donde:

$$\frac{1}{1-y^2} \;=\; 1 + y^2 + y^4 + \mathrm{O}(|y|^6) \;=\; 1 + y^2 + y^4 + \mathrm{O}\big(\|(x,y)\|^6\big) \;.$$

Por lo tanto:

$$\begin{aligned} f(x,y) \;&=\; \left[1 + xy + \frac{1}{2}\, x^2y^2 + \mathrm{O}\big(\|(x,y)\|^6\big) \right] \cdot \big[\, 1 + y^2 + y^4 + \mathrm{O}\big(\|(x,y)\|^6\big) \,\big] \;= \\ &=\; 1 + xy + y^2 + \frac{1}{2}\, x^2y^2 + xy^3 + y^4 + \mathrm{O}\big(\|(x,y)\|^6\big) \;= \\ &=\; 1 + xy + y^2 + \frac{1}{2}\, x^2y^2 + xy^3 + y^4 + \mathrm{o}\big(\|(x,y)\|^5\big) \;. \end{aligned}$$

Entonces la parte 2.$\Longrightarrow$1. del teorema 76 nos dice que el polinomio de Taylor de orden 5 de f en $(0,0)$ es el siguiente:

$$P_5(x,y) \;=\; 1 + xy + y^2 + \frac{1}{2}\, x^2y^2 + xy^3 + y^4 \;,$$

y ahora sabemos que los valores en $(0,0)$ de la derivadas de f de órdenes de 0 a 5 son los arriba indicados, sin necesidad de haber calculado ninguna derivada.

Aviso. La hipótesis $f \in \mathcal{C}^k$ en el enunciado del teorema 76 es muy importante. Esto ya se ve con funciones de una variable. Por ejemplo la función $f(x) : \mathbb{R} \to \mathbb{R}$ dada por

$$f(x) \;=\; \begin{cases} x^4 \operatorname{sen} \dfrac{1}{x^4} & \text{si} \quad x \neq 0 \\[2ex] \qquad 0 & \text{si} \quad x = 0 \end{cases}$$

satisface $f(x) = \mathrm{o}(|x|^3)$, pero no podemos deducir que $f(0) = f'(0) = f''(0) = f'''(0) = 0$ por la sencilla razón de que la función $f'(x)$ existe pero es discontinua en $x = 0$ (de hecho f' es no acotada cerca de 0), con lo cual f'' y f''' ni siquiera existen en ese punto.