2 Diferenciabilidad y extremos locales

2.1 Diferenciabilidad

En esta sección vamos a estudiar la generalización del concepto de derivada. Lo haremos primero para espacios normados cualesquiera; después nos concentraremos en el caso de $\mathbb{R}^n$.

2.1.1 Oes de Landau

Empezamos con dos espacios normados E, F y un punto $x_0 \in E$. En un entorno $U \subseteq E$ de x_0, tenemos definidas una aplicación $f : U \to F$ y una función escalar $\varphi : U \to [0, +\infty)$ con $\varphi(x) > 0$ para $x \neq x_0$.

Definición 66. *Decimos que* **f es una o grande de Landau de $\boldsymbol{\varphi}$,** *y lo indicamos escribiendo* $\boxed{f(x) = \mathrm{O}\big(\varphi(x)\big)}$ *si existe una constante $C \geq 0$ tal que*

$$\big\| \, f(x) \, \big\| \;\leq\; C \, \varphi(x) \quad , \quad \textit{para todo } x \in U \; .$$

Decimos que **f es una o pequeña de Landau de $\boldsymbol{\varphi}$ cuando x tiende a $\mathbf{x_0}$,** *y se indica escribiendo* $\boxed{f(x) = \mathrm{o}\big(\varphi(x)\big)}$ *si se tiene*

$$\lim_{\substack{x \to x_0 \\ x \neq x_0}} \frac{\big\| f(x) \big\|}{\varphi(x)} \;=\; 0 \; .$$

Es obvio que si $f = o(\varphi)$ entonces $f = O(\varphi)$, pero el recíproco suele ser falso.

Ejemplos con $E = F = \mathbb{R}$:

$\operatorname{sen} x = \mathrm{O}(|x|)$, $x_0 = 0$, $U = \mathbb{R}$.

$x^3 \sqrt{x^2 + 10x^4} = \mathrm{O}(|x|^4)$, $x_0 = 0$, $U = (-3, 3)$.

$e^x = \mathrm{O}(1)$, $U = (a, b)$ intervalo finito.

$e^x - 1 = \mathrm{O}(|x|)$, $x_0 = 0$, $U = (-1, 1)$.

$e^x - 1 - x - (x^2/2) - (x^3/6) = \mathrm{O}(|x|^4)$, $x_0 = 0$, $U = (-1, 1)$.

$|x|^{1/2} x = \mathrm{o}(|x|)$, $x_0 = 0$.

$|x|^{1/2} \operatorname{sen} x = \mathrm{o}\big(|x|^{4/3}\big)$, $x_0 = 0$.

Ejemplos con $E = \mathbb{R}^2$, $F = \mathbb{R}$ y $\| \cdot \|$ *cualquier* norma en $\mathbb{R}^2$:

$x^3 - 7x^5 y + 4y^3 = \mathrm{O}\big(\|(x, y)\|^3\big)$, $x_0 = (0, 0)$, $U = B(x_0, 1)$.

$\sqrt[14]{x^2 + 3y^2}\, \log(x + 1) = \mathrm{o}\big(\|(x, y)\|\big)$, $x_0 = (0, 0)$.

Si $\alpha < \beta$, entonces:

$$\|x\|^\beta \;=\; \mathrm{o}\big(\|x\|^\alpha\big) \quad , \quad x_0 \;=\; \mathbf{0} \; .$$

En particular $1 < \beta \implies \|x\|^\beta = \mathrm{o}(\|x\|)$, luego una condición *suficiente* (no necesaria) para que sea $f(x) = \mathrm{o}(\|x\|)$ cuando $x \to \mathbf{0}$ es que sea $f(x) = \mathrm{O}\big(\|x\|^\beta\big)$ para algún $\beta > 1$.

2.1.2 Diferenciabilidad en un punto

Definición 67. *Sean: E, F espacios normados, un punto $x_0 \in E$ y un abierto $U \subseteq E$ entorno de x_0. Decimos que una aplicación $f : U \to F$ es* **diferenciable en $\mathbf{x_0}$** *si existe $T : E \to F$, lineal acotada, tal que:*

$$f(x_0 + h) - f(x_0) - T(h) \;=\; \mathrm{o}\big(\|h\|\big) \quad , \quad \text{cuando } \; h \to \vec{0}_E \; . \tag{15}$$

Es decir, que para todo $\varepsilon > 0$ existe un $\delta > 0$ tal que

$$0 < \|h\| < \delta \implies \frac{\|f(x_0 + h) - f(x_0) - T(h)\|}{\|h\|} < \varepsilon \; , \tag{16}$$

o bien, de manera equivalente: $\|h\| < \delta \;\Rightarrow\; \| \, f(x_0 + h) - f(x_0) - T(h) \, \| \leq \varepsilon \, \|h\|$.
Decimos que f es **diferenciable en U** *si es diferenciable en cada punto de U.*

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 (15). Definimos $T = T_1 - T_2$, que es lineal y tal que $\lim_{h \to \vec{0}_E} \|T(h)\|/\|h\| = 0$. De esto se deduce que la norma de operador $\|T\|$ es nula, luego $T \equiv 0$ y $T_1 = T_2$.

Demostración de b). Sea $T = (df)_{x_0}$. Tomamos el valor δ_1 tal que se cumple (16) 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. □

Importante. El concepto de diferenciable sólo lo definimos para aplicaciones definidas en abiertos. Esto permite que nos podamos mover, a patir del punto x_0, una cierta distancia r *en todas las direcciones,* haciendo más útiles y significativas las fórmulas (15) y (16).

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 ocurre en F. El que una función $R(h)$, de un entorno de x_0 en E a F, sea un $\mathrm{o}(\|h\|)$ es algo que no depende de las normas elegidas en E y en F. Tanto la diferenciabilidad de f en x_0 como la diferencial $(df)_{x_0}$ son, pues, independientes de esas normas.

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 $\displaystyle \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$ existe y es igual a m. Esto equivale a que f sea derivable en x_0 con derivada finita $f'(x_0) = m$. Como hay muchas funciones continuas no derivables $\mathbb{R} \to \mathbb{R}$, queda claro que continuidad no implica diferenciabilidad.

2.1.3 Primeras reglas

Proposición 68. *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 69. (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:*

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

Proposición 70. (Regla de Leibniz). *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*

$$\boxed{d(fg)_{x_0}(v) = (df)_{x_0}(v)g(x_0) + f(x_0)(dg)_{x_0}(v) \quad , \quad \text{para todo } v \in E}$$

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

El caso escalar de esta proposición se demuestra en el apartado 2.1.10.

2.1.4 Regla de la cadena para diferenciales

Veamos lo que ocurre cuando efectuamos la composición de dos aplicaciones diferenciables.

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.

Proposición 71. (Regla de la cadena). *La compuesta* $g \circ f$ *es diferenciable en* x_0 *y su diferencial* $d(g \circ f)_{x_0}$ *es 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. Sean $L_1 = (df)_{x_0}$ y $L_2 = (dg)_{y_0}$. Definimos los restos $R_1 : E \to F$ y $R_2 : F \to G$ por las identidades:

$$R_1(h) \;=\; f(x_0 + h) - f(x_0) - L_1(h) \quad , \quad R_2(k) \;=\; g(y_0 + k) - g(y_0) - L_2(k)\ .$$

Entonces $R_1(h) = \mathrm{o}(\|h\|)$ y $R_2(k) = \mathrm{o}(\|k\|)$. Tenemos:

$$f(x_0 + h) \;=\; f(x_0) + L_1(h) + R_1(h) \;=\; y_0 + k\ , \quad \text{donde} \left\{ \begin{array}{rcl} y_0 & = & f(x_0) \\ k & = & L_1(h) + R_1(h) \end{array} \right.$$

que llevado al desarollo $g(y_0 + k) = g(y_0) + L_2(k) + R_2(k)$ nos da:

$$\begin{array}{rcl} (g \circ f)(x_0 + h) & = & g(y_0 + k) \;=\; g(y_0) + L_2(k) + R_2(k) \;=\; g \circ f(x_0) + L_2(k) + R_2(k) \;= \\ & = & (g \circ f)(x_0) + L_2\big(\, L_1(h) + R_1(h)\,\big) + R_2\big(\, L_1(h) + R_1(h)\,\big)\ , \end{array}$$

es decir:

$$(g \circ f)(x_0 + h) \;=\; (g \circ f)(x_0) + (L_2 \circ L_1)(h) + R(h)\ , \tag{17}$$

siendo

$$R(h) \;=\; L_2\big(R_1(h)\big) + R_2\big(\, L_1(h) + R_1(h)\,\big)\;. \tag{18}$$

Estimamos el primero de los dos sumandos que forman $R(h)$:

$$\big\| L_2\big(R_1(h)\big)\big\| \;\le\; \|L_2\|\,\big\|R_1(h)\big\| = \mathrm{cte}\cdot \mathrm{o}(\|h\|) \;=\; \mathrm{o}(\|h\|)\;. \tag{19}$$

Ahora vamos a estimar $R_2\big(L_1(h)+R_1(h)\big)$. Para $\|h\|$ pequeña se tiene $\big\|R_1(h)\big\| \le \|h\|$, luego para $k = L_1(h)+R_1(h)$ se cumple:

$$\|k\| \;=\; \big\|L_1(h)+R_1(h)\big\| \;\le\; \big(\|L_1\|+1\big)\,\|h\|\;. \tag{20}$$

Dado $\varepsilon > 0$, existe $\delta > 0$ tal que $\|k\| < \delta \implies \big\|R_2(k)\big\| < \varepsilon\|k\|$, y entonces (20) nos permite escribir:

$$\begin{aligned} \|h\| \quad <\quad \delta' := \frac{\delta}{\|L_1\|+1} \quad &\Longrightarrow \|k\| \;=\; \big\|L_1(h)+R_1(h)\big\| \;<\; \delta \;\;\Longrightarrow \\ \Longrightarrow\;\; \big\|R_2\big(\,L_1(h)+R_1(h)\,\big)\big\| \;&=\; \big\|R_2(k)\big\| \;<\; \varepsilon\|k\| \;\le\; \varepsilon\,\big(\|L_1\|+1\big)\,\|h\|\;, \end{aligned}$$

y, como $\varepsilon\,(\|L_1\|+1)$ se hace arbitrariamente pequeño cuando $\varepsilon \searrow 0$, deducimos:

$$R_2\big(L_1(h)+R_1(h)\big) \;=\; \mathrm{o}(\|h\|)\;. \tag{21}$$

Las fórmulas (18), (19) y (21) nos dan $R(h) = \mathrm{o}(\|h\|)$, convirtiendo (17) en la siguiente igualdad:

$$(g\circ f)(x_0+h) \;=\; (g\circ f)(x_0) + (L_2\circ L_1)(h) + \mathrm{o}(\|h\|)\;,$$

que nos dice que $g\circ f$ es diferenciable en x_0 con diferencial $L_2\circ L_1$. □

2.1.5 Derivada respecto de un vector

Definición 72. *Sean E, F espacios normado y $U\subseteq E$ entorno de $x_0\in E$. Sean $f: U\to F$ y $v\in E$. La* **derivada en $\mathbf{x_0}$ de f respecto de v** *es el siguiente límite, si es que existe:*

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

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: $D_v f(x_0)$ existe $\Rightarrow$ $D_{cv} f(x_0)$ existe y es $c\cdot D_v f(x_0)$.

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

$$D_v f(x_0) \;=\; (df)_{x_0}(v) \quad,\quad \text{para todo } v\in E\;.$$

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:

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

□

2.1.6 Dimensiones finitas: matriz jacobiana

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 $\mathbb{R}^m$.

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 $\mathbf{x_0}$.** 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 los $D_{\mathbf{e}_i} f(x_0)$ son vectores de $\mathbb{R}^m$ (vectores columna de altura m).

Si f es diferenciable en x_0, entonces $(df)_{x_0}$ es la única aplicación lineal $L : \mathbb{R}^n \to \mathbb{R}^m$ tal que $L(\mathbf{e}_i) = D_{\mathbf{e}_i} f(x_0) = f_{x_i}(x_0)$ para $i = 1, \ldots, n$. Esto significa que $L(x) \equiv Ax$, siendo

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

la matriz $m \times n$ cuyas columnas son $f_{x_1}(x_0), f_{x_2}(x_0), \ldots, f_{x_n}(x_0)$. Esta matriz se llama **matriz jacobiana de f en $\mathbf{x_0}$** y se la denota Df_{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* $f_1, \ldots, f_m$:

$$D\begin{bmatrix} f_1 \\ \vdots \\ f_m \end{bmatrix} \;=\; \begin{bmatrix} Df_1 \\ \hline Df_2 \\ \hline \vdots \\ \hline Df_m \end{bmatrix}_{m\times n}\;.$$

Sea como sea, si f es diferenciable en x_0 entonces su diferencial en ese punto viene dada por

$$(df)_{x_0}(x) \;=\; (Df_{x_0})\, x \quad , \quad \text{para todo } x \in \mathbb{R}^n\;.$$

2.1.7 Regla de la cadena para jacobianas

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

$$\boxed{D(g\circ f)_{x_0} \;=\; (Dg)_{f(x_0)}\, Df_{x_0}} \tag{22}$$

Veamos primero el caso particular más importante de esta regla: cuando f es un camino.

Definición 73. *Sea $V \subseteq \mathbb{R}^m$ un abierto. Un* **camino diferenciable en V** *viene dado por un intervalo $I \subseteq \mathbb{R}$ y una aplicación $\alpha(t) : I \to \mathbb{R}^m$ diferenciable en cada $t \in I$ y tal que $\alpha(I) \subseteq V$.*

Si denotamos por $(y_1, \ldots, y_m)$ las coordenadas estándar en $\mathbb{R}^m$, entonces cada camino tiene una descripción $\alpha(t) \equiv \big(y_1(t)\,,\, \ldots\,,\, y_m(t)\big)$, siendo $y_1(t), \ldots, y_m(t) : I \to \mathbb{R}$ funciones escalares diferenciables en todo $t \in I$. Este caso es excepcional por dos razones:

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

– Dado $t_0 \in I$, que existan las derivadas $y_1'(t_0), \ldots, y_m'(t_0)$ es suficiente para que $\alpha(t)$ sea diferenciable en t_0 (no es tan sencillo para funciones de dos o más variables).

La matriz jacobiana es una *columna:* $D\alpha_{t_0} = \alpha'(t_0) = \begin{bmatrix} y_1'(t_0) \\ \vdots \\ y_m'(t_0) \end{bmatrix}$, es decir un vector de $\mathbb{R}^m$, que llamamos **vector derivada** o **vector velocidad.** Los **vectores tangentes** al camino en $t = t_0$ son los múltiplos $c\,\alpha'(t_0)$ con $c \in \mathbb{R}$.

Sea ahora $g(y_1, \ldots, y_m) : V \to \mathbb{R}^k$ una función de m variables, diferenciable en cada punto imagen del camino $\alpha(t) \in \alpha(I)$. La compuesta $g \circ \alpha : I \to \mathbb{R}^k$ es un camino diferenciable en $\mathbb{R}^k$ y la regla de la cadena $d(g \circ \alpha)_{t_0} = (dg)_{\alpha(t_0)} \circ (d\alpha)_{t_0}$ se traduce en la igualdad matricial

$$(g \circ \alpha)'(t_0) \;=\; D(g \circ \alpha)_{t_0} \;=\; (Dg)_{\alpha(t_0)}\, D\alpha_{t_0} \;=\; (Dg)_{\alpha(t_0)}\, \alpha'(t_0)\ ,$$

que es del tipo "columna = rectángulo · columna"

$$\mathbb{R}^k \ni (g \circ \alpha)'(t_0) \;=\; \Big[g_{y_1}\big(\alpha(t_0)\big) \,\Big|\, g_{y_2}\big(\alpha(t_0)\big) \,\Big|\, \cdots \,\Big|\, g_{y_m}\big(\alpha(t_0)\big) \Big]_{k\times m} \begin{pmatrix} y_1'(t_0) \\ \vdots \\ y_m'(t_0) \end{pmatrix}_{m\times 1} ,$$

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

$$\boxed{\frac{d}{dt}\, g\big(\alpha(t)\big) \;=\; \frac{d}{dt}\, g\big(\, y_1(t)\,,\, \ldots\,,\, y_m(t)\,\big) \;=\; y_1'(t)\, g_{y_1}\big(\alpha(t)\big) + \cdots + y_m'(t)\, g_{y_m}\big(\alpha(t)\big) \;\in \mathbb{R}^k} \tag{23}$$

Ahora mostraremos que el caso general de la regla de la cadena consiste en realidad en varios casos particulares de (23). 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 un punto $x_0 \in U$ y con sus valores en V. Fijados x_0 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 (23) en $t = 0$ 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 pedir que sea $(g \circ f)_{x_i}(x_0) = (Dg)_{y_0}\, f_{x_i}(x_0)$ para $i = 1, \ldots, n$, y esto 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 (22), 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 $\mathbf{i}$ toma cualquier valor entre 1 y n:

$$\boxed{\left.\frac{\partial}{\partial x_{\mathbf{i}}}\right|_{x_0} g\big(f_1(x)\,,\, \ldots\,,\, f_m(x)\big) \;=\; g_{y_1}\big(f(x_0)\big)\, \frac{\partial f_1}{\partial x_{\mathbf{i}}}(x_0) + \cdots + g_{y_m}\big(f(x_0)\big)\, \frac{\partial f_m}{\partial x_{\mathbf{i}}}(x_0)} \tag{24}$$

fórmulas que resultan de sustituir en (23) el operador $\dfrac{d}{dt}$ por los operadores $\left.\dfrac{\partial}{\partial x_i}\right|_{x_0}$.

2.1.8 Estudio de las condiciones de diferenciabilidad

Sean un abierto $V \subseteq \mathbb{R}^m$ y una aplicación $g : V \to \mathbb{R}^k$. Dado $y_0 \in V$, para que g sea diferenciable en y_0 es necesario y suficiente que se cumplan las tres condiciones siguientes:

dif1: Para todo $v \in \mathbb{R}^m$ debe existir la derivada $D_v g(y_0)$.

dif2: La aplicación $\mathbb{R}^m \to \mathbb{R}^k$ dada por $v \longmapsto D_v g(y_0)$ debe ser lineal.

dif3: La expresión $g(y_0 + h) - g(y_0) - D_h g(y_0)$ debe ser un $\mathrm{o}(\|h\|)$.

Supongamos que g satisface la condición **dif1.** Satisfará también la condición **dif2** si para todo $v = (a_1, \ldots, a_m) \in \mathbb{R}^m$ se cumple

$$D_{a_1 \mathbf{e}_1 + \cdots + a_m \mathbf{e}_m}\, g\,(y_0) \;=\; a_1\, D_{\mathbf{e}_1} g(y_0) + \cdots + a_m\, D_{\mathbf{e}_m} g(y_0)\,. \tag{25}$$

Considerando el camino rectilíneo $\alpha(t) \equiv y_0 + tv$, de velocidad constante v, la fórmula (25) se convierte en

$$\left.\frac{d}{dt}\right|_{t=0} g\big(\alpha(t)\big) \;=\; a_1\, g_{y_1}\big(\alpha(0)\big) + \cdots + a_m\, g_{y_m}\big(\alpha(0)\big)\,,$$

y llegamos a esta conclusión:

> 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 (23) **para caminos rectilíneos** en un tiempo t en el que pasen por y_0.

Para caminos diferenciables cualesquiera, y g diferenciable en y_0, la regla de la cadena afirma que g cumple dos condiciones:

cadena1: Fijado v, para todos los infinitos caminos $\alpha(t)$ que en el instante $t = 0$ pasan por y_0 con velocidad v la derivada $(g \circ \alpha)'(0)$ existe y es la misma. Esto incluye al camino rectilíneo.

cadena2: Para los caminos $\alpha(t)$ con $\alpha(0) = y_0$, el vector $(g \circ \alpha)'(0)$ depende *linealmente* de $\alpha'(0)$.

Resulta que hay un recíproco de la regla de la cadena:

Teorema 74. *Sean un abierto $V \subseteq \mathbb{R}^m$, una aplicación $g : V \to \mathbb{R}^k$ y un punto $y_0 \in V$. Son equivalentes:*

– *g es diferenciable en y_0.*

– *g cumple las condiciones* **dif1, dif2** *y* **dif3.**

– *g cumple las condiciones* **cadena1** *y* **cadena2.**

No se puede comprobar directamente si g cumple la regla de la cadena para la infinidad de caminos que pasan por y_0: si g es diferenciable en y_0, la manera de demostrarlo es con las condiciones **dif 1, dif2** y **dif 3.** Pero si conocemos un camino particular $\alpha(t)$, con $\alpha(0) = y_0$, para el cual g no cumple la regla de la cadena en $t = 0$, entonces ya sabemos que g no es diferenciable en y_0.

2.1.9 Ejemplos especiales

El propósito de este apartado es aportar evidencia de que, para funciones de dos o más variables, la diferenciabilidad de f en un punto x_0 es una propiedad muy exigente. Los tres ejemplos cumplen **dif1.**

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:

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

De hecho f cumple la condición **cadena1,** porque se puede demostrar que para caminos diferenciables cualesquiera $\alpha(t)$ con $\alpha(0) = (0, 0)$ esta función satisface:

$$\left.\frac{d}{dt}\right|_{t=0} f\big(\alpha(t)\big) \;=\; f\big(\alpha'(0)\big) \;=\; \text{valor que sólo depende de } \alpha'(0)\,.$$

Observa esto: la condición **dif2** es la restricción de **cadena2** a caminos rectilíneos, luego **dif2** es más débil que **cadena2.**
Pues bien: $f(x,y)$ ni siquiera cumple **dif2** (luego tampoco **cadena 2**), porque $f(v) \equiv D_v f(0,0)$ no es una función lineal. Luego $f(x,y)$ no es diferenciable en $(0,0)$.
Este ejemplo nos sirve también para ver que **cadena1** $\not\Longrightarrow$ **cadena 2,** luego ambas condiciones hacían falta en el enunciado del teorema 74.

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

La derivada $D_v g(\mathbf{0})$ 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 $v = (a,b) \in \mathbb{R}^2$. O sea que g cumple la regla de la cadena a lo largo de rectas pasando por $(0,0)$.
Si $a = 0$ o $b = 0$, 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:

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

Veamos, sin embargo, que 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 $t \searrow 0$, ni siquiera existe el *límite de dos variables* $\lim_{(x,y)\to(0,0)} g(x,y)$. Al ser g discontinua en $(0,0)$, no es diferenciable en este punto.

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. Es decir que 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 (desigualdad de Young para $p = 2$), luego $|h(x,y)| \leq |x|/2 \leq \|(x,y)\|_\infty/2$ en todo el plano $\mathbb{R}^2$. Al ser $h = \mathrm{O}(\|(x,y)\|_\infty)$, se tiene $\lim_{(x,y)\to(0,0)} h(x,y) = 0 = h(0,0)$.
Pero h no cumple **cadena1,** luego no es diferenciable en $(0,0)$. Los caminos $\alpha(t) = (t,0)$ y $\gamma(t) = (t,t^2)$ pasan por $(0,0)$ cuando $t = 0$ con la misma velocidad $v = (1,0)$. Como α es rectilíneo, se tiene $(h \circ \alpha)'(0) = D_{(1,0)} h(0,0) = 0$, en cambio $(h \circ \gamma)'(0) = \dfrac{d}{dt}\Big|_{t=0} \dfrac{t}{2} = \dfrac{1}{2} \neq 0$.
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, por ejemplo $\gamma(t) \equiv \big(x(t), y(t)\big)$ nos ha dado:

$$(h \circ \gamma)'(0) = \frac{1}{2} \neq 0 = x'(0) \cdot 0 + y'(0) \cdot 0 = x'(0)\, h_x\big(\gamma(0)\big) + y'(0)\, h_y\big(\gamma(0)\big)\,.$$

Este ejemplo nos sirve también para ver que **dif1+dif2** $\not\Longrightarrow$ **cadena1.**

2.1.10 Funciones de clase $\mathcal{C}^1$

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 75. *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} y 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:

$$|x| \;=\; \|(x,0)\| \;\le\; \|(x,y)\| \;\ge\; \|(0,y)\| \;=\; |y| \quad \text{para todo} \quad (x,y) \in \mathbb{R}^2\ . \tag{26}$$

Esta propiedad la tienen muchas normas, entre otras las normas p. Una vez que se cumple (26), 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:

$$f(x'') - f(x_0) \;=\; \big(\, f(x') - f(x_0)\,\big) + \big(\, f(x'') - f(x')\,\big) \;=\; (c-a)\, f_{x_1}(z_1) + (d-b)\, f_{x_2}(z_2)\ . \tag{27}$$

Escribamos ahora:

$$f_{x_1}(z_1) \;=\; f_{x_1}(x_0) + \text{error}_1 \quad , \quad f_{x_2}(z_2) \;=\; f_{x_2}(x_0) + \text{error}_2\ . \tag{28}$$

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 (27) con (28) y definiendo $\text{error} = \text{error}_1\,(c-a) + \text{error}_2\,(d-b)$ sale:

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

Ya sólo nos falta ver que $\text{error} = \text{o}(\|x'' - x_0\|)$. Ahora bien, por la condición (26) se tiene:

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

de donde:

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

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.

La función $f(x,y) = \begin{cases} (x^2+y^2)\,\text{sen}\,\dfrac{1}{x^2+y^2} & \text{si} \quad (x,y)\neq(0,0) \\ 0 & \text{si} \quad (x,y)=(0,0) \end{cases}$ 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. $\square$

Definición 76. *Decimos que $f: U \to \mathbb{R}^m$ es* **diferenciable 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 70 (regla de Leibniz), caso escalar. La función producto $\text{prod}: \mathbb{R}^2 \to \mathbb{R}$, dada por $(x,y) \mapsto xy$, tiene jacobiana $D\,\text{prod} = [y \;\; x]$, claramente continua, luego $\text{prod} \in \mathcal{C}^1(\mathbb{R}^2)$ 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 $fg \equiv \text{prod} \circ \begin{bmatrix} f \\ g \end{bmatrix}$ y podemos aplicar la regla de la cadena:

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

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

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 el apartado 2.2.2 demostramos un resultado del que (4) es un caso particular.

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 2.1.9 son de clase $\mathcal{C}^1$ en $\mathbb{R}^2 \setminus \{(0,0)\}$, e igual el ejemplo que acabamos de dar $(x^2+y^2)\,\text{sen}\big(1/(x^2+y^2)\big)$, luego son diferenciables en cada punto de $\mathbb{R}^2 \setminus \{(0,0)\}$ sin que haga falta analizarlos más.

La función $f \equiv \sqrt[3]{x^3+y^3}$ 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 ningún punto de L. Para el punto $(0,0)$ ya lo hemos visto en el apartado 2.1.9. 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.

2.2 Derivadas de orden mayor

Hay funciones diferenciables cuyas derivadas, a su vez, son también diferenciables. Estas funciones tienen *derivadas segundas.* También hay funciones con derivadas terceras, etc. En esta sección exponemos los hechos más básicos relativos a todas estas derivadas.

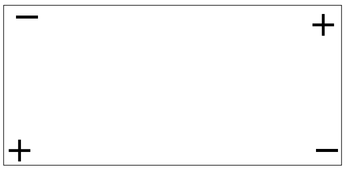
2.2.1 Derivadas cruzadas

Teorema 77. (Karl Hermann Amandus Schwarz). *Sea $f(x_1, x_2)$ definida en un entorno de $x_0 = (a, b)$ y tal que:*
(1) Las funciones f_{x_1}, f_{x_2} y $f_{x_1 x_2} = \dfrac{\partial}{\partial x_2} f_{x_1}$ existen cerca de x_0.
(2) La función $f_{x_1 x_2}$ es continua en x_0.
Entonces existe $f_{x_2 x_1}(x_0) = \dfrac{\partial}{\partial x_1}\Big|_{x_0} f_{x_2}$ y es igual a $f_{x_1 x_2}(x_0)$.

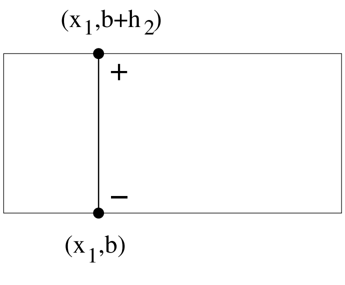
Demostración. Aquí utilizaremos la norma euclídea estándar, denotada $\|\cdot\|$.
Para cada $h = (h_1, h_2)$ consideramos el rectángulo de lados paralelos a los ejes cuyos vértices son $x_0 = (a, b), (a + h_1, b), (a, b + h_2), (a + h_1, b + h_2) = x_0 + h$.
En el caso $h_1, h_2 > 0$, el dibujo es así:

Para h pequeño, todo el rectángulo está contenido en el entorno de x_0 donde existen $f_{x_1}, f_{x_2}, f_{x_1 x_2}$.
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 $f_{x_1 x_2}$ existe en todo el rectángulo, hay un número $\eta = \eta(h)$ entre b y $b + h_2$, tal que

$$f_{x_1}(\xi, b + h_2) - f_{x_1}(\xi, b) \;=\; h_2 \cdot f_{x_1 x_2}(\xi, \eta) \; ,$$

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:

$$\lim_{\substack{h\to(0,0)\\ h_1\neq 0\\ h_2>0}} \frac{\Sigma(h)}{h_1h_2} \;=\; f_{x_1x_2}(x_0)\,. \tag{29}$$

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, (29) 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 $\displaystyle\left.\frac{\partial}{\partial x_1}\right|_{x_0} f_{x_2}$ y es igual a $f_{x_1x_2}(x_0)$, que es lo afirmado por el teorema. $\square$

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

2.2.2 Derivadas de orden 3, 4, etc.

Definición 78. *Sea $U\subseteq\mathbb{R}^n$ un abierto. Se dice que $f:U\to\mathbb{R}^m$ es* **diferenciable de clase $\boldsymbol{\mathcal{C}^k}$ 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 $\boldsymbol{\mathcal{C}^\infty}$ en U** *o que* **es suave en U,** *si es $\mathcal{C}^k$ en U 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 cuyo orden es la **longitud del multíndice** $|\alpha| \stackrel{\text{def}}{=} \alpha_1 + \cdots + \alpha_n$ 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) $f \equiv (f_1, \ldots, f_m)$ 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) Todo polinomio es $\mathcal{C}^\infty$ (sus derivadas de todos los órdenes existen y son polinomios).
(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).

Podemos decir para la clase $\mathcal{C}^k$ algo que ya dijimos 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 2.1.9 y $(x^2+y^2) \operatorname{sen}\big(1/(x^2+y^2)\big)$ definen funciones $\mathcal{C}^\infty$ en $\mathbb{R}^2 \setminus \{(0,0)\}$. La función vista al final del apartado 2.2.1 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.

2.3 Desarrollo cuadrático de Taylor

Definición 79. *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:*

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

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:*

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

Teorema 80. *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 $f(a+h)-f(a)=g'(0)+(1/2)\,g''(\theta)=(df)_ah+(1/2)\,g''(\theta)$. Calculamos:

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

luego $g''(\theta)=\sum_{1\le i,j\le n}h_ih_j\,f_{x_ix_j}(z)$, 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 $R=(1/2)\,\mathrm{Hess}(f)_z(h)-(1/2)\,\mathrm{Hess}(f)_a(h)=(1/2)\,\sum_{1\le i,j\le n}h_ih_j\,\big(f_{x_ix_j}(z)-f_{x_ix_j}(a)\big)$. 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:

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

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:

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

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

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

donde $\widetilde{R} = (1/6)\ \sum_{1 \leq i,j,k \leq n} h_i\, h_j\, h_k\, f_{x_i x_j x_k}(\widetilde{z})$ y $\widetilde{z} = a + \widetilde{\theta} h$ está en $\overline{B}(a, \|h\|)$.
Para cada terna ijk tenemos $|h_i h_j h_k| \leq \|h\|^3$, luego $|\widetilde{R}| \leq (1/6)\,\|h\|^3 \sum_{1 \leq i,j,k \leq n} |f_{x_i x_j x_k}(z)|$. 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: $|f_{x_i x_j x_k}(x)| \leq M$ 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 $|f_{x_i x_j x_k}(\widetilde{z})| \leq M$ para ijk cualesquiera, con lo cual:

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

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

2.4 Extremos locales

Definición 81. *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 $f(x) \leq f(a)$ 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 $f(x) \geq f(a)$ 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 82. *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 83. *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$(f)_a(v) \leq 0$ 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$(f)_a(v) \geq 0$ 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 $\text{Hess}(f)_{(0,0)}$ 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 84. *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. De nuevo $\|\cdot\|$ es la norma euclídea estándar. Escribamos $Q(\cdot) = \text{Hess}(f)_a(\cdot)$ 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:

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

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 $f(a+h) = f(a) + 0 + (1/2)\, Q(h) + R(h)$ 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.

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. □

2.5 Polinomios de Taylor de grado arbitrario

Dados los desarrollos:

$$\begin{aligned} f(a+h) &= f(a) + (Df_a)h + R_1(h) \; , \\ f(a+h) &= f(a) + (Df_a)h + \frac{1}{2}\, \text{Hess}(f)_a(h) + R_2(h) \; , \end{aligned}$$

que hemos estudiado en las secciones anteriores, el primero es una aproximación de $f(a+h)$ por un polinomio de primer grado en h, con error R_1, mientras que el segundo es una aproximación de $f(a+h)$ por un polinomio de segundo grado en h, con error R_2.
En esta sección describimos las aproximaciones análogas por polinomios de grado mayor, así como métodos para calcular tales aproximaciones.

2.5.1 Definición de los polinomios de Taylor

Definición 85. *Una* **forma de grado k** *o* **polinomio homogéneo de grado k** *es un polinomio que se escribe como una combinación lineal de monomios (en las variables $x_1, \ldots, x_n$) todos de grado k.*
Si $k=0$, es una constante. Si $k=1$, es una **forma lineal.** *Si $k=2$, es una* **forma cuadrática.** *Si $k=3$, es una* **forma cúbica,** *etc.*

En los apartados 2.1.6 y 2.1.10 hemos visto que a una función $f \in \mathcal{C}^1(U)$ y un punto $a \in U$ se les asocia una forma lineal $\varphi_1(h)$ tal que

$$\text{para todo } h \in \mathbb{R}^n \quad \left.\frac{d}{dt}\right|_{t=0} f(a+th) \;=\; \varphi_1(h)\,,$$

y de hecho esta forma lineal es la diferencial de f en a: $\varphi_1(h) = (df)_a(h) = (Df_a)h$.

En la sección 2.3 hemos visto que a una función $f \in \mathcal{C}^2(U)$ y un punto $a \in U$ se les asocia una forma cuadrática $\varphi_2(h)$ tal que

$$\text{para todo } h \in \mathbb{R}^n \quad \left.\frac{d^2}{dt^2}\right|_{t=0} f(a+th) \;=\; \varphi_2(h)\,,$$

y de hecho $\varphi_2(h) = h^t\big(\mathrm{Hess}(f)_a\big)h$ es la forma hessiana de f en a.

A una $f \in \mathcal{C}^3(U)$ y un punto $a \in U$ se les asocia una forma cúbica $\varphi_3(h)$ tal que

$$\text{para todo } h \in \mathbb{R}^n \quad \left.\frac{d^3}{dt^3}\right|_{t=0} f(a+th) \;=\; \varphi_3(h)\,.$$

Podemos escribir $\varphi_3(h)$ de varias maneras; una de ellas es la siguiente:

$$\varphi_3(h) \;=\; \sum_{1 \le i_1, i_2, i_3 \le n} h_{i_1} h_{i_2} h_{i_3}\, f_{x_{i_1} x_{i_2} x_{i_3}}(a)\,.$$

En general, para una $f \in \mathcal{C}^k(U)$ y un punto $a \in U$ calculamos fácilmente unas formas $\varphi_0, \varphi_1(h), \ldots, \varphi_k(h)$ tales que cada $\varphi_j(h)$ es una forma de grado j y además:

$$\text{para } j = 0, 1, \ldots k \quad \text{y todo } h \in \mathbb{R}^n\,, \quad \text{se tiene} \quad \left.\frac{d^j}{dt^j}\right|_{t=0} f(p+th) \;=\; \varphi_j(h)\,.$$

En particular, la forma φ_0 es la constante $f(a)$.

Definición 86. *Sean un abierto $U \subseteq \mathbb{R}^n$, una función escalar $f \in \mathcal{C}^k(U)$ y un punto $a \in U$. El* **polinomio de Taylor de orden k de f en a** *es*

$$P_k(x) \;\equiv\; \varphi_0 + \frac{1}{1!}\,\varphi_1(x-a) + \frac{1}{2!}\,\varphi_2(x-a) + \cdots + \frac{1}{k!}\,\varphi_k(x-a)\,,$$

es decir:

$$P_k(a+h) \;\equiv\; f(a) + (Df_a)h + \frac{1}{2}\,h^t\left(\mathit{Hess}(f)_a\right)h + \frac{1}{3!}\,\varphi_3(h) + \frac{1}{4!}\,\varphi_4(h) + \cdots + \frac{1}{k!}\,\varphi_k(h)\,,$$

luego $P_k(x)$ es un polinomio cuyo grado en x no es mayor que k.

2.5.2 Cálculo de los polinomios de Taylor

El siguiente teorema enuncia las propiedades básicas de este polinomio y nos dice cómo calcularlo.

Teorema 87. *Sean un abierto $U \subseteq \mathbb{R}^n$, una función escalar $f(x) \in \mathcal{C}^k(U)$ y un punto $a \in U$. Sea $P(x)$ un polinomio de grado no mayor que k. Son equivalentes:*

1. *P es el polinomio de Taylor de orden k de f en a.*
2. *Se tiene $D^\alpha f(a) = D^\alpha P(a)$ para todo multíndice $\alpha = (\alpha_1, \dots \alpha_n)$ con $0 \leq |\alpha| \leq k$.*
3. *La diferencia $f(x) - P(x)$ es un $o(\|x-a\|^k)$.*

Si además es $f \in \mathcal{C}^{k+1}(U)$, entonces $f(x) - P_k(x) = O\big(\|x-a\|^{k+1}\big)$.

Idea de la demostración. Sea $j = |\alpha| \leq k$. En $x = a$ se anulan todas las derivadas parciales de $\varphi_j(x-a)$ de orden distinto de j, tanto si es un orden menor que j como si es un orden mayor que j. Luego $D^\alpha P_k(x)|_{x=a} = (1/j!)\, D^\alpha \varphi_j(h)|_{h=0}$, valor que coincide con $D^\alpha f(x)|_{x=a}$ por la construcción de las formas $\varphi_j(h)$.

Se demuestra que 2.$\Longrightarrow$3. de la misma manera que se demostró la primera parte del teorema 80.

Veamos ahora que 3.$\Longrightarrow$1. Sea $P_k(x)$ el verdadero polinomio de Taylor de orden k de f en a. Como acabamos de decir que $P_k(x)$ difiere de $f(x)$ en un $\mathrm{o}\big(\|x-a\|^k\big)$, si $P(x)$ cumple 3. entonces:

$$P(x) - P_k(x) \;=\; \mathrm{o}\big(\|x-a\|^k\big)\,. \tag{30}$$

Tomando la descomposición en partes homogńeas:

$$P(a+h) - P_k(a+h) \;\equiv\; H_0 + H_1(h) + H_2(h) + \cdots + H_k(h)\,,$$

de (30) deducimos primero que $H_0 = 0$. Entonces tenemos:

$$H_1(h) \;=\; -H_2(h) - \cdots - H_k(h) + \mathrm{o}\big(\|h\|\big) \;=\; \mathrm{O}(\|h\|^2) + \mathrm{o}\big(\|h\|^k\big) \;=\; \mathrm{o}\big(\|h\|^1\big)\,,$$

que implica $H_1(h) \equiv 0$. Esto, a su vez, nos permite deducir $H_2(h) \equiv 0$ por un argumento análogo, etc., hasta llegar a $H_k(h) \equiv 0$. En definitiva $P(x) - P_k(x) \equiv 0$, es decir $P(x) \equiv P_k(x)$. Queda demostrada la equivalencia de 1., 2. y 3.

Supongamos ahora que $f(x) \in \mathcal{C}^{k+1}(U)$ y sea $P_{k+1}(x)$ el polinomio de Taylor de orden $k+1$ de f en a. Por lo ya explicado, se tiene

$$f(x) \;=\; P_{k+1}(x) + \mathrm{o}\big(\|x-a\|^{k+1}\big)\,,$$

y también:

$$P_{k+1}(x) \;=\; P_k(x) + \frac{1}{(k+1)!}\,\varphi_{k+1}(x-a) \;=\; P_k(x) + \mathrm{O}\big(\|x-a\|^{k+1}\big)\,,$$

luego:

$$f(x) \;=\; P_k(x) + \mathrm{O}\big(\|x-a\|^{k+1}\big) + \mathrm{o}\big(\|x-a\|^{k+1}\big) \;=\; P_k(x) + \mathrm{O}\big(\|x-a\|^{k+1}\big)\,.$$

□

Si calculamos $P_k(x)$ utilizando la equivalencia entre 1. y 2., tenemos el **método de los coeficientes indeterminados.** Por ejemplo, intentemos hallar $P_3(x)$ para una función $f(x)$ en $a = 1$. Un método muy malo (que nunca se utiliza) sería expresar P_3 en términos de las potencias de x: $P_3(x) = c_0 + c_1 x + c_2 x^2 + c_3 x^3$. Imponer las condiciones:

$$P_3^{(j)}(1) \;=\; f^{(j)}(1) \quad , \quad j = 0,1,2,3, \tag{31}$$

equivale al sistema:

$$\left.\begin{array}{ccccccccc} c_0 & + & c_1 & + & c_2 & + & c_3 & = & f(1) \\ & & c_1 & + & 2c_2 & + & 3c_3 & = & f'(1) \\ & & & & 2c_2 & + & 6c_3 & = & f''(1) \\ & & & & & & 6c_3 & = & f'''(1) \end{array}\right\}$$

que tendríamos que resolver. El método bueno consiste en expresar P_3 en términos de las potencias de $x-1$: $P_3(x) = b_0 + b_1(x-1) + b_2(x-1)^2 + b_3(x-1)^3$ y entonces (31) equivale a:

$$b_0 = f(1) \quad , \quad b_1 = f'(1) \quad , \quad 2!b_2 = f''(1) \quad , \quad 3!b_3 = f'''(1) \ ,$$

que es muchísimo más fácil de resolver.
En general, se debe expresar $P_k(x_1, \ldots, x_n)$ como una suma de monomios:

$$\text{coeficiente} \cdot (x_1 - a_1)^{m_1} \, (x_2 - a_2)^{m_2} \, \cdots \, (x_n - a_n)^{m_n} \quad , \quad m_1 + \cdots + m_n \leq k \ .$$

Cuando $f(x)$ es una combinación de funciones sencillas, interesa combinar los desarrollos de Taylor de esas funciones sencillas en vez de hallar derivadas de $f(x)$. Este método se basa en la equivalencia entre 1. y 3. y lo llamamos **método de las oes de Landau.** Veamos un ejemplo.
Consideramos la función $f(x,y) = \dfrac{e^{xy}}{1-y^2}$, que es $\mathcal{C}^\infty$ cerca de $(0,0)$.
Empezamos por recordar el desarrollo $e^t = 1 + t + (1/2)\, t^2 + \mathrm{O}(|t|^3)$, que nos da:

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

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:

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

Entonces la parte $3. \Longrightarrow 1.$ del teorema 87 nos dice que el polinomio de Taylor de orden 5 de f en $(0,0)$ es el siguiente:

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

y ahora sabemos que las derivadas no nulas de f en $(0,0)$, de órdenes de 0 a 5, son:

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

y todas las demás derivadas de órdenes de 0 a 5 de f en $(0,0)$ son nulas. Todo esto sin necesidad de haber hecho ninguna derivación.

Aviso. La hipótesis $f \in \mathcal{C}^k$ en el enunciado del teorema 87 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

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

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.