

Capítulo 1

Cálculo en varias variables

1.1 Límites de funciones de varias variables

El manejo de límites para funciones de varias variables es a la vez muy importante, y difícil. Es importante para el material que presentaremos, pues todo se desarrollará alrededor del concepto de diferenciabilidad, y en dicho concepto jugará un papel importante el límite de una magnitud que depende de varias variables (el residuo).

Es difícil, porque hay una gran diferencia entre los límites de funciones de una variable (ya de por sí un concepto sutil) y los de varias variables. Para verificar la existencia ó no de límites de funciones de una variable, sólo necesitamos analizar dos posibilidades : a qué (límite) tiende la función por la izquierda y por la derecha del punto en cuestión. En cambio, para estudiar un límite de una función de dos variables ó más, en principio están involucradas infinitas maneras de acercarse al punto en cuestión : todas las posibles trayectorias que nos llevan a él.

Ejemplo 1.1-1. Sea $f(x, y) = y/x$. Con dominio

$$\{(x, y) \in \mathbb{R}^2 \mid y \neq 0\}.$$

Estudiamos el límite (si existe) cuando $(x, y) \rightarrow (0, 0)$. Una forma de acercarse al origen es a través de una recta de pendiente m que pasa por el origen: $y = mx$, y hacemos $x \rightarrow 0$. Lo que ocurre es que

$$\lim_{x \rightarrow 0} f(x, mx) = \lim_{x \rightarrow 0} \frac{x}{mx} = \lim_{x \rightarrow 0} \frac{1}{m} = \frac{1}{m},$$

de manera que el valor de límite depende de m , la pendiente de la recta. Si nos acercamos al origen por la parábola $y = x^2$, nos queda

$$\lim_{x \rightarrow 0} f(x, x^2) = \lim_{x \rightarrow 0} \frac{x}{x^2} = \lim_{x \rightarrow 0} \frac{1}{x},$$

el cual, como sabemos, no existe. Consideremos por último la trayectoria $x = y^2$, tenemos que

$$\lim_{y \rightarrow 0} f(y^2, y) = \lim_{y \rightarrow 0} \frac{y^2}{y} = \lim_{y \rightarrow 0} y = 0.$$

¿Cuál es la conclusión de todo esto? Por supuesto, el límite no existe. Cuando nos acercamos al origen por diferentes trayectorias, obtenemos diferentes límites e incluso el límite puede no existir a lo largo de ciertas trayectorias. El que el límite exista equivale a que se obtiene un mismo valor al aproximarse por *cualquier* trayectoria.

Ejemplo 1.1-2. Sea

$$f(x, y) = \frac{\sin(x) \sin(y)}{xy}.$$

Queremos ver si $\lim_{(x,y) \rightarrow (0,0)} f(x, y)$ existe.

Hagamos algunas observaciones sobre el dominio D de f . Primeramente la función no está definida en $(0, 0)$ y para analizar el límite es necesario estudiar valores de $f(x, y)$ en puntos (x, y) *arbitrariamente cercanos a* $(0, 0)$ sin ser $(0, 0)$. Técnicamente $(0, 0)$ debe ser punto de acumulación del dominio D , vgr. para cualquier $\delta > 0$ debe suceder que $B'(0, 0; \delta) \cap D \neq \emptyset$ donde la bola con centro en $(0, 0)$ y radio δ se define como

$$B_\delta(0, 0) = \{(x, y) \mid \sqrt{x^2 + y^2} < \delta\}$$

y la notación $B'_\delta(0, 0)$ significa que se excluye el centro. Tomemos por ejemplo $D = B'_1(0, 0)$. Es claro entonces que $(0, 0)$ es punto de acumulación de D .

La siguiente es una tabla de valores (x, y) tomados al azar y los correspondientes valores de $f(x, y)$

x	y	$\sqrt{x^2 + y^2}$	$f(x, y)$
0.624545	0.929048	1.11946	0.807257
0.538409	0.61957	0.820823	0.892609
0.0235492	0.772784	0.773143	0.903314
0.387152	0.800026	0.888779	0.874456
0.0740555	0.369479	0.376827	0.976509
0.381591	0.544371	0.664794	0.928417
0.167903	0.287027	0.332529	0.981698
0.180206	0.815687	0.835356	0.887916
0.712596	0.927343	1.16951	0.791529
0.459414	0.866261	0.980545	0.848927
0.695937	0.998536	1.21713	0.775578
0.380652	0.53238	0.654465	0.930569
0.0713921	0.0694885	0.0996267	0.998347
0.842243	0.91281	1.24201	0.767895
0.047843	0.296704	0.300537	0.985016

Si seleccionamos los valores de $f(x, y)$ con (x, y) lo más cercano a $(0, 0)$, es decir con norma más pequeña, tenemos

x	y	$\sqrt{x^2 + y^2}$	$f(x, y)$
0.0740555	0.369479	0.376827	0.976509
0.167903	0.287027	0.332529	0.981698
0.0713921	0.0694885	0.0996267	0.998347
0.047843	0.296704	0.300537	0.985016

Comprobamos que los valores de $f(x, y)$ parecen tender a 1

Definición 1. Sea $f: D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$, definida en una bola con centro en \mathbf{p} , $D = B_r(\mathbf{p})$. Decimos que $\lim_{\mathbf{x} \rightarrow \mathbf{p}} f(\mathbf{x}) = L$ si y sólo si para todo número positivo ϵ existe $\delta > 0$ tal que si

$$\mathbf{x} \in D \quad y \quad 0 < \|\mathbf{x} - \mathbf{p}\| < \delta$$

entonces

$$\|f(\mathbf{x}) - L\| < \epsilon.$$

Observación: la primera norma involucrada es la de \mathbb{R}^n y la segunda la de \mathbb{R}^m . Como referencia futura, si $\mathbf{x} = (x_1, x_2, \dots, x_p) \in \mathbb{R}^p$ escribimos

$$\|\mathbf{x}\| = (x_1^2 + x_2^2 + \dots + x_p^2)^{1/2}$$

la cual se llamará en lo sucesivo la norma euclídeana en \mathbb{R}^p . La mayoría de las veces, si no hay lugar a confusión, utilizaremos simplemente $\|\mathbf{x}\|$ para la norma del vector x .

Para probar que un límite existe, debe establecerse que la función se aproxima al candidato a límite, independientemente de la trayectoria que se siga. En la práctica, esto se hace estableciendo las desigualdades e implicaciones de la definición, sin suponer ninguna trayectoria en particular (ver ejemplos 2 y 3). Para demostrar que un límite no existe, basta con mostrar dos trayectorias concretas tales que la función tiende a diferentes valores a través de ellas, o bien que por una trayectoria concreta el límite (de una variable) no existe (ver ejemplos 1 y 4). Con más precisión:

Proposición 1. *Sea $f: D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$, definida en una bola con centro en \mathbf{p} , $D = B_r(\mathbf{p})$. Si γ_1, γ_2 son dos curvas continuas que pasan por \mathbf{p} y si*

$$\lim_{\mathbf{x} \in \gamma_1, \mathbf{x} \rightarrow \mathbf{p}} f(\mathbf{x}) \neq \lim_{\mathbf{x} \in \gamma_2, \mathbf{x} \rightarrow \mathbf{p}} f(\mathbf{x}),$$

entonces $\lim_{\mathbf{x} \rightarrow \mathbf{p}} f(\mathbf{x})$ no existe

Naturalmente la notación $\mathbf{x} \in \gamma, \mathbf{x} \rightarrow \mathbf{p}$ significa que debe parametrizarse la curva γ , digamos por s de forma tal que $\lim_{s \rightarrow 0^+} \gamma(s) = \mathbf{p}$, y así $\lim_{\mathbf{x} \in \gamma, \mathbf{x} \rightarrow \mathbf{p}} f(\mathbf{x}) = \lim_{s \rightarrow 0^+} f(\gamma(s))$, tal como lo hicimos en los ejemplos 1 y 2.

Ejemplo 1.1-3. Sea $f(x, y) = x^2 + y^2$. Probemos que $\lim_{(x,y) \rightarrow (a,b)} f(x, y) = a^2 + b^2$. Comenzamos por estimar

$$|x^2 + y^2 - a^2 - b^2| \leq |x^2 - a^2| + |y^2 - b^2| = |x + a||x - a| + |y + b||y - b|$$

y notamos que $|x + a|, |y + b|$ se pueden acotar superiormente si (x, y) está suficientemente cercano a (a, b) en tanto que $|x - a|, |y - b|$ se pueden hacer arbitrariamente pequeños si (x, y) es suficientemente cercano a (a, b) . Precisemos: Tomemos $(x, y) \in B_1(a, b)$ con ello

$$|x - a| \leq \sqrt{(x - a)^2 + (y - b)^2} < 1, \quad \text{de donde } |x| \leq |a| + 1, \text{ luego}$$

$$|x + a| \leq |x| + |a| \leq 2|a| + 1.$$

similarmente $|y + b| \leq 2|b| + 1$ y hemos conseguido acotar $|x + a|$ y $|y + b|$. Por otro lado si $(x, y) \in B_\delta(a, b)$ entonces

$$|x - a|, |y - b| \leq \sqrt{(x - a)^2 + (y - b)^2} < \delta$$

juntando lo anterior tenemos que

$$|x+a||x-a| + |y+b||y-b| \leq (2|a|+1)\delta + (2|b|+1)\delta = 2(|a|+|b|+1)\delta$$

que podemos hacer menor que ϵ si tomamos

$$\delta < \frac{\epsilon}{2(|a|+|b|+1)}.$$

Concluimos que $\lim_{(x,y) \rightarrow (a,b)} f(x,y) = a^2 + b^2$. Observe que el límite se obtiene sustituyendo (a,b) en $f(x,y)$. Cuando

$$\lim_{(x,y) \rightarrow (a,b)} f(x,y) = f(a,b)$$

se dice que la función f es *continua* en (a,b) .

Ejemplo 1.1-4. Sea

$$f(x,y) = \frac{x^2y}{x^2+y^2}$$

podemos considerar que el dominio es alguna bola $B_r(0,0)$ (por ejemplo si $r = \infty$ el dominio es todo el plano menos el origen). Se observa que el numerador es un polinomio homogéneo de grado 3 en tanto que el denominador es homogéneo de grado 2, así que cuando $(x,y) \rightarrow (0,0)$ en numerador tiende a cero más rápido que en denominador. Proponemos entonces que $\lim_{(x,y) \rightarrow (a,b)} f(x,y) = 0$.

En efecto, ya que $x^2 + y^2 \geq x^2$ entonces

$$\left| \frac{x^2y}{x^2+y^2} \right| \leq \frac{|x|^2|y|}{|x|} = |x||y| \leq (\sqrt{x^2+y^2})(\sqrt{x^2+y^2}) = x^2 + y^2$$

y ahora podemos hacer $x^2 + y^2 < \epsilon$ si $\sqrt{x^2 + y^2} < \epsilon^{1/2} \equiv \delta$

Ejemplo 1.1-5. Se pide analizar si existe $\lim_{(x,y) \rightarrow (0,0)} f(x,y)$.

$$f(x,y) = \frac{x^2y}{x^2+y^6}.$$

Observe que el denominador ya no es homogéneo pero se puede usar la misma estimación que antes

$$\left| \frac{x^2y}{x^2+y^6} \right| \leq \frac{|x|^2|y|}{|x|} = |x||y|.$$

Ejemplo 1.1-6.

Ejemplo 1.1-7. Se pide analizar si existe $\lim_{(x,y) \rightarrow (0,0)} f(x, y)$.

$$f(x, y) = \frac{xy}{x^2 + y^2}.$$

Ninguna de las estimaciones $x^2 + y^2 \geq x^2, y^2$ funciona pues se obtiene respectivamente

$$\left| \frac{xy}{x^2 + y^2} \right| \leq \frac{|x|}{|y|}, \frac{|y|}{|x|}$$

Si nos acercamos al origen por rectas $y = mx$

$$f(x, mx) = \frac{mx^2}{x^2 + m^2x^2} = \frac{m}{1 + m^2}$$

así que no obtenemos un mismo límite. Por (??) el límite no existe.

Ejemplo 1.1-8. Se pide analizar si existe $\lim_{(x,y,z) \rightarrow (0,0,0)} f(x, y, z)$ para

$$f(x, y, z) = \frac{xy^2z}{x^4 + y^2 + z^2}.$$

Tratemos que “cancelar” la y^2 den numerador

$$\left| \frac{xy^2z}{x^4 + y^2 + z^2} \right| \leq \frac{|x||y|^2|z|}{|y|^2} = |x||z|$$

de donde se ve que $\lim_{(x,y,z) \rightarrow (0,0,0)} f(x, y, z) = 0$.

Ejemplo 1.1-9. Analizar la existencia de

$$\lim_{\mathbf{h} \rightarrow 0} \frac{\mathbf{h}}{\|\mathbf{h}\|}.$$

El límite no existe porque si hacemos tender a $\mathbf{h} = (h_1, h_2, \dots, h_n)$ por los, por ejemplo $\mathbf{h} = (h_1, 0, \dots, 0)$ tenemos

$$\lim_{h_1 \rightarrow 0} \frac{\mathbf{h}}{\|\mathbf{h}\|} = \lim_{h_1 \rightarrow 0} \frac{h_1}{|h_1|}$$

que no existe.

En muchas ocasiones estaremos interesados en calcular límites de funciones que no están definidas en el punto donde se calcula el límite. Un ejemplo típico es el de la derivada

$$\lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h}$$

Aquí la función de interés es el cociente incremental (fijos f y a)

$$\frac{f(a+h) - f(a)}{h}$$

que no está definido para $h = 0$. Es necesario “simplificar” el factor h en el denominador.

Para límites de combinaciones de funciones, se cumple un teorema similar al de una variable,

Teorema 1. Sean f, g funciones en $B_r(\mathbf{p}) \subset \mathbb{R}^n \rightarrow \mathbb{R}$. Si $\lim_{\mathbf{x} \rightarrow \mathbf{p}} f(\mathbf{x}) = L$ y $\lim_{\mathbf{x} \rightarrow \mathbf{p}} g(\mathbf{x}) = M$, entonces los límites siguiente existen y son iguales a los valores indicados

$$(i) \lim_{\mathbf{x} \rightarrow \mathbf{p}} [f(\mathbf{x}) \pm g(\mathbf{x})] = L \pm M,$$

$$(ii) \lim_{\mathbf{x} \rightarrow \mathbf{p}} [f(\mathbf{x})g(\mathbf{x})] = LM, \text{ y}$$

$$(iii) \lim_{\mathbf{x} \rightarrow \mathbf{p}} \left[\frac{f(\mathbf{x})}{g(\mathbf{x})} \right] = \frac{L}{M}, \text{ si } M \neq 0.$$

Observación Es importante señalar que el teorema anterior afirma dos cosas, que los límites de las combinaciones existen y por otro lado sus valores.

1.2 Ejercicios

En todos los casos que siguen determinar si el límite existe ó no y argumentar por qué.

$$1. \lim_{(x,y) \rightarrow (x_0,y_0)} x^n + y^m, \text{ con } m, n \text{ enteros positivos fijos.}$$

$$2. \lim_{(x,y) \rightarrow (x_0,y_0)} x \sin(x^2 + y^2).$$

$$3. \lim_{(x,y) \rightarrow (x_0,y_0)} \frac{x^2 y^2}{(x^2 + y^2)^2}.$$

4. $\lim_{(x,y) \rightarrow (x_0,y_0)} \frac{x^2 y^{2+a}}{(x^2 + y^2)^2}$, con $a > 0$ fijo.
5. $\lim_{(x,y) \rightarrow (x_0,y_0)} \frac{x^2 + y}{(x^2 + y^2)^{1/2}}$.
6. $\lim_{(x,y) \rightarrow (x_0,y_0)} \frac{x \sin y}{|x| + |y|}$.
7. $\lim_{(x,y,z) \rightarrow (x_0,y_0,z_0)} f(x, y, z) = \frac{x^2 y z}{(x^2 + y^2 + z^2)^{3/2}}$.

1.3 Derivada direccional, parcial y gradiente

Las derivadas parciales para una función de varias variables, se definen como la deriva usual respecto de alguna de las variables, manteniendo constante el resto. Se podría pensar ingenuamente que este concepto generaliza la idea de derivada como aproximación lineal para funciones de una variable, pero veremos más adelante que para funciones de varias variables, es necesaria una definición más fuerte con el fin de mantener la idea geométrica de aproximación lineal o de plano tangente. Comencemos por plantearnos la siguiente pregunta:

¿Cómo estudiar el comportamiento (continuidad, tasa de cambio, etc.) de una función de dos variables $z = f(x, y)$ en un entorno de un punto (a, b) de su dominio?

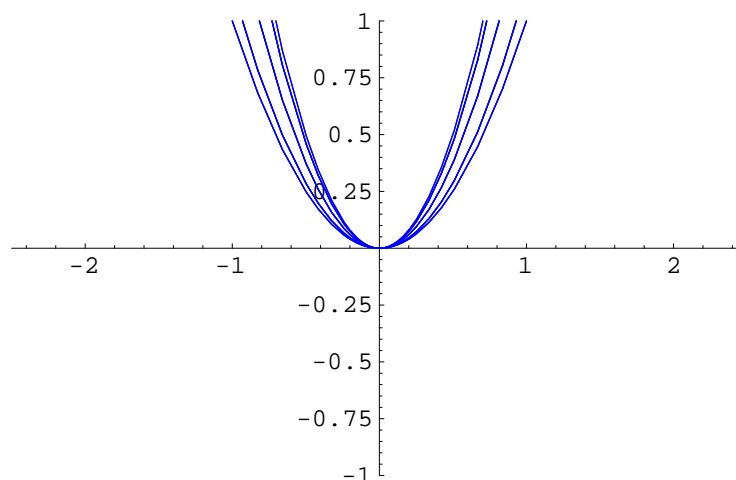
Existen al menos dos formas:

1. Analizando la función de una variable que se obtiene restringiendo la función a una recta que pase por el punto (a, b) y dirección arbitraria.
2. Por medio de sus curvas de nivel $f(x, y) = \text{constante}$ en un entorno del punto (a, b)

Siguiendo este tren de ideas, vamos a discutir con más detalle cada una.

1.3.1 La derivada direccional

Consideremos una función real de dos variables definida por $z = f(x, y)$. Supongamos que la función está definida en un disco $f: B_r(a, b) \rightarrow \mathbb{R}$. La

Figure 1.1: Gráficos de algunas funciones $f_{\hat{\mathbf{u}}}$

recta que pasa por el punto (a, b) y tiene dirección $\hat{\mathbf{u}} = (\alpha, \beta)$ tiene por ecuación paramétrica

$$(x, y) = (a, b) + t(\alpha, \beta)$$

Naturalmente, cuando nos referimos a una dirección suponemos que $\hat{\mathbf{u}}$ es un vector unitario, i.e., $\alpha^2 + \beta^2 = 1$.

La restricción a dicha recta es la función de una variable

$$f_{\hat{\mathbf{u}}}(t) = f(a + t\alpha, b + t\beta)$$

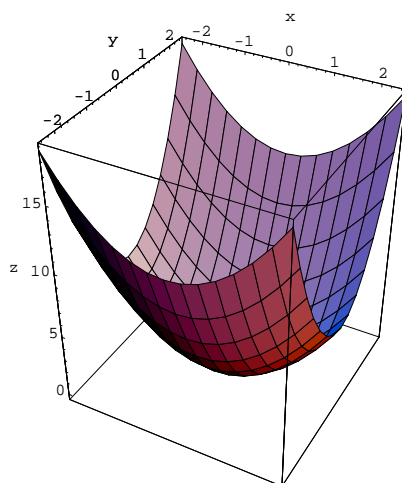
y está definida para t suficientemente pequeña (para $|t| < r$).

Ejemplo 1.3-10. Sea $f(x, y) = x^3 + 2y^2$ queramos estudiar la función en un entorno de $(0, 0)$.

En la figura 1.1 se muestran las gráficas de algunas funciones $f_{\hat{\mathbf{u}}}$ para diversas direcciones. Se puede notar la prevalencia de un mínimo en el origen. De hecho la gráfica de la función $z = x^3 + 2y^2$ comprueba esta conjetura.

Probemos que en cualquier dirección se tiene un mínimo. Cualquier vector unitario en el plano se puede escribir como $(\cos \theta, \sin \theta)$, luego la restricción de f a lo largo de la dirección θ es

$$f_{\theta}(t) = (t \cos \theta)^2 + 2(t \sin \theta)^2 = (\cos^2 \theta + 2 \sin^2 \theta)t^2 = (1 + \sin^2 \theta)t^2$$

Figure 1.2: Gráfico de $z = x^3 + 2y^2$

como el coeficiente de t^2 es siempre positivo entonces se tiene un mínimo en $t = 0$.

Ejemplo 1.3–11. Haga el mismo análisis con la función $z = x^2 - y^2$.

Los gráficos de algunas funciones $f_\theta(t)$ y del gráfico $z = x^2 - y^2$ se muestran en la figuras (1.3,1.4).

Se puede verificar que la función no puede tener un máximo o mínimo en el origen. En efecto a lo largo de la dirección $(1, 0)$, $f_{(1,0)} = t^2$ tiene un máximo y en la dirección $(0, 1)$, $f_{(0,1)} = -t^2$ tiene un mínimo en $t = 0$.

Vamos ahora a precisar algunos términos. Como siempre, supondremos que las funciones están definidas en alguna bola con centro en el punto en cuestión. En la segunda parte de estas notas relajaremos esta condición e introduciremos el concepto de *punto interior*.

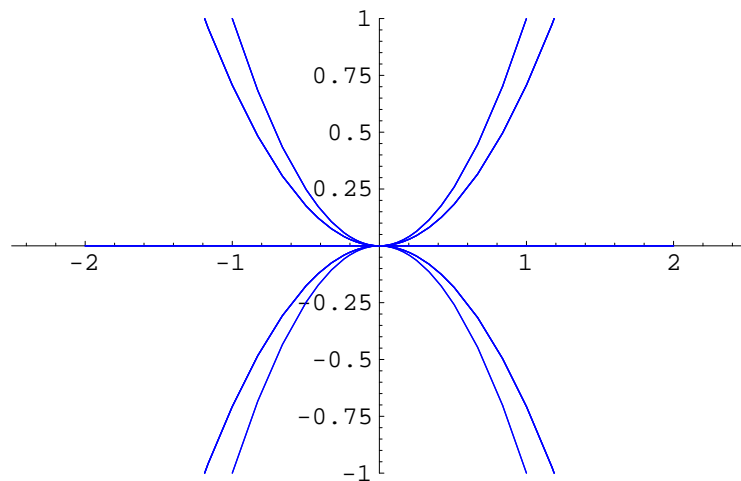
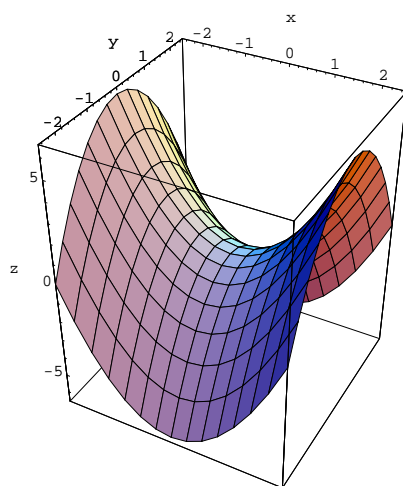
Definición 2. Sea $f: B_r(a, b) \subset \mathbb{R}^n \rightarrow \mathbb{R}$, $\hat{\mathbf{u}} = (\alpha, \beta)$ un vector unitario. La derivada de f en la dirección (α, β) en el punto (a, b) se define como

$$D_{\hat{\mathbf{u}}}f(a, b) = \lim_{t \rightarrow 0} \frac{f(a + t\alpha, b + t\beta) - f(a, b)}{t}$$

si tal límite existe.

El límite $D_{\hat{\mathbf{u}}}f(a, b)$ se llama la derivada direccional en la dirección (α, β) . Observe que la derivada direccional no es sino

$$D_{\hat{\mathbf{u}}}f(a, b) = f'_{\hat{\mathbf{u}}}(0)$$

Figure 1.3: Gráficos de algunas funciones $f_\theta(t)$ Figure 1.4: Gráfico de $z = x^2 - y^2$

y describe la razón de cambio de $f_{\hat{\mathbf{u}}}(t)$ en el punto (a, b) .

En particular si tomamos la dirección $(1, 0)$ o $(0, 1)$ tenemos lo que se conoce como las derivadas parciales:

Definición 3. Sea $f: B_r(a, b) \subset \mathbb{R}^n \rightarrow \mathbb{R}$. La derivada parciales se definen como

$$D_1f(a, b) = \lim_{h \rightarrow 0} \frac{f(a+h, b) - f(a, b)}{h}, \quad D_2f(a, b) = \lim_{h \rightarrow 0} \frac{f(a, b+h) - f(a, b)}{h}$$

si tales límites existen.

Cuando sea conveniente se usarán cualquiera de las siguientes notaciones

$$D_1f(a, b) = \frac{\partial f}{\partial x}(a, b), \quad D_2f(a, b) = \frac{\partial f}{\partial y}(a, b)$$

y se omite el punto (a, b) cuando está sobreentendido.

En las figuras (1.5, 1.6, 1.7) se muestra la superficie $z = x^2 - y^2$ alrededor del punto $\mathbf{p} = (a, b, f(a, b)) = (1.5, -1.5, 0)$ la derivada parcial $D_1f(a, b)$ es la pendiente de la curva que se obtiene como intersección del plano $x = a$ con la superficie $z = f(x, y)$ (figura 1.5) en el punto; en tanto que $D_2f(a, b)$ es la pendiente de la curva que se obtiene como intersección del plano $y = b$ con la superficie $z = f(x, y)$ (figura 1.6) en el punto. En la figura 1.7 se muestra la superficie cortada por ambos planos.

El siguiente resultado muestra la relación entre derivada direccional y derivadas parciales.

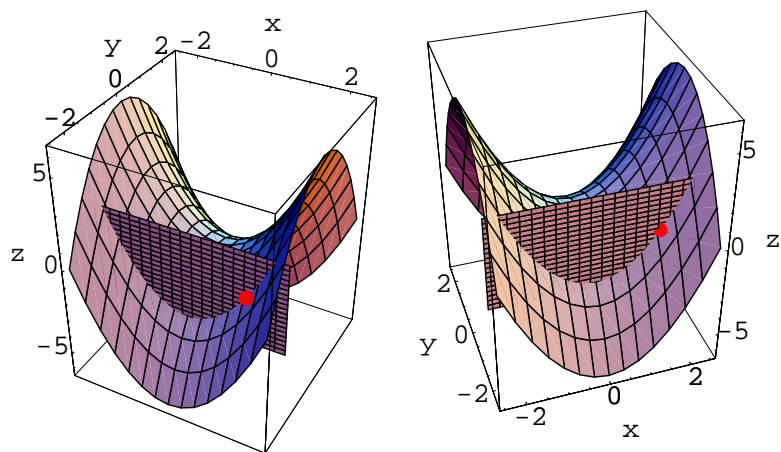
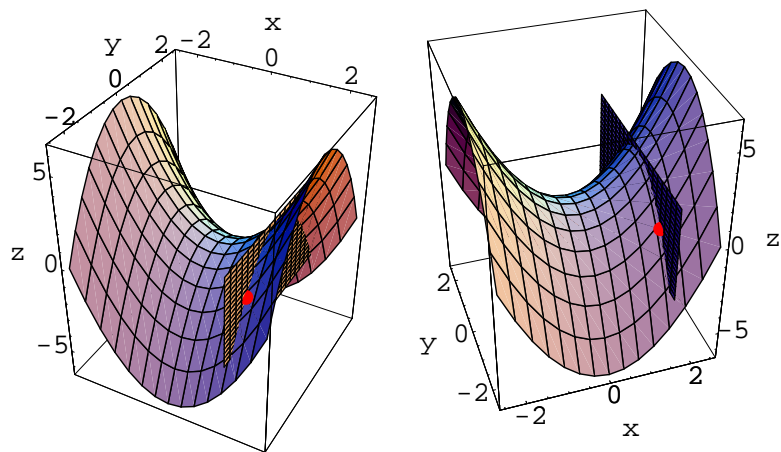
Proposición 2. Sea $f: B_r(a, b) \subset \mathbb{R}^n \rightarrow \mathbb{R}$, $\hat{\mathbf{u}} = (\alpha, \beta)$ un vector unitario. Si existe la derivada de f en la dirección (α, β) en el punto (a, b) , junto con las derivadas parciales, entonces

$$D_{\hat{\mathbf{u}}}f(a, b) = \alpha D_1f(a, b) + \beta D_2f(a, b)$$

Demostración.

Escribamos

$$\frac{f(a + t\alpha, b + t\beta) - f(a, b)}{t} = \left(\frac{f(a + t\alpha, b + t\beta) - f(a + t\alpha, b)}{t} \right) + \left(\frac{f(a + t\alpha, b) - f(a, b)}{t} \right)$$

Figure 1.5: Interpretación geométrica de la derivada parcial $D_1f(a, b)$ Figure 1.6: Interpretación geométrica de la derivada parcial $D_2f(a, b)$

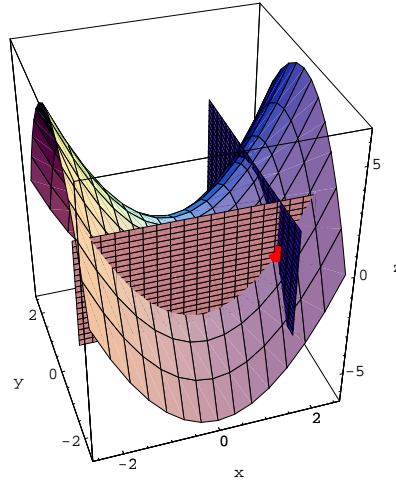


Figure 1.7: Interpretación geométrica de las derivadas parciales $D_1f(a, b), D_2f(a, b)$

Por definición el límite cuando $t \rightarrow 0$ del lado izquierdo existe y es $D_{\hat{\mathbf{u}}}f(a, b)$. En el lado derecho, obtenemos

$$\begin{aligned} \lim_{t \rightarrow 0} \left(\frac{f(a + t\alpha, b) - f(a, b)}{t} \right) &= \alpha \lim_{t \rightarrow 0} \left(\frac{f(a + t\alpha, b) - f(a, b)}{t\alpha} \right) \\ &= \alpha \lim_{h \rightarrow 0} \left(\frac{f(a + h, b) - f(a, b)}{h} \right) = \alpha D_1f(a, b) \end{aligned}$$

por lo tanto el límite del segundo término existe y es:

$$\begin{aligned} \lim_{t \rightarrow 0} \left(\frac{f(a + t\alpha, b + t\beta) - f(a + t\alpha, b)}{t} \right) &= \beta \lim_{t \rightarrow 0} \left(\frac{f(a + t\alpha, b + t\beta) - f(a + t\alpha, b)}{t} \right) = \\ &\equiv \beta B \end{aligned}$$

existe, luego

$$D_{\hat{\mathbf{u}}}f(a, b) = \alpha D_1f(a, b) + \beta B. \quad (1.1)$$

Análogamente escribiendo

$$\begin{aligned} \frac{f(a + t\alpha, b + t\beta) - f(a, b)}{t} &= \left(\frac{f(a + t\alpha, b + t\beta) - f(a, b + t\beta)}{t} \right) + \\ &\quad \left(\frac{f(a, b + t\beta) - f(a, b)}{t} \right) \end{aligned}$$

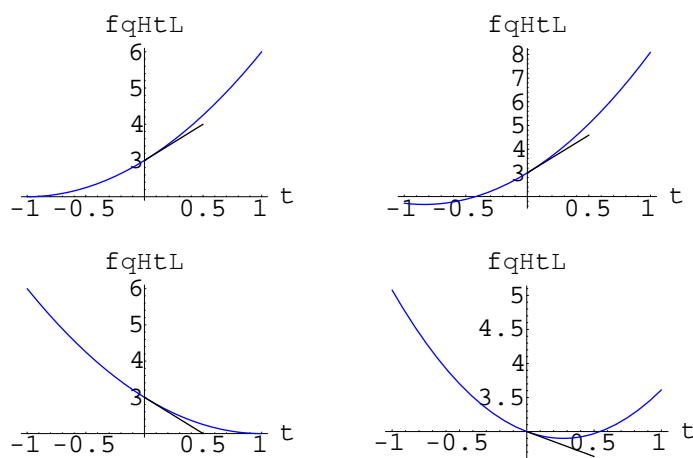


Figure 1.8: Gráfico de las funciones direccionales $f_\theta(t)$, para la función $z = x^2 + 2y^2$

y haciendo $t \rightarrow 0$ obtenemos

$$D_{\hat{u}}f(a, b) = \alpha A + \beta D_2f(a, b) \quad (1.2)$$

donde

$$A \equiv \lim_{t \rightarrow 0} \left(\frac{f(a + t\alpha, b + t\beta) - f(a, b + t\beta)}{t} \right)$$

en resumen

$$D_{\hat{u}}f(a, b) = \alpha D_1f(a, b) + \beta B = \alpha A + \beta D_2f(a, b).$$

Como α y β son arbitrarios de (1.1,1.2) se sigue que

$$D_{\hat{u}}f(a, b) = \alpha D_1f(a, b) + \beta D_2f(a, b).$$

Q.E.D.

La figura (1.11) muestra algunas de las funciones $f_\theta(t)$ para $f(x, y) = x^2 + 2y^2$ y el punto $(1, 1)$.

Se puede observar que en algunas direcciones la función crece y en otras decrece. Nos preguntamos en qué dirección es máximo el crecimiento instantáneo?

Veamos con detalle. Si $z = f(x, y)$ está definida en un disco $: B_r(a, b) \subset \mathbb{R}^2$ y $\hat{\mathbf{u}} = (\cos \theta, \sin \theta)$ es un vector unitario al igual que en el ejemplo (1.3.1), entonces la función $f_\theta(t) = f((a, b) + t(\cos \theta, \sin \theta))$ nos da la variación de la función en la dirección θ a partir de (a, b) .

El crecimiento instantáneo en la dirección θ de acuerdo a la proposición (2) es

$$m'(\theta) = D_1f(a, b) \cos \theta + D_2f(a, b) \sin \theta$$

Por lo tanto la dirección de máximo crecimiento se obtiene maximizando la tasa de crecimiento instantánea m respecto de θ . Igualando la derivada a cero obtenemos

$$0 = m'(\theta) = -D_1f(a, b) \sin \theta + D_2f(a, b) \cos \theta.$$

Vemos que cualquier solución es de la forma

$$(\cos \theta, \sin \theta) = \lambda(D_1f(a, b), D_2f(a, b))$$

en donde λ es un escalar, siempre que el vector $(D_1f(a, b), D_2f(a, b))$ sea distinto de cero. Este se llama el vector gradiente

$$\nabla f(a, b) = (D_1f(a, b), D_2f(a, b)),$$

y determina una dirección θ_0 de crecimiento crítica

$$(\cos \theta_0, \sin \theta_0) = \frac{\nabla f(a, b)}{\|\nabla f(a, b)\|}$$

En efecto,

$$m''(\theta_0) = -\|\nabla f(a, b)\|^2$$

y la dirección crítica θ_0 es de *máximo* crecimiento.

Enunciemos esta propiedad importante

Proposición 3. Sea $: B_r(a, b) \subset \mathbb{R}^2$ y $\hat{\mathbf{u}} = (\cos \theta, \sin \theta)$ un vector unitario en la dirección θ . La dirección de máximo crecimiento de f a partir del punto (a, b) es

$$(\cos \theta_0, \sin \theta_0) = \frac{\nabla f(a, b)}{\|\nabla f(a, b)\|}$$

Ejemplo 1.3-12. Encuentre el ángulo de la dirección de máximo crecimiento de la función $f(x, y) = x^2 + 2y^2$ en el punto $(1, 1)$.

Tenemos

$$f_\theta(t) = (1 + t \cos \theta)^2 + 2(1 + t \sin \theta)^2$$

de donde la tasa de crecimiento en la dirección θ es

$$m(\theta) = 2 \cos \theta + 4 \sin \theta$$

Resolviendo la ecuación

$$0 = m'(\theta) = 4 \cos \theta - 2 \sin \theta$$

obtenemos

$$\tan \theta = \pm \arccos(2)$$

o bien $\theta = \arctan(2)$. Existe una solución en $(-\pi/2, \pi/2)$ $[(\cos \theta, \sin \theta)$ en el primer cuadrante] y otra en $(\pi/2, 3\pi/2)$ $[(\cos \theta, \sin \theta)$ en el tercer cuadrante], que difiere por π . Usando la fórmula

$$\cos \theta = \frac{1}{\sqrt{1 + \tan^2 \theta}}, \quad \sin \theta = \frac{\tan \theta}{\sqrt{1 + \tan^2 \theta}}$$

para la primera solución, obtenemos

$$\cos \theta = \frac{1}{\sqrt{5}}, \quad \sin \theta = \frac{2}{\sqrt{5}} \quad (1.3)$$

La segunda solución solo difiere en signo. Es fácil ver que $m''(\theta_0) < 0$ para $(\cos \theta, \sin \theta)$ en el primer cuadrante de modo que la dirección de máximo crecimiento es (1.3).

Calculando directamente el vector unitario en la dirección del gradiente en $(1, 1)$ obtenemos

$$\nabla f(x, y) = (2x, 4y), \quad \text{de donde} \quad \frac{\nabla f(x, y)}{\|\nabla f(x, y)\|} = \frac{(2, 4)}{2\sqrt{5}} = \left(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}} \right)$$

que es consistente con

Otra propiedad importante del gradiente es en relación a las curvas de nivel $f(x, y) = \text{constante}$

Proposición 4. Sea $f: B_r(a, b) \subset \mathbb{R}^2$, Si $\nabla f(a, b) \neq \mathbf{0}$, entonces $\nabla f(a, b)$ es ortogonal a la curva de nivel $f(x, y) = c$ que pasa por (a, b) .

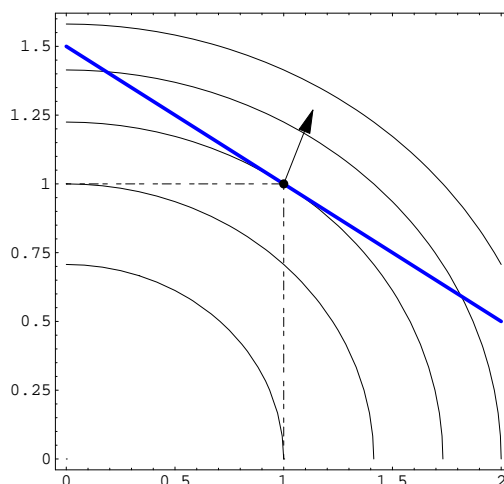


Figure 1.9: Curva de nivel $f(x, y) = c$ y el vector gradiente

Demostración.

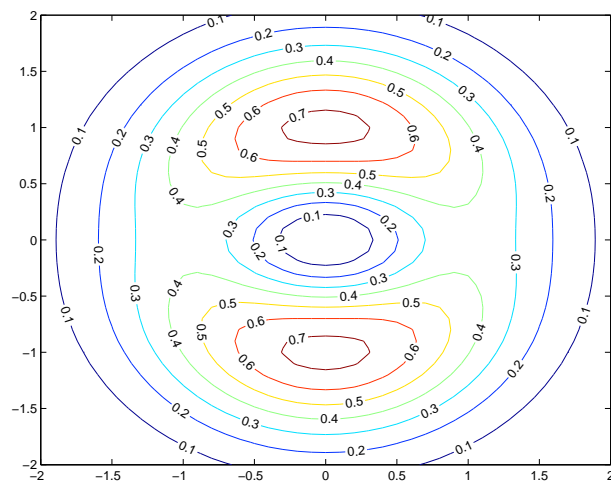
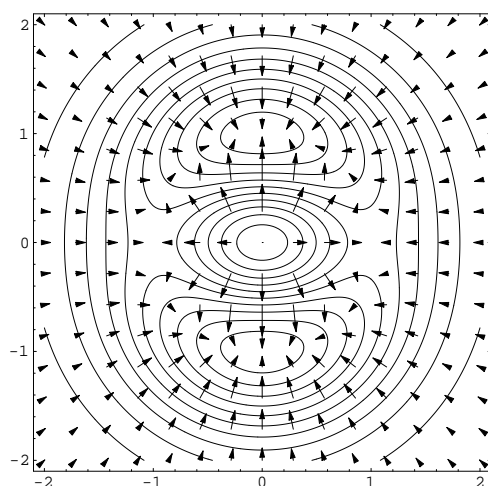
(Véase la figura 1.9). Porque supongamos que podemos parametrizar la curva de nivel por $y = g(x)$ de modo que $g(a) = b$, luego $f(x, g(x)) = c$ para x próximo a $x = a$. Derivando respecto a x en $x = a$ obtenemos

$$0 = \frac{\partial f}{\partial x}(a, b) + \frac{\partial f}{\partial y}(a, b)g'(a) = \left(\frac{\partial f}{\partial x}(a, b), \frac{\partial f}{\partial y}(a, b)\right) \cdot (1, m)$$

donde $(1, m)$ es el vector tangente a la curva de nivel en el punto (a, b) . Lo anterior nos dice que la curva de nivel es ortogonal al gradiente en ese punto. Si la curva de nivel se puede parametrizar como $x = h(y)$ el argumento es similar. El teorema de la función implícita garantiza que si $\nabla f(a, b) \neq \mathbf{0}$ alguna de las parametrizaciones anteriores es posible. □

Ejemplo 1.3–13. En la figura (1.10) se muestran las curvas de nivel de una función $f(x, y)$. Estime la dirección de máximo crecimiento de la función en puntos sobre la curva del nivel 1.

El *campo gradiente* asociado a una función $z = f(x, y)$ es al campo vectorial $X = \nabla f$. En la figura 1.11 se muestra el campo gradiente junto con las curvas de nivel para la función $f(x, y) = (x^2 + y^2) \exp(-x^2 - y^2)$. Observe la propiedad mencionada en la proposición (4). a la curv

Figure 1.10: Curvas de nivel $f(x, y) = const..$ Figure 1.11: Curvas de nivel de la función $f(x, y) = (x^2 + y^2) \exp(-x^2 - y^2)$ y su campo gradiente.

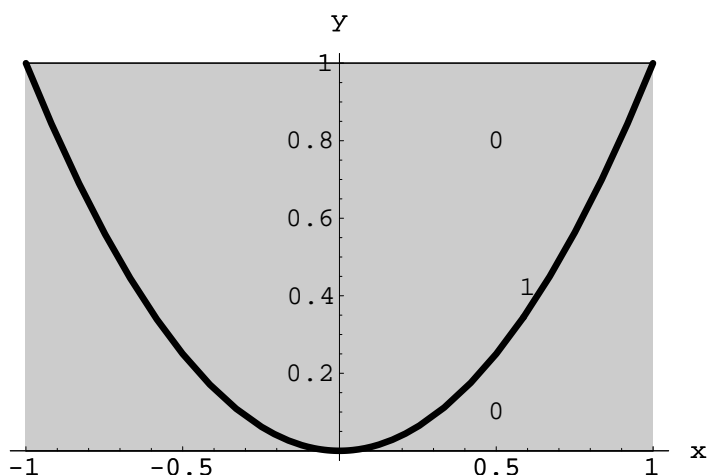


Figure 1.12: Ejemplo 1.4–14

1.4 Discusión previa del concepto de derivada

El concepto de la derivada de una función real como la mejor aproximación lineal en la vecindad de un punto es de fundamental importancia en el Cálculo. Ingenuamente podría suponerse que para una función de dos variables la generalización de la derivada serían las derivadas parciales o las derivadas direccionales. En los siguientes ejemplos vamos a mostrar por qué es necesaria una definición más astringente.

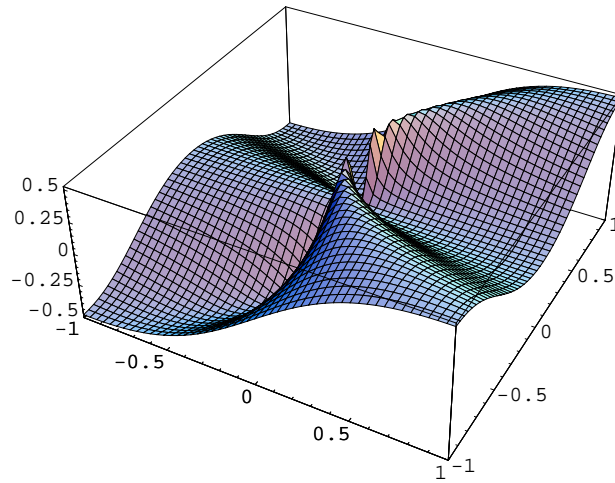
Ejemplo 1.4–14. Sea

$$f(x, y) = \begin{cases} 1 & \text{si } y = x^2 \text{ y } (x, y) \neq (0, 0), \\ 0 & \text{si } y \neq x^2 \text{ o } (x, y) = (0, 0) \end{cases}$$

Véase la figura 1.4. Ya que f se anula a lo largo de los ejes y el origen, se verifica de inmediato que

$$\frac{\partial f}{\partial x}(0) = \frac{\partial f}{\partial y}(0) = 0.$$

Por otro lado no tendría ningún sentido la aproximación lineal de $f(h, k)$ por el plano “tangente“ $z = 0$ ya que la gráfica de la superficie $z = f(x, y)$ tiene una gran “zanja” a lo largo de la curva $y = x^2$. Observe de paso que el límite de f no existe cuando (x, y) tiende a cero.

Figure 1.13: Gráfico de $z = f(x, y)$ para el ejemplo

El siguiente es un ejemplo más sutil aún.

Ejemplo 1.4–15. Sea

$$f(x, y) = \begin{cases} \frac{xy^2}{x^2+y^4} & \text{si } (x, y) \neq (0, 0), \\ 0 & \text{si } (x, y) = (0, 0) \end{cases}$$

Igual que en el ejemplo anterior,

$$\frac{\partial f}{\partial x}(0) = \frac{\partial f}{\partial y}(0) = 0.$$

Analicemos las derivadas direccionales. Para ello consideremos $\hat{\mathbf{u}} = (\alpha, \beta)$. un vector unitario, como $f(0, 0) = 0$ tenemos

$$D_{\hat{\mathbf{u}}}f(0, 0) = \lim_{t \rightarrow 0} \frac{t^3 \alpha \beta^2}{t^2 \alpha^2 + t^4 \beta^4} = \lim_{t \rightarrow 0} t \left(\frac{\alpha \beta^2}{\alpha^2 + t^2 \beta^4} \right)$$

por lo tanto si $\alpha \neq 0$ entonces $D_{\hat{\mathbf{u}}}f(0, 0) = 0$. Si $\alpha = 0$ se verifica directamente que $f(t\alpha, t\beta) = 0$, por lo tanto $D_{\hat{\mathbf{u}}}f(0, 0) = 0$ para cualquier dirección $\hat{\mathbf{u}}$.

Un ojo a la gráfica 1.13 en un entorno de $(0, 0)$ nos sugiere que la función puede no ser continua ahí.

Si nos acercamos al origen a lo largo de rectas $y = mx$ obtenemos

$$\lim_{x \rightarrow 0} f(x, mx) = \lim_{x \rightarrow 0} \frac{xm^2x^2}{x^2 + m^4x^4} = \lim_{x \rightarrow 0} \frac{m^2x}{1 + m^4x^2} = 0$$

lo cual evidencia que el límite podría ser cero. Sin embargo al acercarse al origen a lo largo de $x = y^2$ obtenemos

$$\lim_{x \rightarrow 0} f(y^2, y) = \lim_{y \rightarrow 0} \frac{y^4}{y^4 + y^4} = \lim_{y \rightarrow 0} \frac{y^4}{2y^4} = \lim_{y \rightarrow 0} \frac{1}{2} = \frac{1}{2}$$

que es distinto de $f(0, 0) = 0$. Por lo tanto la función no es continua en $(0, 0)$. Si miramos con más atención gráfica 1.13 vemos que no tiene sentido alguna aproximación lineal en un entorno del origen.

1.5 La derivada

Con el fin generalizar la definición de derivada para funciones de varias variables vamos a re-elaborar la definición usual de derivada para funciones de una variable.

Recordemos la definición de derivada para funciones de una variable

Definición 4. Sea $f: (a - r, a + r) \subset \mathbb{R} \rightarrow \mathbb{R}$. Decimos que la función tiene derivada en a si el límite

$$\lim_{h \rightarrow 0} \frac{f(x + h) - f(a)}{h}$$

existe. En tal caso, el límite $f'(a)$ es la derivada de f en a .

Observe que la existencia del límite anterior es equivalente a escribir

$$f(a + h) - f(a) = f'(a)h + r(h) \tag{1.4}$$

(en realidad definimos $r(h) \equiv f(a + h) - f(a) - f'(a)h$) en donde

$$\lim_{h \rightarrow 0} \frac{r(h)}{h} = 0 \tag{1.5}$$

Diremos que $r(h)$ es un residuo de primer orden y escribiremos¹ $r(h) = o(h)$. Concretamente, (1.4) es una definición equivalente de derivada:

¹Precisamente: si $\delta(h)$ es una función real definida para h próximo de cero, diremos que $\delta(h) = o(h)$ si $\lim_{h \rightarrow 0} \frac{\delta(h)}{h} = 0$.

Proposición 5. Sea $f: (a - r, a + r) \subset \mathbb{R} \rightarrow \mathbb{R}$. Entonces f es derivable en a si y solo si existe un número λ con la propiedad

$$f(a + h) - f(a) = \lambda h + r(h)$$

en donde

$$\lim_{h \rightarrow 0} \frac{r(h)}{h} = 0$$

Demostración.

Hemos probado ya que si $f'(a)$ existe entonces se puede escribir como en (1.4) con la propiedad (1.5). Ahora supongamos que es posible escribir

$$f(a + h) - f(a) = \lambda h + r(h)$$

para cierto λ , donde $r(h)$ satisface (1.5). Veamos que la derivada $f'(a)$ existe,

$$\lim_{h \rightarrow 0} \frac{f(a + h) - f(a)}{h} = \lim_{h \rightarrow 0} \frac{\lambda h + r(h)}{h} = \lambda + \lim_{h \rightarrow 0} \frac{r(h)}{h} = \lambda$$

es decir la derivada existe y es precisamente $f'(a) = \lambda$. □

Observe que de la igualdad (1.4) se sigue que

$$\lim_{h \rightarrow 0} f(a + h) = f(a)$$

equivalentemente $\lim_{x \rightarrow a} f(x) = f(a)$, lo que significa que la función *debe ser continua en a* .

Teorema 2. Sea $f: (a - r, a + r) \subset \mathbb{R} \rightarrow \mathbb{R}$. Si f es derivable en a entonces es continua en a .

Con el fin de tener una notación más cómoda introducimos la noción de incremento funcional

Definición 5. Si $z = f(x, y)$ está definida en una bola $B_r(a, b) \subset \mathbb{R}^2$, definimos el incremento funcional en el punto (a, b) con incrementos independientes (h, k) como

$$\Delta f(a, b; (h, k)) = f(a + h, b + k) - f(a, b)$$

Adoptaremos (1.4,1.5) como definición, pues tiene varias ventajas sobre la original, en particular que puede extenderse al caso general de funciones definidas entre espacios euclidianos de cualquier dimensión y aún entre espacios de dimensión infinita.

Definición 6 (funciones reales de variable real). Sea $f: (a-r, a+r) \subset \mathbb{R} \rightarrow \mathbb{R}$. Decimos que es diferenciable en a si y sólo si existe un número λ tal que es posible escribir

$$f(a+h) - f(a) = \lambda h + r(h)$$

en donde

$$\lim_{h \rightarrow 0} \frac{r(h)}{h} = 0$$

Comentario Podemos re-escribir (1.4) como

$$f(a+h) - (f'(a)h + f(a)) = r(h)$$

y como la ecuación de la recta tangente a la curva $y = f(x)$ en el punto $(a, f(a))$ es precisamente $y = f'(a)x + f(a)$, la igualdad anterior nos dice que la diferencia entre los valores de la función y los de la recta tangente vienen dados por $r(h)$, un residuo de orden uno. También se dice que la gráfica de la función y la recta tangente *tienen contacto de primer orden en el punto* $(a, f(a))$.

Podemos ahora generalizar las ideas anteriores para obtener el concepto de derivada de una función $f: \mathbb{R}^2 \rightarrow \mathbb{R}$.

Definición 7. Sea $f: B_r(a, b) \subset \mathbb{R}^2 \rightarrow \mathbb{R}$. Decimos que f es diferenciable en (a, b) si existen dos números λ, μ tales que

$$f(a+h, b+k) = f(a, b) + \lambda h + \mu k + r(h, k)$$

donde

$$\lim_{(h,k) \rightarrow (0,0)} \frac{r(h, k)}{\|(h, k)\|} = 0,$$

en tal caso, la derivada es la función lineal $Df(a, b): \mathbb{R}^2 \rightarrow \mathbb{R}$, $Df(a, b) \cdot (h, k) = \lambda h + \mu k$.

Observación 1. ¿Por qué definir así la derivada? Podríamos proponer de inmediato a la pareja $[\lambda \ \mu]$, pero por varias razones que se irán aclarando a medida que avance la exposición, es mucho más conveniente considerar a la derivada como una transformación lineal. De cualquier forma, en la base canónica, la matriz 1×2 , asociada a la derivada es $[Df(a, b)] = [\lambda \ \mu]$ y no perdemos nada.

Observación 2. En cuanto al residuo, más explícitamente, se cumple que

$$\lim_{(h,k) \rightarrow (0,0)} \frac{r(h,k)}{\sqrt{h^2 + k^2}} = 0$$

Por un resultado del Análisis que se verá en la segunda parte de estas notas, todas las normas en \mathbb{R}^n son equivalentes, es decir dan el mismo límite, cuando existe. En particular podemos utilizar cualquier norma conveniente en el denominador de la expresión anterior. Así, la propiedad de residuo es equivalente a, p.ej.

$$\lim_{(h,k) \rightarrow (0,0)} \frac{r(h,k)}{|h| + |k|} = 0$$

y en casos concretos podemos usar la norma que más convenga.

Observación 3. En la definición de derivada (7) podemos escribir

$$\Delta f(a, b; (h, k)) = Df(a, b) \cdot (h, k) + r(h, k)$$

y en el mismo sentido que para funciones de una variable la derivada es la mejor aproximación lineal a la gráfica de la superficie $z = f(x, y)$ en un entorno del punto $(a, b, f(a, b))$.

Observación 4. Que la función sea derivable en (a, b) es más fuerte sólo tener derivadas parciales y aún direccionales. Con más precisión:

Proposición 6. *Sea $f: B_r(a, b) \subset \mathbb{R}^2 \rightarrow \mathbb{R}$. Si f es derivable en (a, b) entonces existen las derivadas direccionales en (a, b) en cualquier dirección.*

Demostración.

Porque si (α, β) es un vector unitario, tomando en la definición de derivada (??) los incrementos independientes como $(h, k) = (t\alpha, t\beta)$ se tiene

$$f(a + t\alpha, b + t\beta) - f(a, b) = \lambda(t\alpha) + \mu(t\beta) + r(t\alpha, t\beta)$$

Si ahora calculamos el cociente que define la derivada direccional

$$\begin{aligned} \frac{f(a + t\alpha, b + t\beta) - f(a, b)}{t} &= \lambda\alpha + \mu\beta + \frac{r(t\alpha, t\beta)}{t} \\ &= \lambda\alpha + \mu\beta + \frac{r(t\alpha, t\beta)}{\|(t\alpha, t\beta)\|} \frac{\|(t\alpha, t\beta)\|}{t} \\ &= \lambda\alpha + \mu\beta + \frac{r(t\alpha, t\beta)}{\|(t\alpha, t\beta)\|} \frac{|t|}{t} \end{aligned}$$

aunque $|t|/t$ no tiene límite cuando t tiende a cero, está acotado y por la propiedad del residuo el límite del último término es cero. Por lo tanto la derivada direccional existe y es

$$D_{(\alpha,\beta)}f(a,b) = \lambda\alpha + \mu\beta.$$

□

En particular se tiene para las derivadas parciales

$$D_1f(a,b) = \lambda, \quad D_2f(a,b) = \mu$$

es decir

$$D_{(\alpha,\beta)}f(a,b) = D_1f(a,b)\alpha + D_2f(a,b)\beta.$$

siempre que f sea derivable en (a,b) , y en tal caso se tiene la fórmula muy útil para calcular la derivada

$$Df(a,b) \cdot (h,k) = D_1f(a,b)h + D_2f(a,b)k. \quad (1.6)$$

Observación 5. Remarcamos que la sola existencia de las derivadas parciales no implica la existencia de la derivada, aunque la fórmula (1.6) sea tentadora. Para poder aplicarla es necesario saber por otros argumentos que la función es derivable (véase por ejemplo la sección ??)

Observación 6. En notación clásica, (1.6) se escribe como

$$dz = \frac{\partial z}{\partial x}dx + \frac{\partial z}{\partial y}dy$$

y dx, dy, dz se conocen como *diferenciales*. Por esta razón es que se usa indistintamente el término “derivable” o diferenciable.

Ejemplos En muchos casos, incluidas algunas demostraciones, la mejor estrategia para obtener la diferenciabilidad es calcular el incremento $\Delta f(a,b; (h,k)) = f(a+h, b+k) - f(a,b)$ y expandir en términos lineales en h, k más un residuo, el candidato natural para la derivada serán los coeficientes de la parte lineal en h, k . Después debe verificarse cuidadosamente que² $r(h,k)$ es $o(h,k)$.

²Se puede extender naturalmente la noción de residuo de primer orden: Si $\delta(h,k)$ está definido para h, k próximo de cero decimos que $\delta(h,k) = o(h,k)$ si $\lim_{(h,k) \rightarrow (0,0)} \frac{\delta(h,k)}{\|(h,k)\|} = 0$.

Ejemplo 1.5-16. Calcular la derivada de $f(x, y) = x^2 + y^3$.

Calculamos el incremento funcional y expandemos en h, k

$$\begin{aligned}\Delta f(a, b; (h, k)) &= (a + h)^2 + (b + k)^3 - (a^2 + b^3) \\ &= \underbrace{(2ah + 3bk)}_{Df(a,b)(h,k)} + \underbrace{(h^2 + 2bk^2 + k^3)}_{r(h,k)}\end{aligned}$$

Ahora debemos ver que $r(h, k) = o(h, k)$. Observe que el grado mínimo del numerador es 2, así

$$\begin{aligned}\lim_{(h,k) \rightarrow (0,0)} \frac{r(h, k)}{\|(h, k)\|} &= \lim_{(h,k) \rightarrow (0,0)} \frac{h^2 + 2bk^2 + k^3}{\sqrt{h^2 + k^2}} \\ &= \lim_{(h,k) \rightarrow (0,0)} \frac{h^2}{\sqrt{h^2 + k^2}} + \lim_{(h,k) \rightarrow (0,0)} \frac{2bk^2}{\sqrt{h^2 + k^2}} + \\ &\quad \lim_{(h,k) \rightarrow (0,0)} \frac{k^3}{\sqrt{h^2 + k^2}} +\end{aligned}$$

usando alguna de las desigualdades $h^2 + k^2 \geq h^2$, $h^2 + k^2 \geq k^2$ se puede probar que cada término tiene límite cero.

Ejemplo 1.5-17. Investigar la diferenciabilidad de $f(x, y) = \sin x + \sin y$.

$$\begin{aligned}\Delta f(a, b; (h, k)) &= \sin(a + h) + \sin(b + k) - \sin(a) - \sin(b) \\ &= \sin a \cos h + \sin b \cos k - \sin(a) - \sin(b)\end{aligned}$$

En este caso la expansión en términos lineales en h, k podría hacerse usando el teorema del valor medio

$$\begin{aligned}\sin a \cos h + \sin b \cos k - \sin(a) - \sin(b) \\ &= \sin(a) + \cos(a^*)h + \sin(b) - \cos(b^*)k - \sin(a) - \sin(b) \\ &= \cos(a^*)h - \cos(b^*)k\end{aligned}$$

en donde a^* está entre a y $a + h$ y b^* está entre b y $b + k$. Por ello a^* depende de a y h , b^* depende de b y k . Necesitamos coeficientes lineales en h y k que no dependan de h, k . Ello puede subsanarse notando que para (h, k) próximo a cero, (a^*, b^*) es próximo a (a, b) , luego podemos escribir

$$\begin{aligned}\Delta f(a, b; (h, k)) &= (\cos(a^*) \pm \cos(a))h - (\cos(b^*) \pm \cos(b))k \\ &= \cos(a)h - \cos(b)k + r(h, k)\end{aligned}$$

donde

$$r(h, k) \equiv (\cos(a^*) - \cos(a))h + (\sin(b^*) - \sin(b))k$$

Veamos ahora el residuo. Como queremos probar el límite $r(h, k)/\|(h, k)\|$ es cero, tomamos valor absoluto para poder estimar

$$\begin{aligned} & \left| \frac{(\cos(a^*) - \cos(a))h + (\sin(b^*) - \sin(b))k}{\sqrt{h^2 + k^2}} \right| \\ & \leq |\cos(a^*) - \cos(a)| \frac{|h|}{\sqrt{h^2 + k^2}} \\ & \quad + |\sin(b^*) - \sin(b)| \frac{|k|}{\sqrt{h^2 + k^2}} \end{aligned}$$

Los términos $|\cos(a^*) - \cos(a)|$, $|\sin(b^*) - \sin(b)|$ tienden a cero cuando $(h, k) \rightarrow (0, 0)$ por *continuidad de las funciones* \cos, \sin , en tanto que los factores

$$\frac{|h|}{\sqrt{h^2 + k^2}} \rightarrow 0, \quad \frac{|k|}{\sqrt{h^2 + k^2}} \rightarrow 0.$$

En conclusión f es diferenciable en todo punto (a, b) y

$$Df(a, b) \cdot (h, k) = \cos(a)h - \cos(b)k$$

1.6 Derivación de funciones en $\mathbb{R}^n \rightarrow \mathbb{R}^m$

En lo que sigue supondremos, salvo pocas excepciones, que las funciones a considerar están definidas en todo un espacio euclideo, para simplificar la notación y presentar ms claramente los conceptos básicos. Sin embargo, todos los enunciados pueden fácilmente re-escribirse en términos de los dominios apropiados. Volveremos a este punto al establecer la Regla de la Cadena. Por último, en algunos temas de carácter local, tales como los Teoremas de Función Implícita e Inversa es indispensable considerar que las funciones están sólo definidas en ciertos subconjuntos de los espacios euclideos respectivos.

Definición 8. Sea $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$, $\mathbf{p} \in \mathbb{R}^n$. Decimos que f es diferenciable en \mathbf{p} si existe una transformación lineal $\Lambda: \mathbb{R}^n \rightarrow \mathbb{R}^m$ tal que

$$f(\mathbf{p} + \mathbf{h}) - f(\mathbf{p}) = \Lambda \cdot \mathbf{h} + r(\mathbf{h})$$

donde

$$\lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{r(\mathbf{h})}{\|\mathbf{h}\|} = \mathbf{0}$$

Ejemplo 1.6–18. [Derivada de una constante]. Sea $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$, $f(\mathbf{z}) = \mathbf{c}$, una constante, entonces $\Delta f(\mathbf{p}; \mathbf{h}) = \mathbf{0}$, por lo que $Df(\mathbf{p}) \equiv 0$ (la función lineal cero).

Ejemplo 1.6–19. [Derivada de una transformación lineal]. Si $L: \mathbb{R}^n \rightarrow \mathbb{R}^m$ es lineal, entonces $DL(\mathbf{p}) = L$ para todo $\mathbf{p} \in \mathbb{R}^n$. pues $L(\mathbf{p} + \mathbf{h}) - L(\mathbf{p}) = L(\mathbf{p}) + L(\mathbf{h}) - L(\mathbf{p}) = L(\mathbf{h})$ así que la transformación lineal que funciona es la misma L con residuo idénticamente igual a cero.

Observación 7. Los ejemplos anteriores muestran que es necesario distinguir entre *em la derivada en un punto*, la cual es una transformación lineal, y *la función derivada*, la cual es una función que asocia a cada punto \mathbf{x} —donde la derivada exista— una transformación lineal. Así, para $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$, tenemos que

$$Df(\mathbf{x}) \in L(\mathbb{R}^n, \mathbb{R}^m) \quad \text{es una transformación lineal,}$$

en tanto que

$$Df: U \subset \mathbb{R}^n \rightarrow L(\mathbb{R}^n, \mathbb{R}^m),$$

donde U es el conjunto de puntos donde la función es derivable, es una función que no es necesariamente lineal. En el ejemplo 1.5, $Df(x, y) \cdot (h, k) = 2xh + 3y^2k$ no es lineal con respecto a x o y . Desde luego, fijado el punto (x, y) la transformación correspondiente es lineal en (h, k) .

Ejemplo 1.6–20. Podemos incluir el caso de funciones $f: \mathbb{R} \rightarrow \mathbb{R}$ en nuestra definición general. En principio, la derivada de una tal función en un punto es un número: $f'(x)$. Pero podemos interpretar tal número como la transformación lineal que se obtiene por multiplicación por la constante. Así $\lambda \cdot h = f'(a)h$. Así, la definición (8) es un caso particular de (1.4).

1.7 Algunas propiedades teóricas

Denotemos por $L(\mathbb{R}^n, \mathbb{R}^m)$ el conjunto de las transformaciones lineales $\mathbb{R}^n \rightarrow \mathbb{R}^m$. Este conjunto es en sí mismo un espacio vectorial, pues están definidas la

suma y producto de una transformación lineal por un escalar puntualmente, es decir.

$$(f + g)(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x}), \quad (\lambda f)(\mathbf{x}) = \lambda f(\mathbf{x}).$$

(acostubraremos en lo sucesivo indicar por $f \cdot \mathbf{x}$ la aplicación de la función en el punto \mathbf{x} , cuando f sea una transformación lineal; la notación es reminiscente de la operación de producto de una matriz por un vector. Las propiedad de nuestro intereés aquí es que se puede definir una noción de tamaño o má precisamente de *norma* de una transformación lineal como sigue: Sean $\|\cdot\|_{\mathbb{R}^n}$, $\|\cdot\|_{\mathbb{R}^m}$ las normas 1 respectivas en \mathbb{R}^n y \mathbb{R}^m , es decir $\|\mathbf{x}\|_{\mathbb{R}^n} = \sum_{i=1}^n |x_i|$ para $\mathbf{x} \in \mathbb{R}^n$ y de manera similar en \mathbb{R}^m . Recurriendo a una base cualquiera \mathbf{e}_i , $i = 1, 2, \dots, n$ de \mathbb{R}^n , obtenemos

$$\begin{aligned} \|L(\mathbf{x})\|_{\mathbb{R}^m} &= \left\| L\left(\sum_{i=1}^n x_i \mathbf{e}_i\right) \right\|_{\mathbb{R}^m} \\ &= \left\| \sum_{i=1}^n x_i L(\mathbf{e}_i) \right\|_{\mathbb{R}^m} \\ &\leq M \sum_{i=1}^n |x_i| = M \|\mathbf{x}\|_{\mathbb{R}^n} \end{aligned}$$

para alguna constante $M > 0$, ya que el conjunto $\{\|L(\mathbf{e}_i)\|_{\mathbb{R}^m} \mid i = 1, 2, \dots, n\}$ es finito. En otras palabras, para toda \mathbf{x} distinto de $\mathbf{0}$ el cociente

$$\frac{\|L(\mathbf{x})\|_{\mathbb{R}^m}}{\|\mathbf{x}\|_{\mathbb{R}^n}}$$

está acotado, por lo tanto podemos considerar

$$\|L\|_{\mathbb{R}^n, \mathbb{R}^m} \equiv \sup\left\{ \frac{\|L(\mathbf{x})\|_{\mathbb{R}^m}}{\|\mathbf{x}\|_{\mathbb{R}^n}} \mid \mathbf{x} \neq \mathbf{0} \right\}$$

Se puede verificar que $\|L\|_{\mathbb{R}^n, \mathbb{R}^m}$ define una norma en el espacio vectorial $L(\mathbb{R}^n, \mathbb{R}^m)$ (véase la parte II de estas notas). En particular, si en lo sucesivo omitimos los subíndices en las normas involucradas,

$$\|L\mathbf{x}\| \leq \|L\| \|\mathbf{x}\| \tag{1.7}$$

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

Se tiene en particular

Proposición 7. Sea $L \in L(\mathbb{R}^n, \mathbb{R})$ entonces L es continua en $\mathbf{0}$. Además son equivalentes la siguientes afirmaciones:

1. L es continua en $\mathbf{0}$.
2. L es continua en \mathbf{x} para toda $\mathbf{x} \in \mathbb{R}^n$.

Demostración.

De la desigualdad (1.7) se sigue que $\|L \cdot \mathbf{x}\| < \epsilon$ si $\|\mathbf{x}\| < \epsilon/\|L\|$, por lo tanto L es continua en $\mathbf{0}$.

Suponga ahora que L es continua en $\mathbf{0}$. Para $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ se tiene

$$\|L \cdot \mathbf{x} - L \cdot \mathbf{y}\| = \|L \cdot (\mathbf{x} - \mathbf{y})\| \leq \|L\| \|\mathbf{x} - \mathbf{y}\|$$

lo que prueba que L es continua en \mathbf{x} . □

Análogamente, una función $B: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^m$ se dice *bilineal* si para $\mathbf{x} \in \mathbb{R}^n$ fijo la función $B(\mathbf{x}, \cdot): \mathbb{R}^n \rightarrow \mathbb{R}^m$ es lineal y para $\mathbf{y} \in \mathbb{R}^n$ fijo la función $B(\cdot, \mathbf{y}): \mathbb{R}^n \rightarrow \mathbb{R}^m$ es lineal. Al igual que para las funciones lineales recurriendo a una base $\mathbf{e}_i, i = 1, 2, \dots, n$ de \mathbb{R}^n , podemos escribir

$$\begin{aligned} \|B(\mathbf{x}, \mathbf{y})\| &= \left\| \sum_{i,j=1}^n x_i y_j B(\mathbf{e}_i, \mathbf{e}_j) \right\| \\ &\leq \sum_{i=1}^n |x_i| |y_j| \|B(\mathbf{e}_i, \mathbf{e}_j)\| \\ &\leq n \|\mathbf{x}\| \sum_{i=1}^n |y_j| \|B(\mathbf{e}_i, \mathbf{e}_j)\| \\ &\leq n^2 \|\mathbf{x}\| \|\mathbf{y}\| \sum_{i=1}^n \|B(\mathbf{e}_i, \mathbf{e}_j)\| \\ &\leq M \|\mathbf{x}\| \|\mathbf{y}\| \end{aligned}$$

para alguna constante $M > 0$, de donde el cociente

$$\frac{\|B(\mathbf{x}, \mathbf{y})\|}{\|\mathbf{x}\| \|\mathbf{y}\|}$$

es siempre acotado, si $\mathbf{x}, \mathbf{y} \neq \mathbf{0}$. Por lo tanto se puede definir la norma de la aplicación bilineal B , como

$$\|B\| = \sup \left\{ \frac{\|B(\mathbf{x}, \mathbf{y})\|}{\|\mathbf{x}\| \|\mathbf{y}\|} \mid \mathbf{x}, \mathbf{y} \neq \mathbf{0} \right\}$$

y naturalmente se tiene

$$\|B(\mathbf{x}, \mathbf{y})\| \leq \|B\| \|\mathbf{x}\| \|\mathbf{y}\|$$

para toda $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

Proposición 8. Si $\Lambda, \Lambda' \in L(\mathbb{R}^n, \mathbb{R})$ satisfacen ambas

$$\Delta f(\mathbf{p}; \mathbf{h}) = \Lambda \cdot \mathbf{h} + o(\mathbf{h}), \quad \Delta f(\mathbf{p}; \mathbf{h}) = \Lambda' \cdot \mathbf{h} + o(\mathbf{h}),$$

entonces $\Lambda = \Lambda'$

Demostración.

Restando ambas igualdades tenemos

$$S \cdot \mathbf{h} = o(\mathbf{h})$$

donde $S = \Lambda - \Lambda'$, es decir

$$\lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{S \cdot \mathbf{h}}{\|\mathbf{h}\|} = \mathbf{0}$$

Tomemos una base de vectores unitarios \mathbf{e}_i , $i = 1, 2, \dots, n$ y de manera sucesiva en i , $\mathbf{h} = \delta \mathbf{e}_i$, con $\delta \rightarrow 0+$ luego,

$$\begin{aligned} \mathbf{0} &= \lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{S \cdot \mathbf{h}}{\|\mathbf{h}\|} \\ &= \lim_{\delta \rightarrow 0} \frac{\delta S \cdot \mathbf{e}_i}{|\delta|} \\ &= \frac{\delta}{|\delta|} S \cdot \mathbf{e}_i \\ &= S \cdot \mathbf{e}_i \end{aligned}$$

para $i = 1, 2, \dots, n$, es decir $S \equiv \mathbf{0}$ como transformación lineal. Por lo tanto $\Lambda = \Lambda'$. \square

El resultado anterior justifica hablar de la derivada en el punto \mathbf{p} y escribirla como $Df(\mathbf{p})$.

Teorema 3. Si la derivada de $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ en \mathbf{p} existe, entonces la función es continua en \mathbf{p} .

Demostración.

Como $Df(\mathbf{p})$ es lineal, entonces es continua en $\mathbf{0}$, de

$$\Delta f(\mathbf{x}; \mathbf{h}) = f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) = Df(\mathbf{p}) \cdot \mathbf{h} + o(\mathbf{h})$$

se sigue que si $\mathbf{h} \rightarrow \mathbf{0}$, entonces $\Delta f(\mathbf{x}; \mathbf{h}) \rightarrow \mathbf{0}$ lo que prueba la continuidad. \square

La recíproca no es cierta. Ya en una variable la función $|x|$ es continua pero no es derivable en el origen.

La noción de derivada parcial se puede también extender a funciones en $\mathbb{R}^n \rightarrow \mathbb{R}^m$.

Definición 9. Sea $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$. Sea $\{\boldsymbol{\delta}_i, i = 1, 2, \dots, n\}$ la base canónica en \mathbb{R}^n . La derivada parcial i -ésima en el punto $\mathbf{p} \in \mathbb{R}^n$ se define por

$$D_i f(\mathbf{p}) = \lim_{t \rightarrow 0} \frac{f(\mathbf{p} + t\boldsymbol{\delta}_i) - f(\mathbf{p})}{t}$$

si tal límite existe.

Observe que la derivada parcial es un vector $D_i f(\mathbf{p}) \in \mathbb{R}^m$.

Proposición 9. Sea $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$. Si f es diferenciable en $\mathbf{p} \in \mathbb{R}^n$, entonces existen todas las derivadas parciales y

$$D_i f(\mathbf{p}) = Df(\mathbf{p}) \cdot \boldsymbol{\delta}_i$$

Demostración.

En la definición de derivada, tomando el incremento de la forma $\mathbf{h} = t\boldsymbol{\delta}_i$ con $t \rightarrow 0$ tenemos

$$\begin{aligned} & \frac{f(\mathbf{p} + t\boldsymbol{\delta}_i) - f(\mathbf{p})}{t} \\ &= \frac{1}{t} Df(\mathbf{p}) \cdot (t\boldsymbol{\delta}_i) + \frac{1}{t} r(t\boldsymbol{\delta}_i) \\ &= Df(\mathbf{p}) \cdot \boldsymbol{\delta}_i + \frac{1}{t} r(t\boldsymbol{\delta}_i) \rightarrow Df(\mathbf{p}) \cdot \boldsymbol{\delta}_i, \quad \text{cuando } t \rightarrow 0. \end{aligned}$$

\square

1.8 Matriz Jacobiana, gradiente

La proposición anterior se puede interpretar como sigue. Recordemos que la matriz asociada a $Df(\mathbf{p}): \mathbb{R}^n \rightarrow \mathbb{R}^m$ relativa a la bases canónicas $\boldsymbol{\delta}_i$, $i = 1, 2, \dots, n$ y $\boldsymbol{\delta}'_j$, $j = 1, 2, \dots, m$ en \mathbb{R}^n y \mathbb{R}^m es una matriz $m \times n$ cuya i -ésima columna es el vector $Df(\mathbf{p}) \cdot \boldsymbol{\delta}_i$ expresado en la base $\boldsymbol{\delta}'_j$ o sea

$$\begin{aligned} Df(\mathbf{p}) \cdot \boldsymbol{\delta}_i &= \sum_{j=1}^m (\boldsymbol{\delta}'_j \cdot Df(\mathbf{p}) \cdot \boldsymbol{\delta}_i) \boldsymbol{\delta}'_j \\ &= \sum_{j=1}^m (Df_j(\mathbf{p}) \cdot \boldsymbol{\delta}_i) \boldsymbol{\delta}'_j \\ &= \sum_{j=1}^m D_i f_j(\mathbf{p}) \boldsymbol{\delta}'_j \end{aligned}$$

En otras palabras la matriz asociada a la derivada es la *matriz Jacobiana*

$$\begin{bmatrix} D_1 f_1 & D_2 f_1 & \cdots & D_n f_1 \\ D_1 f_2 & D_2 f_2 & \cdots & D_n f_2 \\ \cdots & \cdots & \cdots & \cdots \\ D_1 f_m & D_2 f_m & \cdots & D_n f_m \end{bmatrix}$$

Ejemplo 1.8-21. Calcule la matriz Jacobiana de $f: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ dada por $f(x, y) = (x^2 - y^2, 2xy)$.

Tenemos $f_1(x, y) = x^2 - y^2$, $f_2(x, y) = 2xy$.

$$[Df(, xy)] = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{bmatrix} = \begin{bmatrix} 2x & -2y \\ 2y & -2x \end{bmatrix}$$

Ejemplo 1.8-22. En el caso de funciones $f: \mathbb{R}^n \rightarrow \mathbb{R}$ la matriz Jacobiana se representa por una matriz $1 \times n$ o un vector renglón que se conoce como *vector gradiente en coordenadas cartesianas*

$$[Df(x)] = [D_1 f(x), D_2 f(x), \dots, D_n f(x)] \equiv \nabla f(x)$$

Si $\mathbf{h} \in \mathbb{R}^n$ entonces

$$Df(\mathbf{x}) \cdot \mathbf{h} = \nabla f(\mathbf{x}) \cdot \mathbf{h}$$

Es sólo en este caso que el gradiente se puede identificar con la derivada.

En general, si $\langle \cdot, \cdot \rangle$ es un producto escalar y q_1, q_2, \dots, q_n son coordenadas arbitrarias el vector gradiente se define por la igualdad anterior, es decir

$$Df(\mathbf{q}) \cdot \mathbf{h} = \langle \nabla f(x), \mathbf{h} \rangle$$

1.8.1 Coordenadas curvilíneas

Suponga que se da un producto escalar en cada punto $\langle \cdot, \cdot \rangle_{\mathbf{q}}$ (ésto es lo que se llama una métrica Riemanniana). El campo gradiente se define por la identidad

$$Df(\mathbf{q}) \cdot \mathbf{h} = \langle \nabla f(\mathbf{q}), \mathbf{h} \rangle_{\mathbf{q}} \quad (1.8)$$

Esta construcción es usual cuando se usan coordenadas *curvilíneas*, o *generalizadas*. En tal caso sea $\mathbf{r} = \mathbf{r}(q_1, q_2, \dots, q_n)$ el cambio a coordenadas cartesianas $\mathbf{r} = (x_1, x_2, \dots, x_n)$. La condición de que sea un cambio de coordenadas significa que en cada punto \mathbf{q} los vectores

$$\mathbf{e}_i(\mathbf{q}) \equiv \frac{\partial \mathbf{r}}{\partial q_i}, \quad i = 1, 2, \dots, n$$

formen un conjunto linealmente independiente. Equivalentemente el determinante de matriz Jacobiana

$$\frac{\partial(x_1, x_2, \dots, x_n)}{\partial(q_1, q_2, \dots, q_n)}$$

es siempre distinto de cero. Defina entonces

$$\langle \mathbf{v}, \mathbf{v} \rangle_{\mathbf{q}} = \sum_{i,j=1}^n (\mathbf{e}_i \cdot \mathbf{e}_j) v_i w_j \quad (1.9)$$

donde

$$\mathbf{v} = \sum_{i=1}^n v_i \mathbf{e}_i, \quad \mathbf{w} = \sum_{j=1}^n w_j \mathbf{e}_j$$

definen las coordenadas v_i, w_j en términos de la base \mathbf{e}_j . Es costumbre denotar la relación métrica 1.9 en la forma

$$ds^2 = \sum_{i,j=1}^n g_{ij} dq_i dq_j$$

donde

$$g_{ij} = (\mathbf{e}_i \cdot \mathbf{e}_j)$$

son los coeficientes métricos y dependen en general de \mathbf{q}

En el caso de una base ortogonal se cumple

$$\mathbf{e}_i \cdot \mathbf{e}_j = 0, \quad \text{si } i \neq j$$

por lo que solo los coeficientes métricos diagonales $g_{11}, g_{22}, \dots, g_{nn}$ son distintos de cero (ninguno es cero por hipótesis). En tal caso se puede escribir

$$ds^2 = \sum_i^n g_{ii} dq_i^2 = \sum_i^n h_i^2 dq_i^2,$$

que equivale a escribir

$$\mathbf{e}_i = h_i \hat{\mathbf{e}}_i$$

donde $\hat{\mathbf{e}}_i$ es un vector unitario en la dirección \mathbf{e}_i , es decir tanto $h_i = \|\mathbf{e}_i\|$.

Ejemplo 1.8-23. Calcule los coeficientes métricos h_r, h_θ en coordenadas polares y el gradiente de una función $f(r, \theta)$

El cambio a coordenadas cartesianas es

$$x_1 = r \cos \theta, \quad x_2 = r \sin \theta$$

por lo que la matriz Jacobiana es

$$\begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix}$$

de donde

$$\mathbf{e}_r = (\cos \theta, \sin \theta), \quad \mathbf{e}_\theta = r(-\sin \theta, \cos \theta)$$

Es claro que el sistema de coordenadas es ortogonal, pues $\mathbf{e}_r \cdot \mathbf{e}_\theta = 0$. Como \mathbf{e}_r es unitario entonces $h_r = 1$, por otro lado $\|\mathbf{e}_\theta\| = r$ de donde $h_\theta = r$.

Sea $\nabla f(r, \theta) = A\mathbf{e}_r + B\mathbf{e}_\theta$. De la definición del gradiente (1.8) tenemos

$$\langle \nabla f(r, \theta), (h, k) \rangle = h_r Ah + h_\theta B = A + rBk = Df(r, \theta)(h, k) = \frac{\partial f}{\partial r} h + \frac{\partial f}{\partial \theta} k$$

para cualesquiera (h, k) por lo tanto

$$A = \frac{\partial f}{\partial r}, \quad \text{y} \quad B = \frac{1}{r} \frac{\partial f}{\partial \theta}$$

1.8.2 Ejemplos de no diferenciabilidad

En esta sección veremos algunos ejemplos “exóticos” que manera de contraejemplos muestran que la existencia de derivadas parciales o continuidad no son suficientes para tener la diferenciabilidad.

Ejemplo 1.8-24. Considérese $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ definida por

$$f(x, y) = \begin{cases} \frac{xy}{x^2+y^2} & \text{si } (x, y) \neq (0, 0) \\ 0 & \text{si } (x, y) = (0, 0) \end{cases}$$

Como la función se anula a lo largo de los ejes, las derivadas parciales existen en el origen y

$$\frac{\partial f}{\partial x}(0, 0) = \frac{\partial f}{\partial y}(0, 0) = 0$$

de manera que si esta función fuese diferenciable en el origen, se cumpliría

$$\frac{hk}{h^2 + k^2} - 0 = 0h + 0k + r(h, k)$$

En particular $\lim_{(h,k) \rightarrow (0,0)} r(h, k) = 0$, sin embargo, en el ejercicio (1.1) probamos que

$$\lim_{(h,k) \rightarrow (0,0)} \frac{hk}{h^2 + k^2} - 0 = 0h + 0k + r(h, k)$$

NO existe, por lo que llegamos a una contradicción. Por lo tanto f no es diferenciable en el origen.

Otra forma de verlo: esta función no es continua, entonces por la Proposición 3, no puede ser diferenciable.

Este ejemplo también prueba que la recíproca de la Proposición dif-pder no es vlida, pues la función es no diferenciable, pero existen las derivadas parciales inclusive en el origen.

Ejemplo 1.8-25. Considere la función

$$f(x, y) = \begin{cases} \frac{x^2y}{x^2+y^2} & \text{si } (x, y) \neq (0, 0) \\ 0 & \text{si } (x, y) = (0, 0) \end{cases}$$

Esta función es continua en el origen, y tiene derivadas parciales en todo punto (iguales a cero en el origen), pero no es diferenciable en el origen. Veamos por qué: por un argumento similar al usado en el ejemplo 1.8.2 tendríamos

$$\frac{h^2k}{h^2 + k^2} = r(h, k)$$

con $r(h, k) = o(h, k)$. Sin embargo,

$$\lim_{(h,k) \rightarrow (0,0)} \frac{h^2k}{h^2 + k^2} \frac{1}{\sqrt{h^2 + k^2}}$$

no existe pues tomando $k = h$ tenemos

$$\begin{aligned} \lim_{(h,k) \rightarrow (0,0)} \frac{h^2k}{(h^2 + k^2)^{3/2}} &= \lim_{h \rightarrow 0} \frac{h^3}{(h^2 + h^2)^{3/2}} = \lim_{h \rightarrow 0} \\ &= \frac{h^3}{(2h^2)^{3/2}} = \frac{1}{(2)^{3/2}} \neq 0 \end{aligned}$$

El teorema que sigue es muy importante en la práctica, pues nos permite establecer la diferenciableidad de una función que tiene derivadas parciales más una propiedad adicional.

Teorema 4. Si $f: \mathbb{R}^n \rightarrow \mathbb{R}$ tiene derivadas parciales continuas en una bola $B_r(\mathbf{p}) \subset \mathbb{R}^n$, entonces la función es diferenciable en \mathbf{p} .

Demostración.

Haremos la demostración para dimensión $n = 2$. Los argumentos pueden generalizarse fácilmente para n arbitrario.

Tomando (h, k) suficientemente pequeños y por el teorema del valor medio para funciones de una variable,

$$\begin{aligned} f(a+h, b+k) - f(a, b) &= f(a+h, b+k) - f(a, b+k) + f(a, b+k) - f(a, b) \\ &= \frac{\partial f}{\partial x}(a^*, b+k)h + \frac{\partial f}{\partial y}(a, b^*)k \\ &= \frac{\partial f}{\partial x}(a, b)h + \frac{\partial f}{\partial y}(a, b)k + r(h, k) \end{aligned}$$

en donde

$$r(h, k) = \left(\frac{\partial f}{\partial x}(a^*, b+k) - \frac{\partial f}{\partial x}(a, b) \right) h + \left(\frac{\partial f}{\partial y}(a, b^*) - \frac{\partial f}{\partial y}(a, b) \right) k$$

en donde a^* está entre a y $a+h$, b^* está entre b y $b+k$. Por la continuidad de las derivadas parciales los factores entre paréntesis tienden a cero, en tanto que

$$\frac{h}{\sqrt{h^2 + k^2}}, \quad \frac{k}{\sqrt{h^2 + k^2}}$$

son acotados cuando $(h, k) \rightarrow (0, 0)$ luego

$$\lim_{(h,k) \rightarrow (0,0)} \frac{r(h, k)}{\sqrt{h^2 + k^2}} = 0.$$

□

La recíproca del teorema anterior es falsa, como lo muestra el siguiente ejemplo en una variable: Sea $f(x) = x^2 \sin(1/x)$, $f(0) = 0$. Su derivada es $f'(x) = 2x \sin(1/x) - \cos(1/x)$ para $x \neq 0$, y $f'(0) = 0$ (esta última debe calcularse a partir de la definición de derivada). Pero

$$\lim_{x \rightarrow 0} f'(x) \neq f'(0)$$

Así que la derivada, que es la única derivada parcial en este caso, no es continua en el origen, a pesar de que la función es diferenciable allí.

Definición 10. Si $f: \mathbb{R}^n \rightarrow \mathbb{R}$ tiene derivadas parciales continuas en $D \subset \mathbb{R}^n$ decimos que es continuamente diferenciable en D o bien, que es una función de clase $C^1(D)$.

Ejemplo 1.8-26. Analice la diferenciabilidad de la función del ejemplo 1.8.2, fuera del origen.

Las derivadas parciales

$$\frac{\partial f}{\partial x} = \frac{y}{x^2 + y^2} - \frac{2x^2 y}{(x^2 + y^2)^2}, \quad \frac{\partial f}{\partial y} = \frac{x}{x^2 + y^2} - \frac{2xy^2}{(x^2 + y^2)^2}$$

son funciones continuas en todo punto $(x, y) \neq (0, 0)$ en particular dado $(a, b) \neq (0, 0)$ ambas son continuas en una bola $B_r(a, b)$ de radio suficientemente pequeño. Por el teorema (??) la función es diferenciable en (a, b)

1.9 Reglas de derivación

Las reglas de derivación para combinaciones algebraicas de funciones diferenciables, junto con la regla de la cadena nos proporcionan la metodología más general para el cálculo explícito de derivadas.

Teorema 5. Sean $f, g: \mathbb{R}^n \rightarrow \mathbb{R}^m$, $h: \mathbb{R}^m \rightarrow \mathbb{R}$ diferenciables entonces también lo son $f \pm g$, hf , f/h ésta última siempre que $h \neq 0$. Además,

$$D(f \pm g)(\mathbf{x}) = Df(\mathbf{x}) \pm Dg(\mathbf{x}) \quad (1.10)$$

$$D(hf)(\mathbf{x}) = h(\mathbf{x})Df(\mathbf{x}) + f(\mathbf{x})Dh(\mathbf{x}) \quad (1.11)$$

$$D\left(\frac{f}{h}\right)(\mathbf{x}) = \frac{h(\mathbf{x})Df(\mathbf{x}) - f(\mathbf{x})Dh(\mathbf{x})}{h(\mathbf{x})^2} \quad (1.12)$$

Demostración.

La parte (1.10) es fácil de establecer. Probaremos aquí la fórmula para el producto y la del cociente queda como ejercicio. Tenemos que (convenimos en escribir primero los escalares)

$$\begin{aligned} \Delta(hf)(\mathbf{x}; \mathbf{h}) &= h(\mathbf{x} + \mathbf{h})f(\mathbf{x} + \mathbf{h}) - h(\mathbf{x})f(\mathbf{x}) \\ &= h(\mathbf{x} + \mathbf{h})f(\mathbf{x} + \mathbf{h}) \pm h(\mathbf{x})f(\mathbf{x} + \mathbf{h}) - h(\mathbf{x})f(\mathbf{x}) \\ &= [h(\mathbf{x} + \mathbf{h}) - h(\mathbf{x})]f(\mathbf{x} + \mathbf{h}) + g(\mathbf{x})[f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x})] \\ &= [Dh(\mathbf{x}) \cdot \mathbf{h} + r_1(\mathbf{h})]f(\mathbf{x} + \mathbf{h}) + g(\mathbf{x})[Df(\mathbf{x}) \cdot \mathbf{h} + r_2(\mathbf{h})] \\ &= [Dh(\mathbf{x}) \cdot \mathbf{h}]f(\mathbf{x} + \mathbf{h}) + g(\mathbf{x})[Df(\mathbf{x}) \cdot \mathbf{h}] + r_1(\mathbf{h})f(\mathbf{x} + \mathbf{h}) + g(\mathbf{x})r_2(\mathbf{h}) \\ &= [Dh(\mathbf{x}) \cdot \mathbf{h}]f(\mathbf{x}) + g(\mathbf{x})[Df(\mathbf{x}) \cdot \mathbf{h}] + \\ &\quad r_1(\mathbf{h})f(\mathbf{x} + \mathbf{h}) + g(\mathbf{x})r_2(\mathbf{h}) + [Dh(\mathbf{x}) \cdot \mathbf{h}][f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x})] \\ &= [Dh(\mathbf{x})f(\mathbf{x})] \cdot \mathbf{h} + [g(\mathbf{x})Df(\mathbf{x})] \cdot \mathbf{h} + \\ &\quad r_1(\mathbf{h})f(\mathbf{x} + \mathbf{h}) + g(\mathbf{x})r_2(\mathbf{h}) + [Dh(\mathbf{x}) \cdot \mathbf{h}]\Delta f(\mathbf{x}; \mathbf{h}) \end{aligned}$$

donde $r_1(\mathbf{h}), r_2(\mathbf{h}) = o(\mathbf{h})$. Observe también que $Dh(\mathbf{x})f(\mathbf{x}) + g(\mathbf{x})Df(\mathbf{x}) \in L(\mathbb{R}^n, \mathbb{R}^m)$ es una transformación lineal. Así, debemos probar que la última expresión es $o(\mathbf{h})$. Analicemos cada término:

Como f es continua en \mathbf{x} entonces $f(\mathbf{x} + \mathbf{h})$ es acotada cuando $\mathbf{h} \rightarrow \mathbf{0}$, luego $r_1(\mathbf{h})f(\mathbf{x} + \mathbf{h}) = o(\mathbf{h})$. Trivialmente $g(\mathbf{x})r_2(\mathbf{h}) = o(\mathbf{h})$. En cuanto al último término

$$\|[Dh(\mathbf{x}) \cdot \mathbf{h}]\Delta f(\mathbf{x}; \mathbf{h})\| \leq \|Dh(\mathbf{x})\| \|\mathbf{h}\| \|\Delta f(\mathbf{x}; \mathbf{h})\|,$$

ya que $\|\Delta f(\mathbf{x}; \mathbf{h})\|$ tiende a cero entonces este término es también $o(\mathbf{h})$.

1.10 Ejercicios

1. Para una función de una variable $y = f(x)$ considere el siguiente problema: De todas las rectas que pasan por $(a, f(a))$ encuentre aquella que mejor aproxime a $f(x)$ en un entorno de $(a, f(a))$ en el sentido de que si $y - f(a) = m(x - a)$ es tal recta y

$$f(a + h) - (mh + f(a)) = r(h)$$

entonces $r'(0) = 0$.

2. Sea $z = f(x, y)$ definida en $B_r(a, b)$ considere el siguiente problema: De todas los planos que pasan por $(a, f(a))$ $z - f(a, b) = m(x - a) + n(y - b)$ encuentre aquél (i.e. m y n) que mejor aproxime a $z = f(x, y)$ en un entorno de $(a, f(a))$ en el sentido de que si

$$\Delta f(a, b) \cdot (h, k) - (mh + nk) = r(h, k)$$

entonces r tiene derivada en $(0, 0)$ y $r'(0, 0) = 0$ (como transformación lineal).

3. Comprobar la diferenciabilidad de $f: \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = x^n$, con n natural. Obtener la derivada y el residuo.
4. Lo mismo para $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, $f(x, y) = \sin(x) + \cos(y)$.
5. Lo mismo para $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $f(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$ donde A es una matriz $n \times n$.
6. Lo mismo que en el ejercicio anterior con A una matriz simétrica: $A^T = A$.
7. 10. Suponga que $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$ es bilineal, es decir B es lineal respecto del primer factor: $B(\mathbf{x} + \mathbf{y}, \mathbf{z}) = B(\mathbf{x}, \mathbf{z}) + B(\mathbf{y}, \mathbf{z})$ y $B(\alpha \mathbf{x}, \mathbf{z}) = \alpha B(\mathbf{x}, \mathbf{z})$, y es lineal en el segundo factor. Encuentre $DB(\mathbf{x})(\mathbf{h}, \mathbf{k})$
Hint. La derivada $DB(\mathbf{x})$ es una función lineal $\mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n$, por lo que $DB(\mathbf{x})(\mathbf{h}, \mathbf{k})$ debe ser lineal en (\mathbf{h}, \mathbf{k}) .
8. Dar un ejemplo de una función derivable que no tenga derivadas parciales continuas, usando la función $f: \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = x^2 \sin(x)$, para $x \neq 0$, $f(0) = 0$.

9. Probar que una transformación lineal $\Lambda \in L(\mathbb{R}^n, \mathbb{R}^n)$ es continua, i.e. satisface

$$\lim_{\mathbf{h} \rightarrow \mathbf{0}} \Lambda \cdot \mathbf{h} = \mathbf{0}$$

10. Sea $\langle \cdot, \cdot \rangle$ un producto interior. Encuentre la derivada de $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $f(\mathbf{x}) = \langle \mathbf{x}, \mathbf{x} \rangle$.
11. Encuentre la derivada de $f: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$, dada por $f(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle$.
12. ¿Por qué no es derivable $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $f(x) = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$ en $\mathbf{x} = \mathbf{0}$?
13. Establecer la equivalencia de nuestra definición con la siguiente (v. gr. Spivak, "Cálculo en Variedades"):

Definición. Decimos que $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$, es diferenciable en $\mathbf{a} \in \mathbb{R}^n$ si existe $\Lambda \in L(\mathbb{R}^n, \mathbb{R}^m)$ que satisface

$$\lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) - \Lambda \cdot \mathbf{h}}{\|\mathbf{h}\|} = \mathbf{0}$$

14. Si $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$ es $f(\mathbf{x}) = \lambda \mathbf{x}$ donde λ es un escalar, encontrar $Df(\mathbf{x})$.
15. Encuentre $Df(\mathbf{x})$ para:

(a) $f: \mathbb{R}^n \setminus S^{n-1} \rightarrow \mathbb{R}^n$,

$$f(\mathbf{x}) = \frac{\mathbf{x}}{(1 - \|\mathbf{x}\|)}.$$

(b) $f: \mathbb{R}^n \rightarrow \mathbb{R}$,

$$f(\mathbf{x}) = \begin{cases} \exp(-1/\|\mathbf{x}\|^2) & \text{si } \mathbf{x} \neq \mathbf{0} \\ 0 & \text{si } \mathbf{x} = \mathbf{0} \end{cases}$$

16. Las siguientes funciones están definidas en el origen como $f(0, 0) = 0$. Determinar si en el origen las siguientes funciones (a) son continuas, (b) tienen derivadas direccionales, y (c) son diferenciables.

(a) $f(x, y) = xy/\sqrt{x^2 + y^2}$

(b) $f(x, y) = (x^{1+a}y^2)/(x^2 + y^2)$.

(c) $f(x, y) = x^2 \sin y/\sqrt{x^2 + y^2}$.

1.11 La Regla de la Cadena

El resultado más importante para combinar funciones diferenciables es el

Teorema 6 (Regla de la Cadena). : Si $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ es diferenciable en $\mathbf{x} \in \mathbb{R}^n$ $g: \mathbb{R}^m \rightarrow \mathbb{R}^p$ es diferenciable en $\mathbf{y} = f(\mathbf{x})$, entonces $g \circ f$ es diferenciable en \mathbf{x} , y

$$D(g \circ f)(\mathbf{x}) = Dg(f(\mathbf{x})) \circ Df(\mathbf{x})$$

En otras palabras el siguiente diagrama es *comutativo* en el sentido de que ambos caminos dan lo mismo

$$\begin{array}{ccc} \mathbb{R}^m & & \\ Df(\mathbf{x}) \downarrow & \searrow D(g \circ f)(\mathbf{x}) & \\ \mathbb{R}^n & \xrightarrow{Dg(f(\mathbf{x}))} & \mathbb{R}^p \end{array}$$

Demostración.

Calculemos el incremento finito de la composición

$$\Delta(g \circ f)(\mathbf{x}; \mathbf{h}) = g(f(\mathbf{x} + \mathbf{h})) - g(f(\mathbf{x})) = g(\mathbf{y} + \mathbf{v}) - g(\mathbf{y})$$

donde $\mathbf{y} = f(\mathbf{x})$ y $\mathbf{v} = f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) = \Delta f(\mathbf{x}, \mathbf{h})$. Note que \mathbf{v} depende de \mathbf{h} . Por ser g diferenciable en \mathbf{y}

$$\begin{aligned} \Delta(g \circ f)(\mathbf{x}; \mathbf{h}) &= Dg(\mathbf{y}) \cdot \mathbf{v} + r(\mathbf{v}) \\ &= Dg(\mathbf{y}) \cdot (f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x})) + r(\mathbf{v}) \\ &= Dg(\mathbf{y}) \cdot (Df(\mathbf{x}) \cdot \mathbf{h} + s(\mathbf{h})) + r(\mathbf{v}) \quad [\text{por ser } f \text{ diferenciable}] \\ &= Dg(\mathbf{y}) \cdot (Df(\mathbf{x}) \cdot \mathbf{h}) + Dg(\mathbf{y}) \cdot s(\mathbf{h}) + r(\mathbf{v}) \quad [\text{por linealidad}] \end{aligned}$$

Aquí $s(\mathbf{h})$ $r(\mathbf{v})$ son residuos que satisfacen

$$s(\mathbf{h}) = o(\mathbf{h}), \quad r(\mathbf{v}) = o(\mathbf{v})$$

Proveamos que el residuo total es

$$Dg(\mathbf{y}) \cdot s(\mathbf{h}) + r(\mathbf{v}) = o(\mathbf{h})$$

Para ello analicemos cada término. Por la desigualdad de Banach

$$\frac{\|Dg(\mathbf{y}) \cdot s(\mathbf{h})\|}{\|\mathbf{h}\|} \leq \frac{\|Dg(\mathbf{y})\| \|s(\mathbf{h})\|}{\|\mathbf{h}\|} = \|Dg(\mathbf{y})\| \left(\frac{\|s(\mathbf{h})\|}{\|\mathbf{h}\|} \right) \rightarrow \mathbf{0}$$

cuando $\mathbf{h} \rightarrow \mathbf{0}$. Como el residuo $r(\mathbf{v}) = o(\mathbf{v})$

$$\lim_{\mathbf{v} \rightarrow \mathbf{0}} \frac{\|r(\mathbf{v})\|}{\|\mathbf{v}\|} = \mathbf{0}$$

pero $\mathbf{v} = \Delta f(\mathbf{x}, \mathbf{h})$ y por la continuidad de f en \mathbf{x} , luego $\mathbf{v} \rightarrow \mathbf{0}$ cuando $\mathbf{h} \rightarrow \mathbf{0}$. Así,

$$\begin{aligned} \lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{\|r(\mathbf{v})\|}{\|\mathbf{h}\|} &= \lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{\|\mathbf{v}\|}{\|\mathbf{h}\|} \frac{\|r(\mathbf{v})\|}{\|\mathbf{v}\|} \\ &= \lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{\|Df(\mathbf{x}) \cdot \mathbf{h} + s(\mathbf{h})\|}{\|\mathbf{h}\|} \frac{\|r(\mathbf{v})\|}{\|\mathbf{v}\|} \\ &\leq \lim_{\mathbf{h} \rightarrow \mathbf{0}} \left(\|Df(\mathbf{x})\| + \frac{\|s(\mathbf{h})\|}{\|\mathbf{h}\|} \right) \frac{\|r(\mathbf{v})\|}{\|\mathbf{v}\|} \\ &= \lim_{\mathbf{h} \rightarrow \mathbf{0}} \left(\|Df(\mathbf{x})\| + \frac{\|s(\mathbf{h})\|}{\|\mathbf{h}\|} \right) \frac{\|r(\mathbf{v})\|}{\|\mathbf{v}\|} \end{aligned}$$

Como el término entre paréntesis permanece acotado, se sigue que $r(\mathbf{v}) = o(\mathbf{h})$. \square

Observación 8. Recordemos que en realidad, muchas funciones están definidas sólo en algún subconjunto de un espacio euclideo. Por ejemplo, la función $f(x, y) = \sqrt{xy}$ no está definida en todo \mathbb{R}^2 , sino únicamente en el conjunto

$$D = \{x, y \mid xy \geq 0\}$$

es decir, el plano euclideo menos el segundo y cuarto cuadrantes, incluyendo los ejes coordenados.

Si $f: D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$, Para cuestionarse sobre la diferenciabilidad f en un punto $\mathbf{x} \in D$, debe tener sentido evaluar $f(\mathbf{x} + \mathbf{h})$ para $\|\mathbf{h}\|$ suficientemente pequeño, es decir deberá existir una bola de radio r suficientemente pequeño tal que $B_r(\mathbf{x}) \subset D$ —lo que se llama un *punto interior* de D .

En la regla de la cadena, si además $g: E \subset \mathbb{R}^m \rightarrow \mathbb{R}^p$, el punto $\mathbf{y} \in f(D)$ deberá ser un punto interior de E —el dominio de g — relativo a $f(D)$ para poder evaluar $g(\mathbf{y} + \mathbf{v})$ con $\|\mathbf{v}\|$ pequeño, pero sólo para puntos de la forma $\mathbf{y} + \mathbf{v} = f(\mathbf{x} + \mathbf{h}) \in f(D)$, es decir deberá existir una bola de radio suficientemente pequeño δ tal que $B_\delta(\mathbf{y}) \subset f(D) \cap E$.

Cuando nos cuestionamos sobre la diferenciabilidad *global* es decir sobre todo un conjunto entonces los dominios de las funciones necesitan ser conjuntos abiertos, es decir todos sus puntos deben ser interiores. En este contexto, un enunciado más preciso para la regla de la cadena sería:

Teorema 7 (Regla de la cadena global). Sean $U \subset \mathbb{R}^n$, $V \subset \mathbb{R}^m$ abiertos, Si $f: U \rightarrow \mathbb{R}^m$ es diferenciable en U , $g: V \rightarrow \mathbb{R}^p$ es diferenciable en V y $f(U) \subset V$, entonces $g \circ f$ es diferenciable en U , y además

$$D(g \circ f)(\mathbf{x}) = Dg(f(\mathbf{x})) \circ Df(\mathbf{x})$$

para toda $\mathbf{x} \in U$.

Ejemplo 1.11–27. La función $f(x, y) = \cos(xy)$ es diferenciable en todo \mathbb{R}^2 , pues es la composición de las funciones diferenciables

$$\begin{array}{c} \mathbb{R}^2 \longrightarrow \mathbb{R} \longrightarrow \mathbb{R} \\ \\ (x, y) \longrightarrow \boxed{xy = z} \longrightarrow \cos(z) \end{array}$$

1.12 La versión matricial de la regla de la cadena

De acuerdo a la regla de la cadena, si f, g son diferenciables, la derivada de la composición $g \circ f$ es la composición de las derivadas $Dg(\mathbf{y}) \circ Df(\mathbf{x})$ en puntos correspondientes: $\mathbf{y} = f(\mathbf{x})$. En relación al *cálculo* sabemos que toda transformación lineal tiene una matriz asociada, una vez elegida una base y que la composición de transformaciones lineales corresponde a la multiplicación matricial. Específicamente, si en el diagrama del Teorema 7 se eligen las bases canónicas en $\mathbb{R}^n, \mathbb{R}^m, \mathbb{R}^p$, se tiene

$$[D(g \circ f)(\mathbf{x})] = [Dg(\mathbf{y})][Df(\mathbf{x})]$$

por otro lado, en las bases canónica, $[Dg(\mathbf{y})], [Df(\mathbf{x})]$ son las matrices de derivadas parciales, así,

$$\begin{bmatrix} \frac{\partial g^j}{\partial x_i} \end{bmatrix} = \begin{bmatrix} \frac{\partial g^j}{\partial y_k} \end{bmatrix} \begin{bmatrix} \frac{\partial f^k}{\partial x_i} \end{bmatrix}$$

o sea

$$\frac{\partial g^j}{\partial x_i} = \sum_{k=1}^m \frac{\partial g^j}{\partial y_k} \frac{\partial f^k}{\partial x_i}$$

para $i = 1, 2, \dots, n$ y $j = 1, 2, \dots, p$.

Ejemplo 1.12–28. Sean

$$\begin{aligned}w &= u^2 - v^2, \\z &= 2uv\end{aligned}$$

$$\begin{aligned}u &= x^3 + xy + y^3, \\v &= x - y\end{aligned}$$

Se pide calcular las derivas parciales $\frac{\partial w}{\partial x}$, $\frac{\partial w}{\partial y}$.

Por la regla de la cadena

$$\begin{aligned}\frac{\partial w}{\partial x} &= \frac{\partial w}{\partial u} \frac{\partial u}{\partial x} + \frac{\partial w}{\partial v} \frac{\partial v}{\partial x}, \\&= 2u(3x^2 + y) - 2v(1) = 2(u(3x^2 + y) - v) \\ \frac{\partial w}{\partial y} &= \frac{\partial z}{\partial u} \frac{\partial u}{\partial y} + \frac{\partial w}{\partial v} \frac{\partial v}{\partial y}, \\&= 2u(x + 3y^2) - 2v(-1) = 2(u(x + 3y^2) + v)\end{aligned}$$

De forma análoga se calculan

$$\begin{aligned}\frac{\partial z}{\partial x} &= 2(u + v(3x^2 + y)) \\ \frac{\partial z}{\partial y} &= 2(-u + v(x + 3y^2))\end{aligned}$$

Alternativamente, y este no es el camino más usual, se puede calcular la composición

$$\begin{aligned}w &= (x^3 + xy + y^3)^2 - (x - y)^2 \\z &= 2(x^3 + xy + y^3)(x - y)\end{aligned}$$

$$\begin{aligned}\frac{\partial w}{\partial x} &= 2(3x^2 + y)(x^3 + xy + y^3) - 2(x - y) \\ \frac{\partial w}{\partial y} &= 2(x - y) + 2(x + 3y^2)(x^3 + xy + y^3) \\ \frac{\partial z}{\partial x} &= 2(x - y)(3x^2 + y) + 2(x^3 + xy + y^3) \\ \frac{\partial z}{\partial y} &= 2(x - y)(x + 3y^2) - 2(x^3 + xy + y^3)\end{aligned}$$

1.12.1 Casos particulares de la regla de la cadena

Ya que este resultado es posiblemente el que más se utiliza de todo el cálculo diferencial, conviene que aprendamos a manejarlo correctamente y a aprovechar muchas de sus potencialidades. Vamos a presentar en detalle los casos particulares más relevantes.

1. El caso $n = m = p = 1$. Aquí se trata de funciones reales de una variable real.

$$\mathbb{R} \xrightarrow{f} \mathbb{R} \xrightarrow{g} \mathbb{R}$$

El enunciado usual es

$$(g \circ f)'(x) = g'(f(x))f'(x)$$

o bien en la notación clásica, si hacemos $u = f(x)$, $y = g(u)$, tenemos

$$\frac{dy}{dx} = \frac{dy}{du} \frac{du}{dx}.$$

2. El caso $n = 1, m \in \mathbb{N}, p = 1$.

$$\mathbb{R} \xrightarrow{f} \mathbb{R}^m \xrightarrow{g} \mathbb{R}$$

Aquí conviene usar la variable t en vez de x , para recordar que es una variable real. Tenemos

$$u(t) = g(f(t)) = g(f_1(t), f_1(t), \dots, f_m(t))$$

de donde

$$u'(t) = \sum_{k=1}^m D_k g(f(t)) f_k'(t) = \nabla g \cdot f'(t)$$

3. El caso $n \in \mathbb{N}, m = p = 1$. Aquí

$$\mathbb{R}^n \xrightarrow{f} \mathbb{R} \xrightarrow{g} \mathbb{R}$$

Con $u(\mathbf{x}) = g(f(\mathbf{x}))$, podemos escribir,

$$Du(\mathbf{x}) = g'(f(\mathbf{x}))Df(\mathbf{x})$$

donde $g'(f(\mathbf{x}))$ es un escalar (i.e. una matriz 1×1) o bien por la definición de gradiente

$$\begin{aligned} \nabla u(\mathbf{x}) \cdot \mathbf{h} &= Du(\mathbf{x}) \cdot \mathbf{h} = g'(f(\mathbf{x})) (Df(\mathbf{x}) \cdot \mathbf{h}) \\ &= g'(f(\mathbf{x})) (\nabla(\mathbf{x}) \cdot \mathbf{h}) = [g'(f(\mathbf{x}))\nabla(\mathbf{x})] \cdot \mathbf{h} \end{aligned}$$

y como $\mathbf{h} \in \mathbb{R}^n$ es arbitrario

$$\nabla u(\mathbf{x}) = g'(f(\mathbf{x}))\nabla f(\mathbf{x}).$$

Ejemplo 1.12-29. Sea $f = (f_1, f_2): \mathbb{R}^2 \rightarrow \mathbb{R}$, $g: \mathbb{R}^2 \rightarrow \mathbb{R}$. En este caso,

$$u(x, y, z) = g(f_1(x, y, z), f_2(x, y, z)) = g(u_1, u_2)$$

donde $u_1 = f_1(x, y)$, $u_2 = f_2(x, y)$. Tenemos

$$Du(x, y, z) = D_1g(u_1, u_2)Df_1(x, y, z) + D_2g(u_1, u_2)Df_2(x, y, z)$$

Evaluando en $\mathbf{h} \in \mathbb{R}^3$ arbitrario y usando la definición de gradiente

$$\begin{aligned} Du(x, y, z) \cdot \mathbf{h} &= D_1g(u_1, u_2)Df_1(x, y, z) \cdot \mathbf{h} + D_2g(u_1, u_2)Df_2(x, y, z) \cdot \mathbf{h} \\ \nabla u(x, y, z) \cdot \mathbf{h} &= D_1g(u_1, u_2)Df_1(x, y, z) \cdot \mathbf{h} + D_2g(u_1, u_2)Df_2(x, y, z) \cdot \mathbf{h} \end{aligned}$$

en donde $D_1g(u_1, u_2)$, $D_2g(u_1, u_2)$ son escalares, luego

$$\nabla u(x, y, z) = D_1g(u_1, u_2)\nabla f_1(x, y, z) + D_2g(u_1, u_2)\nabla f_2(x, y, z)$$

1.13 Aplicaciones de la regla de la cadena

La fórmula de Euler. Decimos que $g: \mathbb{R}^m \rightarrow \mathbb{R}$ es homogénea de grado si $g(t\mathbf{x}) = t^s g(\mathbf{x})$ para toda $t \in \mathbb{R}$, y $\mathbf{x} \in \mathbb{R}^m$. Las funciones homogéneas son muy importantes, e incluyen en particular a los polinomios homogéneos, por ejemplo

$$g(x, y, z) = xy^2 + x^3 + xz^2 - xyz$$

es una función polinomial $\mathbb{R}^3 \rightarrow \mathbb{R}$ homogénea de grado 3. Otro ejemplo, es

$$g(x, y) = \exp(y/x)$$

que es una función $\mathbb{R}^2 \rightarrow \mathbb{R}$ homogénea de grado 0.

Vamos a demostrar la llamada fórmula de Euler para funciones homogéneas

$$\nabla g(\mathbf{x}) \cdot \mathbf{x} = sg(\mathbf{x})$$

o en notación clásica poniendo $u = g(x_1, x_2, \dots, x_m)$

$$\sum_{i=1}^m x_i \frac{\partial u}{\partial x_i} = su.$$

En efecto, sea $\mathbf{x} \in \mathbb{R}^n$ fijo y sea $u(t) = g(t\mathbf{x})$. Por la regla de la cadena tenemos

$$u'(1) = \nabla g(\mathbf{x}) \cdot \mathbf{x},$$

por otro lado $u(t) = t^s g(\mathbf{x})$, de donde $u'(1) = s g(\mathbf{x})$. Comparando ambos resultados se prueba la fórmula.

El gradiente y el hiperplano tangente Consideremos una función real de varias variables reales, vgr. $g: \mathbb{R}^m \rightarrow \mathbb{R}$. De la definición de derivada y gradiente, tenemos que si g es diferenciable en $\mathbf{p} \in \mathbb{R}^m$ entonces para $\mathbf{h} \in \mathbb{R}^m$ arbitrario

$$g(\mathbf{p} + \mathbf{h}) - g(\mathbf{p}) = Dg(\mathbf{p}) \cdot \mathbf{h} + r(\mathbf{h}) = \nabla g(\mathbf{p}) \cdot \mathbf{h} + r(\mathbf{h})$$

donde $r(\mathbf{h}) = o(\mathbf{h})$. Haciendo $z = g(\mathbf{p} + \mathbf{h})$, $z_0 = g(\mathbf{p})$ y tomando $\mathbf{h} = \mathbf{x} - \mathbf{p}$, para \mathbf{x} próximo de \mathbf{p} se puede despreciar el término $o(\mathbf{h})$ dando la aproximación

$$z - z_0 = \nabla g(\mathbf{x})(\mathbf{x} - \mathbf{p})$$

la cual es la ecuación de un hiperplano en \mathbb{R}^m cuyo vector normal es

$$\mathbf{n} = (\nabla g(\mathbf{x}), -1).$$

Dicho hiperplano es tangente a la superficie $M \subset \mathbb{R}^m$ definida por la gráfica de la función

$$\text{graf}(g) = \{(\mathbf{x}, z) \mid z = g(\mathbf{x})\}.$$

como habremos de probar en un momento. Lo que se quiere resaltar aquí que esta es la idea intuitiva que hemos estado persiguiendo desde un principio para el concepto de función diferenciable: la existencia de un plano tangente a la gráfica.

Verifiquemos entonces la afirmación anterior. Para ello consideremos una curva diferenciable arbitraria contenida en $\text{graf}(g)$, que pase por el punto \mathbf{p} , digamos $\gamma: \mathbb{R} \rightarrow \mathbb{R}^m$ con $\gamma(0) = \mathbf{p}$. Debemos entonces verificar que $\gamma'(0)$,

el vector tangente a la curva, pertenece el plano que tiene por normal $\nabla g(\mathbf{p})$. Pero escribiendo $\gamma(t) = (\mathbf{x}(t), z(t))$ entonces

$$z(t) = g(\mathbf{x}(t))$$

por estar la curva sobre $\text{graf}(g)$, luego

$$z'(0) = \nabla g(\mathbf{p}) \cdot \mathbf{x}'(0) \quad \text{o bien} \quad (\mathbf{x}'(0), z'(0)) \cdot (\nabla g(\mathbf{p}), -1) = 0$$

que era lo que se quería probar.

Las ideas anteriores se pueden generalizar como sigue (véase también la sección § 1.3). Consideremos la hipersuperficie de nivel

$$M_c = F^{-1}(c)$$

que pasa por un punto $\mathbf{y} \in \mathbb{R}^{m+1}$, i.e. $F(\mathbf{y}) = c$, donde $F: \mathbb{R}^{m+1} \rightarrow \mathbb{R}$. En este caso $\nabla F(\mathbf{y})$ es *normal a la superficie* M_c . En efecto, si $\nu: \mathbb{R} \rightarrow \mathbb{R}^{m+1}$ es una curva en M_c por definición $F(\nu(t)) = c$ para toda t de donde, por la regla de la cadena,

$$0 = \nabla F(\mathbf{y}) \cdot \nu'(0)$$

que muestra que la superficie M_c es perpendicular al vector gradiente. Veamos que ésta es una generalización del caso anterior. En efecto, si $g: \mathbb{R}^m \rightarrow \mathbb{R}$ es diferenciable en \mathbf{p} y $z_0 = g(\mathbf{p})$, podemos considerar a $F: \mathbb{R}^m \times \mathbb{R} = \mathbb{R}^{m+1} \rightarrow \mathbb{R}$ definida como $F(\mathbf{x}, z) = g(\mathbf{x}) - z$. Entonces F es diferenciable en (\mathbf{p}, z_0) y el gradiente en (\mathbf{p}, z_0) es

$$\nabla F(\mathbf{p}, z_0) = (\nabla g(\mathbf{p}), -1).$$

Así, si $\gamma: \mathbb{R} \rightarrow \text{graf}(g) \subset \mathbb{R}^m$ entonces $z(t) = g(\gamma(t))$ satisface que $\nu(t) = (\gamma(t), z(t))$ es una curva $\nu: \mathbb{R} \rightarrow \mathbb{R}^{m+1}$ contenida en M_0 cuyo vector tangente en $t = 0$ es

$$(\gamma'(0), -1).$$

1.14 Ejercicios

1. Enuncie con precisión y muestre que se sigue de la regla de la cadena, las siguientes afirmaciones típicas en libros de cálculo:

- (A) “Sustituyendo las coordenadas cartesianas x, y por las coordenadas polares $x = r \cos \theta$, $y = r \sin \theta$ en la función $w = f(x, y)$ se obtiene

$$\frac{\partial w}{\partial r} = \frac{\partial f}{\partial x} \cos \theta + \frac{\partial f}{\partial y} \sin \theta, \quad \frac{1}{r} \frac{\partial w}{\partial \theta} = -\frac{\partial f}{\partial x} \sin \theta + \frac{\partial f}{\partial y} \cos \theta.”.$$

- (B) “Sean $w = w(u, v)$, $u = u(x, y, z)$, $v = v(x, y, z)$ funciones diferenciables. Considerando a w como función de x, y, z , su gradiente en cada punto (x_0, y_0, z_0) está en el mismo plano que los gradientes de u y v de en ese punto”.
2. ¿Por qué está equivocado el siguiente argumento? Si a $w = f(x, y, z)$ y $z = g(x, y)$ le aplicamos la Regla de la Cadena,

$$\frac{\partial w}{\partial x} = \frac{\partial w}{\partial x} \frac{\partial x}{\partial x} + \frac{\partial w}{\partial y} \frac{\partial y}{\partial x} + \frac{\partial w}{\partial z} \frac{\partial z}{\partial x} = \frac{\partial w}{\partial x} + \frac{\partial w}{\partial z} \frac{\partial z}{\partial x},$$

por lo tanto

$$\frac{\partial w}{\partial z} \frac{\partial z}{\partial x} = 0.$$

de modo que alguna de las derivadas parciales

$$\frac{\partial w}{\partial z}, \quad \frac{\partial z}{\partial x}$$

es cero.

3. Use la regla de la cadena para mostrar que bajo las bajo hipótesis adecuadas (enunciarlas!):

$$\frac{d}{dx} \int_0^x f(x, y) dy = f(x, x) + \int_0^x \frac{\partial f}{\partial x}(x, y) dy$$

4. (El teorema de transporte en dimensión 1) Haga las hipótesis necesarias para calcular la derivada

$$\frac{d}{dt} \int_{a(t)}^{b(t)} f(t, y) dy$$

Se denomina teorema de transporte pues el dominio $D_t = [a(t), b(t)]$ “se mueve” con el tiempo t y $f(t, y)$ es una cantidad que depende del tiempo y de la variable espacial $y \in D_t$.

5. (La diferencial parcial) Sea $F: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^p$ y $(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^n \times \mathbb{R}^m$. Fijando \mathbf{b} podemos considerar la diferencial de la función $F_{\mathbf{b}}: \mathbb{R}^n \rightarrow \mathbb{R}^p$ en el punto \mathbf{a} :

$$DF_{\mathbf{b}}(\mathbf{a}) \in L(\mathbb{R}^n, \mathbb{R}^p)$$

que si existe, se llama *la diferencial de F respecto de la primera variable* y se denota por

$$D_1F(\mathbf{a}, \mathbf{b}) = DF_{\mathbf{b}}(\mathbf{a}) \in L(\mathbb{R}^n, \mathbb{R}^p)$$

Análogamente *la diferencial de F respecto de la segunda variable*, si existe, se denota por

$$D_2F(\mathbf{a}, \mathbf{b}) = DF_{\mathbf{a}}(\mathbf{b}) \in L(\mathbb{R}^m, \mathbb{R}^p).$$

Muestre que si la función $F: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^p$ es diferenciable en (\mathbf{a}, \mathbf{b}) entonces existen las diferenciales parciales $D_1F(\mathbf{a}, \mathbf{b})$, $D_2F(\mathbf{a}, \mathbf{b})$ y si $(\mathbf{h}_1, \mathbf{h}_2) \in \mathbb{R}^n \times \mathbb{R}^m$ entonces

$$DF(\mathbf{a}, \mathbf{b}) \cdot (\mathbf{h}_1, \mathbf{h}_2) = D_1F(\mathbf{a}, \mathbf{b}) \cdot \mathbf{h}_1 + D_2F(\mathbf{a}, \mathbf{b}) \cdot \mathbf{h}_2. \quad (1.13)$$

Muestre el recíproco: Si

$$D_1F(\mathbf{x}, \mathbf{y}) \in L(\mathbb{R}^n, \mathbb{R}^p), \quad D_2F(\mathbf{x}, \mathbf{y}) \in L(\mathbb{R}^m, \mathbb{R}^p)$$

son continuas como funciones de $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^n \times \mathbb{R}^m$, para (\mathbf{x}, \mathbf{y}) en una bola con centro en (\mathbf{a}, \mathbf{b}) y radio r , $B_r(\mathbf{a}, \mathbf{b}) \subset \mathbb{R}^n \times \mathbb{R}^m$, entonces F es diferenciable en (\mathbf{a}, \mathbf{b}) y es válida (1.13).

*Note que los espacios lineales $L(\mathbb{R}^n, \mathbb{R}^p)$, $L(\mathbb{R}^m, \mathbb{R}^p)$ son en sí mismos espacios vectoriales normados, así que tiene sentido decir que

$$\begin{aligned} D_1F: B_r(\mathbf{a}, \mathbf{b}) \subset \mathbb{R}^n \times \mathbb{R}^m &\rightarrow L(\mathbb{R}^n, \mathbb{R}^p), \\ D_2F: B_r(\mathbf{a}, \mathbf{b}) \subset \mathbb{R}^n \times \mathbb{R}^m &\rightarrow L(\mathbb{R}^m, \mathbb{R}^p) \end{aligned}$$

sean funciones continuas.

6. Considere la ecuación diferencial

$$\frac{d\mathbf{x}}{dt} = f(\mathbf{x}) \quad (1.14)$$

donde $f: D \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ es de clase $C^1(D)$ — es decir f tiene derivadas continuas en D —. Sea $\mathbf{x} = \varphi(\mathbf{u}, t)$ la solución de (1.14) que tiene por condición inicial $\mathbf{x}(0) = \mathbf{u}$. Muestre que la función $D_1\varphi(\mathbf{x}, t) \in L(\mathbb{R}^n, \mathbb{R}^n)$ definida para (\mathbf{x}, t) donde esté definida $\varphi(\mathbf{x}, t)$, satisface la ecuación diferencial “matricial”

$$\frac{d}{dt}D_1\varphi(\mathbf{x}, t) = Df(\varphi(\mathbf{x}, t))D_1\varphi(\mathbf{x}, t)$$

llamada también *ecuación variacional a lo largo de la solución* $\varphi(\mathbf{x}, t)$.

7. Este ejercicio muestra que la regla de la cadena no necesariamente se cumple si alguna de las dos funciones en la composición no es diferenciable. Considere

$$f(x, y) = \begin{cases} \frac{xy^2}{x^2+y^2} & \text{si } (x, y) \neq (0, 0), \\ 0 & \text{si } (x, y) = (0, 0) \end{cases}$$

Pruebe que: (a) Existen las dos derivadas parciales de f en el origen. (b) Si $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ está dada por $g(t) = (at, bt)$ donde $a, b \in \mathbb{R}$ son constantes no ambas nulas, entonces $f \circ g$ es diferenciable y

$$D(f \circ g)(0) = \frac{ab^2}{a^2 + b^2},$$

pero

$$Df(0, 0) \circ Dg(0) = 0.$$

8. La ecuación en derivadas parciales de Black-Scholes

$$\frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 s^2 \frac{\partial^2 C}{\partial s^2} + rs \frac{\partial C}{\partial s} - rC = 0, \quad s > 0, \quad 0 < t < T$$

y la condición de frontera $C(s, T) = \max\{s - K, 0\}$ gobiernan el comportamiento de los valores $C(s, t)$ de una opción de compra (*call europeo*). Muestre que esta ecuación puede reducirse a la ecuación de calor mediante dos cambios de variables sucesivos.

- (a) Sea $s = K \exp(x)$, $t = T - 2\tau/\sigma^2$ y $C(s, t) = Kv(x, \tau)$. La ecuación queda

$$\frac{\partial v}{\partial \tau} = \frac{\partial^2 v}{\partial x^2} + (\lambda - 1) \frac{\partial v}{\partial x} - \lambda v, \quad -\infty < x < \infty, \quad \tau > 0.$$

y $v(x, 0) = \max\{\exp(x) - 1, 0\}$, con $\lambda = 2\tau/\sigma^2$.

(b) Hacemos

$$v(x, \tau) = \exp\left(-\frac{1}{2}(\lambda - 1)x - \frac{1}{4}(\lambda + 1)^2\tau\right) w(x, \tau)$$

y la ecuación anterior se transforma en

$$\frac{\partial w}{\partial \tau} = \frac{\partial^2 w}{\partial x^2}, \quad -\infty < x < \infty, \quad \tau > 0.$$

con

$$w(x, 0) = \max\left\{\exp\left(\frac{1}{2}(\lambda + 1)x\right) - \exp\left(\frac{1}{2}(\lambda - 1)x\right), 0\right\}$$

1.15 El teorema del valor medio

Recordemos primero el caso de una variable.

Teorema 8 (TVM en una variable). *Si $g: [a, b] \rightarrow \mathbb{R}$ es continua y diferenciable en (a, b) , entonces existe $c \in (a, b)$ tal que*

$$g(b) - g(a) = f'(c)(b - a).$$

Mediante la Regla de la Cadena y el TVM en una variable podemos probar la versión más general

Teorema 9 (TVM para funciones $\mathbb{R}^n \rightarrow \mathbb{R}$). *Sea $f: D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ diferenciable en un disco $D \subset \mathbb{R}^n$, entonces dados $\mathbf{p}, \mathbf{q} \in \mathbb{R}^n$ existe \mathbf{c} perteneciente al segmento que une \mathbf{p} con \mathbf{q} tal que*

$$f(\mathbf{p}) - f(\mathbf{q}) = Df(\mathbf{c}) \cdot (\mathbf{p} - \mathbf{q})$$

Demostración.

Ya que D es un disco, entonces si $\mathbf{p}, \mathbf{q} \in \mathbb{R}^n$ entonces el segmento $(1-t)\mathbf{p} + t\mathbf{q}$, $0 \leq t \leq 1$, pertenece también al disco D . Considérese entonces $g: [0, 1] \rightarrow \mathbb{R}$ definida por

$$g(t) = f((1-t)\mathbf{p} + t\mathbf{q}),$$

que es diferenciable por ser composición de funciones diferenciables. Además por el TVM en una variable real, existe $c \in (0, 1)$ tal que

$$g(1) - g(0) = g'(c).$$

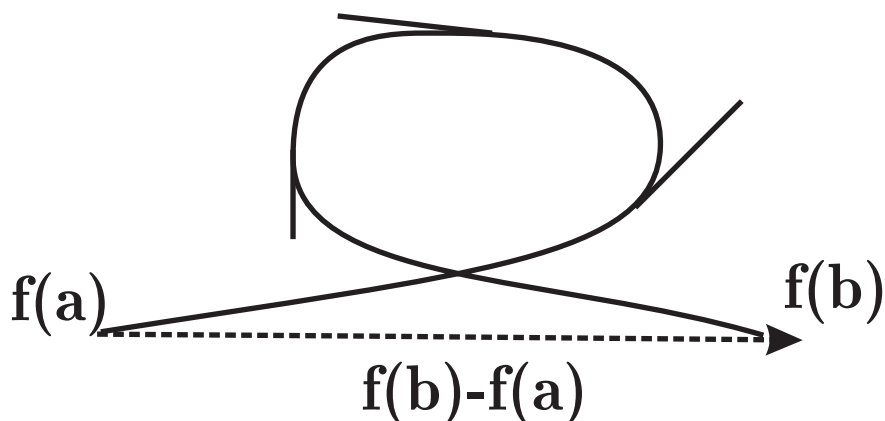


Figure 1.14: Contraejemplo al TVM para funciones a valores vectoriales.

Calculando la última derivada por la regla de la cadena nos queda

$$g'(c) = Df((1-c)\mathbf{p} + c\mathbf{q}) \cdot (\mathbf{p} - \mathbf{q}).$$

Como $g(1) = f(\mathbf{p})$, $g(0) = f(\mathbf{q})$, tomando $\mathbf{c} = (1-c)\mathbf{p} + c\mathbf{q}$ se obtiene el resultado. □

Observe que por tratarse de una función a valores reales, se puede escribir

$$f(\mathbf{p}) - f(\mathbf{q}) = \nabla f(\mathbf{c}) \cdot (\mathbf{p} - \mathbf{q})$$

Observación 9. Desafortunadamente, el TVM que establecimos para funciones a valores reales no puede extenderse a funciones a valores vectoriales. Considere como ejemplo, la curva mostrada en la figura 1.14: ningún punto de la curva el vector tangente es horizontal.

En el caso general de funciones a valores vectoriales, ya no se tiene una igualdad del valor medio como en el caso de funciones a valores reales. Sin embargo, existen algunas variantes que mencionaremos a continuación.

Teorema 10. Sea $f: D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciable en un disco $D \subset \mathbb{R}^n$, entonces dados $\mathbf{p}, \mathbf{q} \in \mathbb{R}^n$

$$f(\mathbf{p}) - f(\mathbf{q}) = \left(\int_0^1 Df((1-t)\mathbf{p} + t\mathbf{q}) dt \right) \cdot (\mathbf{p} - \mathbf{q})$$

Este resultado se puede interpretar como un TVM “promedio”

Demostración.

Consideremos de manera análoga al caso de funciones a valores reales $g: [0, 1] \rightarrow \mathbb{R}$ definida por

$$g(t) = f((1-t)\mathbf{p} + t\mathbf{q}),$$

por el Teorema Fundamental del Cálculo,

$$g(1) - g(0) = f(\mathbf{p}) - f(\mathbf{q}) = \int_0^1 g'(t) dt = \int_0^1 Df((1-t)\mathbf{p} + t\mathbf{q}) \cdot (\mathbf{p} - \mathbf{q}) dt$$

de donde se sigue el resultado. □

Corolario 1. Sea $f: D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciable en un disco $D \subset \mathbb{R}^n$. Dados $\mathbf{p}, \mathbf{q} \in \mathbb{R}^n$, si

$$\|Df((1-t)\mathbf{p} + t\mathbf{q})\| \leq M, \quad 0 \leq t \leq 1$$

para cierta constante $M \geq 0$, entonces

$$\|f(\mathbf{p}) - f(\mathbf{q})\| \leq M\|\mathbf{p} - \mathbf{q}\|$$

Demostración.

Basta aplicar la desigualdad de Banach

$$\|Df((1-t)\mathbf{p} + t\mathbf{q}) \cdot (\mathbf{p} - \mathbf{q})\| \leq \|Df((1-t)\mathbf{p} + t\mathbf{q})\| \|\mathbf{p} - \mathbf{q}\|$$

□

Proposición 10. Sea $f: D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciable en un disco $D \subset \mathbb{R}^n$, $\mathbf{k} \in \mathbb{R}^m$. Dados $\mathbf{p}, \mathbf{q} \in \mathbb{R}^n$, entonces existe \mathbf{c} perteneciente al segmento que une \mathbf{p} con \mathbf{q} tal que

$$(f(\mathbf{p}) - f(\mathbf{q})) \cdot \mathbf{k} = (Df(\mathbf{c}) \cdot (\mathbf{p} - \mathbf{q})) \cdot \mathbf{k}$$

Demostración.

La función a valores reales $f_{\mathbf{k}}(\mathbf{x}) = f(\mathbf{x}) \cdot \mathbf{k}$, donde (\cdot) denota el producto interior usual, cumple con las hipótesis del Teorema 9, luego

$$f_{\mathbf{k}}(\mathbf{p}) - f_{\mathbf{k}}(\mathbf{q}) = Df_{\mathbf{k}}(\mathbf{c}) \cdot (\mathbf{p} - \mathbf{q})$$

pero (verifique)

$$Df_{\mathbf{k}}(\mathbf{c}) \cdot \mathbf{h} = (Df(\mathbf{c}) \cdot \mathbf{h}) \cdot \mathbf{k}$$

–observe el distinto significado de (\cdot) –. De aquí que se sigue el resultado □

Corolario 2. Sea $f: D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciable en un disco $D \subset \mathbb{R}^n$. Dados $\mathbf{p}, \mathbf{q} \in \mathbb{R}^n$, entonces existe una transformación lineal $\Lambda \in L(\mathbb{R}^n, \mathbb{R}^m)$ tal que

$$f(\mathbf{p}) - f(\mathbf{q}) = L \cdot (\mathbf{p} - \mathbf{q})$$

Demostración.

Consideremos la base canónica \mathbf{e}_j , $j = 1, 2, \dots, m$ de \mathbb{R}^m y apliquemos la proposición anterior tomando sucesivamente $\mathbf{k} = \mathbf{e}_j$. Observe que $f_{\mathbf{e}_j} = f_j$, la j -ésima función componente, luego existe \mathbf{c}_j perteneciente al segmento que une \mathbf{p} con \mathbf{q} tal que

$$f_j(\mathbf{p}) - f_j(\mathbf{q}) = Df_j(\mathbf{c}_j) \cdot (\mathbf{p} - \mathbf{q})$$

La transformación lineal $\Lambda = (\Lambda_1, \Lambda_2, \dots, \Lambda_m) \in L(\mathbb{R}^n, \mathbb{R}^m)$ se puede tomar como $\Lambda_j = Df_j(\mathbf{c}_j) \in L(\mathbb{R}^n, \mathbb{R})$.

Observación 10. En el último corolario, observe que la transformación lineal Λ no concide necesariamente con $Df(\mathbf{c})$, para algún \mathbf{c} entre \mathbf{p} y \mathbf{q} , ya que los \mathbf{c}_i no son en general iguales.

1.16 El Teorema de la función inversa

Los teoremas de la función inversa e implícita constituyen uno de los pilares del cálculo diferencial. Ambos tienen múltiples aplicaciones y conexiones con otras ramas de la matemática. En esta sección enunciaremos el teorema de la función inversa en una y varias variables. Comenzaremos por motivar la versión en una variable y gradualmente con ejemplos en varias –pocas– variables.

Consideremos una función de una variable $y = f(x)$. Nos preguntamos bajo qué condiciones podemos “despejar a y en función de x ”. En la figura 1.15 se muestran dos ejemplos. En ambos casos la función es invertible ya que $y = f(x)$ es creciente, por lo cual la inversa $x = g(y)$ está bien definida, lo cual es el significado preciso de “despejar³ a y ”; sin embargo

³El término “despejar” sigue siendo ambiguo, por ejemplo, ¿se puede despejar a x de la ecuación $y = xe^x$? La respuesta es no en términos de funciones elementales, como polinomios, cocientes de polinomios (funciones racionales), raíces (aún de exponentes fraccionarios) de funciones racionales, logaritmos o exponenciales. Puede uno simplemente darle un nombre a la función inversa, digamos $x = \text{LogExp}(y)$ y mostrar sus propiedades.

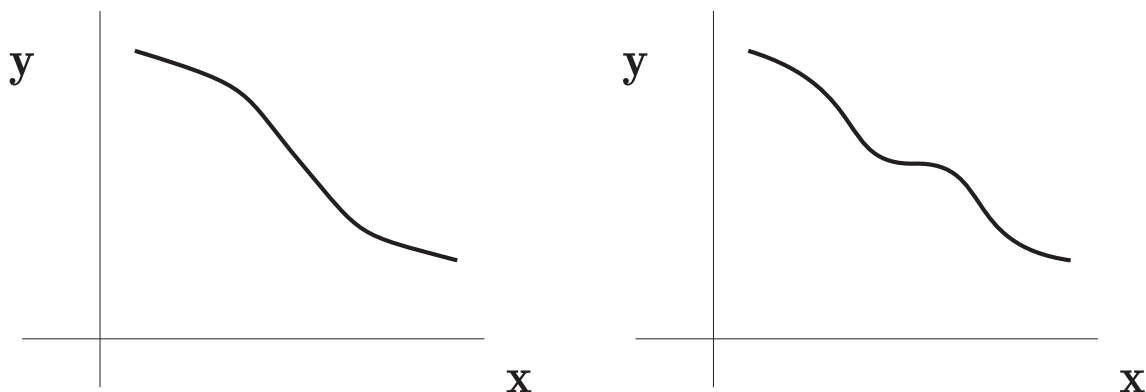


Figure 1.15: La función inversa

en el segundo caso la inversa no resulta una función diferenciable ya que la derivada de la función original se anula en un punto.

Teorema 11 (TFINV en $\mathbb{R} \rightarrow \mathbb{R}$). Sea $f: \mathbb{R} \rightarrow \mathbb{R}$ de clase $C^1(\mathbb{R})$. Si $f'(x_0) \neq 0$ y $y_0 = f(x_0)$, entonces existen vecindades $I_{x_0} = (x_0 - \delta, x_0 + \delta)$ de x_0 , $I_{y_0} = (y_0 - \delta', y_0 + \delta')$ de y_0 , y una función $g: I_{y_0} \rightarrow I_{x_0}$ tal que $x = h(f(x))$ para toda $x \in I_{x_0}$, y $y = f(h(y))$ para toda $y \in I_{y_0}$. Además, la función g es diferenciable en y_0 y

$$h'(y_0) = \frac{1}{f'(x_0)}.$$

Ejemplo 1.16-30. Sea $y = ax$, entonces si $a \neq 0$, entonces se puede despejar $x = a^{-1}y$. Observe que $a = dy/dx$.

Antes de proceder al siguiente ejemplo, mencionemos el siguiente resultado útil.

Lema 1. Sea $f: \mathbb{R} \rightarrow \mathbb{R}$ de clase C^1 en una vecindad de 0 y $f(0) = 0$. Entonces $f(x) = xh(x)$ con h continua en 0

Finalmente queremos saber el comportamiento de la función, si es monótona, si es continua, sus derivadas, el rango de valores, etc. En este sentido, en el enunciado del teorema de la función inversa, la existencia y la diferenciable de la función inversa es lo realmente importante, y no como se hace énfasis en los libros elementales, que la derivada de la inversa es $\frac{dx}{dy} = \frac{1}{\frac{dy}{dx}}$.

Demostración.

Por el teorema del valor medio

$$f(x) = xf'(x^*)$$

para algún x^* entre 0 y x , luego $x^* \rightarrow 0$ si $x \rightarrow 0$. Sea $h(x) = f'(x^*)$, como f' es continua entonces $\lim_{x \rightarrow 0} h(x) = f'(0) = h(0)$, así que h es continua en 0.

Ejemplo 1.16-31. La siguiente es una generalización *no lineal* del ejemplo anterior: Sea

$$y = ax + r(x),$$

donde $r(x) = xm(x)$, m es de clase C^1 y $m(0) = 0$. Entonces

$$x = a^{-1}y + k(y),$$

donde $k(y) = yn(y)$, n es de clase C^1 y $n(0) = 0$.

Primeramente (justifique!) $r(0) = r'(0) = 0$, es decir, r comienza con términos cuadráticos.

Sea $f(x) = ax + r(x)$ entonces f es de clase C^1 en una vecindad de 0. Si $a \neq 0$, entonces existe una función $x = g(y)$ definida para $|y| < \delta'$ con rango $|x| < \delta$, tal que

$$y = f(g(y)) = ag(y) + r(g(y))$$

o bien, si $x = g(y)$

$$x = a^{-1}y + k(y)$$

donde

$$k(y) = a^{-1}r(g(y))$$

Como $g(0) = 0$, entonces $k(0) = r(g(0)) = 0$ luego $k(y) = yn(y)$ para cierta $n(y)$ continuamente diferenciable. Además

$$n(0) = k'(0) = r'(g(0))g'(0) = r'(0)g'(0) = 0.$$

Ejemplo 1.16-32. El siguiente ejemplo generaliza el anterior a más dimensiones. Sean $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ y considere la ecuación

$$\mathbf{y} = A\mathbf{x} + r(\mathbf{x}),$$

donde $A \in \mathbb{R}^{n \times n}$ es no singular; $r(\mathbf{x}) = \mathbf{x}m(\mathbf{x})$, m es de clase C^1 y $m(\mathbf{0}) = \mathbf{0}$. Entonces

$$\mathbf{x} = A^{-1}\mathbf{y} + k(\mathbf{y}),$$

donde $k(\mathbf{y}) = \mathbf{y}n(\mathbf{y})$, n es de clase C^1 y $n(\mathbf{0}) = \mathbf{0}$.

El lector puede verificar que el lema 1, se generaliza a \mathbb{R}^n , es decir, si $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$ es de clase C^1 en una vecindad de $\mathbf{0}$ y $f(\mathbf{0}) = \mathbf{0}$; entonces $f(\mathbf{x}) = \mathbf{x}h(x)$ con $h: \mathbb{R}^n \rightarrow \mathbb{R}$ es continua en $\mathbf{0}$.

En cierto sentido, el último ejemplo muestra la esencia del TFIN: Si la parte lineal es no singular en un punto, entonces la función es localmente invertible en un entorno del punto.

En la dirección de dar una demostración del TFINV y del TFIM para funciones de varias variables, probaremos primeramente el TFin y basado en ello daremos una prueba del TFIm. Primeramente veremos una propiedad en norma, de una función *lineal* invertible, también llamada *coercitividad*

Proposición 11. *Sea $\Lambda \in L(\mathbb{R}^n, \mathbb{R}^n)$. Si Λ es invertible entonces existe $k > 0$ tal que para toda $\mathbf{x} \in \mathbb{R}^n$,*

$$k\|\mathbf{x}\| \leq \|\Lambda \cdot \mathbf{x}\|$$

Demostración.

Defina

$$k = \inf\{\|\Lambda \cdot \mathbf{x}\| \mid \|\mathbf{x}\| = 1\}$$

Se afirma que $k > 0$, si Λ es invertible. En caso contrario, podríamos encontrar una sucesión $\{\mathbf{x}_n\}$ en \mathbb{R}^n de vectores unitarios, tales que $\|\Lambda \cdot \mathbf{x}_n\| < 1/n$. Pasando a una subsucesión si fuese necesario, podemos suponer⁴ que \mathbf{x}_n converge a algún punto \mathbf{x}^* de norma unitaria. Por continuidad

$$\Lambda \cdot \mathbf{x}^* = \lim_{n \rightarrow \infty} \Lambda \cdot \mathbf{x}_n = \mathbf{0}$$

así hemos encontrado $\mathbf{x}^* \neq \mathbf{0}$ tal que $\Lambda \cdot \mathbf{x}^* = \mathbf{0}$, una contradicción.

Ahora, para $\mathbf{x} \in \mathbb{R}^n$ arbitrario, naturalmente

$$k \leq \left\| \Lambda \cdot \left(\frac{\mathbf{x}}{\|\mathbf{x}\|} \right) \right\|$$

⁴Aquí usamos el siguiente resultado del Análisis que probaremos más adelante: Toda sucesión en un conjunto compacto, tiene una subsucesión convergente. En este caso el compacto es la esfera unitaria en \mathbb{R}^n .

de donde por linealidad

$$k\|\mathbf{x}\| \leq \|\Lambda \cdot \mathbf{x}\|.$$

□

Recuerde que hemos definido una función de clase $C^1(D)$ si sus derivadas parciales existen y son funciones continuas en D . La siguiente es una forma alternativa de ver a las funciones de clase C^1 , independientes de coordenadas.

Definición 11. Una función $f: D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$, definida en un disco abierto D , se dice de clase C^1 en D , si la diferencial $Df: D \rightarrow L(\mathbb{R}^n, \mathbb{R}^m)$ es continua.

Observación 11. Se puede dotar al espacio $L(\mathbb{R}^n, \mathbb{R}^m)$ de diversas normas, de modo que tiene sentido hablar de la continuidad de la diferencial vista como función del punto. Ante la elección de bases dicho espacio no es sino \mathbb{R}^{mn} donde todas las normas son equivalentes.

Observación 12. Sea \mathbf{e}_i , $i = 1, 2, \dots, n$ una base de \mathbb{R}^n . Notemos que la función $ev_{\mathbf{e}_i}: L(\mathbb{R}^n, \mathbb{R}^m) \rightarrow \mathbb{R}^m$, dada por $L \rightarrow L \cdot \mathbf{e}_i$ es continua, ya que evidentemente es lineal y por la desigualdad de Banach,

$$\|ev_{\mathbf{e}_i} \cdot L\| = \|L \cdot \mathbf{e}_i\| \leq \|L\| \|\mathbf{e}_i\|$$

se sigue la continuidad. La Derivada parcial se obtiene como la composición $D_i f = ev_{\mathbf{e}_i} \circ Df$, es decir, $D_i f(\mathbf{p}) = Df(\mathbf{p}) \cdot \mathbf{e}_i$, así recuperamos la definición anterior de función de clase C^1 . El recíproco es también cierto ya que $Df(\mathbf{p}) \cdot \mathbf{h} = \sum_{i=1}^n D_i f(\mathbf{p}) h_i$, es decir

$$Df = \sum_{i=1}^n D_i f \pi_i$$

donde $\pi: \mathbb{R}^n \rightarrow \mathbb{R}$ son las proyecciones: $\pi(\mathbf{h}) = h_i$, para $\mathbf{h} \in \mathbb{R}^n$, y $D_i f \pi_i$ es la multiplicación de funciones a valores reales.

El siguiente resultado muestra cómo la diferencial aproxima linealmente a la función en toda una vecindad donde la función se a de clase C^1 .

Lema 2. Si $f: D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ es de clase C^1 en el disco $B_r(\mathbf{p}) \subset D$, entonces dado $\epsilon > 0$ existe $\delta > 0$ tal que

$$\|f(\mathbf{x}_1) - f(\mathbf{x}_2) - Df(\mathbf{p}) \cdot (\mathbf{x}_1 - \mathbf{x}_2)\| \leq \epsilon \|\mathbf{x}_1 - \mathbf{x}_2\|$$

Demostración.

Como f es de clase C^1 , escoja $\delta > 0$ tal que

$$\|Df(\mathbf{x}) - Df(\mathbf{p})\|_{L(\mathbb{R}^n, \mathbb{R}^m)} \leq \epsilon/m$$

De acuerdo al corolario 2 existe una transformación lineal $\Lambda \in L(\mathbb{R}^n, \mathbb{R}^m)$ tal que

$$f(\mathbf{x}_1) - f(\mathbf{x}_2) = \Lambda \cdot (\mathbf{x}_1 - \mathbf{x}_2)$$

luego

$$\begin{aligned} & \|f(\mathbf{x}_1) - f(\mathbf{x}_2) - Df(\mathbf{p}) \cdot (\mathbf{x}_1 - \mathbf{x}_2)\| \\ &= \|(\Lambda - Df(\mathbf{p})) \cdot (\mathbf{x}_1 - \mathbf{x}_2)\| \leq \|\Lambda - Df(\mathbf{p})\| \|\mathbf{x}_1 - \mathbf{x}_2\| \end{aligned}$$

en la demostración del corolario 2, la transformación lineal Λ se contruye como

$$\Lambda = (Df_2(\mathbf{c}_1), Df_2(\mathbf{c}_2), \dots, Df_m(\mathbf{c}_m))$$

donde \mathbf{c}_j son puntos contenidos en el segmento que une \mathbf{x}_1 con \mathbf{x}_2 , en particular están contenidos en $B_r(\mathbf{p})$ donde f es de clase C^1 , luego considerando la norma 1 en $L(\mathbb{R}^n, \mathbb{R}^m)$,

$$\begin{aligned} \|\Lambda - Df(\mathbf{p})\|_{L(\mathbb{R}^n, \mathbb{R}^m)} &= \sum_{j=1}^m \|Df_j(\mathbf{c}_j) - Df_j(\mathbf{p})\|_{L(\mathbb{R}^n, \mathbb{R})} \\ &\leq \sum_{j=1}^m \|Df(\mathbf{c}_j) - Df(\mathbf{p})\|_{L(\mathbb{R}^n, \mathbb{R}^m)} \\ &\leq \sum_{j=1}^m \frac{\epsilon}{m} \leq \epsilon \end{aligned}$$

finalmente

$$\|f(\mathbf{x}_1) - f(\mathbf{x}_2) - Df(\mathbf{p}) \cdot (\mathbf{x}_1 - \mathbf{x}_2)\| \leq \|\Lambda - Df(\mathbf{p})\| \|\mathbf{x}_1 - \mathbf{x}_2\| \leq \epsilon \|\mathbf{x}_1 - \mathbf{x}_2\|$$

□

Teorema 12 (TFINV débil). Si $f: D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ es de clase C^1 en el disco $B_r(\mathbf{p}) \subset D$, y $Df(\mathbf{p})$ es inyectiva, entonces existe $\delta > 0$ tal que la restricción de f al disco cerrado $\overline{B_\delta(\mathbf{p})}$ tiene inversa continua definida en la imagen $f(\overline{B_\delta(\mathbf{p})})$.

Demostración.

Como $Df(\mathbf{p})$ es invertible, existe $r > 0$ tal que

$$r\|\mathbf{h}\| \leq \|Df(\mathbf{p}) \cdot \mathbf{h}\| \quad (1.15)$$

Por otra parte del lema 2 con $\epsilon = r/2$,

$$\|f(\mathbf{x}_1) - f(\mathbf{x}_2) - Df(\mathbf{p}) \cdot (\mathbf{x}_1 - \mathbf{x}_2)\| \leq \frac{r}{2}\|\mathbf{x}_1 - \mathbf{x}_2\|$$

de la desigualdad del triángulo

$$\begin{aligned} & \|Df(\mathbf{p}) \cdot (\mathbf{x}_1 - \mathbf{x}_2)\| - \|f(\mathbf{x}_1) - f(\mathbf{x}_2)\| \\ & \leq \|f(\mathbf{x}_1) - f(\mathbf{x}_2) - Df(\mathbf{p}) \cdot (\mathbf{x}_1 - \mathbf{x}_2)\| \\ & \leq \frac{r}{2}\|\mathbf{x}_1 - \mathbf{x}_2\| \end{aligned}$$

por lo tanto

$$\|Df(\mathbf{p}) \cdot (\mathbf{x}_1 - \mathbf{x}_2)\| \leq \frac{r}{2}\|\mathbf{x}_1 - \mathbf{x}_2\| + \|f(\mathbf{x}_1) - f(\mathbf{x}_2)\|,$$

y aplicando la desigualdad 1.15 con $\mathbf{h} = \mathbf{x}_1 - \mathbf{x}_2$, obtenemos

$$r\|\mathbf{x}_1 - \mathbf{x}_2\| \leq \|Df(\mathbf{p}) \cdot (\mathbf{x}_1 - \mathbf{x}_2)\| \leq \frac{r}{2}\|\mathbf{x}_1 - \mathbf{x}_2\| + \|f(\mathbf{x}_1) - f(\mathbf{x}_2)\|,$$

de donde

$$\frac{r}{2}\|\mathbf{x}_1 - \mathbf{x}_2\| \leq \|f(\mathbf{x}_1) - f(\mathbf{x}_2)\|$$

lo cual muestra que la función es inyectiva, ya que si $\mathbf{x}_1 \neq \mathbf{x}_2$ entonces $f(\mathbf{x}_1) \neq f(\mathbf{x}_2)$.

La continuidad de la inversa se sigue de las siguientes propiedades que serán justificadas en la segunda parte de estas notas:

Proposición 12. *Sea $f: K \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ continua con K compacto. Si f tiene inversa $g: f(K) \rightarrow \mathbb{R}^n$ entonces la inversa es una función continua.*

□

El teorema anterior muestra que el sistema de ecuaciones

$$\begin{aligned} y_1 &= f_1(x_1, x_2, \dots, x_n), \\ y_2 &= f_2(x_1, x_2, \dots, x_n), \\ &\vdots \\ y_n &= f_m(x_1, x_2, \dots, x_n), \end{aligned}$$

y donde $\mathbf{q} = f(\mathbf{p})$, tiene una solución única

$$\begin{aligned} x_1 &= h_1(y_1, y_2, \dots, y_n), \\ x_2 &= h_2(y_1, y_2, \dots, y_n), \\ &\dots \\ x_n &= h_n(y_1, y_2, \dots, y_n), \end{aligned} \tag{1.16}$$

siempre que el determinante del jacobiano

$$\left| \frac{\partial(y_1, y_2, \dots, y_n)}{\partial(x_1, x_2, \dots, x_n)} \right|$$

sea distinto de cero en $\mathbf{p} = (p_1, p_2, \dots, p_n)$, y para (x_1, x_2, \dots, x_n) suficientemente próximo de (p_1, p_2, \dots, p_n) . Además la solución 1.16 depende continuamente de (y_1, y_2, \dots, y_n) en un entorno de \mathbf{q} . El teorema, sin embargo, *no dice nada acerca de la estructura de los puntos imagen*, es decir en que rango es posible variar (y_1, y_2, \dots, y_n) , por ejemplo los puntos imagen podrían estar en algún conjunto que no fuese abierto.

Ejemplo 1.16-33. La función $f: \mathbb{R}^2 \rightarrow \mathbb{R}^2$, $f(x_1, x_2) = (x_1 + x_2, e^{x_1+x_2})$, tiene por imagen

$$\{(y_1, y_2) \mid y_2 = e^{y_1}\}$$

que no es abierto en \mathbb{R}^2 .

Ejemplo 1.16-34. Considere la función en $\mathbb{R}^2 \rightarrow \mathbb{R}$

$$f(x, y) = \begin{cases} \frac{(x^2 - y^2)(x + y)}{x^2 + y^2} & \text{si } (x, y) \neq (0, 0), \\ 0 & \text{si } (x, y) = (0, 0) \end{cases}$$

Las derivadas parciales en el origen son:

$$\begin{aligned} D_1 f(0, 0) &= \lim_{h \rightarrow 0} \frac{1}{h} f(h, 0) = \frac{1}{h} \frac{h^3}{h^2} = 1, \\ D_2 f(0, 0) &= \lim_{k \rightarrow 0} \frac{1}{k} f(0, k) = \frac{1 - k^3}{k^2} = -1, \end{aligned}$$

Considere ahora la función $F: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ definida como $F(x, y) = (f(x, y), x + y)$. La matriz Jacobiana en el origen es

$$\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

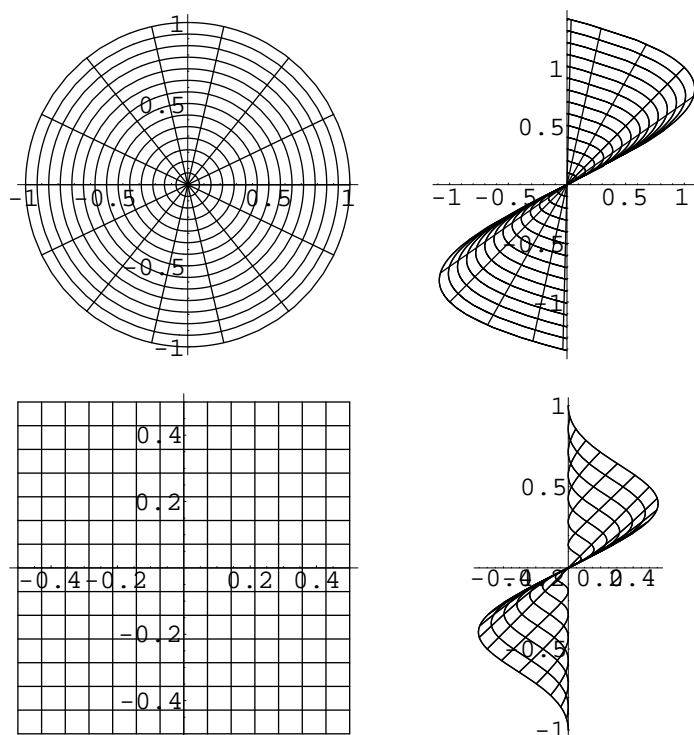


Figure 1.16: Imagen del disco y el cuadrado unitario bajo la función F del ejemplo 1.16

que evidentemente es no singular, pues el determinante es 2. Sin embargo en la figura ?? se muestra, en la columna izquierda el disco y el rectángulo de lado 1 con centro en el origen y en la columna derecha sus imágenes bajo F , como se puede apreciar las imágenes no son vecindades de $F(0, 0) = (0, 0)$, pues no existe un disco con centro en $(0, 0)$ contenido en la imagen, en ningún caso.

Otro inconveniente del teorema anterior, es que se busca tener resultados sobre la diferenciabilidad de la función inversa. La primera dificultad es que la imagen $f(\overline{B_\delta(\mathbf{p})})$ debería ser al menos una vecindad de $f(\mathbf{p})$ para poder hablar de la diferencial de la inversa; la segunda dificultad es probar que en efecto la inversa es diferenciable.

Debido a que la demostración de este resultado es un tanto más complicada que la demostración del teorema más débil 12, dejaremos la de-

mostración para el siguiente capítulo que puede ser opcional en una primera lectura, sin embargo daremos ahora el enunciado preciso. En lo que sigue preferiremos desarrollar con más detalles ejemplos concretos e ideas más intuitivas.

Teorema 13 (TFINV). *Sea $f: D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ de clase C^1 en un disco $B_r(\mathbf{p}) \subset D$. Si $Df(\mathbf{p})$ es invertible, entonces:*

1. *Existe un disco $B_\delta(\mathbf{p}) \subset B_r(\mathbf{p})$ tal que la imagen $V = f(B_\delta(\mathbf{p}))$ es una vecindad de $f(\mathbf{p})$.*
2. *La función inversa $g: V \rightarrow \mathbb{R}^n$ existe y es continua*
3. *Si $\mathbf{y} \in V$ es un punto interior de V , con $\mathbf{y} = f(\mathbf{x})$, entonces g es diferenciable en \mathbf{y} con derivada*

$$Dg(\mathbf{y}) = Df(\mathbf{x})^{-1}.$$

En la práctica, el análisis de una transformación $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$ comienza con calcular el *conjunto singular* donde la diferencial no sea inyectiva y su imagen, los *valores críticos*. Así, fuera del conjunto singular y de los valores críticos, el TFIN garantiza la existencia de una inversa local que además es diferenciable. Vamos a precisar

Definición 12. *Sea $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$ diferenciable. Se dice que $\mathbf{x} \in \mathbb{R}^n$ es un punto crítico de f , si $Df(\mathbf{x}) \in L(\mathbb{R}^n, \mathbb{R}^n)$ no es invertible⁵. Un punto $\mathbf{y} \in \text{Im}(f) \subset \mathbb{R}^n$ es un valor crítico, si es la imagen de un punto crítico, en caso contrario se dice que es un valor regular. Los puntos donde $Df(\mathbf{x})$ es invertible se llaman puntos regulares.*

El conjunto de puntos críticos de f se llama el conjunto de singularidades de f , o conjunto singular y se denota por

$$\Sigma f = \{\mathbf{x} \in \mathbb{R}^n \mid Df(\mathbf{x}) \text{ es singular}\}.$$

Ejemplo 1.16–35. El cambio de coordenadas polares a cartesianas

$$\begin{aligned} x &= r \cos \theta \\ y &= r \sin \theta \end{aligned} \tag{1.17}$$

⁵Recuerde que para una transformación lineal son equivalentes las propiedades de ser: (a) inyectiva \Leftrightarrow , (b) suprayectiva, \Leftrightarrow (c) invertible, \Leftrightarrow (d) determinante distinto de cero

puede considerarse como una función $T: \mathbb{R}^2(r, \theta) \rightarrow \mathbb{R}^2(x, y)$. Como es usual, distinguimos el dominio y el contradominio por las coordenadas y llamamos a (r, θ) el plano polar y (x, y) el plano cartesiano.

El determinante de la matriz jacobiana es

$$\begin{vmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{vmatrix} = r$$

de donde el conjunto de singularidades es el eje $r = 0$, en el plano polar. La imagen del conjunto singular es el origen en el plano cartesiano, ya que $x^2 + y^2 = r^2$.

Fuera del conjunto de singularidades la función es localmente invertible con inversa diferenciable. En la figura 1.17 se muestra la imagen de los rectángulos $[0, 1] \times [0, \pi]$ y $[0, 1] \times [0, 2\pi]$. Ambos tienen una arista sobre el conjunto singular $r = 0$ del plano polar la cual se mapea en el origen, la arista derecha $r = 1$, se mapea en cada caso a la semicircunferencia exterior o la circunferencia completa, respectivamente. Observe también que las aristas horizontales $[0, 1] \times \{0\}$ y $[0, 1] \times \{2\pi\}$ del rectángulo mayor, se aplican ambas sobre el rayo $0 \leq x \leq 1, y = 0$ del plano cartesiano. Esto muestra una dificultad de las coordenadas polares para representar de manera única puntos en el plano cartesiano. En el ejercicio (?) se muestra una manera de lidiar con esta dificultad, esencialmente identificando θ con $\theta + 2\pi k$, cuando k sea cualquier entero. Esta manera de pensar obliga a pensar a la transformación como una función

$$f: [0, \infty) \times S^1 \rightarrow \mathbb{R}^2$$

donde $[0, \infty) \times S^1$ es un *cilindro*, el producto cartesiano de un intervalo con la circunferencia de radio 1. Esta manera de pensar nos lleva a rehacer el cálculo para funciones de la forma $f: M \rightarrow \mathbb{R}^m$ donde M es alguna *superficie* en \mathbb{R}^p de dimensión n .

En la figura 1.18 se muestra la imagen bajo la transformación polar en la vecindad un punto $(r, \theta) = (0, \pi/4)$ sobre el conjunto singular en cambio en la figura 1.19 se muestra la imagen de un pequeño rectángulo en una vecindad de un punto regular $(1, \pi/4)$ en el plano polar y su imagen en el plano cartesiano. En la figura 1.20 se muestra la imagen de un rectángulo más grande que toca al conjunto singular, su imagen en el plano cartesiano revela que la función no es biyectiva en esta vecindad.

Ésto último muestra que el TFIV garantiza bajo las hipótesis adecuadas, la existencia de la inversa pero sólo *localmente*.

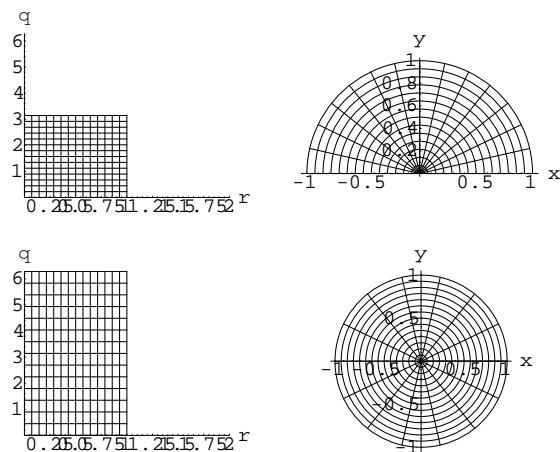


Figure 1.17: Imagen bajo la transformación polar 1.17 de dos rectángulos de base unitaria y alturas $\theta = \pi/2, 2\pi$.

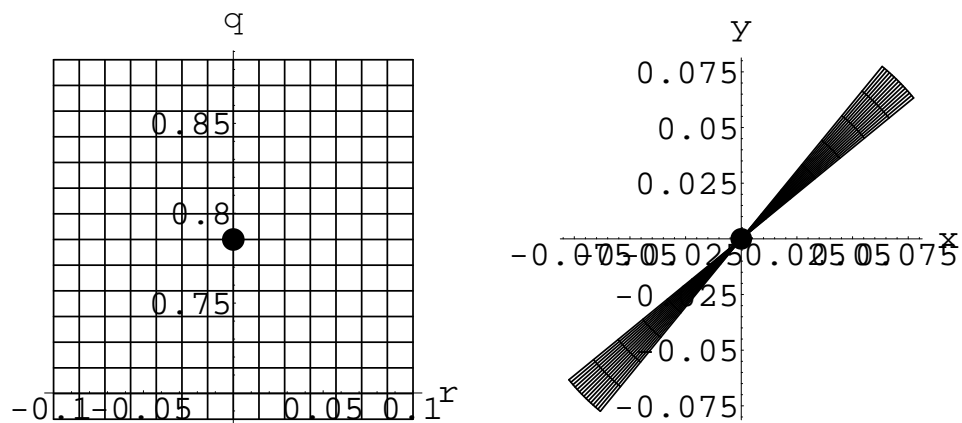


Figure 1.18: Imagen bajo la transformación polar 1.17 de un rectángulo alrededor de $(0, \pi/4)$ en el plano polar y su imagen bajo la transformación 1.17.

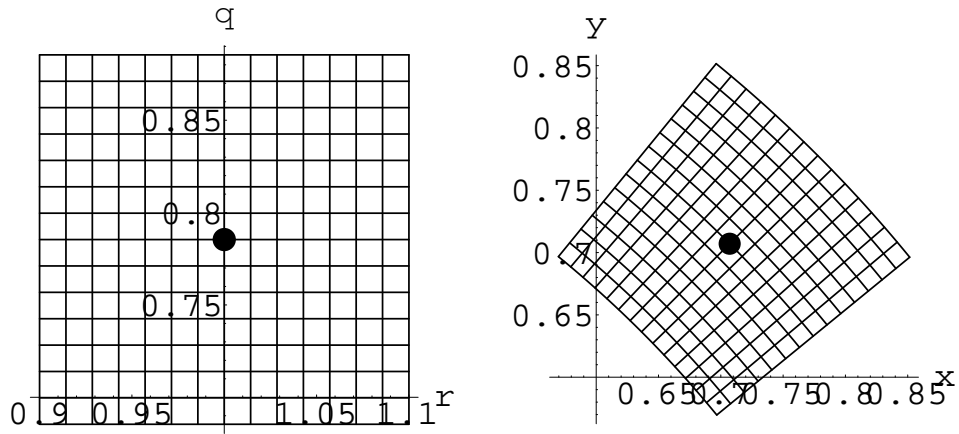


Figure 1.19: Imagen bajo la transformación polar 1.17 de un pequeño rectángulo mayor alrededor de $(1, \pi/4)$ en el plano polar y su imagen bajo la transformación 1.17.

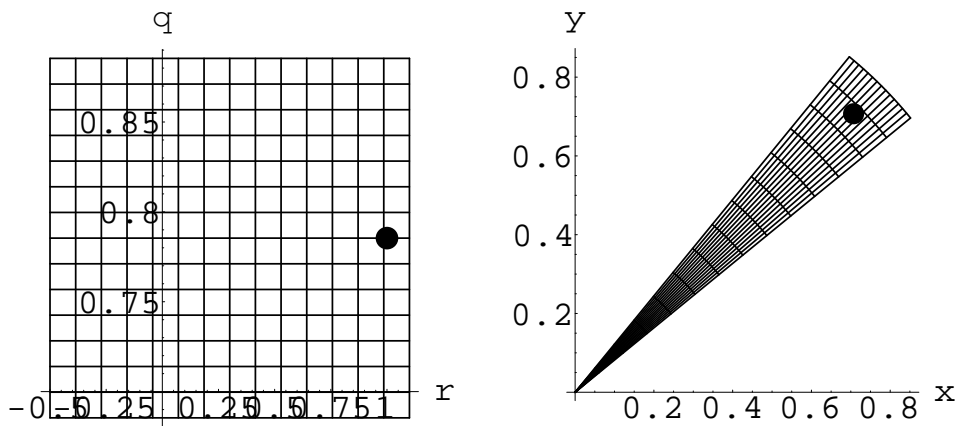


Figure 1.20: Imagen bajo la transformación polar 1.17 de un rectángulo mayor alrededor de $(0, \pi/4)$ en el plano polar y su imagen bajo la transformación 1.17.

Ejemplo 1.16–36. El cambio de coordenadas esféricas a cartesianas

$$\begin{aligned}x &= r \sin \theta \cos \varphi \\y &= r \sin \theta \sin \varphi \\z &= r \cos \theta\end{aligned}$$

puede considerarse como una función $T: \mathbb{R}^3(r, \theta, \varphi) \rightarrow \mathbb{R}^3(x, y, z)$. Analicemos la invertibilidad de la derivada $DT(r, \theta, \varphi)$ mediante la matriz Jacobiana

$$\begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} & \frac{\partial x}{\partial \varphi} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} & \frac{\partial y}{\partial \varphi} \\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial \theta} & \frac{\partial z}{\partial \varphi} \end{pmatrix} = \begin{pmatrix} \sin \theta \cos \varphi & r \cos \theta \cos \varphi & -r \sin \theta \sin \varphi \\ \sin \theta \sin \varphi & r \cos \theta \sin \varphi & r \sin \theta \cos \varphi \\ \cos \theta & -r \sin \theta & 0 \end{pmatrix}$$

cuyo determinante es $r^2 \sin \theta$, por lo tanto si $r \neq 0$ y $\theta \neq k\pi$ para algún k entero, la diferencial es invertible. Observe que los valores críticos constituyen el eje z en el espacio cartesiano, ya que

$$x^2 + y^2 = r^2 \sin^2 \theta$$

Por el teorema de la función inversa, fuera del conjunto singular es posible despejar a (r, θ, φ) en función de (x, y, z) . En este caso esto es posible con algún esfuerzo pues hay que manipular fórmulas no lineales, sin embargo *la derivada de la función inversa* se puede calcular aún cuando no se conozca explícitamente la función inversa:

$$\begin{pmatrix} \sin \theta \cos \varphi & r \cos \theta \cos \varphi & -r \sin \theta \sin \varphi \\ \sin \theta \sin \varphi & r \cos \theta \sin \varphi & r \sin \theta \cos \varphi \\ \cos \theta & -r \sin \theta & 0 \end{pmatrix}$$

1.17 El teorema de la función implícita

Consideremos la curva de nivel $f(x, y) = c$ y un punto sobre ésta: $f(x_0, y_0) = c$. Nos preguntamos bajo qué condiciones es posible ver esta curva como la gráfica de una función $y = \phi(x)$.

Ejemplo 1.17–37. Considere la curva $x^2 + y^2 = 1$. Alrededor del punto $(1, 0)$ sobre la curva se puede despejar a y :

$$y = \pm \sqrt{1 - x^2}$$

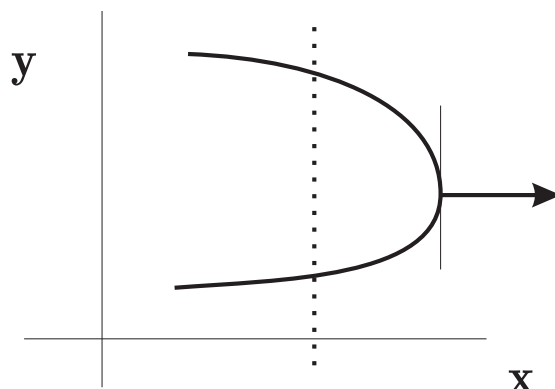


Figure 1.21: La curva $f(x, y)$ no es la gráfica de una función $y = \phi(x)$.

donde el signo \pm significa que debemos elegir una rama de la raíz cuadrada, en este caso elegimos el signo más pues queremos que la gráfica de $y = \phi(x)$ satisfaga $1 = \phi(0)$. Claramente la función $\phi(x) = \sqrt{1 - x^2}$ sólo está definida y es diferenciable en $(0, 1)$. Esto muestra que no podemos esperar que la función ϕ esté definida sino en un entorno de x_0 .

Geoméricamente, será posible despejar a y en función de x cuando toda recta vertical cercana a $x = x_0$ corte a la curva de nivel $f(x, y) = c$ en un solo punto, es decir la curva no se doble sobre sí misma como en la Figura 1.21

Observe que si ocurre un doblez, la tangente a la curva es vertical y por lo tanto el gradiente es horizontal. Como

$$\nabla f(x, y) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right)$$

la tangente a la curva es vertical precisamente cuando $\frac{\partial f}{\partial y} = 0$. Este resultado está contenido en el Teorema de la Función Implícita (TFIm) para funciones de una variable real.

Teorema 14 (TFIM en $\mathbb{R} \rightarrow \mathbb{R}$). Sea $f(x, y): B_r(x_0, y_0) \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ de clase C^1 en $B_r(x_0, y_0)$, y $f(x_0, y_0) = c$. Si

$$\frac{\partial f}{\partial y}(x_0, y_0) \neq 0$$

entonces existe una función diferenciable

$$\phi: (x_0 - \delta, x_0 + \delta) \rightarrow (y_0 - \delta, y_0 + \delta)$$

tal que

$$f(x, \phi(x)) = c$$

y viceversa,

$$f(x, y) = c, \quad |x - x_0| < \delta, \quad |y - y_0| < \delta$$

implica que $y = \phi(x)$. Además

$$\phi'(x) = -\frac{\frac{\partial f}{\partial x}(x, \phi(y))}{\frac{\partial f}{\partial y}(x, \phi(y))}.$$

Ejemplo 1.17–38. Considere el sistema lineal

$$\begin{aligned} ax + by + cz &= 0, \\ dx + ey + fz &= 0 \end{aligned}$$

el cual “tiene una variable libre”. Con más precisión, si

$$\Delta_1 = ae - bd \neq 0$$

entonces se puede escribir

$$\begin{aligned} ax + by &= -cz, \\ dx + ey &= -fz \end{aligned}$$

y para cada z se obtienen valores únicos de x y y ; en otras palabras existen funciones $x = \varphi_1(z)$, $y = \psi_1(z)$ que satisfacen el sistema

$$\begin{aligned} a\varphi_1(z) + b\psi_1(z) + cz &= 0, \\ d\varphi_1(z) + e\psi_1(z) + fz &= 0 \end{aligned}$$

para toda $z \in \mathbb{R}$. Las funciones φ_1 , ψ_1 se encuentran explícitamente, usando la regla de Cramer, como

$$x = \varphi_1(z) = \frac{(-ce + bf)z}{ae - bd}, \quad y = \psi_1(z) = \frac{(-af + cd)z}{ae - bd}.$$

Observe que $\varphi_1(0) = 0$ y $\psi_1(0) = 0$

Ejemplo 1.17-39. Considere el mismo sistema lineal que en el ejemplo anterior

$$\begin{aligned} ax + by + cz &= 0, \\ dx + ey + fz &= 0 \end{aligned}$$

pero suponga ahora que

$$\Delta_2 = af - cd \neq 0$$

entonces se despejar a x, z en función de y ,

$$x = \varphi_2(y) = \frac{(-bf + ce)y}{af - cd}, \quad z = \psi_2(y) = \frac{(-ae + bd)y}{af - cd}.$$

es decir

$$\begin{aligned} a\varphi_2(y) + by + c\psi_2(y) &= 0, \\ d\varphi_2(y) + ey + f\psi_2(y) &= 0 \end{aligned}$$

para toda $y \in \mathbb{R}$.

Ejemplo 1.17-40. Para el mismo sistema que en el ejemplo anterior, si

$$\Delta_3 = bf - ce \neq 0$$

entonces es posible despejar a y, z en función de x .

Ejemplo 1.17-41. Considere la versión no lineal del ejemplo (1.17),

$$\begin{aligned} ax + by + cz + r(x, y, z) &= 0, \\ dx + ey + fz + s(x, y, z) &= 0 \end{aligned}$$

donde $ae - bd \neq 0$, y donde r, s son funciones continuamente diferenciables tales que $r(x, y, z) \leq k(|x|^2 + |y|^2 + |z|^2)$, $s(x, y, z) \leq k'(|x|^2 + |y|^2 + |z|^2)$, para ciertas constantes k, k' y para (x, y, z) suficientemente pequeño. Inspirados por el caso lineal y notando que $x = y = z = 0$ satisface el sistema, podríamos conjeturar que existen funciones $\varphi_1(z), \psi_1(z)$ diferenciables, definidas para $|z| < \delta$, tales que $\varphi_1(0) = 0, \psi_1(0) = 0$ y

$$\begin{aligned} a\varphi_1(z) + b\psi_1(z) + cz + r(\varphi_1(z), \psi_1(z), z) &= 0, \\ d\varphi_1(z) + e\psi_1(z) + fz + s(\varphi_1(z), \psi_1(z), z) &= 0 \end{aligned}$$

para toda z . Si tal es el caso, observe que derivando respecto de z ,

$$\begin{aligned} a\varphi_1'(z) + b\psi_1'(z) + cz + \\ r_x(\varphi_1(z), \psi_1(z), z)\varphi_1'(z) + r_y(\varphi_1(z), \psi_1(z), z)\psi_1'(z) + r_z(\varphi_1(z), \psi_1(z), z) &= 0, \\ d\varphi_1'(z) + e\psi_1'(z) + fz + \\ s_x(\varphi_1(z), \psi_1(z), z)\varphi_1'(z) + s_y(\varphi_1(z), \psi_1(z), z)\psi_1'(z) + s_z(\varphi_1(z), \psi_1(z), z) &= 0, \end{aligned}$$

y evaluando en $z = 0$ y recordando que $\varphi_1(0) = 0$, $\psi_1(0) = 0$, obtenemos

$$\begin{aligned} a\varphi_1'(0) + b\psi_1'(0) &= 0, \\ d\varphi_1'(0) + e\psi_1'(0) &= 0, \end{aligned}$$

de donde podemos obtener

$$\varphi_1'(0) = \frac{(-ce + bf)}{ae - bd}, \quad \psi_1'(0) = \frac{(-af + cd)}{ae - bd}.$$

es decir

$$x = \varphi_1(z) = \frac{(-ce + bf)}{ae - bd}z + \dots, \quad y = \psi_1(z) = \frac{(-af + cd)}{ae - bd}z + \dots.$$

donde (\dots) significa términos de orden 2 en adelante.

Teorema 15 (Teorema de la función implícita). *Sea $F: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^m$, de clase C^1 en un disco alrededor de $(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^n \times \mathbb{R}^m$. Si $F(\mathbf{a}, \mathbf{b}) = \mathbf{c}$ y $D_2F(\mathbf{a}, \mathbf{b}) \in L(\mathbb{R}^m, \mathbb{R}^m)$ es no singular, entonces existe $\varphi: B_\delta(\mathbf{a}) \subset \mathbb{R}^n \rightarrow V \subset \mathbb{R}^m$ tal que V es una vecindad de \mathbf{b} , $\varphi(\mathbf{a}) = \mathbf{b}$ y*

$$F(\mathbf{x}, \varphi(\mathbf{x})) = \mathbf{c}$$

para toda $\mathbf{x} \in B_\delta(\mathbf{a})$. Además φ es diferenciable en $B_\delta(\mathbf{a})$ y su derivada es

$$D\varphi(\mathbf{x}) = -D_2F(\mathbf{x}, \varphi(\mathbf{x}))^{-1} \circ D_1F(\mathbf{x}, \varphi(\mathbf{x})) \quad (1.18)$$

En otras palabras, las soluciones de la ecuación

$$F(\mathbf{x}, \mathbf{y}) = \mathbf{c}$$

cercanas de una solución (\mathbf{a}, \mathbf{b}) se pueden parametrizar diferenciablemente como $\mathbf{y} = \varphi(\mathbf{x})$.

Observación 13. Observe la estructura de la composición en (1.18) (omitimos el punto de evaluación):

$$\begin{array}{ccc}
 & \mathbb{R}^m & \\
 D_1F \nearrow & & \searrow -D_2F^{-1} \\
 \mathbb{R}^n & \xrightarrow{D\varphi} & \mathbb{R}^m
 \end{array}$$

Observación 14. Un resultado similar se aplica para $F: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$, de clase C^1 en un disco alrededor de $(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^n \times \mathbb{R}^m$. Bajo la condición de que $F(\mathbf{a}, \mathbf{b}) = \mathbf{c}$ y $D_1F(\mathbf{a}, \mathbf{b}) \in L(\mathbb{R}^n, \mathbb{R}^n)$ sea no singular, entonces existe $\psi: B_\delta(\mathbf{b}) \subset \mathbb{R}^m \rightarrow U \subset \mathbb{R}^n$ tal que U es vecindad de \mathbf{a} , $\psi(\mathbf{b}) = \mathbf{a}$ y

$$F(\psi(\mathbf{y}), \mathbf{y}) = \mathbf{c}$$

para toda $\mathbf{y} \in B_\delta(\mathbf{b})$. Además ψ es diferenciable y

$$D\psi(\mathbf{y}) = -D_1F(\psi(\mathbf{y}), \mathbf{y})^{-1} \circ D_2F((\psi(\mathbf{y}), \mathbf{y})).$$

en otras palabras, el siguiente diagrama conmuta:

$$\begin{array}{ccc}
 & \mathbb{R}^n & \\
 D_2F \nearrow & & \searrow -D_1F^{-1} \\
 \mathbb{R}^m & \xrightarrow{D\psi} & \mathbb{R}^n
 \end{array}$$

Ejemplo 1.17-42. Retomemos el ejemplo (1.17).

$$\begin{aligned}
 ax + by + cz + r(x, y, z) &= 0, \\
 dx + ey + fz + s(x, y, z) &= 0
 \end{aligned}$$

donde r, s son funciones de clase C^1 de orden cuadrático en el origen, i.e. $|r(x, y, z)| \leq k(x^2 + y^2 + z^2)$ para (x, y, z) próximo de cero.

Para aplicar el TFIM, reformulemos el problema como sigue: Sean

$$\begin{aligned}
 F_1(z, (x, y)) &= ax + by + cz + r(x, y, z), \\
 F_2(z, (x, y)) &= dx + ey + fz + s(x, y, z)
 \end{aligned}$$

es decir, $F: \mathbb{R}(z) \times \mathbb{R}^2(x, y) \rightarrow \mathbb{R}^2$, con $F = (F_1, F_2)$ y $F_i: \mathbb{R}^2 \rightarrow \mathbb{R}$, para $i = 1, 2$ [suponemos sin pérdida de generalidad que dichas funciones están definidas para toda $(z, (x, y)) \in \mathbb{R} \times \mathbb{R}^2$, aunque solo es necesario que estén

definidas en un entorno de $(0, (0, 0))$. Como r, s son de orden cuadrático, $(0, (0, 0))$ es solución del sistema

$$\begin{aligned} F_1(z, (x, y)) &= 0, \\ F_2(z, (x, y)) &= 0 \end{aligned}$$

Claramente F es de clase C^1 en un entorno de $(0, (0, 0))$. Procedamos con el análisis, por un lado $D_2F(z_0, (x_0, y_0)) \in L(\mathbb{R}^2, \mathbb{R}^2)$, luego

$$D_2F(z_0, (x_0, y_0)) \cdot (dx, dy) = (D_2F_1(0, (0, 0)) \cdot (dx, dy), D_2F_2(0, (0, 0)) \cdot (dx, dy))$$

donde⁶

$$\begin{aligned} &D_2F_1(0, (0, 0)) \cdot (dx, dy) \\ &= \frac{\partial F_1}{\partial x}(0, 0, 0)dx + \frac{\partial F_1}{\partial y}(0, 0, 0)dy \\ &= a dx + b dy + \frac{\partial r}{\partial x}(0, 0, 0)dx + \frac{\partial r}{\partial y}(0, 0, 0)dy \\ &= a dx + b dy \end{aligned}$$

y

$$\begin{aligned} &D_2F_2(0, (0, 0)) \cdot (dx, dy) \\ &= \frac{\partial F_2}{\partial x}(0, 0, 0)dx + \frac{\partial F_2}{\partial y}(0, 0, 0)dy \\ &= k dx + m dy + \frac{\partial s}{\partial x}(0, 0, 0)dx + \frac{\partial s}{\partial y}(0, 0, 0)dy \\ &= k dx + m dy \end{aligned}$$

por ser r, s de orden cuadrático. Por lo tanto

$$D_2F(z_0, (x_0, y_0)) \cdot (dx, dy) = (a dx + b dy, k dx + m dy) = \begin{pmatrix} a & b \\ k & m \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix}$$

es decir la matriz asociada a $D_2F(z_0, (x_0, y_0))$ en la base canónica de \mathbb{R}^2 es

$$[D_2F(z_0, (x_0, y_0))] = \begin{pmatrix} a & b \\ k & m \end{pmatrix}$$

⁶Observe como identificamos $F_i: \mathbb{R} \times \mathbb{R}^2 \rightarrow \mathbb{R}$ con una función $F_i: \mathbb{R}^3 \rightarrow \mathbb{R}$, es decir identificamos $(z, (x, y))$ con (z, x, y) .

por lo cual, la condición para poder despejar a x, y en función de z es precisamente

$$\begin{vmatrix} a & b \\ k & m \end{vmatrix} \neq 0.$$

Ejemplo 1.17-43. Sea $F: \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}^2$ definida para $\mathbf{x} = (x_1, x_2)$, $\mathbf{y} = (y_1, y_2)$ por

$$F(\mathbf{x}, \mathbf{y}) = (x_1^3 + x_2x_1 + y_2, x_1y_2 + x_2^2 - y_1).$$

¿Para cuáles puntos (\mathbf{x}, \mathbf{y}) se puede resolver la ecuación $F(\mathbf{x}, \mathbf{y}) = \mathbf{0}$ para \mathbf{x} en términos de \mathbf{y} ?

Sean

$$\begin{aligned} F_1(\mathbf{x}, \mathbf{y}) &= x_1^3 + x_2x_1 + y_2, \\ F_2(\mathbf{x}, \mathbf{y}) &= x_1y_2 + x_2^2 - y_1 \end{aligned}$$

Necesitamos calcular la diferencial parcial

$$\frac{\partial F}{\partial \mathbf{x}}(\mathbf{x}_0, \mathbf{y}_0) \in L(\mathbb{R}^2, \mathbb{R}^2)$$

que se representa la matriz 2×2 ,

$$\begin{pmatrix} \frac{\partial F_1}{\partial x_1} & \frac{\partial F_1}{\partial x_2} \\ \frac{\partial F_2}{\partial x_1} & \frac{\partial F_2}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 3x_1^2 + x_2 & x_1 \\ y_2 & 2x_2 \end{pmatrix}$$

Por lo tanto, si $2x_2(3x_1^2 + x_2) - x_1y_1 \neq 0$ se puede despejar a $\mathbf{x} = (x_1, x_2)$ en función de $\mathbf{y} = (y_1, y_2)$ en un entorno del punto.

1.17.1 Aplicaciones

Mecánica

El estado de movimiento de una partícula en un potencial gravitacional Newtoniano se puede describir por su energía y momento angular

$$e = \frac{1}{2} \|\mathbf{v}\|^2 - \frac{1}{\|\mathbf{x}\|}, \quad \mathbf{h} = \mathbf{x} \times \mathbf{v}$$

que son constantes de movimiento, i.e. permanecen constantes todo el tiempo.

La pregunta es cómo describir todos los posibles estados de movimiento con el menor número de parámetros.

Para comenzar definamos

$$F(x_1, x_2, x_3, v_1, v_2, v_3) = (e, h_1, h_2, h_3)$$

donde e y h_i son la energía y las componentes del momento angular. Por ejemplo, si fijamos la energía $e = e_0$ y el momento angular $\mathbf{h} = \mathbf{h}_0$, solo podemos esperar despejar cuatro variables del total de seis: $(x_1, x_2, x_3, v_1, v_2, v_3)$. Veamos cuándo se puede despejar la velocidad \mathbf{v} y x_3 en función de x_1, x_2 : Para ello debemos verificar que el siguiente discriminante

$$\Delta_3 = \begin{vmatrix} \frac{\partial e}{\partial v_1} & \frac{\partial e}{\partial v_2} & \frac{\partial e}{\partial v_3} & \frac{\partial e}{\partial v_4} \\ \frac{\partial h_1}{\partial v_1} & \frac{\partial h_1}{\partial v_2} & \frac{\partial h_1}{\partial v_3} & \frac{\partial h_1}{\partial v_4} \\ \frac{\partial h_2}{\partial v_1} & \frac{\partial h_2}{\partial v_2} & \frac{\partial h_2}{\partial v_3} & \frac{\partial h_2}{\partial v_4} \\ \frac{\partial h_3}{\partial v_1} & \frac{\partial h_3}{\partial v_2} & \frac{\partial h_3}{\partial v_3} & \frac{\partial h_3}{\partial v_4} \end{vmatrix} \neq 0$$

El lector deberá verificar que el determinant anterior es

$$\Delta_3 = -h_3(\mathbf{x} \cdot \mathbf{v})$$

Por lo tanto, si la componente en la dirección z del momento angular es distinta de cero y la velocidad no es perpendicular al vector de posición, entonces es posible despejar a \mathbf{v} y a x_3 en función de x_1, x_2 . Una pregunta de interés es entonces determinar la región de Hill, es decir la región en el plano x_1, x_2 que sea compatible con los valores de energía y momento angular dados. Por ejemplo, de la ecuación de la energía se deduce que

$$\frac{1}{2} \|\mathbf{v}\|^2 = e + \frac{1}{\|\mathbf{x}\|} \geq 0$$

por lo tanto, si $e < 0$ la región de Hill en \mathbf{x} es la esfera

$$\|\mathbf{x}\| < -\frac{1}{e}$$

menos el origen. Si $e \geq 0$ la región de Hill en \mathbf{x} es todo $\mathbb{R}^2 \setminus \{\mathbf{0}\}$. En el plano x_1-x_2 sucede que

$$\sqrt{x_1^2 + x_2^2} \leq \|\mathbf{x}\| < -\frac{1}{e}$$

y la región de Hill en el plano x_1-x_2 , para $e < 0$, está contenida en el disco de radio $-e^{-1}$ menos el origen. En la frontera del disco $\mathbf{v} = \mathbf{0}$ por lo tanto no podemos aplicar el teorema de la función implícita.

1.18 Derivadas parciales de orden superior

En esta sección estudiaremos las diferenciales de orden superior. Para ello consideremos una función $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciable en un disco $B_r(\mathbf{x}_0) \subset \mathbb{R}^n$. La diferencial es una función

$$Df: B_r(\mathbf{x}_0) \subset \mathbb{R}^n \rightarrow L(\mathbb{R}^n, \mathbb{R}^m).$$

Mencionamos anteriormente que $L(\mathbb{R}^n, \mathbb{R}^m)$ constiuye un espacio vectorial en sí mismo, por lo que tiene sentido preguntarse sobre la diferencial de Df en un punto, digamos \mathbf{x}_0 . En caso de existir tendríamos

$$D(Df)(\mathbf{x}_0) \in L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^m))$$

Definición 13. Sea $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciable en un disco $B_r(\mathbf{x}_0) \subset \mathbb{R}^n$. Decimos que f es dos veces diferenciable en \mathbf{x}_0 , si $Df: B_r(\mathbf{x}_0) \rightarrow \mathbb{R}^m$ es diferenciable en \mathbf{x}_0 .

El espacio vectorial $L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^m))$ parece tener una estructura complicada, por ello vamos a hacer una digresión sobre las funciones multilíneales que al final nos permitirá identificar la diferencial de orden k en un punto con una función multilíneal.

El espacio de funciones multilíneales

Definición 14. El espacio de funciones multilíneales en k argumentos vectoriales en \mathbb{R}^n a valores en \mathbb{R}^m , o brevemente k -multilíneales vectoriales, es

$$L_k(\mathbb{R}^n, \mathbb{R}^m) = \left\{ B: (\mathbb{R}^n)^k \rightarrow \mathbb{R} \mid B \text{ es multilíneal} \right\}$$

Los espacios $L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^m))$ y $L_2(\mathbb{R}^n, \mathbb{R}^m)$ se pueden identificar de manera canónica, vgr. existe. En efecto, si $B \in L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^m))$ definamos \hat{B} como,

$$(B \cdot \mathbf{x}) \cdot \mathbf{y} = \hat{B} \cdot (\mathbf{x}, \mathbf{y})$$

Dejamos al lector probar que la aplicación

$$L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^m)) \xrightarrow{B \rightarrow \hat{B}} L_2(\mathbb{R}^n, \mathbb{R}^m)$$

es un isomorfismo lineal (i.e. es lineal y biyectiva).

De manera análoga, los espacios $L(\mathbb{R}^n, L(\mathbb{R}^n, \dots, L(\mathbb{R}^n, \mathbb{R}^m)))$ y $L_k(\mathbb{R}^n, \mathbb{R}^m)$ se pueden indentificar de manera canónica mediante el isomorfismo

$$L(\mathbb{R}^n, \overset{k\text{-factores}}{L(\mathbb{R}^n, \dots, L(\mathbb{R}^n, \mathbb{R}^m))}) \xrightarrow{B \rightarrow \hat{B}} L_k(\mathbb{R}^n, \mathbb{R}^m)$$

donde,

$$(\dots((B \cdot \mathbf{x}_1) \cdot \mathbf{x}_2) \dots) \mathbf{x}_k = \hat{B} \cdot (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k)$$

Observación 15. En consonancia con la notación $\Lambda \cdot \mathbf{h}$ para denotar la evaluación de una transformación lineal Λ en un vector \mathbf{h} , denotaremos por $B \cdot (\mathbf{h}, \mathbf{k})$ la evaluación de la transformación bilineal B en la pareja de vectores (\mathbf{h}, \mathbf{k}) . Análogamente $B(\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_k)$ denota la evaluación de la función multilínea B .

En el conjunto $L_k(\mathbb{R}^n, \mathbb{R}^m)$ están definidas las operaciones de suma y producto por un escalar, de modo que es un espacio vectorial sobre los reales.

Se puede también dotar de una norma a $L_k(\mathbb{R}^n, \mathbb{R}^m)$ definiendo

$$\|B\| = \sup\{B(\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_k) \mid \|\mathbf{x}_1\| = \|\mathbf{x}_2\| = \dots = \|\mathbf{x}_k\| = 1\} \quad (1.19)$$

con lo cual se satisface la desigualdad de Banach:

$$\|B(\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_k)\| \leq \|\mathbf{h}_1\| \|\mathbf{h}_2\| \dots \|\mathbf{h}_k\|$$

Por supuesto se puede probar que el supremo en (1.19) existe por tratarse de funciones multilíneas en espacios vectoriales *de dimensión finita*.

Definición 15. Una función $B(\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_k)$ se dice *simétrica*, si

$$B(\mathbf{h}_{\sigma_1}, \mathbf{h}_{\sigma_2}, \dots, \mathbf{h}_{\sigma_k}) = B(\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_k)$$

para cualquier permutación $(\sigma_1, \sigma_2, \dots, \sigma_k)$ se los índices $(1, 2, \dots, k)$.

Por ejemplo, una función bilínea es simétrica si $B(\mathbf{x}, \mathbf{y}) = B(\mathbf{y}, \mathbf{x})$.

Proposición 13. Si $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ es 2 veces diferenciable en \mathbf{x}_0 entonces $D(\widehat{Df})(\mathbf{x}_0)$ es una función bilínea simétrica.

1.18.1 Derivadas de orden superior

Definición 16. Sea $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciable en un disco $B_r(\mathbf{x}_0) \subset \mathbb{R}^n$, si $Df: B_r(\mathbf{x}_0) \rightarrow \mathbb{R}^m$ es diferenciable en \mathbf{x}_0 , la diferencial de orden dos se define como la función bilínea

$$D^2 f(\mathbf{x}_0) = D(\widehat{Df})(\mathbf{x}_0).$$

Si

$$D(D(D(\cdots D f) \cdots)): B_r(\mathbf{x}_0) \rightarrow L(\mathbb{R}^n, \cdots L(\mathbb{R}^n, \mathbb{R}^m))$$

es diferenciable en \mathbf{x}_0 , entonces la diferencial de orden k se define como la función k -multilineal

$$D^k f(\mathbf{x}_0) = D(D(D(\widehat{\cdots D f} \cdots)))(\mathbf{x}_0).$$

En otras palabras

$$\begin{aligned} (D(Df)(\mathbf{x}_0) \cdot \mathbf{h}) \cdot \mathbf{k} &= D^2 f(\mathbf{x}_0) \cdot (\mathbf{h}, \mathbf{k}) \\ ((D(D(Df)))(\mathbf{x}_0) \cdot \mathbf{h}) \cdot \mathbf{k} \mathbf{m} &= D^3 f(\mathbf{x}_0) \cdot (\mathbf{h}, \mathbf{k}, \mathbf{m}), \end{aligned}$$

etc.

El siguiente resultado nos permite calcular de manera efectiva la segunda diferencial

Proposición 14. Sea $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ continuamente diferenciable en un disco $B_r(\mathbf{x}_0)$ y dos veces diferenciable en \mathbf{x}_0 . Para $\mathbf{h} \in \mathbb{R}^n$ defina $f'_\mathbf{h}: B_r(\mathbf{x}_0) \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ como

$$f'_\mathbf{h}(\mathbf{x}) = Df(\mathbf{x}) \cdot \mathbf{h}$$

entonces $f'_\mathbf{h}$ es diferenciable en \mathbf{x}_0 y para $\mathbf{k} \in \mathbb{R}^n$ se cumple que

$$Df'_\mathbf{h}(\mathbf{x}_0) \cdot \mathbf{k} = D^2 f(\mathbf{x}_0) \cdot (\mathbf{h}, \mathbf{k})$$

Corolario 3. Sea $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$, $\mathbf{r}: \mathbb{R} \rightarrow \mathbb{R}^n$. Bajo las hipótesis apropiadas,

$$\frac{d}{dt} Df(\mathbf{r}) \cdot \mathbf{h} = D^2 f(\mathbf{r}(t)) \cdot (\mathbf{r}'(t), \mathbf{h})$$

Demostración.

Por la regla de la cadena

$$\begin{aligned} \frac{d}{dt} Df(\mathbf{r}(t)) \cdot \mathbf{h} &= \left(\frac{d}{dt} Df(\mathbf{r}(t)) \right) \cdot \mathbf{h} \\ &= (D(Df)(\mathbf{r}(t)) \cdot \mathbf{r}'(t)) \cdot \mathbf{h} \\ &= D^2 f(\mathbf{r}(t)) \cdot (\mathbf{r}'(t), \mathbf{h}) \end{aligned}$$

□

Corolario 4. Sea $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$, $\mathbf{r}: \mathbb{R} \rightarrow \mathbb{R}^n$. Bajo las hipótesis apropiadas,

$$\frac{d}{dt} D^2 f(\mathbf{r}) \cdot (\mathbf{h}, \mathbf{k}) = D^3 f(\mathbf{r}(t)) \cdot (\mathbf{r}'(t), \mathbf{h}, \mathbf{k})$$

Ejemplo 1.18-44. Considere la función $f: \mathbb{R}^3 \rightarrow \mathbb{R}$, $f(x_1, x_2, x_3) = x_1 x_2 - x_2 x_3 + x_3 x_1$. Sean $\mathbf{h} = (h_1, h_2, h_3)$, $\mathbf{k} = (k_1, k_2, k_3)$, entonces

$$f'_{\mathbf{h}} = \frac{\partial f}{\partial x_1} h_1 + \frac{\partial f}{\partial x_2} h_2 + \frac{\partial f}{\partial x_3} h_3 = h_3(x_1 - x_2) + h_2(x_1 - x_3) + h_1(x_2 + x_3)$$

de donde

$$D^2 f(a, b, c) \cdot ((h_1, h_2, h_3), (k_1, k_2, k_3)) = (h_2 + h_3)k_1 + (h_1 - h_3)k_2 + (h_1 - h_2)k_3$$

(note que $D^2 f(a, b, c)$ no depende de (a, b, c))

Ejemplo 1.18-45. Sea $f: \mathbb{R}^3 \rightarrow \mathbb{R}^2$, $f(x, y, z) = (3xy^2, x^2 - z^3)$. Sean $\mathbf{h} = (h_1, h_2, h_3)$, $\mathbf{k} = (k_1, k_2, k_3)$ entonces

$$f'_{\mathbf{h}} = (3y^2 h_1 + 6xy h_2, 3x^2 h_1 - 3z^2 h_3)$$

de donde

$$D^2 f(a, b, c) \cdot ((h_1, h_2, h_3), (k_1, k_2, k_3)) = (6bh_2 k_1 + (6bh_1 + 6ah_2)k_2, 6(ah_1 k_1 - ch_3 k_3))$$

Proposición 15. Sea $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ de clase C^2 en un disco $B_r(\mathbf{x}_0) \subset \mathbb{R}^n$. Sea $f_{\mathbf{h}}(t) = f(\mathbf{x}_0 + t\mathbf{h})$, entonces en el desarrollo de Taylor de $f_{\mathbf{h}}(t)$,

$$f_{\mathbf{h}}(t) = \mathbf{a} + t\mathbf{b} + \frac{t^2}{2}\mathbf{c} + R_3,$$

$$\mathbf{a} = f(\mathbf{x}_0), \quad \mathbf{b} = Df(\mathbf{x}_0) \cdot \mathbf{h}, \quad \mathbf{c} = D^2 f(\mathbf{x}_0) \cdot (\mathbf{h}, \mathbf{h}).$$

Demostración.

Evaluando $f_{\mathbf{h}}(0)$ y $f'_{\mathbf{h}}(t)$ obtenemos

$$\begin{aligned} f_{\mathbf{h}}(0) &= f(\mathbf{x}_0), \\ f'_{\mathbf{h}}(t) &= Df(\mathbf{x}_0 + t\mathbf{h}) \cdot \mathbf{h} \end{aligned} \tag{1.20}$$

en particular,

$$f'_{\mathbf{h}}(0) = Df(\mathbf{x}_0) \cdot \mathbf{h}$$

derivando una vez más (1.20) respecto de t , obtenemos del Corolario (3),

$$f''_{\mathbf{h}}(0) = D^2 f(\mathbf{x}_0) \cdot (\mathbf{h}, \mathbf{h})$$

El resultado anterior nos permite calcular la forma cuadrática $D^2 f(\mathbf{x}_0)(\mathbf{h}, \mathbf{h})$. El siguiente resultado nos permite recuperar la forma bilineal

Proposición 16. Sea $B: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^m$ bilineal simétrica, $Q(\mathbf{x}) = B(\mathbf{x}, \mathbf{x})$, entonces

$$B(\mathbf{x}, \mathbf{y}) = \frac{1}{2} (Q(\mathbf{x} + \mathbf{y}) - Q(\mathbf{y}) - Q(\mathbf{x})).$$

Ejemplo 1.18-46. Calcule la segunda diferencial de $f: \mathbb{R}^3 \rightarrow \mathbb{R}$, $f(x_1, x_2, x_3) = x_1 x_2 - x_2 x_3 + x_3 x_1$ en el punto (a, b, c) .

Tenemos

$$f_{(h_1, h_2, h_3)}(t) = (a + h_1 t)(b + h_2 t) + (a + h_1 t)(c + h_3 t) - (b + h_2 t)(c + h_3 t)$$

Expandiendo en serie de Taylor a orden 2 se obtiene

$$\begin{aligned} f_{(h_1, h_2, h_3)}(t) &= (a(b+c) - bc) + (a(h_1 + h_3) + b(h_1 - h_3) + c(h_1 - h_2))t \\ &\quad + (h_1(h_2 + h_3) - h_2 h_3)t^2 + O(t^3) \end{aligned}$$

si denotamos por B la forma bilineal $D^2 f(a, b, c)$ y por Q la correspondiente forma cuadrática, entonces

$$Q(h_1, h_2, h_3) = 2(h_1(h_2 + h_3) - h_2 h_3)$$

luego

$$B((h_1, h_2, h_3), (h_1, h_2, h_3)) = (h_2 + h_3)k_1 + (h_1 - h_3)k_2 + (h_1 - h_2)k_3.$$

ahora podemos generalizar la

1.18.2 El teorema de Taylor

Las diferenciales de orden superior permiten aproximar los valores de una función en un entorno de un punto \mathbf{x}_0 y es la generalización para funciones de una variable real.

Teorema 16 (Teorema de Taylor). *Suponga que $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ tiene diferenciales continuas de todos los órdenes hasta el k y que además posea diferencial de orden $k + 1$ en un disco $B_r(\mathbf{x}_0)$. Entonces*

$$\begin{aligned} f(\mathbf{x}_0 + \mathbf{h}) = & \\ & f(\mathbf{x}_0) + Df(\mathbf{x}_0) + \frac{1}{2!}D^2f(\mathbf{x}_0) \cdot (\mathbf{h}, \mathbf{h}) + \frac{1}{3!}D^3f(\mathbf{x}_0) \cdot (\mathbf{h}, \mathbf{h}, \mathbf{h}) \\ & + \cdots + \frac{1}{k!}D^k f(\mathbf{x}_0) \cdot (\mathbf{h}, \mathbf{h}, \dots, \mathbf{h}) + R_k \end{aligned}$$

donde el residuo R_k depende de \mathbf{x}_0 y \mathbf{h} . Específicamente

$$R_k(\mathbf{x}_0, \mathbf{h}) = \frac{1}{(k+1)!}D^{k+1}f(\mathbf{x}_0^*) \cdot (\mathbf{h}, \mathbf{h}, \dots, \mathbf{h})$$

para algún punto \mathbf{x}_0^* contenido en el segmento que une \mathbf{x}_0 con \mathbf{h} .

1.19 Máximos y mínimos de funciones de varias variables

En esta sección abordaremos el problema de determinar cuándo una función real de varias variables posee un mínimo o un máximo (valor extremo). Obviamente habremos de considerar funciones a valores reales pues habremos de comparar distintos valores de la función, cosa que es imposible en funciones a valores vectoriales. Existe sin embargo un concepto de optimalidad que en cierta forma balancea valores antagónicos llamada optimización de Pareto pero que no habremos de considerar en estas notas, sino de pasada.

Para funciones de una variable real, (a) un punto extremo en un punto interior del dominio es un punto crítico, i.e. la primera derivada se anula, además, tratándose de un punto crítico, existen dos criterios para determinar si tal punto crítico es extremo: (b) el cambio de signo de la primera derivada o, (c) el signo de la segunda derivada en el punto crítico, si esta existe.

Para funciones de dos o más variables, el criterio (a) se generaliza para más variables.

Definición 17. *Sea $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$ decimos que f tiene un máximo (resp. mínimo) local en \mathbf{x}_0 si $f(\mathbf{x}_0) \geq f(\mathbf{x})$ (resp. $f(\mathbf{x}_0) \leq f(\mathbf{x})$) para toda \mathbf{x} en un disco $B_r(\mathbf{x}_0) \subset \mathbb{R}^n$.*

Ejemplo 1.19–47. La función lineal $f(x) = x + 1$, $x \in [0, 1]$ tiene un mínimo y un máximo en los extremos del intervalo y la derivada es siempre positiva. En cambio la función $f(x) = x(1 - x)$ tiene un máximo en un punto interior $x = 1/2$. Observe que en este caso $f'(1/2) = 0$.

Definición 18. Un punto $\mathbf{x} \in D \subset \mathbb{R}^n$ se dice interior, si existe un disco $B_r(\mathbf{x}) \subset D$.

Proposición 17. Sea $f: U \rightarrow \mathbb{R}$ una función diferenciable en $U \subset \mathbb{R}^n$. Si f tiene un extremo en un punto interior de U , \mathbf{x}_0 , entonces $Df(\mathbf{x}_0) = 0$.

Demostración.

Sea r tal $B_r(\mathbf{x}) \subset U$ y sea $\hat{\mathbf{u}}$ un vector unitario cualquiera. El segmento $\mathbf{r}(t) = \mathbf{x}_0 + t\hat{\mathbf{u}}$ está contenido en D para $|t|$ suficientemente pequeño, pues \mathbf{x}_0 es punto interior. De la hipótesis se sigue que la función real de una variable real $\phi(t) = f(\mathbf{r}(t))$ tiene un máximo en $t = 0$, por lo que su derivada se anula en $t = 0$. Por la regla de la cadena:

$$0 = \phi'(0) = Df(\mathbf{r}(0)) \cdot \mathbf{r}'(0) = Df(\mathbf{x}_0) \cdot \hat{\mathbf{u}}$$

y como $\hat{\mathbf{u}}$ es arbitrario, entonces $Df(\mathbf{x}_0) = 0$ □

El teorema anterior nos da *condiciones necesarias* para la existencia de un máximo (o mínimo) en un punto interior, lo que queremos es un *criterio suficiente* para determinar si un punto crítico es un máximo o mínimo.

Para comenzar, la situación es un poco más complicada que para funciones de una variable real pues a partir de un punto \mathbf{x}_0 , la función puede aumentar en unas direcciones o disminuir en otras o de plano mantenerse constante. En la Figura 1.22 se muestran las gráficas de dos funciones $z = f(x, y)$ en un entorno del punto crítico $(0, 0)$.

1.19.1 Clasificación de formas cuadráticas

Comencemos por clasificar funciones $f: \mathbb{R}^n \rightarrow \mathbb{R}$ de orden cuadrático con un punto crítico en el origen, vgr.,

$$f(\mathbf{x}) = A\mathbf{x} + \mathcal{B}(\mathbf{x}, \mathbf{x})$$

donde \mathcal{B} es una función bilineal y por lo tanto $\mathcal{B}(\mathbf{h}, \mathbf{h}) = O(\mathbf{h})$. Por definición $A = Df(\mathbf{0}) = 0$, así que estamos considerando una forma cuadrática $f(\mathbf{x}) =$

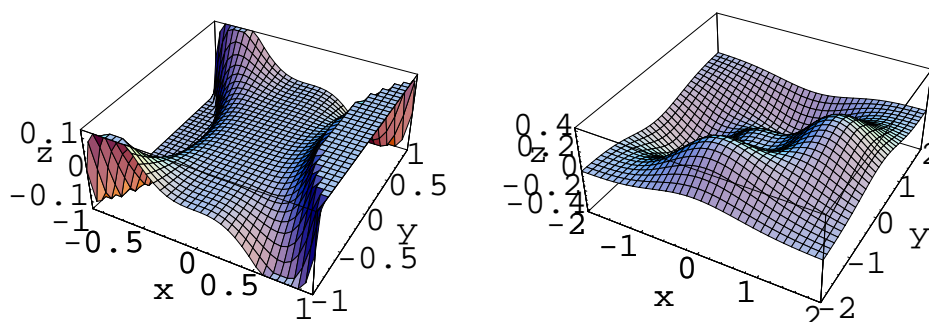


Figure 1.22: Puntos críticos de funciones $z = f(x, y)$.

$\mathcal{B}(\mathbf{x}, \mathbf{x})$. En términos de la base canónica δ_i , $i = 1, 2, \dots, n$ podemos escribir por bilinealidad

$$\mathcal{B}(\mathbf{x}, \mathbf{x}) = \sum_{i,j=1}^n x_i x_j \mathcal{B}(\delta_i, \delta_j) \equiv \sum_{i=1}^n x_i \left(\sum_{j=1}^n B_{ij} x_j \right)$$

donde la matrix de coeficientes $B = (B_{ij})$ es simétrica. Si los elementos $\mathbf{x} \in \mathbb{R}^n$ se representan por vectores columna, entonces se puede escribir en forma compacta

$$\mathcal{B}(\mathbf{x}, \mathbf{x}) = \mathbf{x}^T B \mathbf{x}.$$

El siguiente resultado de álgebra lineal será muy útil para nuestro análisis.

Teorema 17. *Sea $B \in \mathbb{R}^{n \times n}$ una matrix simétrica. Entonces existe una matrix ortogonal $Q \in \mathbb{R}^{n \times n}$ tal que*

$$Q^T B Q = D$$

donde D es una matrix diagonal. Las columnas de Q forman un conjunto ortonormal de vectores propios de B .

Ejemplo 1.19-48. La matrix simétrica

$$B = \begin{pmatrix} 2 & \sqrt{2} \\ \sqrt{2} & 1 \end{pmatrix}$$

se diagonaliza mediante la matriz ortogonal

$$Q = \frac{1}{\sqrt{3}} \begin{pmatrix} -1 & \sqrt{2} \\ \sqrt{2} & 1 \end{pmatrix}$$

en efecto

$$\begin{aligned} Q^T B Q &= \frac{1}{3} \begin{pmatrix} -1 & \sqrt{2} \\ \sqrt{2} & 1 \end{pmatrix} \begin{pmatrix} 2 & \sqrt{2} \\ \sqrt{2} & 1 \end{pmatrix} \begin{pmatrix} -1 & \sqrt{2} \\ \sqrt{2} & 1 \end{pmatrix} \\ &= \frac{1}{3} \begin{pmatrix} -1 & \sqrt{2} \\ \sqrt{2} & 1\sqrt{2} \end{pmatrix} \begin{pmatrix} 0 & 3\sqrt{2} \\ 0 & 3 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 \\ 0 & 3 \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \end{aligned}$$

Por supuesto, si conocemos solo la matriz simétrica la pregunta es cómo calcular Q . El siguiente resultado muestra que vectores propios correspondientes a valores propios distintos son ortogonales:

Proposición 18. *Sea B simétrica. Si λ_1, \mathbf{e}_1 , λ_2, \mathbf{e}_2 son parejas de valores y vectores propios asociadas a valores propios distintos, entonces \mathbf{e}_1 y \mathbf{e}_2 son ortogonales.*

Demostración.

Multipliquemos las ecuaciones

$$\begin{aligned} B\mathbf{e}_1 &= \lambda_1 \mathbf{e}_1, \\ B\mathbf{e}_2 &= \lambda_2 \mathbf{e}_2, \end{aligned}$$

la primera escalarmente por \mathbf{e}_2 y la segunda por \mathbf{e}_1 ; al restar se obtiene

$$\mathbf{e}_2^T B \mathbf{e}_1 - \mathbf{e}_1^T B \mathbf{e}_2 = (\lambda_1 - \lambda_2) \mathbf{e}_2^T \mathbf{e}_1$$

El lado izquierdo es cero, por simetría, luego si $\lambda_1 \neq \lambda_2$ entonces $\mathbf{e}_2^T \mathbf{e}_1 = 0$. \square

En otras palabras, si B tiene n valores propios distintos λ_i , $i = 1, 2, \dots, n$, entonces la base de vectores propios se obtiene resolviendo los sistemas de ecuaciones

$$B\mathbf{e}_i = \lambda_i \mathbf{e}_i.$$

Ejemplo 1.19–49. Diagonalice la matriz simétrica

$$\begin{pmatrix} 3 & 0 & 0 \\ 0 & 4 & \sqrt{3} \\ 0 & \sqrt{3} & 6 \end{pmatrix}$$

El polinomio característico $p(\lambda) = (\lambda - 3)^2(7 - \lambda)$ tiene $\lambda_{1,2} = 3$ como raíz (doble) y $\lambda_3 = 7$. No es difícil ver que para $\lambda_3 = 7$ el sistema lineal a resolver es

$$\begin{pmatrix} -4 & 0 & 0 \\ 0 & -3 & \sqrt{3} \\ 0 & \sqrt{3} & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Claramente la segunda y tercera ecuación son linealmente dependientes así que el sistema a resolver se reduce a

$$\begin{aligned} -4x &= 0 \\ -3y + \sqrt{3}z &= 0 \end{aligned}$$

cuya solución general es

$$\left(0, \frac{z}{\sqrt{3}}, z \right)$$

tomando $z = \sqrt{3}/2$ obtenemos un valor propio unitario

$$\mathbf{e}_3 = \left(0, \frac{1}{2}, \frac{\sqrt{3}}{2} \right).$$

Para $\lambda_{1,2} = 3$ el sistema

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & \sqrt{3} \\ 0 & \sqrt{3} & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

se reduce a una sola ecuación independiente

$$y + \frac{z}{\sqrt{3}} = 0$$

cuya solución general depende de dos parámetros, x y z :

$$(x, -\sqrt{3}z, z)$$

Tomando $x = 1, z = 0$ obtenemos el vector propio unitario

$$\mathbf{e}_1 = (1, 0, 0).$$

haciendo $x = 0, z = 1$ otro vector propio *linealmente independiente del primero*

$$(0, -\sqrt{3}, 1).$$

un vector unitario es

$$\mathbf{e}_2 = \left(0, \frac{\sqrt{3}}{2}, -\frac{1}{2}\right).$$

En resumen, la matriz ortogonal que diagonaliza la forma cuadrática es

$$Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{\sqrt{3}}{2} & \frac{1}{2} \\ 0 & -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}$$

y se comprueba directamente que

$$Q^T B Q = \text{diag}(3, 3, 7).$$

Observación 16. Aún cuando haya valores propios repetidos, digamos que λ_k tenga multiplicidad $k < n$, el sistema lineal

$$(A - \lambda_k I)\mathbf{x} = \mathbf{0}$$

tiene rango k , es decir existen k soluciones linealmente independientes. Por el proceso de ortogonalización de Gram-Schmidt se pueden obtener k vectores propios asociados al *mismo valor propio* que sean ortogonales entre sí, $\mathbf{e}_1^{(k)}, \mathbf{e}_2^{(k)}, \dots, \mathbf{e}_k^{(k)}$. Estos vectores propios se completan con el resto hasta completar una base.

Definición 19. Una matriz simétrica B se dice:

1. *Definida, si*

$$\mathbf{x}^T B \mathbf{x} \neq 0$$

para toda $\mathbf{x} \neq \mathbf{0}$;

2. *No negativa, si*

$$\mathbf{x}^T B \mathbf{x} \geq 0;$$

3. Positiva definida, si

$$\mathbf{x}^T B \mathbf{x} > 0$$

y $\mathbf{x}^T B \mathbf{x} = 0$, solo si, $\mathbf{x} = 0$;

4. Positiva (negativa) semidefinida, si es no negativa (no positiva), y para cada \mathbf{x} existe \mathbf{x}' tal que

$$\mathbf{x}'^T B \mathbf{x}' = 0.$$

El siguiente criterio permite decidir si una forma cuadrática es definida positiva o no.

Teorema 18. Una matriz simétrica B es:

1. Definida si y solo si todos los valores propios son distintos de cero.
2. Positiva (negativa) definida si y solo si todos los valores propios son positivos (negativos).
3. Positiva semidefinida si y solo si todos los valores propios son no negativos y al menos uno es cero.

Ejemplo 1.19–50.

La matriz asociada a la forma cuadrática

$$\mathcal{B}(\mathbf{x}, \mathbf{x}) = 2x_1^2 + 2\sqrt{2}x_1x_2 + x_2^2$$

es

$$B = \begin{pmatrix} 2 & \sqrt{2} \\ \sqrt{2} & 1 \end{pmatrix}$$

cuyo polinomio característico es

$$p(\lambda) = \lambda(\lambda - 3)$$

de donde los valores propios son $\lambda_1 = 0$ y $\lambda_2 = 3$. Se sigue que la forma cuadrática es positiva semidefinida. El lector puede comprobar la matriz

$$Q = \frac{1}{\sqrt{3}} \begin{pmatrix} -1 & \sqrt{2} \\ \sqrt{2} & 1 \end{pmatrix}$$

reduce la matriz B a su forma diagonal

$$D = Q^T B Q = \text{diag}(0, 3)$$

por lo tanto si en la forma bilineal

$$\mathcal{B}(\mathbf{x}, \mathbf{x}) = \mathbf{x}^T Q \mathbf{x}$$

se efectúa el cambio de variable $\mathbf{x} = Q\mathbf{y}$, entonces la forma cuadrática $\mathcal{B}(\mathbf{x}, \mathbf{x})$ se transforma en la forma cuadrática

$$\mathcal{B}'(\mathbf{y}, \mathbf{y}) = \mathbf{x}^T Q \mathbf{x} (Q\mathbf{y})^T Q (Q\mathbf{y}) = \mathbf{y}^T (Q^T B Q) \mathbf{y} = \mathbf{y}^T D \mathbf{y} = 3y_2^2.$$

1.19.2 Aplicación a la determinación de máximos y mínimos

Nuestro interés en diagonalizar una forma cuadrática es que la forma diagonal es tan más simple que se puede determinar el carácter de máximo o mínimo.

Teorema 19. *Sea $f: \mathbb{R}^n \rightarrow \mathbb{R}$ una función diferenciable en $B_r(\mathbf{x}_0) \subset \mathbb{R}^n$ que posee diferencial de orden dos en \mathbf{x}_0 . Sea B la matriz asociada a la forma cuadrática $D^2 f(\mathbf{x}_0)$. Entonces*

1. *Si B es positiva definida, entonces \mathbf{x}_0 es un máximo local.*
2. *Si B es negativa definida, entonces \mathbf{x}_0 es un mínimo local.*
3. *Si B es definida, pero no es negativa ni positiva, entonces \mathbf{x}_0 es un punto silla*