Capítulo 4

La derivada de funciones de $\mathbb{R}^n$ en $\mathbb{R}$

El contenido de este capítulo será sin duda el primero en el que aparecerán conceptos realmente novedosos para el lector. Y aun cuando los primeros conceptos de derivación que definiremos para este tipo de funciones son muy "parecidos" al concepto de derivada de funciones de $\mathbb{R}$ en $\mathbb{R}$, la definición general de la derivada de funciones de $\mathbb{R}^n$ en $\mathbb{R}$ abrirá un camino hacia una visión más general de este concepto.

Antes de iniciar nuestro recorrido por estos temas, vamos a revisar algunos aspectos relacionados con los conjuntos en donde se encuentran las variables de las funciones con las que vamos a trabajar y la relación entre las varias representaciones que podemos hacer de éstos por medio del conjunto $\mathbb{R}^n$.

4.1 Un interludio de Álgebra Lineal

Como vimos en el capítulo 1, los objetos matemáticos que sirven para representar a las variables (independientes o dependientes) de las funciones que nos ocupan pertenecen a conjuntos que suelen estar dotados de una estructura algebraica que hace de ellos un cierto tipo de espacio que se conocen con el nombre de *espacios vectoriales de dimensión finita sobre los números reales*.

Como ya también vimos, una de las características más importantes de este tipo de espacios con el que vamos a trabajar, y que por ahora denotaremos en general con la letra V, es que siempre hay varias maneras de elegir una colección finita de elementos de V, digamos $\{\hat{v}_1, \ldots, \hat{v}_n\}$, con la propiedad de que para cualquier otro elemento $\hat{v} \in V$ existen $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$, únicos, tales que

$$\hat{v} = \lambda_1 \hat{v}_1 + \cdots + \lambda_n \hat{v}_n.$$

Como mencionamos en su momento (y como seguramente el lector ya sabe), este tipo de colecciones reciben el nombre de *base de* V, y la propiedad que las define es la que permite identificar al conjunto V con el conjunto $\mathbb{R}^n$. Esta identificación de V con $\mathbb{R}^n$ tiene la ventaja de que hace corresponder adecuadamente la estructura algebraica de V con la estructura algebraica que definimos para $\mathbb{R}^n$, estructura que por cierto, ¡convierte a $\mathbb{R}^n$ en un espacio vectorial!

Además de lo anterior, en el capítulo 1 también vimos algunos ejemplos específicos de espacios vectoriales, como es el caso de las flechas (del plano o el espacio) que parten de un punto fijo. Estos espacios particulares están dotados de estructuras geométricas (tales como el concepto de distancia o de ángulo), las cuales pudimos "trasladar" a $\mathbb{R}^n$, y consecuentemente a cualquier otro espacio vectorial que se pueda identificar con éste.

Observemos que si $\{\hat{v}_1, \ldots, \hat{v}_n\}$ es una base del espacio vectorial V, la $n-$ada con la que se identifica al vector $\hat{v}_i$ es aquella que tiene casi todas sus coordenadas 0, salvo por la $i-$ésima, que deberá ser 1; a esta $n-$ada la denotaremos por $\hat{e}_i$. Es decir que

$$\hat{e}_i := (0, \ldots, 0, 1, 0, \ldots, 0),$$

en donde el 1 está en la coordenada i.

Como seguramente el lector ya sabe (y si no, lo podrá probar muy fácilmente), la colección $\{\hat{e}_1, \ldots, \hat{e}_n\}$ es una base para $\mathbb{R}^n$ y se le conoce con el nombre de *base canónica*.

De acuerdo con los conceptos de magnitud (norma) y de ángulo (producto punto) que definimos en el capítulo 1, todos los elementos de esta base tienen magnitud 1 y son mutuamente perpendiculares. Es decir, $\|\hat{e}_i\| = 1$ para cada $i \in \{1, \dots, n\}$ y $\hat{e}_i \cdot \hat{e}_j = 0$ para cada $i, j \in \{1, \dots, n\}$, con $i \neq j$, o lo que es lo mismo,

$$\hat{e}_i \cdot \hat{e}_j = \begin{cases} 1 & \text{si } i = j \\ \\ 0 & \text{si } i \neq j \end{cases}$$

para cada $i, j \in \{1, \dots, n\}$. En general, cuando los elementos de una base de un espacio vectorial se indentifican con $n-$adas que cumplen con estas dos características, decimos que esta base es *ortonormal*.

Como ya habíamos mencionado en el capítulo 1, este tipo de bases son las que se utilizan para construir (en el caso del plano o del espacio) un sistema de referencia cartesiano. Es por esta razón que en estos casos, cuando se dibujen los vectores (o flechas) que representen a una base de este tipo, habrá que hacerlo con estas características (mutuamente perpendiculares, y de la misma longitud). No sin cierto abuso de notación, usaremos los mismos $\hat{e}_1, \dots, \hat{e}_n$ para nombrar a los vectores (o flechas) que representen geométricamente a la base canónica de $\mathbb{R}^n$.

Una vez establecidos los conceptos y notaciones anteriores, la situación principal que deseamos abordar en esta sección es la siguiente: supongamos que la variable independiente de una cierta función pertenece a un espacio vectorial V, y que las colecciones $\{\hat{v}_1, \dots, \hat{v}_n\} \subset V$ y $\{\hat{v}'_1, \dots, \hat{v}'_n\} \subset V$, ambas, son bases de V. Si identificamos al espacio V con $\mathbb{R}^n$ a través de la base $\{\hat{v}_1, \dots, \hat{v}_n\}$ (en cuyo caso cada vector $\hat{v}_i$ se corresponderá con la n-ada $\hat{e}_i$), entonces, usando esta misma manera de identificar a V con $\mathbb{R}^n$, cada vector $\hat{v}'_i$ se corresponderá con alguna n-ada, que denotaremos por $\hat{e}'_i$, y se tendrá que la colección de n-adas $\{\hat{e}'_1, \dots, \hat{e}'_n\}$ será una base (o sistema coordenado) de $\mathbb{R}^n$ (afirmación que seguramente el lector ya habrá probado en su curso de Álgebra Lineal).

Algo que es importante destacar es que, aun cuando la colección $\{\hat{e}'_1, \dots, \hat{e}'_n\}$ sea una base de $\mathbb{R}^n$, ésta no tiene que ser necesariamente ortonormal. *Para los fines de este texto, supondremos que las colecciones* $\{\hat{v}_1, \dots, \hat{v}_n\} \subset V$ *y* $\{\hat{v}'_1, \dots, \hat{v}'_n\} \subset V$ *se elegirán de tal manera que la colección* $\{\hat{e}'_1, \dots, \hat{e}'_n\}$ *siempre sea una base ortonormal (o sistema coordenado cartesiano) de* $\mathbb{R}^n$.

De esta forma, dadas las colecciones $\{\hat{v}_1, \dots, \hat{v}_n\} \subset V$ y $\{\hat{v}'_1, \dots, \hat{v}'_n\} \subset V$, la situación que tendremos (como ya habíamos mencionado desde el capítulo 1) es que habrá dos formas diferentes de identificar al espacio vectorial V con el conjunto (o espacio vectorial) $\mathbb{R}^n$. Esta doble identificación se traducirá en lo siguiente: así como para cada $\hat{v} \in V$ existirán dos formas de expresar a $\hat{v}$ como combinación lineal de las bases $\{\hat{v}_1, \dots, \hat{v}_n\}$ y $\{\hat{v}'_1, \dots, \hat{v}'_n\}$, para cada $\hat{x} \in \mathbb{R}^n$ (pensando a $\mathbb{R}^n$ más como un "representante" del espacio vectorial V, que como un conjunto de n-adas), consideraremos dos sistemas coordenados cartesianos en los cuales expresaremos a $\hat{x}$. Es decir, un vector $\hat{x} \in \mathbb{R}^n$ tendrá coordenadas $(x_1, \dots, x_n)$ en el sistema coordenado cartesiano determinado por la base canónica $\{\hat{e}_1, \dots, \hat{e}_n\}$, y coordenadas $(x'_1, \dots, x'_n)$ en el otro sistema coordenado cartesiano determinado por la base ortonormal $\{\hat{e}'_1, \dots, \hat{e}'_n\}$. La figura 4.1 ilustra esta situación para el caso particular del espacio vectorial V formado por las flechas que parten de un punto fijo O, razón por la cual las bases $\{\hat{e}_1, \hat{e}_2\}$ y $\{\hat{e}'_1, \hat{e}'_2\}$ se "fijan" en el mismo punto.

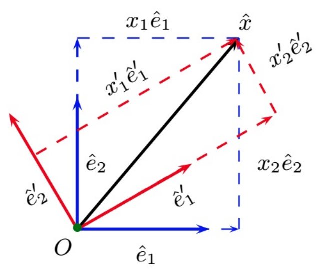


Figura 4.1: El vector $\hat{x}$ y sus correspondientes coordenadas (x_1, x_2) y (x'_1, x'_2) en las bases ortonormales $\{\hat{e}_1, \hat{e}_2\}$ y $\{\hat{e}'_1, \hat{e}'_2\}$, respectivamente.

En virtud de lo anterior, tendremos que una misma función f definida para los elementos de un subconjunto A del espacio vectorial V (y que en un abuso de lenguaje diremos que está definida para los elementos de un subconjunto A del espacio vectorial $\mathbb{R}^n$), ésta se podrá poner en términos de unas ciertas variables (o coordenadas) $x_1, \ldots, x_n$, o de otras variables (o coordenadas) $x'_1, \ldots, x'_n$.

Aun cuando los conceptos de derivabilidad con los que vamos a trabajar en este capítulo serán definidos sin tener que recurrir a las coordenadas determinadas por alguna base, para efectos de la realización de cálculos específicos sí será necesario recurrir a alguna de éstas. Lo que nos proponemos en este capítulo es dejar clara la relación que existirá entre los cálculos realizados con unas cooredenadas $x_1, \ldots, x_n$ o con otras coordenadas $x'_1, \ldots, x'_n$. Para alcanzar este objetivo, empezaremos por establecer la relación que existe entre dos conjuntos de coordenadas (cartesianas) de un mismo elemento $\hat{x} \in \mathbb{R}^n$.

Si, como dijimos antes, $(x_1, \ldots, x_n)$ son las coordenadas de $\hat{x}$ en el sistema coordenado cartesiano determinado por la base canónica $\{\hat{e}_1, \ldots, \hat{e}_n\}$, y $(x'_1, \ldots, x'_n)$ son otras coordenadas en un sistema coordenado cartesiano determinado por una base ortonormal $\{\hat{e}'_1, \ldots, \hat{e}'_n\}$, nuestro objetivo es mostrar cómo se obtienen unas coordenadas a partir de las otras; es decir, si conocemos las coordenadas $(x_1, \ldots, x_n)$, cómo podemos obtener las coordenadas $(x'_1, \ldots, x'_n)$ y recíprocamente, si conocemos las coordenadas $(x'_1, \ldots, x'_n)$, cómo podemos obtener las coordenadas $(x_1, \ldots, x_n)$.

Como seguramente el lector ya sabrá, cada uno de estos problemas es un problema de *cambio de coordenadas*. También sabrá que se resuelve de la siguiente manera: supongamos primero que tenemos las coordenadas $(x'_1, \ldots, x'_n)$ de $\hat{x} \in \mathbb{R}^n$ en la base ortonormal $\{\hat{e}'_1, \ldots, \hat{e}'_n\}$ y que deseamos conocer sus coordenadas $(x_1, \ldots, x_n)$ en la base canónica. Para ello, bastará con saber cuáles son las coordenadas de cada vector $\hat{e}'_i$ en la base canónica $\{\hat{e}_1, \ldots, \hat{e}_n\}$ (información con la que se cuenta la mayoría de las veces).

Si suponemos que

$$\begin{aligned} \hat{e}'_i &= \left(a_1^{(i)}, \ldots, a_n^{(i)}\right) \\ &= a_1^{(i)}\hat{e}_1 + \cdots + a_n^{(i)}\hat{e}_n \end{aligned}$$

para cada $i \in \{1, \ldots, n\}$, entonces

$$\begin{aligned} \hat{x} &= x'_1\hat{e}'_1 + \cdots + x'_n\hat{e}'_n \\ &= x'_1\left(a_1^{(1)}\hat{e}_1 + \cdots + a_n^{(1)}\hat{e}_n\right) + \\ &\vdots \\ &+ x'_n\left(a_1^{(n)}\hat{e}_1 + \cdots + a_n^{(n)}\hat{e}_n\right) \\ &= \left(a_1^{(1)}, \ldots, a_1^{(n)}\right) \cdot (x'_1, \ldots, x'_n)\hat{e}_1 + \\ &\vdots \\ &+ \left(a_n^{(1)}, \ldots, a_n^{(n)}\right) \cdot (x'_1, \ldots, x'_n)\hat{e}_n, \end{aligned}$$

de donde concluimos que la i-ésima coordenada del vector $\hat{x}$ en la base canónica estará dada por

$$x_i = \left(a_i^{(1)}, \ldots, a_i^{(n)}\right) \cdot (x'_1, \ldots, x'_n)$$

para cada $i \in \{1, \ldots, n\}$, o equivalentemente y escrito usando matrices, que

$$\begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} = \begin{bmatrix} x'_1 & \cdots & x'_n \end{bmatrix} \begin{bmatrix} a_1^{(1)} & \cdots & a_n^{(1)} \\ \vdots & \ddots & \vdots \\ a_1^{(n)} & \cdots & a_n^{(n)} \end{bmatrix}. \tag{4.1}$$

Como el lector se podrá imaginar, para resolver el problema recíproco (obtener las coordenadas $(x'_1, \ldots, x'_n)$ a partir de las coordenadas $(x_1, \ldots, x_n)$), será suficiente con calcular las coordenadas de cada vector $\hat{e}_i$ en el sistema cartesiano determinado por la base $\{\hat{e}'_1, \ldots, \hat{e}'_n\}$.

Es decir, si ahora

$$\hat{e}_i = \left(b_1^{(i)}, \ldots, b_n^{(i)}\right) = b_1^{(i)}\hat{e}'_1 + \cdots + b_n^{(i)}\hat{e}'_n$$

para cada $i \in \{1, \ldots, n\}$, entonces

$$\begin{aligned}
\hat{x} &= x_1 \hat{e}_1 + \cdots + x_n \hat{e}_n \\
&= x_1 \left(b_1^{(1)} \hat{e}'_1 + \cdots + b_n^{(1)} \hat{e}'_n \right) + \\
&\vdots \\
&+ x_n \left(b_1^{(n)} \hat{e}'_1 + \cdots + b_n^{(n)} \hat{e}'_n \right) \\
&= \left(b_1^{(1)}, \ldots, b_1^{(n)} \right) \cdot (x_1, \ldots, x_n) \hat{e}'_1 + \\
&\vdots \\
&+ \left(b_n^{(1)}, \ldots, b_n^{(n)} \right) \cdot (x_1, \ldots, x_n) \hat{e}'_n,
\end{aligned}$$

de donde concluimos que la $i-$ésima coordenada del vector $\hat{x}$ en la base $\{\hat{e}'_1, \ldots, \hat{e}'_n\}$ estará dada por

$$x'_i = \left(b_i^{(1)}, \ldots, b_i^{(n)} \right) \cdot (x_1, \ldots, x_n)$$

para cada $i \in \{1, \ldots, n\}$, o lo que es lo mismo, escrito en forma matricial, que

$$\begin{bmatrix} x'_1 & \cdots & x'_n \end{bmatrix} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} b_1^{(1)} & \cdots & b_n^{(1)} \\ \vdots & \ddots & \vdots \\ b_1^{(n)} & \cdots & b_n^{(n)} \end{bmatrix}. \tag{4.2}$$

Si llamamos

$$M = \begin{bmatrix} a_1^{(1)} & \cdots & a_n^{(1)} \\ \vdots & \ddots & \vdots \\ a_1^{(n)} & \cdots & a_n^{(n)} \end{bmatrix}$$

y

$$M' = \begin{bmatrix} b_1^{(1)} & \cdots & b_n^{(1)} \\ \vdots & \ddots & \vdots \\ b_1^{(n)} & \cdots & b_n^{(n)} \end{bmatrix}$$

se prueba que:

1. cada matriz M y M' es la inversa una de la otra ($M' = M^{-1}$),
2. la inversa de cada una de ellas es su propia transpuesta ($M^{-1} = M^t$), y
3. ambas matrices tienen determinante ± 1.

Nótese que de las propiedades 1 y 2 se deduce que

$$b_j^{(i)} = a_i^{(j)} \tag{4.3}$$

para todas $i, j \in \{1, \ldots, n\}$.

Las matrices que tienen las características anteriores son conocidas con el nombre de *matrices ortonormales* (por razones que seguramente quedan claras). Si el lector aún no conoce la prueba de estas afirmaciones, muy probablemente pronto las verá en su curso de Álgebra Lineal).

Resumiendo la discusión anterior, con base en las identidades 4.1, 4.2 y 4.3, concluimos que: si cada vector $\hat{e}'_i$ tiene coordenadas $\left(a_1^{(i)}, \ldots, a_n^{(i)} \right)$ ($i \in \{1, \ldots, n\}$) en el sistema determinado por la base $\{\hat{e}_1, \ldots, \hat{e}_n\}$, y un vector $\hat{x}$ tiene coordenadas $(x_1,\ldots,x_n)$ en el mismo sistema, entonces las coordenadas $(x'_1,\ldots,x'_n)$ de $\hat{x}$ (en el sistema determinado por la base $\{\hat{e}'_1,\ldots,\hat{e}'_n\}$) están dadas por la identidad (escrita en forma matricial)

$$\begin{bmatrix} x'_1 & \cdots & x'_n \end{bmatrix} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} a_1^{(1)} & \cdots & a_1^{(n)} \\ \vdots & \ddots & \vdots \\ a_n^{(1)} & \cdots & a_n^{(n)} \end{bmatrix}. \tag{4.4}$$

Análogamente, si cada vector $\hat{e}_i$ tiene coordenadas $\left(b_1^{(i)},\ldots,b_n^{(i)}\right)$ $(i\in\{1,\ldots,n\})$ en el sistema determinado por la base $\{\hat{e}'_1,\ldots,\hat{e}'_n\}$, y un vector $\hat{x}$ tiene coordenadas $(x'_1,\ldots,x'_n)$ en el mismo sistema, entonces las coordenadas $(x_1,\ldots,x_n)$ de $\hat{x}$ (en el sistema determinado por la base $\{\hat{e}_1,\ldots,\hat{e}_n\}$) están dadas por la identidad (escrita en forma matricial)

$$\begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} = \begin{bmatrix} x'_1 & \cdots & x'_n \end{bmatrix} \begin{bmatrix} b_1^{(1)} & \cdots & b_1^{(n)} \\ \vdots & \ddots & \vdots \\ b_n^{(1)} & \cdots & b_n^{(n)} \end{bmatrix}. \tag{4.5}$$

Nótese que, de acuerdo con las identidades 4.3, también se tiene que

$$\begin{bmatrix} x'_1 & \cdots & x'_n \end{bmatrix} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} b_1^{(1)} & \cdots & b_n^{(1)} \\ \vdots & \ddots & \vdots \\ b_1^{(n)} & \cdots & b_n^{(n)} \end{bmatrix} \tag{4.6}$$

y

$$\begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} = \begin{bmatrix} x'_1 & \cdots & x'_n \end{bmatrix} \begin{bmatrix} a_1^{(1)} & \cdots & a_n^{(1)} \\ \vdots & \ddots & \vdots \\ a_1^{(n)} & \cdots & a_n^{(n)} \end{bmatrix}, \tag{4.7}$$

identidades que también nos serán muy útiles.

Una consecuencia muy importante de las identidades anteriores tiene que ver con los conceptos de producto punto y norma euclideana, los cuales se definieron en el capítulo 1 en términos de coordenadas. Como se recordará, si $(x_1,\ldots,x_n)$ y $(y_1,\ldots,y_n)$ son las coordenadas de $\hat{x}$ y $\hat{y}$ respectivamente, en el sistema cartesiano determinado por la base canónica $\{\hat{e}_1,\ldots,\hat{e}_n\}$, definimos que

$$\hat{x}\cdot\hat{y} = x_1y_1+\cdots+x_ny_n = \sum_{i=1}^{n} x_iy_i.$$

Si ahora suponemos que $(x'_1,\ldots,x'_n)$ y $(y'_1,\ldots,y'_n)$ son las coordenadas de $\hat{x}$ y $\hat{y}$ respectivamente, en un sistema coordenado cartesiano determinado por otra base ortonormal $\{\hat{e}'_1,\ldots,\hat{e}'_n\}$, de acuerdo con la identidad 4.4 (y denotando por Id_n a la matriz identidad de $n\times n$) se tiene que

$$\begin{aligned}
[x'_1y'_1+\cdots+x'_ny'_n] &= \begin{bmatrix} x'_1 & \cdots & x'_n \end{bmatrix} \begin{bmatrix} y'_1 \\ \vdots \\ y'_n \end{bmatrix} \\
&= \begin{bmatrix} x'_1 & \cdots & x'_n \end{bmatrix} \begin{bmatrix} y'_1 & \cdots & y'_n \end{bmatrix}^t \\
&= \left(\begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} M^t\right)\left(\begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix} M^t\right)^t \\
&= \left(\begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} M^t\right)\left(\left(M^t\right)^t \begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix}^t\right) \\
&= \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}\left(M^tM\right)\begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix}^t \\
&= \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}\left(M^{-1}M\right)\begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix}^t
\end{aligned}$$

$$
\begin{aligned}
&= \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} (Id_n) \begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix}^t \\
&= \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix}^t \\
&= \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \\
&= [x_1 y_1 + \cdots + x_n y_n].
\end{aligned}
$$

Por lo tanto, se tiene que

$$x_1' y_1' + \cdots + x_n' y_n' = x_1 y_1 + \cdots + x_n y_n.$$

De esta última identidad concluimos que el valor de $\hat{x} \cdot \hat{y}$ no depende del sistema coordenado cartesiano que se esté usando para identificar a los vectores $\hat{x}$ y $\hat{y}$, y si recordamos que $\|\hat{x}\| = \sqrt{\hat{x} \cdot \hat{x}}$, también concluimos que el valor de la norma de un vector $\hat{x} \in \mathbb{R}^n$ no depende de las coordenadas (cartesianas) que se usen. Por ahora nos será suficiente con saber lo anterior y conocer las relaciones establecidas en 4.4, 4.5, 4.6 y 4.7.

Para concluir esta sección, daremos un

Ejemplo 4.1 *En el plano, una vez establecido un origen O, considere el sistema coordenado cartesiano determinado por dos vectores $\{\hat{e}_1, \hat{e}_2\}$ (y que por lo mismo tendremos que dibujar mutuamente perpendiculares y de la misma longitud), y el sistema cartesiano determinado por los vectores $\{\hat{e}_\theta, \hat{e}_\theta^\perp\}$, en donde $\hat{e}_\theta$ se obtiene de girar (en el sentido contrario a las manecillas del reloj) θ radianes al vector $\hat{e}_1$, y $\hat{e}_\theta^\perp$ se obtiene de girar (en el sentido contrario a las manecillas del reloj) $\pi/2$ radianes al vector $\hat{e}_\theta$ (ver figura 4.2).*

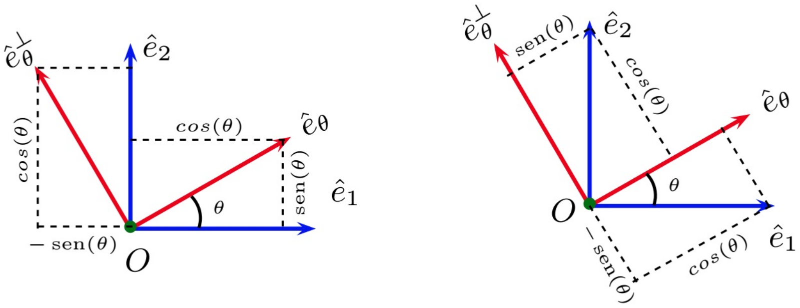


Figura 4.2: Las bases ortonormales $\{\hat{e}_1, \hat{e}_2\}$ y $\{\hat{e}_\theta, \hat{e}_\theta^\perp\}$ del ejemplo 4.1.

Por la manera en que construimos los vectores $\hat{e}_\theta$ y $\hat{e}_\theta^\perp$, sabemos que éstos forman una base ortonormal y que sus coordenadas en el sistema determinado por $\{\hat{e}_1, \hat{e}_2\}$ son $(\cos(\theta), \text{sen}(\theta))$ para $\hat{e}_\theta$, y $(-\text{sen}(\theta), \cos(\theta))$ para $\hat{e}_\theta^\perp$. Es decir, se tiene que

$$
\begin{aligned}
\hat{e}_\theta &= \cos(\theta)\hat{e}_1 + \text{sen}(\theta)\hat{e}_2 \qquad (4.8)\\
\hat{e}_\theta^\perp &= -\text{sen}(\theta)\hat{e}_1 + \cos(\theta)\hat{e}_2.
\end{aligned}
$$

De esta forma, si $\hat{x}$ es un vector del plano que parte del punto O y que tiene coordenadas (x', y') en el sistema coordenado determinado por $\{\hat{e}_\theta, \hat{e}_\theta^\perp\}$, de acuerdo con la identidad 4.1, las coordenadas (x, y) de $\hat{x}$ en el sistema coordenado determinado por la base $\{\hat{e}_1, \hat{e}_2\}$ estarán dadas por la identidad de matrices

$$\begin{bmatrix} x & y \end{bmatrix} = \begin{bmatrix} x' & y' \end{bmatrix} \begin{bmatrix} \cos(\theta) & \text{sen}(\theta) \\ -\text{sen}(\theta) & \cos(\theta) \end{bmatrix}.$$

Es decir

$$
\begin{aligned}
x &= \cos(\theta)x' - \text{sen}(\theta)y' \\
y &= \text{sen}(\theta)x' + \cos(\theta)y'.
\end{aligned}
$$

Recíprocamente, si ahora observamos que

$$
\begin{aligned}
\hat{e}_1 &= \cos(\theta)\hat{e}_\theta - \operatorname{sen}(\theta)\hat{e}_\theta^\perp \\
\hat{e}_2 &= \operatorname{sen}(\theta)\hat{e}_\theta + \cos(\theta)\hat{e}_\theta^\perp
\end{aligned}
$$

(identidades que podemos obtener "despejando" los vectores $\hat{e}_1$ y $\hat{e}_2$ de las identidades 4.8), de acuerdo con la identidad 4.2, si (x, y) son las coordenadas de $\hat{x}$ en el sistema coordenado determinado por la base $\{\hat{e}_1, \hat{e}_2\}$, las coordenadas (x', y') en el sistema coordenado determinado por $\{\hat{e}_\theta, \hat{e}_\theta^\perp\}$ estarán dadas por la identidad de matrices

$$
\begin{bmatrix} x' & y' \end{bmatrix} = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} \cos(\theta) & -\operatorname{sen}(\theta) \\ \operatorname{sen}(\theta) & \cos(\theta) \end{bmatrix}
$$

o lo que es lo mismo, por

$$
\begin{aligned}
x' &= \cos(\theta)x + \operatorname{sen}(\theta)y \\
y' &= -\operatorname{sen}(\theta)x + \cos(\theta)y.
\end{aligned}
$$

Nótese que en este ejemplo se confirma que las matrices

$$
M = \begin{bmatrix} \cos(\theta) & -\operatorname{sen}(\theta) \\ \operatorname{sen}(\theta) & \cos(\theta) \end{bmatrix} \quad \text{y} \quad M' = \begin{bmatrix} \cos(\theta) & \operatorname{sen}(\theta) \\ -\operatorname{sen}(\theta) & \cos(\theta) \end{bmatrix}
$$

tienen las propiedades que habíamos mencionado.

4.2 La derivada direccional

Antes de entrar de lleno en la definición del concepto de derivada direccional, es importante mencionar que de aquí en adelante (salvo que se diga lo contrario) supondremos que las funciones con las que trabajeremos están definidas sobre un conjunto abierto y conexo $U \subset \mathbb{R}^n$.

Sea pues $f : U \subset \mathbb{R}^n \to \mathbb{R}$ y $\hat{x}_0 \in U$. Recordemos que el concepto de derivación de una función f de $\mathbb{R}$ en $\mathbb{R}$ en un punto x_0 de su dominio consiste básicamente en los siguientes pasos: calcular "el cambio" de los valores de f en x y en x_0, es decir, calcular $f(x) - f(x_0)$, en donde $x \neq x_0$; calcular "el cambio" que hay entre x y x_0 es decir, calcular $x - x_0$; calcular la "razón" (o "proporción") entre estas dos cantidades, es decir

$$
\frac{f(x) - f(x_0)}{x - x_0};
$$

y finalmente determinar si estas "razones" (o "proporciones") tienen un valor límite cuando la variable x se "aproxima" (o "tiende") al valor x_0. Como seguramente el lector ya habrá notado, no podemos copiar este procedimiento al caso de las funciones que nos ocupan. En efecto, si $f : U \subset \mathbb{R}^n \to \mathbb{R}$ y $\hat{x}_0, \hat{x} \in U$, $\hat{x} \neq \hat{x}_0$, el "cambio" de los valores de la función en $\hat{x}$ y en $\hat{x}_0$ (es decir, $f(\hat{x}) - f(\hat{x}_0)$) y el "cambio" entre $\hat{x}$ y $\hat{x}_0$, es decir, $\hat{x} - \hat{x}_0$, son de naturaleza distinta; mientras que la primera de estas cantidades es un número real $(f(\hat{x}) - f(\hat{x}_0))$, la otra $(\hat{x} - \hat{x}_0)$ es un vector, de tal forma que no tenemos una manera de calcular la "razón (o proporción)" entre dichas cantidades. Sin embargo, no todo está perdido y podemos adecuar la idea original, de alguna manera, al tipo de funciones que nos ocupa.

Dado que estamos suponiendo que U es un conjunto abierto, si $\hat{x}_0 \in U$ y $\hat{u}$ es un vector fijo que no sea el vector $\hat{0}$, los vectores de la forma $\hat{x} = \hat{x}_0 + h\hat{u}$, con $h \in \mathbb{R}$, tienen la particularidad de seguir perteneciendo a U si h es "suficientemente pequeña", aunque cabe aclarar que esta "pequeñez" de h también dependerá de la magnitud del vector $\hat{u}$. Más específicamente, sabemos que existe $r > 0$ tal que si $\|\hat{x} - \hat{x}_0\| < r$, entonces $\hat{x} \in U$, de modo que si tomamos $\hat{x} = \hat{x}_0 + h\hat{u}$, este punto pertenecerá a U si

$$
\begin{aligned}
r &> \|\hat{x} - \hat{x}_0\| \\
&= \|\hat{x}_0 + h\hat{u} - \hat{x}_0\| \\
&= \|h\hat{u}\| \\
&= |h| \, \|\hat{u}\| ,
\end{aligned}
$$

es decir, si $|h| < r/\|\hat{u}\|$.

Por otra parte, si observamos que

$$\|\hat{x} - \hat{x}_0\| = |h|\,\|\hat{u}\|$$

vemos que la magnitud del "cambio" entre $\hat{x}$ y $\hat{x}_0$ está determinada tanto por h, como por la magnitud del vector $\hat{u}$.

Por las razones que acabamos de exponer, si al tomar $\hat{x} = \hat{x}_0 + h\hat{u}$ pedimos que el vector $\hat{u}$ tenga magnitud 1 lograremos las siguientes ventajas: primera, que la pertenencia del vector $\hat{x}$ al conjunto U sólo dependerá del número h (bastará con que $|h| < r$ para que $\hat{x} = \hat{x}_0 + h\hat{u}$ pertenezca a la bola $B_r(\hat{x}_0)$, la que a su vez está contenida en U); y segunda, el "cambio" entre $\hat{x}$ y $\hat{x}_0$ (es decir el vector $\hat{x} - \hat{x}_0$) estará completamente determinado por el número h (¡no sólo su magnitud! ¡también su dirección!).

Con base en los supuestos anteriores, el cociente

$$\frac{f(\hat{x}_0 + h\hat{u}) - f(\hat{x}_0)}{h} \tag{4.9}$$

será una medida de la "razón (o proporción)" entre el cambio de los valores de la función (en los puntos $\hat{x}_0 + h\hat{u}$ y $\hat{x}_0$) y el cambio entre dichos valores de la variable, el cual está determinado por el número h.

Aquí es importante enfatizar que los puntos sobre los cuales estamos evaluando a f (para compararlos con su valor en $\hat{x}_0$) están restringidos a aquellos que son de la forma $\hat{x}_0 + h\hat{u}$, es decir, sólo estamos evaluando a f sobre los puntos que están en la recta que pasa por $\hat{x}_0$ y cuya dirección está determinada por el vector $\hat{u}$ (¡y que además se quedan contenidos en U!).

Por esta razón, si existe el límite del cociente 4.9 cuando h tiende a 0, diremos que la función f es derivable en el punto $\hat{x}_0$ en la dirección del vector $\hat{u}$, y al valor de este límite le llamaremos *la derivada direccional de f en $\hat{x}_0$ en la dirección del vector $\hat{u}$*. Dejamos plasmada la discusión anterior en la siguiente

Definición 4.2 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$ y $\hat{u} \in \mathbb{R}^n$ tal que $\|\hat{u}\| = 1$. Decimos que f es derivable en el punto $\hat{x}_0$ en la dirección del vector $\hat{u}$, si*

$$\lim_{h\to 0} \frac{f(\hat{x}_0 + h\hat{u}) - f(\hat{x}_0)}{h}$$

existe, en cuyo caso llamaremos al valor de dicho límite la derivada direccional de f en $\hat{x}_0$ en la dirección de $\hat{u}$, y la denotaremos por $D_{\hat{u}}f(\hat{x}_0)$, es decir

$$D_{\hat{u}}f(\hat{x}_0) := \lim_{h\to 0} \frac{f(\hat{x}_0 + h\hat{u}) - f(\hat{x}_0)}{h}$$

Como dijimos anteriormente, en términos generales este concepto de derivada es una medida de la razón de cambio entre el valor de la función f en $\hat{x}_0$ y el valor en puntos de la recta que pasa por $\hat{x}_0$ y cuya dirección está determinada por el vector $\hat{u}$. En algunos casos, a semejanza de lo que sucede con las funciones de $\mathbb{R}$ en $\mathbb{R}$, dicha derivada también tiene una interpretación geométrica importante.

Para el caso particular de $\mathbb{R}^2$, si nos fijamos en la curva que obtenemos al intersecar la gráfica de f con el plano perpendicular al plano XY que contenga a la recta que mencionamos anteriormente (ver figura 4.3), la derivada direccional de f en $\hat{x}_0$ en la dirección de $\hat{u}$ puede interpretarse como la pendiente de la recta tangente a dicha curva en el punto $(\hat{x}_0, f(\hat{x}_0))$.

Sin duda lo siguiente que tendremos que hacer será dar un ejemplo que ilustre este concepto, pero antes de ello convendrá hacer algunas observaciones importantes. Como el lector podrá notar, en la definición que acabamos de dar no fue necesario recurrir a ningún sistema coordenado específico.

Por otra parte, como para hacer cálculos específicos sí será necesario expresar a f en términos de algunas coordenadas, en el siguiente ejemplo no sólo vamos a mostrar cómo se calcula la derivada direccional de f en algún punto $\hat{x}_0$ y en alguna dirección $\hat{u}$, sino que mostraremos que dichos cálculos son idependientes (cuando menos para este ejemplo específico) del sistema coordenado que usemos.

Ejemplo 4.3 *Sean $\hat{x}_0, \hat{u} \in \mathbb{R}^2$ tales que en la base canónica tienen coordenadas (x_0, y_0) y (u_1, u_2), respectivamente. Sea $f : \mathbb{R}^2 \to \mathbb{R}$ tal que, si un punto $\hat{x} \in \mathbb{R}^2$ tiene coordenadas (x, y) en la misma base, entonces el valor de f en $\hat{x}$ está dado por*

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

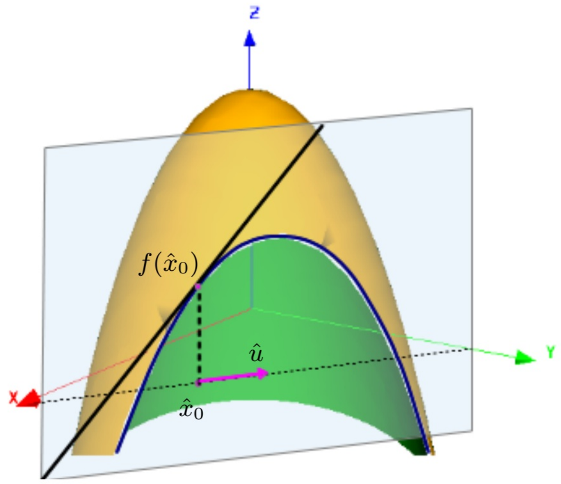


Figura 4.3: En $\mathbb{R}^2$, si intersecamos la gráfica de f con el plano perpendicular al plano XY que contenga a la recta que pasa por $\hat{x}_0$ en la dirección de $\hat{u}$, obtenemos una curva. Podemos interpretar la derivada $D_{\hat{u}}f(\hat{x}_0)$ como la pendiente de la recta tangente a dicha curva en el punto $(\hat{x}_0, f(\hat{x}_0))$.

Primero vamos a calcular $D_{\hat{u}}f(\hat{x}_0)$. De acuerdo con la definición 4.2, tenemos que

$$\begin{aligned}
D_{\hat{u}}f(\hat{x}_0) &= \lim_{h\to 0} \frac{f(\hat{x}_0 + h\hat{u}) - f(\hat{x}_0)}{h} \\
&= \lim_{h\to 0} \frac{f(x_0 + hu_1, y_0 + hu_2) - f(x_0, y_0)}{h} \\
&= \lim_{h\to 0} \frac{(x_0 + hu_1)^2 + (y_0 + hu_2)^2 - (x_0^2 + y_0^2)}{h} \\
&= \lim_{h\to 0} \frac{2x_0hu_1 + (hu_1)^2 + 2y_0hu_2 + (hu_2)^2}{h} \\
&= \lim_{h\to 0} \left(2x_0u_1 + 2y_0u_2 + hu_1^2 + hu_2^2\right) \\
&= 2(x_0u_1 + y_0u_2).
\end{aligned}$$

Supongamos ahora que tenemos otro sistema coordenado determinado por una base $\{\hat{e}_1', \hat{e}_2'\}$ tal que el cambio de coordenadas de este sistema al sistema canónico está dado por la matriz

$$M = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

Es decir, si un punto $\hat{x}$ tiene coordenadas (x', y') en la base $\{\hat{e}_1', \hat{e}_2'\}$, entonces sus coordenadas (x, y) en la base canónica estarán dadas por

$$\begin{bmatrix} x & y \end{bmatrix} = \begin{bmatrix} x' & y' \end{bmatrix} M = \begin{bmatrix} x' & y' \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

o lo que es lo mismo, que

$$\begin{aligned}
x &= a_{11}x' + a_{21}y' \\
y &= a_{12}x' + a_{22}y'.
\end{aligned}$$

De esta forma, el valor de f en $\hat{x}$ expresado en términos de sus coordenadas (x', y'), ahora estará dado por

$$f(x', y') = f(\hat{x})$$

$$
\begin{aligned}
&= f(x,y) \\
&= x^2 + y^2 \\
&= (a_{11}x' + a_{21}y')^2 + (a_{12}x' + a_{22}y')^2 \\
&= \left(a_{11}^2 + a_{12}^2\right)(x')^2 + 2(a_{11}a_{21} + a_{12}a_{22})x'y' + \left(a_{21}^2 + a_{22}^2\right)(y')^2.
\end{aligned}
$$

Si recordamos que la matriz de cambio de base M es tal que

$$
\begin{aligned}
\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} &= MM^{-1} \\
&= MM^t \\
&= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22}, \end{bmatrix}
\end{aligned}
$$

concluimos que $a_{11}^2 + a_{12}^2 = 1 = a_{21}^2 + a_{22}^2$ y que $a_{11}a_{21} + a_{12}a_{22} = 0$, de modo que

$$
\begin{aligned}
f(\hat{x}) &= f(x',y') \\
&= (x')^2 + (y')^2.
\end{aligned}
$$

No debería causar sorpresa al lector que las expresiones de f en términos de las coordenadas (x,y) y de las coordenadas (x',y') sean análogas en este caso ya que, en términos geométricos, f asigna a cada punto $\hat{x}$ su norma al cuadrado (cantidad que se expresa de forma análoga en ambos sistemas coordenados, en virtud de que éstos son ortonormales).

De esta forma, si las coordenadas del punto $\hat{x}_0$ y del vector $\hat{u}$ en el sistema coordenado determinado por la base $\{\hat{e}_1', \hat{e}_2'\}$ están dadas por (x_0', y_0') y (u_1', u_2') respectivamente, al calcular $D_{\hat{u}}f(\hat{x}_0)$ usando la expresión de f en este sistema y repitiendo los cálculos que hicimos al principio, obtendremos nuevamente que

$$
D_{\hat{u}}f(\hat{x}_0) = 2(x_0'u_1' + y_0'u_2').
$$

Lo que ahora mostraremos es que los dos valores que obtuvimos para $D_{\hat{u}}f(\hat{x}_0)$ usando la expresión de f en cada uno de los sistemas coordenados coinciden. Para ello, recordemos que las coordenadas del punto $\hat{x}_0$ y del vector $\hat{u}$ en ambos sistemas también satisfacen las identidades

$$
\begin{bmatrix} x_0 & y_0 \end{bmatrix} = \begin{bmatrix} x_0' & y_0' \end{bmatrix} M
$$

y

$$
\begin{bmatrix} u_1 & u_2 \end{bmatrix} = \begin{bmatrix} u_1' & u_2' \end{bmatrix} M
$$

respectivamente, de tal forma que, usando la notación del producto de matrices, obtenemos que

$$
\begin{aligned}
2(x_0u_1 + y_0u_2) &= 2\begin{bmatrix} x_0 & y_0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \\
&= 2\left(\begin{bmatrix} x_0' & y_0' \end{bmatrix} M\right)\left(\begin{bmatrix} u_1' & u_2' \end{bmatrix} M\right)^t \\
&= 2\begin{bmatrix} x_0' & y_0' \end{bmatrix}(MM^t)\begin{bmatrix} u_1' \\ u_2' \end{bmatrix} \\
&= 2\begin{bmatrix} x_0' & y_0' \end{bmatrix}(MM^{-1})\begin{bmatrix} u_1' \\ u_2' \end{bmatrix} \\
&= 2\begin{bmatrix} x_0' & y_0' \end{bmatrix}\begin{bmatrix} u_1' \\ u_2' \end{bmatrix} \\
&= 2(x_0'u_1' + y_0'u_2').
\end{aligned}
$$

Aun cuando lo que acabamos de hacer sólo muestra para un ejemplo específico que el valor de $D_{\hat{u}}f(\hat{x}_0)$ no depende del sistema coordenado que usemos, parte del material que vamos a desarrollar en este capítulo será útil para probar que este concepto es, en general, independiente de dichos sistemas.

Dada la similitud que existe entre el concepto de derivada direccional y el de derivada para funciones de $\mathbb{R}$ en $\mathbb{R}$, es de suponerse que muchas de las propiedades de esta última también las satisfaga la primera. Una de estas propiedades, que en el caso real resulta ser muy importante, es el hecho de que toda función que sea derivable en un punto tiene que ser continua en ese punto.

Dado que la derivada direccional $D_{\hat{u}}f(\hat{x}_0)$ sólo toma en cuenta la variación de la función sobre la recta que pasa por $\hat{x}_0$ en la dirección del vector $\hat{u}$, lo que podemos probar es que si dicha derivada direccional existe, entonces la función restringida a dicha recta será continua en el punto $\hat{x}_0$.

Este hecho queda plasmado en la siguiente proposición y dejamos su prueba al lector.

Proposición 4.4 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$ y $\hat{u} \in \mathbb{R}^n$ tal que $\|\hat{u}\| = 1$. Si $D_{\hat{u}}f(\hat{x}_0)$ existe, entonces*

$$\lim_{h\to 0} f(\hat{x}_0 + h\hat{u}) = f(\hat{x}_0).$$

Otras propiedades importantes de la derivada direccional son las relacionadas con la aritmética de las funciones. Nuevamente, de forma análoga a lo que sucede con las funciones de $\mathbb{R}$ en $\mathbb{R}$, además de que la existencia de la derivada direccional se preserva bajo dicha aritmética, las fórmulas de derivación de las funciones que se obtienen al realizar esta aritmética, resultarán ser muy útiles.

Proposición 4.5 *Sean $f, g : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$, y $\hat{u} \in \mathbb{R}^n$ tal que $\|\hat{u}\| = 1$. Si $D_{\hat{u}}f(\hat{x}_0)$ y $D_{\hat{u}}g(\hat{x}_0)$ existen, entonces:*

1. *$D_{\hat{u}}(f + g)(\hat{x}_0)$ existe y además*

$$D_{\hat{u}}(f + g)(\hat{x}_0) = D_{\hat{u}}f(\hat{x}_0) + D_{\hat{u}}g(\hat{x}_0)$$

2. *si $\alpha \in \mathbb{R}$, $D_{\hat{u}}(\alpha f(\hat{x}_0))$ existe y además*

$$D_{\hat{u}}(\alpha f)(\hat{x}_0) = \alpha D_{\hat{u}}f(\hat{x}_0)$$

3. *$D_{\hat{u}}(fg)(\hat{x}_0)$ existe y además*

$$D_{\hat{u}}(fg)(\hat{x}_0) = f(\hat{x}_0)D_{\hat{u}}g(\hat{x}_0) + g(\hat{x}_0)D_{\hat{u}}f(\hat{x}_0)$$

4. *si g es continua en $\hat{x}_0$ y $g(\hat{x}_0) \neq 0$, $D_{\hat{u}}(f/g)(\hat{x}_0)$ existe y además*

$$D_{\hat{u}}(f/g)(\hat{x}_0) = \frac{g(\hat{x}_0)D_{\hat{u}}f(\hat{x}_0) - f(\hat{x}_0)D_{\hat{u}}g(\hat{x}_0)}{g^2(\hat{x}_0)}.$$

Aun cuando la prueba de esta proposición se deja al lector, es importante hacer un comentario acerca de la hipótesis de continuidad que se pide en el inciso 4. En realidad, dicha hipótesis se incluye para seguir garantizando la condición que nos impusimos de que las funciones con las que vamos a trabajar a partir de este capítulo estén definidas sobre un conjunto abierto. Nótese que por el inciso (a) del problema 44 del capítulo 2, y del hecho de que el dominio de g sea el abierto U, podemos asegurar que la función f/g está definida en un abierto que contiene al punto $\hat{x}_0$.

Para finalizar con las propiedades de la derivada direccional relacionadas con las operaciones entre funciones, formularemos en una proposición aparte aquella que nos habla de la composición de funciones.

El caso que va a resultar más interesante será cuando compongamos a f con una función g de $\mathbb{R}$ en $\mathbb{R}$. Esta propiedad, que en realidad es una "consecuencia" de un resultado más general conocido como "la regla de la cadena" (que probaremos más adelante), queda plasmada en la siguiente

Proposición 4.6 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$, $\hat{u} \in \mathbb{R}^n$ tal que $\|\hat{u}\| = 1$ y $(a, b) \subset \mathbb{R}$ tal que $f(U) \subset (a, b)$. Si $g : (a, b) \subset \mathbb{R} \to \mathbb{R}$ es derivable en $f(\hat{x}_0)$ y la $D_{\hat{u}}f(\hat{x}_0)$ existe, entonces la $D_{\hat{u}}(g \circ f)(\hat{x}_0)$ existe y además*

$$D_{\hat{u}}(g \circ f)(\hat{x}_0) = g'(f(\hat{x}_0))D_{\hat{u}}f(\hat{x}_0).$$

Demostración. De acuerdo con la definición de $D_{\hat{u}}(g \circ f)(\hat{x}_0)$, debemos probar que

$$\lim_{h \to 0} \frac{(g \circ f)(\hat{x}_0 + h\hat{u}) - (g \circ f)(\hat{x}_0)}{h} = \lim_{h \to 0} \frac{g(f(\hat{x}_0 + h\hat{u})) - g(f(\hat{x}_0))}{h}$$

existe.

Como seguramente el lector recordará de la correspondiente prueba para el caso de funciones de $\mathbb{R}$ en $\mathbb{R}$, si hacemos

$$k_h = f(\hat{x}_0 + h\hat{u}) - f(\hat{x}_0),$$

el límite anterior se podría escribir como

$$\lim_{h \to 0} \frac{g(f(\hat{x}_0 + h\hat{u})) - g(f(\hat{x}_0))}{h} = \lim_{h \to 0} \frac{g(f(\hat{x}_0) + k_h) - g(f(\hat{x}_0))}{h},$$

de tal forma que si $k_h \neq 0$, entonces

$$\begin{aligned}
\lim_{h \to 0} \frac{g(f(\hat{x}_0 + h\hat{u})) - g(f(\hat{x}_0))}{h} &= \lim_{h \to 0} \frac{g(f(\hat{x}_0) + k_h) - g(f(\hat{x}_0))}{h} \\
&= \lim_{h \to 0} \left[\frac{g(f(\hat{x}_0) + k_h) - g(f(\hat{x}_0))}{k_h} \cdot \frac{k_h}{h} \right] \\
&= \lim_{h \to 0} \left[\frac{g(f(\hat{x}_0) + k_h) - g(f(\hat{x}_0))}{k_h} \cdot \frac{f(\hat{x}_0 + h\hat{u}) - f(\hat{x}_0)}{h} \right].
\end{aligned}$$

Si ahora notamos que, por la proposición 4.4, se tiene que $k_h \to 0$ cuando $h \to 0$ ya que

$$\begin{aligned}
\lim_{h \to 0} k_h &= \lim_{h \to 0} (f(\hat{x}_0 + h\hat{u}) - f(\hat{x}_0)) \\
&= 0,
\end{aligned}$$

entonces

$$\lim_{h \to 0} \frac{g(f(\hat{x}_0) + k_h) - g(f(\hat{x}_0))}{k_h} = g'(f(\hat{x}_0))$$

y obtendríamos el resultado deseado.

Como seguramente el lector ya sabe, el problema con el argumento anterior es que éste sólo funciona si $k_h \neq 0$, y como también recordará, la solución a este problema está en definir una función auxiliar.

Definimos $\varphi : (-r, r) \subset \mathbb{R} \to \mathbb{R}$, con $r > 0$ tal que $B_r(\hat{x}_0) \subset U$, de la siguiente manera:

$$\varphi(h) = \begin{cases} \frac{g(f(\hat{x}_0) + k_h) - g(f(\hat{x}_0))}{k_h} & \text{si } k_h \neq 0 \\ \\ g'(f(\hat{x}_0)) & \text{si } k_h = 0 \end{cases}$$

Lo primero que es importante observar (y que es muy fácil de verificar) es que

$$\frac{g(f(\hat{x}_0) + k_h) - g(f(\hat{x}_0))}{h} = \varphi(h) \frac{f(\hat{x}_0 + h\hat{u}) - f(\hat{x}_0)}{h}$$

para toda $h \in (-r, r)$, $h \neq 0$, de tal forma que nuestro problema se reduce a demostrar que

$$\lim_{h \to 0} \varphi(h) = g'(f(\hat{x}_0)).$$

Sea entonces $\varepsilon > 0$; dado que g es derivable en $f(\hat{x}_0)$, sabemos que existe $\delta' > 0$ tal que si $|k| < \delta'$ (y $f(\hat{x}_0) + k \in (a, b)$), entonces

$$\left| \frac{g(f(\hat{x}_0) + k) - g(f(\hat{x}_0))}{k} - g'(f(\hat{x}_0)) \right| < \varepsilon. \tag{4.10}$$

Por otra parte, por la proposición 4.4 sabemos que

$$\lim_{h \to 0} k_h = \lim_{h \to 0} (f(\hat{x}_0 + h\hat{u}) - f(\hat{x}_0)) = 0$$

y por tanto existe $0 < \delta \leq r$ tal que si $|h| < \delta$, entonces $|k_h| < \delta'$. De esta forma, si $|h| < \delta$, independientemente de que $k_h \neq 0$ o $k_h = 0$, por la desigualdad 4.10 o por el valor de φ cuando $k_h = 0$, en ambos casos se tiene que

$$|\varphi(h) - g'(f(\hat{x}_0))| < \varepsilon$$

y por lo tanto que

$$\lim_{h \to 0} \varphi(h) = g'(f(\hat{x}_0)),$$

con lo cual concluimos la prueba. ■

Terminamos esta serie de proposiciones con una en la que se establece una propiedad que bien podría interpretarse como la versión del Teorema del Valor Medio para la derivada direccional. Su formulación es la siguiente, en donde recordamos que $[\hat{a}, \hat{b}]$ representa al segmento de recta que une al punto $\hat{a}$ con el punto $\hat{b}$.

Proposición 4.7 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{a}, \hat{b} \in U$, $\hat{a} \neq \hat{b}$, tales que $\left[\hat{a}, \hat{b}\right] \subset U$ y $\hat{u} = \left(\hat{b} - \hat{a}\right) / \left\|\hat{b} - \hat{a}\right\| \in \mathbb{R}^n$. Si $D_{\hat{u}} f(\hat{x})$ existe para toda $\hat{x} \in \left[\hat{a}, \hat{b}\right]$, entonces existe $\xi \in \left(0, \left\|\hat{b} - \hat{a}\right\|\right)$ tal que si $\hat{\xi} = \hat{a} + \xi\hat{u}$, se cumple que*

$$f(\hat{b}) - f(\hat{a}) = \left\|\hat{b} - \hat{a}\right\| D_{\hat{u}} f\left(\hat{\xi}\right).$$

Demostración. Como es de suponerse, la prueba de esta proposición se basa en el Teorema del Valor Medio para funciones de $\mathbb{R}$ en $\mathbb{R}$. Definimos $g : \left[0, \left\|\hat{b} - \hat{a}\right\|\right] \subset \mathbb{R} \to \mathbb{R}$ como $g(t) = f(\hat{a} + t\hat{u})$. Nótese que g es derivable para toda $t \in \left[0, \left\|\hat{b} - \hat{a}\right\|\right]$, puesto que

$$\begin{aligned} g'(t) &= \lim_{h \to 0} \frac{g(t+h) - g(t)}{h} \\ &= \lim_{h \to 0} \frac{f(\hat{a} + (t+h)\hat{u}) - f(\hat{a} + t\hat{u})}{h} \\ &= \lim_{h \to 0} \frac{f((\hat{a} + t\hat{u}) + h\hat{u}) - f(\hat{a} + t\hat{u})}{h} \\ &= D_{\hat{u}} f(\hat{a} + t\hat{u}), \end{aligned}$$

de tal forma que, por el Teorema del Valor Medio para funciones $\mathbb{R}$ en $\mathbb{R}$, se tiene que existe $\xi \in \left(0, \left\|\hat{b} - \hat{a}\right\|\right)$ tal que

$$\begin{aligned} f\left(\hat{b}\right) - f(\hat{a}) &= g\left(\left\|\hat{b} - \hat{a}\right\|\right) - g(0) \\ &= \left(\left\|\hat{b} - \hat{a}\right\| - 0\right) g'(\xi) \\ &= \left\|\hat{b} - \hat{a}\right\| D_{\hat{u}} f(\hat{a} + \xi\hat{u}) \\ &= \left\|\hat{b} - \hat{a}\right\| D_{\hat{u}} f\left(\hat{\xi}\right). \end{aligned}$$

■

4.2.1 Derivadas parciales

Si una función f está escrita en términos de las coordenadas $x_1, \ldots, x_n$ asociadas a un sistema de referencia determinado por una base ortonormal $\{\hat{e}_1, \ldots, \hat{e}_n\}$ de $\mathbb{R}^n$, calcular la derivada direccional de f en la dirección de estos vectores básicos (en cualquier punto $\hat{x}$ del dominio de f) resultará más sencillo (y más importante) que en cualquier otra dirección. En efecto, si

$$\hat{x} = x_1\hat{e}_1 + \cdots + x_n\hat{e}_n = (x_1, \ldots, x_n)$$

y el valor de f en $\hat{x}$ se puede escribir en términos de las coordenadas $(x_1, \dots, x_n)$, es decir que

$$f(\hat{x}) = f(x_1, \dots, x_n),$$

entonces el cálculo de la derivada direccional de f en $\hat{x}$, en la dirección de cada vector $\hat{e}_i$ se traduce en lo siguiente: de la definición de la derivada direccional $D_{\hat{e}_i} f(\hat{x})$ sabemos que

$$D_{\hat{e}_i} f(\hat{x}) = \lim_{h \to 0} \frac{f(\hat{x} + h\hat{e}_i) - f(\hat{x})}{h}$$

y dado que las coordenadas del vector $\hat{x}+h\hat{e}_i$ son $(x_1, \dots, x_i+h, \dots, x_n)$, el límite anterior escrito en términos de coordenadas se convierte en

$$\begin{aligned} D_{\hat{e}_i} f(\hat{x}) &= \lim_{h \to 0} \frac{f(\hat{x} + h\hat{e}_i) - f(\hat{x})}{h} \\ &= \lim_{h \to 0} \frac{f(x_1, \dots, x_i + h, \dots, x_n) - f(x_1, \dots, x_n)}{h}. \end{aligned}$$

Como seguramente el lector podrá intuir, en el límite anterior la única coordenada en la que se está teniendo un incremento h es en la $i-$ésima, mientras que en las otras coordenadas no hay cambios, es decir, permanecen fijas. En términos más informales, esto significa que para el cálculo de la derivada direccional $D_{\hat{e}_i} f(\hat{x})$ bastaría con derivar la expresión $f(x_1, \dots, x_n)$ tomando como única variable a x_i, y considerando a las restantes $(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$ como si fueran constantes. En el siguiente ejemplo ilustramos este hecho de manera más clara.

Ejemplo 4.8 *Sea $f : \mathbb{R}^2 \to \mathbb{R}$ la función cuyo valor en un punto $\hat{x} \in \mathbb{R}^2$ está dado en términos de sus coordenadas (x, y) (en la base canónica, a cuyos elementos denotaremos (en este caso) por $\hat{e}_x$ y $\hat{e}_y$) por la expresión*

$$f(x, y) = 4x^5y^2.$$

Calcularemos $D_{\hat{e}_y} f(\hat{x})$ para cualquier $\hat{x} \in \mathbb{R}^2$. Si $\hat{x} = (x, y)$, de acuerdo con la definición de la derivada direccional $D_{\hat{e}_y} f(\hat{x})$, se tiene que

$$\begin{aligned} D_{\hat{e}_y} f(\hat{x}) &= \lim_{h \to 0} \frac{f(\hat{x} + h\hat{e}_y) - f(\hat{x})}{h} \\ &= \lim_{h \to 0} \frac{f(x, y + h) - f(x, y)}{h} \\ &= \lim_{h \to 0} \frac{4x^5(y + h)^2 - 4x^5y^2}{h} \\ &= \lim_{h \to 0} 4x^5 \frac{(y + h)^2 - y^2}{h} \\ &= 4x^5 \lim_{h \to 0} \frac{(y + h)^2 - y^2}{h} \\ &= 4x^5(2y) \\ &= 8x^5y. \end{aligned}$$

Como el lector habrá notado en este ejemplo, el cálculo de la derivada direccional $D_{\hat{e}_y} f(\hat{x})$ se redujo a derivar la expresión $4x^5y^2$ considerando sólo como variable a la coordenada y y tratando al resto de la expresión $(4x^5)$ como una constante. Una de las ventajas de lo anterior (¡entre otras más!) es que podemos simplificar el cálculo de las derivadas direccionales usando los métodos de derivación que aprendimos para las funciones de $\mathbb{R}$ en $\mathbb{R}$.

Por estas características, y algunas otras que veremos más adelante, las derivadas direccionales en la dirección de los vectores de una base ortonormal tienen un nombre y una notación propia, las cuales establecemos en la siguiente

Definición 4.9 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$ y $\hat{u} \in \mathbb{R}^n$ tal que $\|\hat{u}\| = 1$. Si $x_1, \ldots, x_n$ denotan las variables (o coordenadas) determinadas por una base ortonormal $\{\hat{e}_1, \ldots, \hat{e}_n\}$ de $\mathbb{R}^n$, definimos la derivada parcial de f con respecto de la variable x_i en $\hat{x}_0$, que denotamos por $\frac{\partial f}{\partial x_i}(\hat{x}_0)$, como*

$$\begin{aligned}\frac{\partial f}{\partial x_i}(\hat{x}_0) &:= D_{\hat{e}_i} f(\hat{x}_0) \\ &:= \lim_{h \to 0} \frac{f(\hat{x}_0 + h\hat{e}_i) - f(\hat{x}_0)}{h}.\end{aligned}$$

Como es de suponerse, dado que cualquier derivada parcial no es más que una cierta derivada direccional, las proposiciones 4.4, 4.5, 4.6, y 4.7, tienen sus correspondientes versiones para derivadas parciales, las cuales formalizaremos a continuación (sin probar). En todas ellas supondremos que $x_1, \ldots, x_n$ son las variables (o coordenadas) determinadas por una base ortonormal $\{\hat{e}_1, \ldots, \hat{e}_n\}$ de $\mathbb{R}^n$.

Proposición 4.10 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$ y $\hat{x}_0 \in U$. Si $\frac{\partial f}{\partial x_i}(\hat{x}_0)$ existe, entonces*

$$\lim_{h \to 0} f(\hat{x}_0 + h\hat{e}_i) = f(\hat{x}_0).$$

Proposición 4.11 *Sean $f, g : U \subset \mathbb{R}^n \to \mathbb{R}$ y $\hat{x}_0 \in U$. Si $\frac{\partial f}{\partial x_i}(\hat{x}_0)$ y $\frac{\partial g}{\partial x_i}(\hat{x}_0)$ existen para alguna $i \in \{1, \ldots, n\}$, entonces:*

1. *$\frac{\partial (f+g)}{\partial x_i}(\hat{x}_0)$ existe y además*

$$\frac{\partial (f+g)}{\partial x_i}(\hat{x}_0) = \frac{\partial f}{\partial x_i}(\hat{x}_0) + \frac{\partial g}{\partial x_i}(\hat{x}_0)$$

2. *si $\alpha \in \mathbb{R}$, $\frac{\partial (\alpha f)}{\partial x_i}(\hat{x}_0)$ existe y además*

$$\frac{\partial (\alpha f)}{\partial x_i}(\hat{x}_0) = \alpha \frac{\partial f}{\partial x_i}(\hat{x}_0)$$

3. *$\frac{\partial (fg)}{\partial x_i}(\hat{x}_0)$ existe y además*

$$\frac{\partial (fg)}{\partial x_i}(\hat{x}_0) = f(\hat{x}_0)\frac{\partial g}{\partial x_i}(\hat{x}_0) + g(\hat{x}_0)\frac{\partial f}{\partial x_i}(\hat{x}_0)$$

4. *si g es continua en $\hat{x}_0$ y $g(\hat{x}_0) \neq 0$, $\frac{\partial (f/g)}{\partial x_i}(\hat{x}_0)$ existe y además*

$$\frac{\partial (f/g)}{\partial x_i}(\hat{x}_0) = \frac{g(\hat{x}_0)\frac{\partial f}{\partial x_i}(\hat{x}_0) - f(\hat{x}_0)\frac{\partial g}{\partial x_i}(\hat{x}_0)}{g^2(\hat{x}_0)}.$$

Proposición 4.12 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$ y $(a,b) \subset \mathbb{R}$ tal que $f(U) \subset (a,b)$. Si $g : (a,b) \subset \mathbb{R} \to \mathbb{R}$ es derivable en $f(\hat{x}_0)$ y $\frac{\partial f}{\partial x_i}(\hat{x}_0)$ existe para alguna $i \in \{1, \ldots, n\}$, entonces $\frac{\partial (g \circ f)}{\partial x_i}(\hat{x}_0)$ existe y además*

$$\frac{\partial (g \circ f)}{\partial x_i}(\hat{x}_0) = g'(f(\hat{x}_0))\frac{\partial f}{\partial x_i}(\hat{x}_0).$$

Proposición 4.13 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{a}, \hat{b} \in U$, $\hat{a} \neq \hat{b}$, tales que $\left[\hat{a}, \hat{b}\right] \subset U$ y $\hat{b} - \hat{a} = (b-a)\hat{e}_i$. Si para alguna $i \in \{1, \ldots, n\}$ se tiene que $\frac{\partial f}{\partial x_i}(\hat{x})$ existe para toda $\hat{x} \in \left[\hat{a}, \hat{b}\right]$, entonces existe $\xi \in \left(0, \left\|\hat{b} - \hat{a}\right\| = |b-a|\right)$ tal que si $\hat{\xi} = \hat{a} + \xi\hat{e}_i \in \left[\hat{a}, \hat{b}\right]$, se cumple que*

$$f(\hat{b}) - f(\hat{a}) = (b-a)\frac{\partial f}{\partial x_i}(\hat{\xi}). \tag{4.11}$$

En esta última proposición es importante hacer notar dos aspectos relevantes: uno, que la hipótesis de que $\hat{a}, \hat{b} \in U$ sean tales que $\hat{b} - \hat{a} = (b-a)\hat{e}_i$ significa que las coordenadas de $\hat{a}$ y $\hat{b}$ (en la base $\{\hat{e}_1, \ldots, \hat{e}_n\}$) sólo difieren en la $i-$ésima coordenada; y dos, que en el caso en que la cantidad $b-a$ sea negativa, entonces el vector $\hat{u}$ que se construye en la proposición 4.7 es tal que

$$\begin{aligned} \hat{u} &= \frac{\hat{b} - \hat{a}}{\left\|\hat{b} - \hat{a}\right\|} \\ &= \frac{b-a}{|b-a|}\hat{e}_i \\ &= -\hat{e}_i. \end{aligned}$$

De esta forma, de acuerdo con la conclusión de dicha proposición (y el fácil resultado que el lector probará en el problema 1 de este capítulo), se tiene que

$$\begin{aligned} f(\hat{b}) - f(\hat{a}) &= \left\|\hat{b} - \hat{a}\right\| D_{\hat{u}} f(\hat{\xi}) \\ &= |b-a|\, D_{-\hat{e}_i} f(\hat{\xi}) \\ &= -(b-a)(-D_{\hat{e}_i} f(\hat{\xi})) \\ &= (b-a)\,\frac{\partial f}{\partial x_i}(\hat{\xi}), \end{aligned}$$

lo que prueba que la identidad 4.11 se cumple independientemente de la relación (de orden) que guarden los números a y b.

4.3 La derivada global

Salvo por el caso de las derivadas direccionales, quedarnos sólo con la idea de que la derivada es una forma de medir "la razón de cambio" de una función, es algo que dificilmente nos ayudará a "extender" el concepto de derivada a funciones de $\mathbb{R}^n$ en $\mathbb{R}$. Por esta razón, en esta sección empezaremos por recordar que la derivabilidad de una función f de $\mathbb{R}$ en $\mathbb{R}$ en un punto x_0, además de proporcionarnos una medida de "la razón de cambio" de f en x_0, también nos ayuda a resolver un problema geométrico: encontrar la recta tangente a la gráfica de la función en el punto $(x_0, f(x_0))$.

Como el lector seguramente recordará, la derivada de una función f de $\mathbb{R}$ en $\mathbb{R}$ en un punto x_0 $(f'(x_0))$ tiene la propiedad de que, si tomamos la recta con esta pendiente que pasa por el punto $(x_0, f(x_0))$, esta recta es tangente a la gráfica de la función en dicho punto.

Específicamente, lo anterior se traduce en lo siguiente: la recta con pendiente $f'(x_0)$ que pasa por el punto $(x_0, f(x_0))$ es aquella cuya ecuación se puede escribir de la forma

$$y = f'(x_0)(x - x_0) + f(x_0).$$

Que esta recta sea tangente a la gráfica de f en $(x_0, f(x_0))$ significa no sólo que pasa por ese punto, sino que además se "parece" mucho a f cerca de x_0, donde eso de "parecerse mucho a f cerca de x_0" se traduce en que

$$\lim_{x \to x_0} \frac{f(x) - (f'(x_0)(x - x_0) + f(x_0))}{x - x_0} = 0,$$

es decir, que la diferencia $f(x) - (f'(x_0)(x - x_0) + f(x_0))$ se va más "rápido" a cero de lo que se va la diferencia $x - x_0$. De hecho, como seguramente el lector también recordará, la existencia de una recta con estas características garantiza la derivabilidad de la función en el punto x_0. Dado que las rectas que pasan por el punto $(x_0, f(x_0))$ tienen una ecuación de la forma

$$y = m(x - x_0) + f(x_0),$$

si una de estas rectas tiene la propiedad de que

$$\lim_{x \to x_0} \frac{f(x) - (m(x - x_0) + f(x_0))}{x - x_0} = 0, \tag{4.12}$$

entonces podemos asegurar que f es derivable en x_0 y que además $f'(x_0) = m$. En efecto, como

$$\begin{aligned} 0 &= \lim_{x \to x_0} \frac{f(x) - (m(x - x_0) + f(x_0))}{x - x_0} \\ &= \lim_{x \to x_0} \left[\frac{f(x) - f(x_0)}{x - x_0} - m \right], \end{aligned}$$

entonces

$$\begin{aligned} m &= \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} \\ &= f'(x_0), \end{aligned}$$

lo que comprueba nuestra afirmación.

De lo anterior concluimos que la derivabilidad de una función f de $\mathbb{R}$ en $\mathbb{R}$ en un punto x_0 es equivalente a la existencia de una función de la forma $r(x) = m(x - x_0) + f(x_0)$ que satisfaga la identidad 4.12. En realidad bastaría considerar las funciones de la forma $L(x) = mx$, puesto que la función $r(x)$ no es más que, en términos de sus gráficas, la "traslación" de $L(x)$ (que pasa por el origen) al punto $(x_0, f(x_0))$ (ver figura 4.4).

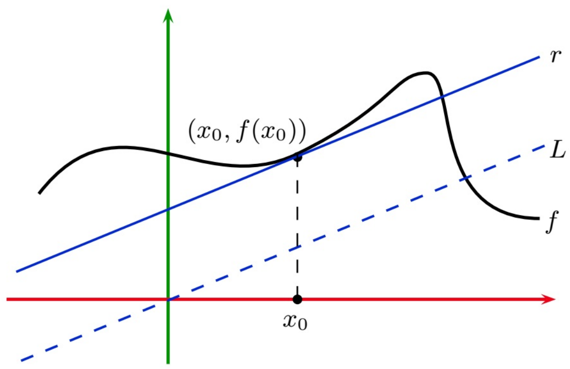


Figura 4.4: La función $r(x) = m(x - x_0) + f(x_0)$ es la "traslación" de la función $L(x) = mx$ (que pasa por el origen) al punto $(x_0, f(x_0))$.

Las funciones de la forma $L(x) = mx$ son justo el tipo de funciones de $\mathbb{R}$ en $\mathbb{R}$ conocidas con el nombre de funciones lineales (las cuales introdujimos en el problema 38 del capítulo 2 para el caso general de funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$), y que son las funciones $L : \mathbb{R} \to \mathbb{R}$ que tienen las siguientes dos propiedades:

1. $L(x + y) = L(x) + L(y)$ para todas $x, y \in \mathbb{R}$, y
2. $L(\alpha x) = \alpha L(x)$ para todas $\alpha, x \in \mathbb{R}$.

Es un problema sencillo mostrar que $L : \mathbb{R} \to \mathbb{R}$ *es una función lineal si y sólo si existe* $m \in \mathbb{R}$ *tal que* $L(x) = mx$ *para toda* $x \in \mathbb{R}$ (en ambas implicaciones se deduce que $m = L(1)$).

Como seguramente el lector ya estará imaginando, la discusión anterior nos sugiere la forma en que podemos dar la definición de derivada (global) de una función de $\mathbb{R}^n$ en $\mathbb{R}$. En términos generales, que una función f de este tipo sea derivable en un punto $\hat{x}_0$ de su dominio significará que existe una función lineal L de $\mathbb{R}^n$ en $\mathbb{R}$ tal que al "trasladarla" para que su valor en $\hat{x}_0$ sea $f(\hat{x}_0)$, esta función lineal "trasladada" $(L(\hat{x} - \hat{x}_0) + f(\hat{x}_0))$, que recibe el nombre de *función (o transformación) afín* se "parece" mucho a la función f "cerca" de $\hat{x}_0$.

Específicamente, nuestra definición de derivabilidad será la siguiente.

Definición 4.14 *Sean* $f : U \subset \mathbb{R}^n \to \mathbb{R}$ *y* $\hat{x}_0 \in U$. *Decimos que* f *es derivable en* $\hat{x}_0$ *si existe una función lineal* $L : \mathbb{R}^n \to \mathbb{R}$ *tal que*

$$\lim_{\hat{x} \to \hat{x}_0} \frac{f(\hat{x}) - (L(\hat{x} - \hat{x}_0) + f(\hat{x}_0))}{\|\hat{x} - \hat{x}_0\|} = 0. \tag{4.13}$$

Sin duda que después de esta definición, sería conveniente dar un ejemplo que la ilustrara. Sin embargo, es importante hacer notar que hasta ahora no tenemos ninguna herramienta que nos permita, dada una función (expresada en algún sistema coordenado), proponer (o intuir) cuál debería de ser la función lineal que satisfaga la condición que se le pide en esta definición.

Por esta razón, primero nos enfocaremos en dar algunas propiedades relacionadas con este concepto de derivada que, entre otras cosas, nos permitirán darnos una idea de cómo encontrar la ya famosa función lineal L.

De hecho, la primera proposición que probaremos es justo la que nos asegura que sólo puede haber una función lineal que satisfaga la propiedad de la definición 4.14.

Proposición 4.15 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$ y $L, \tilde{L} : \mathbb{R}^n \to \mathbb{R}$ funciones lineales. Si L y $\tilde{L}$ satisfacen la condición 4.13 de la definición 4.14, entonces L y $\tilde{L}$ son iguales.*

Demostración. Dado que por hipótesis sabemos que

$$\lim_{\hat{x}\to\hat{x}_0} \frac{f(\hat{x}) - (L(\hat{x}-\hat{x}_0) + f(\hat{x}_0))}{\|\hat{x}-\hat{x}_0\|} = 0$$

y

$$\lim_{\hat{x}\to\hat{x}_0} \frac{f(\hat{x}) - (\tilde{L}(\hat{x}-\hat{x}_0) + f(\hat{x}_0))}{\|\hat{x}-\hat{x}_0\|} = 0,$$

restando ambos límites concluimos que

$$\begin{aligned}\lim_{\hat{x}\to\hat{x}_0} \frac{\tilde{L}(\hat{x}-\hat{x}_0) - L(\hat{x}-\hat{x}_0)}{\|\hat{x}-\hat{x}_0\|} &= \lim_{\hat{x}\to\hat{x}_0} \frac{(\tilde{L}-L)(\hat{x}-\hat{x}_0)}{\|\hat{x}-\hat{x}_0\|} \\ &= 0,\end{aligned}$$

de tal forma que si hacemos $\hat{L} = \tilde{L} - L$ y $\hat{h} = \hat{x} - \hat{x}_0$, entonces $\hat{L}$ es una función lineal para la cual se satisface que

$$\lim_{\hat{h}\to\hat{0}} \frac{\hat{L}(\hat{h})}{\left\|\hat{h}\right\|} = 0.$$

Por los problemas 33 y 38 del capítulo 2, se tiene que $\hat{L}$ es la constante cero, o lo que es lo mismo, que L y $\tilde{L}$ son iguales. ■

Con base en esta proposición ya podemos completar la definición 4.14 de la siguiente forma: si $f : U \subset \mathbb{R}^n \to \mathbb{R}$ es derivable en $\hat{x}_0 \in U$, a la función lineal que satisface la condición 4.13, que por la proposición anterior sabemos que es única, la denotaremos por $Df(\hat{x}_0)$ y diremos que es *la derivada de f en $\hat{x}_0$*. De esta forma, *la derivada de una función de $\mathbb{R}^n$ en $\mathbb{R}$ es una función lineal.*

Más adelante veremos que toda funcion lineal de $\mathbb{R}^n$ en $\mathbb{R}$ se puede representar (dependiendo del sistema coordenado que se elija) por una matriz de $1 \times n$ o por un vector en $\mathbb{R}^n$, de tal forma que en algunas ocasiones diremos (abusando ciertamente del lenguaje) que la derivada de este tipo de funciones también es (o dicho de manera más correcta, que se puede representar por) una de estas matrices (o uno de estos vectores).

Justamente, con el objetivo de poder "calcular" (o conocer) la derivada de una función en un punto $\hat{x}_0$, el siguiente resultado que veremos establece una importante relación entre la función lineal $Df(\hat{x}_0)$ (que es la derivada en el punto $\hat{x}_0$) y las derivadas direccionales en dicho punto.

En términos generales, la siguiente proposición que probaremos nos asegura que si una función es derivable en un punto $\hat{x}_0$, entonces la derivada direccional en $\hat{x}_0$, en la dirección de cualquier vector $\hat{u}$, también existe y además el valor de dicha derivada direccional se obtiene "evaluando" la función lineal $Df(\hat{x}_0)$ en el vector $\hat{u}$.

Proposición 4.16 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$ y $\hat{u} \in \mathbb{R}^n$ tal que $\|\hat{u}\| = 1$. Si f es derivable en $\hat{x}_0$, entonces la derivada direccional de f en $\hat{x}_0$ en la dirección de $\hat{u}$ también existe y además*

$$D_{\hat{u}} f(\hat{x}_0) = Df(\hat{x}_0)(\hat{u}).$$

Demostración. De la definición 4.14 sabemos que

$$\begin{aligned}\lim_{\hat{x}\to\hat{x}_0}\frac{f(\hat{x})-(Df(\hat{x}_0)(\hat{x}-\hat{x}_0)+f(\hat{x}_0))}{\|\hat{x}-\hat{x}_0\|}&=\lim_{\hat{x}\to\hat{x}_0}\frac{f(\hat{x})-f(\hat{x}_0)-Df(\hat{x}_0)(\hat{x}-\hat{x}_0)}{\|\hat{x}-\hat{x}_0\|}\\&=0\end{aligned}$$

de tal forma que, si en particular tomamos $\hat{x}=\hat{x}_0+h\hat{u}$, tenemos entonces que $\hat{x}\to\hat{x}_0$ si $h\to 0$ y por lo tanto se tendrá que

$$\begin{aligned}0&=\lim_{\hat{x}\to\hat{x}_0}\frac{f(\hat{x})-f(\hat{x}_0)-Df(\hat{x}_0)(\hat{x}-\hat{x}_0)}{\|\hat{x}-\hat{x}_0\|}\\&=\lim_{\hat{x}\to\hat{x}_0}\frac{|f(\hat{x})-f(\hat{x}_0)-Df(\hat{x}_0)(\hat{x}-\hat{x}_0)|}{\|\hat{x}-\hat{x}_0\|}\\&=\lim_{h\to0}\frac{|f(\hat{x}_0+h\hat{u})-f(\hat{x}_0)-Df(\hat{x}_0)(h\hat{u})|}{\|h\hat{u}\|}\\&=\lim_{h\to0}\frac{|f(\hat{x}_0+h\hat{u})-f(\hat{x}_0)-Df(\hat{x}_0)(h\hat{u})|}{|h|}\\&=\lim_{h\to0}\left|\frac{f(\hat{x}_0+h\hat{u})-f(\hat{x}_0)-hDf(\hat{x}_0)(\hat{u})}{h}\right|\\&=\lim_{h\to0}\left|\frac{f(\hat{x}_0+h\hat{u})-f(\hat{x}_0)}{h}-Df(\hat{x}_0)(\hat{u})\right|,\end{aligned}$$

lo cual es equivalente a que

$$\lim_{h\to0}\frac{f(\hat{x}_0+h\hat{u})-f(\hat{x}_0)}{h}=Df(\hat{x}_0)(\hat{u}).$$

De aquí concluimos que la derivada direccional de f en $\hat{x}_0$ en la dirección de $\hat{u}$ existe y además que

$$D_{\hat{u}}f(\hat{x}_0)=Df(\hat{x}_0)(\hat{u}).$$

■

Para abordar el problema de calcular la derivada de una función es necesario establecer algunos hechos importantes acerca de las funciones lineales, lo cual haremos a continuación.

Es fácil ver que si $L:\mathbb{R}^n\to\mathbb{R}$ es una función lineal, para conocer el valor de L en cualquier punto $\hat{x}\in\mathbb{R}^n$, es suficiente con saber los valores de L en los elementos de una base de $\mathbb{R}^n$. En efecto, nótese que si $\hat{x}$ tiene coordenadas $(x_1,\ldots,x_n)$ en un sistema coordenado determinado por una base $\beta=\{\hat{v}_1,\ldots,\hat{v}_n\}$, esto significa que

$$\hat{x}=x_1\hat{v}_1+\cdots+x_n\hat{v}_n,$$

de tal forma que, por la "linealidad" de L se tiene que

$$\begin{aligned}L(\hat{x})&=L(x_1\hat{v}_1+\cdots+x_n\hat{v}_n)\\&=x_1L(\hat{v}_1)+\cdots+x_nL(\hat{v}_n).\end{aligned}$$

Esto confirma que para saber el valor de L en $\hat{x}$ es suficiente con conocer las coordenadas de $\hat{x}$ en una base $\beta=\{\hat{v}_1,\ldots,\hat{v}_n\}$ y los valores de L en los elemento de esta base.

Si denotamos por $a_{1i}=L(\hat{v}_i)$, se suele construir la matriz de $1\times n$

$$M_\beta=\begin{bmatrix}a_{11}&\cdots&a_{1n}\end{bmatrix}$$

y decir que la matriz M_β "representa" a la función lineal L en la base $\beta=\{\hat{v}_1,\ldots,\hat{v}_n\}$.

De hecho, esta representación matricial y las operaciones entre matrices son útiles para expresar a $L(\hat{x})$ como

$$L(\hat{x})=L(x_1,\ldots,x_n)$$

$$= \begin{bmatrix} a_{11} & \cdots & a_{1n} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$
$$= M_\beta \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

En el caso particular de funciones lineales de $\mathbb{R}^n$ en $\mathbb{R}$, también podemos “representar” a L por el vector $(L(\hat{v}_1), \ldots, L(\hat{v}_n))$ y usar el producto punto para expresar a $L(\hat{x})$ como

$$\begin{aligned} L(\hat{x}) &= L(x_1, \ldots, x_n) \\ &= (L(\hat{v}_1), \ldots, L(\hat{v}_n)) \cdot (x_1, \ldots, x_n). \end{aligned} \tag{4.14}$$

Con base en lo anterior, si una función f está expresada en términos de las coordenadas $x_1, \ldots, x_n$ determinadas por alguna base ortonormal $\{\hat{e}_1, \ldots, \hat{e}_n\}$ de $\mathbb{R}^n$, para encontrar la matriz que representa (en la misma base $\{\hat{e}_1, \ldots, \hat{e}_n\}$) a la derivada de f en un punto $\hat{x}_0$ de su dominio, es decir $Df(\hat{x}_0)$, bastaría con saber el valor $Df(\hat{x}_0)(\hat{e}_i)$ (para cada $i \in \{1, \ldots, n\}$). De acuerdo con lo probado en la proposición 4.16, se tiene que

$$Df(\hat{x}_0)(\hat{e}_i) = D_{\hat{e}_i} f(\hat{x}_0).$$

Si ahora recordamos, de acuerdo con la definición 4.9, que

$$D_{\hat{e}_i} f(\hat{x}_0) = \frac{\partial f}{\partial x_i}(\hat{x}_0),$$

entonces concluimos que para “calcular” (o “conocer”) a $Df(\hat{x}_0)$ basta con calcular las derivadas parciales de f en $\hat{x}_0$.

En cuanto a la discusión previa, es importante hacer notar que la mera existencia de las derivadas parciales de una función f en un punto $\hat{x}_0$ de su dominio no garantizan que la función sea derivable en dicho punto. Lo que en el fondo está sucediendo, es que en la proposición 4.16 se establece que la existencia de las derivadas direccionales (incluyendo las derivadas parciales) es una condición (o consecuencia) necesaria de la derivabilidad de f en $\hat{x}_0$, pero lo recíproco no es cierto (más adelante daremos un ejemplo que ilustra esta afirmación).

Sin embargo, la importancia de la discusión anterior está en que se muestra con toda precisión cuál es la única función lineal que podría ser la derivada de f en $\hat{x}_0$, lo que sin duda es un hecho de gran valor. Una vez aclarado lo anterior, pasamos a dar el siguiente

Ejemplo 4.17 *Sea $f : \mathbb{R}^2 \to \mathbb{R}$ dada en términos de las coordenadas de la base canónica por la expresión*

$$f(x, y) = x^2 + y^2 - 1.$$

Mostraremos que f es derivable en cualquier punto $\hat{x}_0 = (x_0, y_0) \in \mathbb{R}^2$ y calcularemos $Df(\hat{x}_0)$.

De acuerdo con lo visto anteriormente, hay que calcular las derivadas parciales de f con respecto de las variables x y y,

$$\frac{\partial f}{\partial x}(x, y) = 2x \qquad y \qquad \frac{\partial f}{\partial y}(x, y) = 2y,$$

evaluarlas en $\hat{x}_0 = (x_0, y_0)$

$$\frac{\partial f}{\partial x}(x_0, y_0) = 2x_0 \qquad y \qquad \frac{\partial f}{\partial y}(x_0, y_0) = 2y_0$$

y mostrar que la función lineal (expresada en el mismo sistema coordenado)

$$\begin{aligned} L(\hat{x}) &= L(x, y) \\ &= \begin{pmatrix} \frac{\partial f}{\partial x}(x_0, y_0) & \frac{\partial f}{\partial y}(x_0, y_0) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \end{aligned}$$

$$
\begin{aligned}
&= \begin{pmatrix} 2x_0 & 2y_0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\
&= 2x_0 x + 2y_0 y
\end{aligned}
$$

satisface la condición 4.13 de la definición 4.14.

En efecto, sustituyendo f y L en dicha expresión, tenemos que

$$
\begin{aligned}
\lim_{\hat{x}\to\hat{x}_0} \frac{f(\hat{x}) - (L(\hat{x}-\hat{x}_0) + f(\hat{x}_0))}{\|\hat{x}-\hat{x}_0\|} &= \lim_{\hat{x}\to\hat{x}_0} \frac{x^2+y^2-1-(2x_0(x-x_0)+2y_0(y-y_0)+x_0^2+y_0^2-1)}{\sqrt{(x-x_0)^2+(y-y_0)^2}} \\
&= \lim_{\hat{x}\to\hat{x}_0} \frac{(x^2-2x_0x+x_0^2)+(y^2-2y_0y+y_0^2)}{\sqrt{(x-x_0)^2+(y-y_0)^2}} \\
&= \lim_{\hat{x}\to\hat{x}_0} \frac{(x-x_0)^2+(y-y_0)^2}{\sqrt{(x-x_0)^2+(y-y_0)^2}} \\
&= \lim_{\hat{x}\to\hat{x}_0} \sqrt{(x-x_0)^2+(y-y_0)^2} \\
&= 0,
\end{aligned}
$$

de donde concluimos que f es derivable en $\hat{x}_0$ y que además $Df(\hat{x}_0)$ es la función lineal "representada" (o "asociada"), en el mismo sistema coordenado en que está expresada f, por la matriz

$$
\begin{bmatrix} \frac{\partial f}{\partial x}(x_0,y_0) & \frac{\partial f}{\partial y}(x_0,y_0) \end{bmatrix} = \begin{bmatrix} 2x_0 & 2y_0 \end{bmatrix}.
$$

Es decir, que

$$
\begin{aligned}
Df(\hat{x}_0)(x,y) &= \begin{bmatrix} 2x_0 & 2y_0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\
&= 2x_0x + 2y_0y.
\end{aligned}
$$

Para aprovechar el ejemplo anterior, en el cual podemos "ver" la gráfica de la función f, y confirmar la estrecha relación que existe entre el concepto de derivada y el concepto de tangencia, observemos que la gráfica de la función lineal

$$
Df(\hat{x}_0)(x,y) = 2x_0x + 2y_0y
$$

es un plano (que contiene al origen) y que la gráfica de la función

$$
\begin{aligned}
L(x,y) &= Df(\hat{x}_0)(x-x_0, y-y_0) + f(x_0,y_0) \\
&= 2x_0(x-x_0) + 2y_0(y-y_0) + x_0^2 + y_0^2 - 1
\end{aligned}
$$

es un plano ("trasladado" al punto $(x_0, y_0, x_0^2+y_0^2-1)$) que se ve "tangente" a la gráfica de la función $f(x,y) = x^2+y^2-1$ en este mismo punto (ver figura 4.5).

Con base en las observaciones anteriores, damos la siguiente

Definición 4.18 *Sea $f : U \subset \mathbb{R}^2 \to \mathbb{R}$ derivable en el punto $\hat{x}_0 = (x_0,y_0) \in U$. Decimos que el plano que tiene como ecuación*

$$
z = \frac{\partial f}{\partial x}(x_0,y_0)\,(x-x_0) + \frac{\partial f}{\partial y}(x_0,y_0)\,(y-y_0) + f(x_0,y_0) \tag{4.15}
$$

es el plano tangente a la gráfica de f en el punto $(x_0, y_0, f(x_0,y_0))$.

Nótese que la ecuación 4.15 se puede escribir como

$$
\left(-\frac{\partial f}{\partial x}(x_0,y_0), -\frac{\partial f}{\partial y}(x_0,y_0), 1\right) \cdot (x-x_0, y-y_0, z-f(x_0,y_0)) = 0 \tag{4.16}
$$

de tal forma que el vector

$$
\left(-\frac{\partial f}{\partial x}(x_0,y_0), -\frac{\partial f}{\partial y}(x_0,y_0), 1\right)
$$

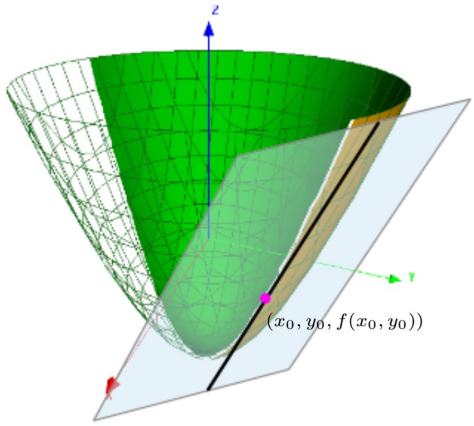


Figura 4.5: El plano "tangente" a la gráfica de la función $f(x,y) = x^2 + y^2 - 1$ en el punto $(x_0, y_0, f(x_0, y_0))$.

es un vector normal a este plano tangente.

De aquí en adelante escribiremos, sin duda abusando de la notación, que la función lineal $Df(\hat{x}_0)$ es igual a la matriz (de $1 \times n$) que la representa en el sistema de referencia que se esté usando, y en cuyas coordenadas está expresada f. Sin embargo, no hay que olvidar que esta matriz depende del sistema de referencia que se esté usando. Parte de lo que haremos a continuación será mostrar la relación que existe entre las matrices asociadas a la función lineal $Df(\hat{x}_0)$ en dos sistemas de referencia (ortonormales) diferentes, y cómo se puede obtener una a partir de la otra.

Una vez dicho lo anterior, si $x_1, \ldots, x_n$ denotan a las "variables" determinadas por una base ortonormal $\{\hat{e}_1, \ldots, \hat{e}_n\}$, entonces escribiremos que

$$Df(\hat{x}_0) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial f}{\partial x_n}(\hat{x}_0) \end{bmatrix}, \tag{4.17}$$

de tal forma que si $\hat{u} \in \mathbb{R}^n$ tiene coordenadas $(u_1, \ldots, u_n)$ en la base $\{\hat{e}_1, \ldots, \hat{e}_n\}$, se tiene que

$$\begin{aligned} Df(\hat{x}_0)(\hat{u}) &= \begin{bmatrix} \frac{\partial f}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial f}{\partial x_n}(\hat{x}_0) \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \\ &= u_1 \frac{\partial f}{\partial x_1}(\hat{x}_0) + \cdots + u_n \frac{\partial f}{\partial x_n}(\hat{x}_0). \end{aligned}$$

Así pues, si ahora $x'_1, \ldots, x'_n$ denotan a las "variables" determinadas por otra base ortonormal $\{\hat{e}'_1, \ldots, \hat{e}'_n\}$ de $\mathbb{R}^n$, y se tiene que

$$\begin{aligned} \hat{e}'_i &= \left(a_1^{(i)}, \ldots, a_n^{(i)}\right) \\ &= a_1^{(i)} \hat{e}_1 + \cdots + a_n^{(i)} \hat{e}_n \end{aligned}$$

para cada $i \in \{1, \ldots, n\}$, entonces

$$\begin{aligned} \frac{\partial f}{\partial x'_i}(\hat{x}_0) &= D_{\hat{e}'_i} f(\hat{x}_0) \\ &= Df(\hat{x}_0)\left(\hat{e}'_i\right) \\ &= a_1^{(i)} \frac{\partial f}{\partial x_1}(\hat{x}_0) + \cdots + a_n^{(i)} \frac{\partial f}{\partial x_n}(\hat{x}_0). \end{aligned}$$

Esto establece la forma de obtener las derivadas parciales de f, con respecto a las variables (o coordenadas) $x'_1, \ldots, x'_n$, en términos de las variables $x_1, \ldots, x_n$. Nótese que esta última identidad también se puede obtener como una cosecuencia de la proposición 4.16, puesto que, como $\hat{e}_i'$ es un vector de norma uno, y

$$\hat{e}_i' = a_1^{(i)}\hat{e}_1 + \cdots + a_n^{(i)}\hat{e}_n,$$

entonces

$$\begin{aligned}
\frac{\partial f}{\partial x_i'}(\hat{x}_0) &= D_{\hat{e}_i'}f(\hat{x}_0) \\
&= Df(\hat{x}_0)\,(\hat{e}_i') \\
&= Df(\hat{x}_0)\left(a_1^{(i)}\hat{e}_1 + \cdots + a_n^{(i)}\hat{e}_n\right) \\
&= a_1^{(i)}Df(\hat{x}_0)\,(\hat{e}_1) + \cdots + a_n^{(i)}Df(\hat{x}_0)\,(\hat{e}_n) \\
&= a_1^{(i)}\frac{\partial f}{\partial x_1}(\hat{x}_0) + \cdots + a_n^{(i)}\frac{\partial f}{\partial x_n}(\hat{x}_0).
\end{aligned}$$

Para terminar de encontrar la relación que existe entre las matrices asociadas a la derivada $Df(\hat{x}_0)$ en dos sistemas de referencia (ortonormales) diferentes, recordemos que en la base $\{\hat{e}_1', \ldots, \hat{e}_n'\}$ esta función lineal estará representada por la matriz

$$\begin{bmatrix} \frac{\partial f}{\partial x_1'}(\hat{x}_0) & \cdots & \frac{\partial f}{\partial x_n'}(\hat{x}_0) \end{bmatrix}$$

de modo que las matrices (de $1 \times n$) que representan a la función lineal $Df(\hat{x}_0)$ (en las correspondientes bases $\{\hat{e}_1', \ldots, \hat{e}_n'\}$ y $\{\hat{e}_1, \ldots, \hat{e}_n\}$) están relacionadas por la identidad

$$\begin{bmatrix} \frac{\partial f}{\partial x_1'}(\hat{x}_0) & \cdots & \frac{\partial f}{\partial x_n'}(\hat{x}_0) \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial f}{\partial x_n}(\hat{x}_0) \end{bmatrix} \begin{bmatrix} a_1^{(1)} & \cdots & a_1^{(n)} \\ \vdots & \ddots & \vdots \\ a_n^{(1)} & \cdots & a_n^{(n)} \end{bmatrix} \tag{4.18}$$

o por la identidad

$$\begin{bmatrix} \frac{\partial f}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial f}{\partial x_n}(\hat{x}_0) \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x_1'}(\hat{x}_0) & \cdots & \frac{\partial f}{\partial x_n'}(\hat{x}_0) \end{bmatrix} \begin{bmatrix} a_1^{(1)} & \cdots & a_n^{(1)} \\ \vdots & \ddots & \vdots \\ a_1^{(n)} & \cdots & a_n^{(n)} \end{bmatrix}. \tag{4.19}$$

Estas expresiones, sin duda alguna, harán que el lector recuerde las identidades 4.4 y 4.7, que son las que obtuvimos cuando analizamos el problema relacionado con el "cambio de coordenadas" (de un mismo vector $\hat{x}$) determinadas por dos bases ortonormales de $\mathbb{R}^n$.

4.3.1 El gradiente

Además de sus implicaciones prácticas para el cálculo explícito de la derivada de una función, la proposición 4.16 tiene otra consecuencia importante, que resultará muy útil para conocer el comportamiento de una función en la vecindad de un punto en el cual sea derivable.

Como asegura la proposición 4.16, si f es derivable en un punto $\hat{x}_0$, entonces la derivada direccional de f en $\hat{x}_0$ en la dirección del vector (unitario) $\hat{u}$ se obtiene evaluando la función lineal $Df(\hat{x}_0)$ (la derivada de f en $\hat{x}_0$) en $\hat{u}$. Si ahora recordamos que toda función lineal es continua (problema 65 del capítulo 2) y que el conjunto

$$S^{n-1} = \{\hat{u} \in \mathbb{R}^n \mid \|\hat{u}\| = 1\}$$

es cerrado (problema 51 del capítulo 2) y claramente acotado, por el corolario 2.52 del capítulo 2 sabemos que deben existir $\hat{u}_1$ y $\hat{u}_2$ en S^{n-1} tales que

$$Df(\hat{x}_0)(\hat{u}_1) \le Df(\hat{x}_0)(\hat{u}) \le Df(\hat{x}_0)(\hat{u}_2)$$

para toda $\hat{u} \in S^{n-1}$, o equivalentemente, que

$$D_{\hat{u}_1} f(\hat{x}_0) \leq D_{\hat{u}} f(\hat{x}_0) \leq D_{\hat{u}_2} f(\hat{x}_0)$$

para toda $\hat{u} \in S^{n-1}$.

Es decir, en términos más coloquiales, siempre existen direcciones $\hat{u}_1$ y $\hat{u}_2$, una en la que la razón de cambio de la función f es mínima ($\hat{u}_1$), y otra en la que la razón de cambio de la función f es máxima ($\hat{u}_2$).

De hecho, lo siguiente que probaremos es que si la $Df(\hat{x}_0)$ no es la función lineal constante cero (en cuyo caso todas las derivadas direccionales valen 0), entonces $\hat{u}_1$ y $\hat{u}_2$ son únicos y además $\hat{u}_1 = -\hat{u}_2$.

Proposición 4.19 *Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}$ y $\hat{x}_0 \in U$. Si f es derivable en $\hat{x}_0$, entonces existe $\hat{u}_0 \in S^{n-1}$ tal que*

$$D_{-\hat{u}_0} f(\hat{x}_0) \leq D_{\hat{u}} f(\hat{x}_0) \leq D_{\hat{u}_0} f(\hat{x}_0)$$

para toda $\hat{u} \in S^{n-1}$. Si la derivada $Df(\hat{x}_0)$ no es la constante cero, entonces $\hat{u}_0$ es único.

Demostración. Como mencionamos anteriormente, dado que toda función lineal es continua y que el conjunto S^{n-1} es cerrado y acotado, por el corolario 2.52 del capítulo 2 sabemos que existe $\hat{u}_0 \in S^{n-1}$ tal que

$$D_{\hat{u}} f(\hat{x}_0) \leq D_{\hat{u}_0} f(\hat{x}_0)$$

para toda $\hat{u} \in S^{n-1}$.

Por otra parte, si $\hat{u} \in S^{n-1}$, entonces $-\hat{u} \in S^{n-1}$, de modo que

$$D_{-\hat{u}} f(\hat{x}_0) \leq D_{\hat{u}_0} f(\hat{x}_0).$$

Por tanto, se tiene que

$$\begin{aligned} D_{-\hat{u}_0} f(\hat{x}_0) &= Df(\hat{x}_0)(-\hat{u}_0) \\ &= -Df(\hat{x}_0)(\hat{u}_0) \\ &= -D_{\hat{u}_0} f(\hat{x}_0) \\ &\leq -D_{-\hat{u}} f(\hat{x}_0) \\ &= -Df(\hat{x}_0)(-\hat{u}) \\ &= D_{\hat{u}} f(\hat{x}_0), \end{aligned}$$

de donde obtenemos la otra desigualdad.

Para probar la unicidad de $\hat{u}_0$, supongamos ahora que $Df(\hat{x}_0)$ no es la constante cero (de modo que $Df(\hat{x}_0)(\hat{u}_0) > 0$) y que existe otro $\hat{u}_0' \in S^{n-1}$, $\hat{u}_0' \neq \hat{u}_0$, tal que $Df(\hat{x}_0)(\hat{u}_0') = Df(\hat{x}_0)(\hat{u}_0)$.

Primero notemos que si $\hat{u}_0' = -\hat{u}_0$, entonces $\hat{u}_0' + \hat{u}_0 = \hat{0}$ y por lo tanto

$$\begin{aligned} 0 &= Df(\hat{x}_0)(\hat{0}) \\ &= Df(\hat{x}_0)(\hat{u}_0' + \hat{u}_0) \\ &= Df(\hat{x}_0)(\hat{u}_0') + Df(\hat{x}_0)(\hat{u}_0) \\ &= 2Df(\hat{x}_0)(\hat{u}_0), \end{aligned}$$

de donde se tendría que $Df(\hat{x}_0)(\hat{u}_0) = 0$. Por tanto, $\hat{u}_0' + \hat{u}_0 \neq \hat{0}$, de modo que si tomamos

$$\tilde{u} = \frac{\hat{u}_0' + \hat{u}_0}{\|\hat{u}_0' + \hat{u}_0\|} \in S^{n-1},$$

entonces

$$\begin{aligned} Df(\hat{x}_0)(\tilde{u}) &= \frac{1}{\|\hat{u}_0' + \hat{u}_0\|}(Df(\hat{x}_0)(\hat{u}_0') + Df(\hat{x}_0)(\hat{u}_0)) \\ &= \frac{2}{\|\hat{u}_0' + \hat{u}_0\|} Df(\hat{x}_0)(\hat{u}_0). \end{aligned}$$

Ahora, dado que $\hat{u}_0', \hat{u}_0 \in S^{n-1}$ y que $\hat{u}_0'$ es diferente de $\hat{u}_0$ y $-\hat{u}_0$, por los incisos (a) y (b) del problema 8 del capítulo 1 se tiene que

$$\begin{aligned} \|\hat{u}_0' + \hat{u}_0\| &< \|\hat{u}_0'\| + \|\hat{u}_0\| \\ &= 2, \end{aligned}$$

de donde

$$1 < \frac{2}{\|\hat{u}_0' + \hat{u}_0\|}.$$

Por lo tanto

$$\begin{aligned} Df(\hat{x}_0)(\hat{u}_0) &< \frac{2}{\|\hat{u}_0' + \hat{u}_0\|} Df(\hat{x}_0)(\hat{u}_0) \\ &= Df(\hat{x}_0)(\tilde{u}), \end{aligned}$$

lo cual contradice la propiedad de que $Df(\hat{x}_0)(\hat{u}_0)$ es el valor máximo de $Df(\hat{x}_0)$ sobre el conjunto S^{n-1}. Esta última contradicción concluye la prueba de que $\hat{u}_0$ es único. ■

Con base en el resultado anterior, podemos dar la siguiente

Definición 4.20 *Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}$ y $\hat{x}_0 \in U$. Si f es derivable en $\hat{x}_0$, definimos el gradiente de f en $\hat{x}_0$, que denotamos por $\nabla f(\hat{x}_0)$, como el vector dado por*

$$\nabla f(\hat{x}_0) := D_{\hat{u}_0} f(\hat{x}_0)\hat{u}_0,$$

en donde $\hat{u}_0 \in \mathbb{R}^n$ es el vector unitario cuya existencia se probó en la proposición 4.19.

Como seguramente el lector ya se estará preguntando, el siguiente problema que abordaremos será el de calcular esta dirección $\hat{u}_0$ en la cual la derivada direccional alcanza su máximo valor. Como siempre, este cálculo depende de que se tenga una expresión explícita de la función en cuestión (en algún sistema coordenado), y de la forma en que se represente a su derivada. Justo en la solución de este problema es que echaremos mano de la otra forma en que podemos representar a una función lineal de $\mathbb{R}^n$ en $\mathbb{R}$. Como se recordará, además de usar el lenguaje de matrices para representar a estas funciones lineales, también podemos usar un vector y el producto punto en $\mathbb{R}^n$, como lo escribimos en la expresión 4.14. Con base en esta expresión, tenemos entonces que, si nuestra función f está expresada en términos de las coordenadas $x_1, \ldots, x_n$ (determinadas por la base $\{\hat{e}_1, \ldots, \hat{e}_n\}$) y el vector (unitario) $\hat{u}$ tiene coordenadas $(u_1, \ldots, u_n)$, entonces podemos escribir que

$$\begin{aligned} D_{\hat{u}} f(\hat{x}_0) &= Df(\hat{x}_0)(\hat{u}) \\ &= (Df(\hat{x}_0)(\hat{e}_1), \ldots, Df(\hat{x}_0)(\hat{e}_n)) \cdot (u_1, \ldots, u_n) \\ &= (D_{\hat{e}_1} f(\hat{x}_0), \ldots, D_{\hat{e}_n} f(\hat{x}_0)) \cdot (u_1, \ldots, u_n) \\ &= \left(\frac{\partial f}{\partial x_1}(\hat{x}_0), \ldots, \frac{\partial f}{\partial x_n}(\hat{x}_0) \right) \cdot (u_1, \ldots, u_n). \end{aligned} \tag{4.20}$$

Si ahora recordamos que

$$\begin{aligned} \left(\frac{\partial f}{\partial x_1}(\hat{x}_0), \ldots, \frac{\partial f}{\partial x_n}(\hat{x}_0) \right) \cdot (u_1, \ldots, u_n) &= \left\| \left(\frac{\partial f}{\partial x_1}(\hat{x}_0), \ldots, \frac{\partial f}{\partial x_n}(\hat{x}_0) \right) \right\| \|(u_1, \ldots, u_n)\| \cos(\theta) \\ &= \left\| \left(\frac{\partial f}{\partial x_1}(\hat{x}_0), \ldots, \frac{\partial f}{\partial x_n}(\hat{x}_0) \right) \right\| \cos(\theta) \end{aligned}$$

(identidad 1.7 del capítulo 1), en donde θ es el ángulo formado por estos vectores, concluimos que el máximo valor de $D_{\hat{u}} f(\hat{x}_0) = Df(\hat{x}_0)(\hat{u})$ (suponiendo que $Df(\hat{x}_0)$ no es la constante cero) se alcanza cuando $\cos(\theta) = 1$, es decir si $\theta = 0$, lo que significa que el vector $\hat{u}_0$ de la proposición anterior estará expresado por

$$\hat{u}_0 = \frac{\left(\frac{\partial f}{\partial x_1}(\hat{x}_0), \ldots, \frac{\partial f}{\partial x_n}(\hat{x}_0) \right)}{\left\| \left(\frac{\partial f}{\partial x_1}(\hat{x}_0), \ldots, \frac{\partial f}{\partial x_n}(\hat{x}_0) \right) \right\|}. \tag{4.21}$$

De esta forma, siempre que tomemos un sistema de coordenadas $x_1, \ldots, x_n$ (inducido por una base ortonormal $\{\hat{e}_1, \ldots, \hat{e}_n\}$), el vector con coordenadas

$$\left(\frac{\partial f}{\partial x_1}(\hat{x}_0), \ldots, \frac{\partial f}{\partial x_n}(\hat{x}_0)\right) \tag{4.22}$$

es el gradiente de f en $\hat{x}_0$, es decir

$$\nabla f(\hat{x}_0) = \left(\frac{\partial f}{\partial x_1}(\hat{x}_0), \ldots, \frac{\partial f}{\partial x_n}(\hat{x}_0)\right). \tag{4.23}$$

Es importante insistir en que, si tomamos otro sistema de coordenadas $x'_1, \ldots, x'_n$ (inducido por otra base ortonormal $\{\hat{e}'_1, \ldots, \hat{e}'_n\}$), el vector con coordenadas

$$\left(\frac{\partial f}{\partial x'_1}(\hat{x}_0), \ldots, \frac{\partial f}{\partial x'_n}(\hat{x}_0)\right)$$

nuevamente es tal que

$$\hat{u}_0 = \frac{\left(\frac{\partial f}{\partial x'_1}(\hat{x}_0), \ldots, \frac{\partial f}{\partial x'_n}(\hat{x}_0)\right)}{\left\|\left(\frac{\partial f}{\partial x'_1}(\hat{x}_0), \ldots, \frac{\partial f}{\partial x'_n}(\hat{x}_0)\right)\right\|},$$

en donde $\hat{u}_0$ es el multicitado vector cuya existencia probamos en la proposición 4.19 (prueba que por cierto, hicimos sin recurrir a un sistema coordenado específico).

En virtud de lo anterior, dado un sistema de coordenadas específico $x_1, \ldots, x_n$, las coordenadas del gradiente de f en $\hat{x}_0$ ($\nabla f(\hat{x}_0)$) estarán dadas por 4.23, y para obtener sus coordenadas en otro sistema $x'_1, \ldots, x'_n$, simplemente habrá que recurrir a la identidad 4.18.

Desde un punto de vista más práctico, el vector gradiente de f en $\hat{x}_0$ tiene dos usos importantes: uno, para representar (y expresar) a la derivada de f en $\hat{x}_0$ (identidad 4.20), es decir que

$$Df(\hat{x}_0)(\hat{x}) = \nabla f(\hat{x}_0) \cdot \hat{x}$$

para toda $\hat{x} \in \mathbb{R}^n$; y dos, para designar la dirección en la que f tiene la máxima razón de cambio, es decir, la dirección en la que su derivada direccional (en $\hat{x}_0$) alcanza su máximo valor (identidad 4.21). Con base en lo anterior, lo siguiente que haremos será ilustrar con un ejemplo estos dos usos.

Ejemplo 4.21 *Sea $f : \mathbb{R}^2 \to \mathbb{R}$ que está dada por la expresión $f(x, y) = 8 - x^2 - y^2 + x - y$ y $\hat{x}_0 = (2, 1)$.*

Tenemos entonces que

$$\frac{\partial f}{\partial x}(x, y) = -2x + 1$$

y

$$\frac{\partial f}{\partial y}(x, y) = -2y - 1,$$

de modo que

$$\begin{aligned}\nabla f(2, 1) &= \left(\frac{\partial f}{\partial x}(2, 1), \frac{\partial f}{\partial y}(2, 1)\right) \\ &= (-3, -3).\end{aligned}$$

Por tanto, la derivada de f en $(2, 1)$ ($Df(2, 1)$) es la función lineal dada por la expresión

$$\begin{aligned}Df(2, 1)(x, y) &= \nabla f(2, 1) \cdot (x, y) \\ &= (-3, -3) \cdot (x, y) \\ &= -3x - 3y,\end{aligned}$$

de tal forma que la función afín

$$Df(2, 1)(x - 2, y - 1) + f(2, 1) = -3(x - 2) - 3(y - 1) + 4$$

es la que debe satisfacer la condición 4.13 de la defnición 4.14.

En efecto, se tiene que

$$\begin{aligned}
&\lim_{(x,y)\to(2,1)} \frac{f(x,y)-(Df(2,1)(x-2,y-1)+f(2,1))}{\|(x,y)-(2,1)\|}\\
&= \lim_{(x,y)\to(2,1)} \frac{8-x^2-y^2+x-y-(-3(x-2)-3(y-1)+4)}{\sqrt{(x-2)^2+(y-1)^2}}\\
&= \lim_{(x,y)\to(2,1)} \frac{-x^2+4x-y^2+2y-5}{\sqrt{(x-2)^2+(y-1)^2}}\\
&= \lim_{(x,y)\to(2,1)} \frac{-(x-2)^2-(y-1)^2}{\sqrt{(x-2)^2+(y-1)^2}}\\
&= \lim_{(x,y)\to(2,1)} -\frac{(x-2)^2+(y-1)^2}{\sqrt{(x-2)^2+(y-1)^2}}\\
&= \lim_{(x,y)\to(2,1)} -\sqrt{(x-2)^2+(y-1)^2}\\
&= 0.
\end{aligned}$$

Por otra parte, se tiene que el vector

$$\begin{aligned}
\hat{u}_0 &= \frac{\nabla f(2,1)}{\|\nabla f(2,1)\|}\\
&= \frac{1}{\sqrt{(-3)^2+(-3)^2}}(-3,-3)\\
&= \left(-\frac{1}{\sqrt{2}},-\frac{1}{\sqrt{2}}\right)
\end{aligned}$$

es el vector que determina la dirección en la que, "parados" en el punto $(2,1)$, obtenemos la máxima razón de cambio de la función f.

Dicho en términos geométricos, si "cortamos" a la gráfica de f por el plano perpendicular al plano XY que contiene a la recta que pasa por el punto $(2,1)$, y cuya dirección está dada por el vector $(-1/\sqrt{2},-1/\sqrt{2})$, entonces obtenemos la curva que en el punto $(2,1,f(2,1))=(2,1,4)$ tiene la recta tangente de máxima pendiente (ver figura 4.6).

4.3.2 Otras propiedades de la derivada

Una vez establecidas estas importantes propiedades del gradiente, continuaremos con el análisis de las condiciones necesarias y suficientes de la derivabilidad de una función. La siguiente propiedad que veremos es una condición necesaria, y por lo tanto nos será muy útil para determinar cuándo una función no es derivable. Esta condición es una generalización (al tipo de funciones que estamos tratando) de la propiedad que ya sabíamos para el caso de funciones de $\mathbb{R}$ en $\mathbb{R}$, y que simplemente nos asegura que, si una función es derivable en un punto, entonces debe ser continua en ese mismo punto.

Proposición 4.22 *Sea $f:U\subset\mathbb{R}^n\to\mathbb{R}$. Si f es derivable en $\hat{x}_0\in U$, entonces f es continua en $\hat{x}_0$.*

Demostración. Dado que f es derivable en $\hat{x}_0$, sabemos que

$$\lim_{\hat{x}\to\hat{x}_0} \frac{f(\hat{x})-(Df(\hat{x}_0)(\hat{x}-\hat{x}_0)+f(\hat{x}_0))}{\|\hat{x}-\hat{x}_0\|}=0$$

Como

$$f(\hat{x})-f(\hat{x}_0)-Df(\hat{x}_0)(\hat{x}-\hat{x}_0)=\frac{f(\hat{x})-(Df(\hat{x}_0)(\hat{x}-\hat{x}_0)+f(\hat{x}_0))}{\|\hat{x}-\hat{x}_0\|}\|\hat{x}-\hat{x}_0\|,$$

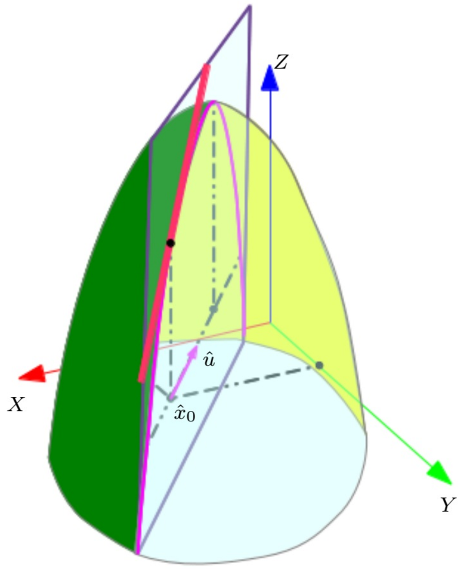


Figura 4.6: En el ejemplo 4.21, el plano perpendicular al plano XY que pasa por el punto $\hat{x}_0 = (2,1)$, y cuya dirección está dada por el vector $\hat{u} = (-1/\sqrt{2}, -1/\sqrt{2})$, contiene a la curva y a su recta tangente, que resulta ser la de máxima pendiente que pasa por el punto $(2,1,f(\hat{x}_0)) = (2,1,4)$.

si $\hat{x} \neq \hat{x}_0$, tenemos entonces que

$$\begin{aligned}\lim_{\hat{x}\to\hat{x}_0} (f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)(\hat{x}-\hat{x}_0)) &= \lim_{\hat{x}\to\hat{x}_0} \left(\frac{f(\hat{x}) - (Df(\hat{x}_0)(\hat{x}-\hat{x}_0) + f(\hat{x}_0))}{\|\hat{x}-\hat{x}_0\|} \|\hat{x}-\hat{x}_0\| \right) \\ &= 0.\end{aligned}$$

Por otra parte, por el problema 65 del capítulo 2 sabemos que toda función lineal es continua, de tal forma que

$$\begin{aligned}\lim_{\hat{x}\to\hat{x}_0} Df(\hat{x}_0)(\hat{x}-\hat{x}_0) &= \lim_{\hat{x}\to\hat{x}_0} (Df(\hat{x}_0)(\hat{x}) - Df(\hat{x}_0)(\hat{x}_0)) \\ &= 0\end{aligned}$$

Por lo tanto

$$\begin{aligned}\lim_{\hat{x}\to\hat{x}_0} (f(\hat{x}) - f(\hat{x}_0)) &= \lim_{\hat{x}\to\hat{x}_0} [(f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)(\hat{x}-\hat{x}_0)) + Df(\hat{x}_0)(\hat{x}-\hat{x}_0)] \\ &= \lim_{\hat{x}\to\hat{x}_0} (f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)(\hat{x}-\hat{x}_0)) + \lim_{\hat{x}\to\hat{x}_0} Df(\hat{x}_0)(\hat{x}-\hat{x}_0) \\ &= 0,\end{aligned}$$

lo que prueba que f es continua en $\hat{x}_0$. ■

Como sucede en el caso real, la condición anterior no es suficiente. Es decir, no es suficiente que una función sea continua en un punto $\hat{x}_0$ para que ésta sea derivable en dicho punto. El ejemplo que vamos a dar de esto es la "generalización" del que se suele dar para ilustrar el mismo hecho en el caso de funciones de $\mathbb{R}$ en $\mathbb{R}$ (la función valor absoluto).

Ejemplo 4.23 *Sea $f : \mathbb{R}^n \to \mathbb{R}$ definida como*

$$f(\hat{x}) = \|\hat{x}\|.$$

Como sabemos, f es continua para toda $\hat{x} \in \mathbb{R}^n$, en particular para $\hat{x} = \hat{0}$. Sin embargo, mostraremos que f no es derivable en $\hat{0}$. Para ello, primero mostraremos que en el $\hat{0}$, para cualquier vector $\hat{u} \in \mathbb{R}^n$ tal que $\|\hat{u}\| = 1$, la derivada direccional $D_{\hat{u}}f(\hat{0})$ no existe.

En efecto, por definición sabemos que

$$\begin{aligned} D_{\hat{u}}f(\hat{0}) &= \lim_{h\to 0} \frac{f(\hat{0}+h\hat{u}) - f(\hat{0})}{h} \\ &= \lim_{h\to 0} \frac{\|h\hat{u}\| - \|\hat{0}\|}{h} \\ &= \lim_{h\to 0} \frac{|h|}{h} \end{aligned}$$

y como el lector recordará, este último límite no existe. Por lo tanto, por la proposición 4.16, f no es derivable en el $\hat{0}$.

Y ya que estamos en los contraejemplos, aprovechando la proposición anterior, mostraremos que el recíproco de la proposición 4.16 no es cierto; es decir, que la existencia de todas las derivadas direccionales de una función f en un punto $\hat{x}_0$ de su dominio no es suficiente para poder asegurar que f sea derivable en dicho punto.

Ejemplo 4.24 *Sea $f : \mathbb{R}^2 \to \mathbb{R}$ definida como*

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

Mostraremos primero que la derivada direccional $D_{\hat{u}}f(\hat{0})$ existe para cualquier vector $\hat{u} = (u_1, u_2) \in \mathbb{R}^n$ tal que $\|\hat{u}\| = 1$.

Supongamos primero que $u_2 \neq 0$. En este caso se tiene que

$$\begin{aligned} D_{\hat{u}}f(\hat{0}) &= \lim_{h\to 0} \frac{f(\hat{0}+h\hat{u}) - f(\hat{0})}{h} \\ &= \lim_{h\to 0} \frac{1}{h}\left(\frac{(hu_1)^2(hu_2)}{(hu_1)^4 + (hu_2)^2}\right) \\ &= \lim_{h\to 0} \frac{1}{h}\left(\frac{h^3u_1^2u_2}{h^2\left(h^2u_1^4 + u_2^2\right)}\right) \\ &= \lim_{h\to 0} \frac{u_1^2u_2}{h^2u_1^4 + u_2^2} \\ &= \frac{u_1^2}{u_2} \end{aligned}$$

y si $u_2 = 0$, es decir que $\hat{u} = (1,0) = \hat{e}_1$, entonces

$$\begin{aligned} D_{\hat{u}}f(\hat{0}) &= \lim_{h\to 0} \frac{f(\hat{0}+h\hat{u}) - f(\hat{0})}{h} \\ &= \lim_{h\to 0} \frac{f(h,0) - f(0,0)}{h} \\ &= \lim_{h\to 0} \frac{0-0}{h} \\ &= 0. \end{aligned}$$

Esto prueba que $D_{\hat{u}}f(\hat{0})$ existe para cualquier vector $\hat{u} \in \mathbb{R}^n$ tal que $\|\hat{u}\| = 1$.

Por otra parte, si evaluamos a f en puntos de la forma (t, t^2), con $t \neq 0$, se tiene que

$$f(t,t^2) = \frac{(t)^2t^2}{(t)^4 + (t^2)^2}$$

$$= \frac{t^4}{t^4 + t^4}$$
$$= \frac{1}{2},$$

de tal forma que, como $(t, t^2) \to (0,0)$ si $t \to 0$, concluimos que f no es continua en el $\hat{0}$ y por la proposición 4.22, que f no es derivable en el $\hat{0}$.

Si bien es cierto que la existencia de las derivadas direccionales de una función f en un punto $\hat{x}_0$ no es suficiente para garantizar su derivabilidad en dicho punto, un análisis más de cerca del ejemplo anterior nos puede arrojar luz sobre cuáles hipótesis adicionales habría que considerar para poder asegurar la derivabilidad de f en $\hat{x}_0$.

Lo primero que haremos será mostrar que la función del ejemplo anterior sí es derivable en cualquier punto $\hat{x} = (x, y) \neq \hat{0}$. Para ello, como ya sabemos, será necesario calcular las derivadas parciales $\partial f / \partial x$ y $\partial f / \partial y$ en $\hat{x}$ y mostrar que la función lineal representada por la matriz

$$\begin{bmatrix} \frac{\partial f}{\partial x}(x,y) & \frac{\partial f}{\partial y}(x,y) \end{bmatrix}$$

satisface la condición establecida en la definición 4.14.

Con base en varios de los incisos de la proposición 4.11 y en los cálculos realizados en el ejemplo anterior, tenemos que

$$\frac{\partial f}{\partial x}(x,y) = \begin{cases} \frac{2xy(y^2 - x^4)}{(x^4+y^2)^2} & \text{si } (x,y) \neq (0,0) \\ \\ 0 & \text{si } (x,y) = (0,0) \end{cases} \tag{4.24}$$

y

$$\frac{\partial f}{\partial y}(x,y) = \begin{cases} \frac{x^2(x^4 - y^2)}{(x^4+y^2)^2} & \text{si } (x,y) \neq (0,0) \\ \\ 0 & \text{si } (x,y) = (0,0) \end{cases} \tag{4.25}$$

Si ahora tomamos $\hat{x}_0 = (x_0, y_0) \neq (0,0)$, para cualquier otra $\hat{x} = (x, y) \in \mathbb{R}^2$ se tiene que

$$\begin{aligned} &|f(\hat{x}) - (Df(\hat{x}_0)(\hat{x} - \hat{x}_0) + f(\hat{x}_0))| \\ &= \left| f(x,y) - f(x_0,y_0) - \frac{\partial f}{\partial x}(\hat{x}_0)(x - x_0) - \frac{\partial f}{\partial y}(\hat{x}_0)(y - y_0) \right| \\ &= \left| f(x,y) - f(x_0,y) + f(x_0,y) - f(x_0,y_0) - \frac{\partial f}{\partial x}(\hat{x}_0)(x - x_0) - \frac{\partial f}{\partial y}(\hat{x}_0)(y - y_0) \right|. \end{aligned}$$

Dado que las derivadas parciales $\partial f / \partial x$ y $\partial f / \partial y$ existen para toda $\hat{x} \in \mathbb{R}^2$, por la proposición 4.13 sabemos que

$$f(x,y) - f(x_0,y) = (x - x_0)\frac{\partial f}{\partial x}(\xi_x, y) \qquad \text{y} \qquad f(x_0,y) - f(x_0,y_0) = (y - y_0)\frac{\partial f}{\partial y}(x_0, \xi_y),$$

y por tanto

$$\begin{aligned} &|f(\hat{x}) - (Df(\hat{x}_0)(\hat{x} - \hat{x}_0) + f(\hat{x}_0))| \\ &\quad = \left| f(x,y) - f(x_0,y) + f(x_0,y) - f(x_0,y_0) - \frac{\partial f}{\partial x}(\hat{x}_0)(x - x_0) - \frac{\partial f}{\partial y}(\hat{x}_0)(y - y_0) \right| \\ &\quad = \left| (x - x_0)\frac{\partial f}{\partial x}(\xi_x, y) + (y - y_0)\frac{\partial f}{\partial y}(x_0, \xi_y) - \frac{\partial f}{\partial x}(\hat{x}_0)(x - x_0) - \frac{\partial f}{\partial y}(\hat{x}_0)(y - y_0) \right| \\ &\quad = \left| \left(\frac{\partial f}{\partial x}(\xi_x, y) - \frac{\partial f}{\partial x}(\hat{x}_0) \right)(x - x_0) + \left(\frac{\partial f}{\partial y}(x_0, \xi_y) - \frac{\partial f}{\partial y}(\hat{x}_0) \right)(y - y_0) \right| \\ &\quad \leq \left| \frac{\partial f}{\partial x}(\xi_x, y) - \frac{\partial f}{\partial x}(\hat{x}_0) \right| |x - x_0| + \left| \frac{\partial f}{\partial y}(x_0, \xi_y) - \frac{\partial f}{\partial y}(\hat{x}_0) \right| |y - y_0| \end{aligned}$$

$$
\begin{aligned}
&\leq \left|\frac{\partial f}{\partial x}(\xi_x, y) - \frac{\partial f}{\partial x}(\hat{x}_0)\right| \|\hat{x} - \hat{x}_0\| + \left|\frac{\partial f}{\partial y}(x_0, \xi_y) - \frac{\partial f}{\partial y}(\hat{x}_0)\right| \|\hat{x} - \hat{x}_0\| \\
&= \left(\left|\frac{\partial f}{\partial x}(\xi_x, y) - \frac{\partial f}{\partial x}(\hat{x}_0)\right| + \left|\frac{\partial f}{\partial y}(x_0, \xi_y) - \frac{\partial f}{\partial y}(\hat{x}_0)\right|\right) \|\hat{x} - \hat{x}_0\|.
\end{aligned}
$$

De aquí obtenemos que

$$
\begin{aligned}
0 &\leq \frac{|f(\hat{x}) - (Df(\hat{x}_0)(\hat{x} - \hat{x}_0) + f(\hat{x}_0))|}{\|\hat{x} - \hat{x}_0\|} \\
&\leq \left|\frac{\partial f}{\partial x}(\xi_x, y) - \frac{\partial f}{\partial x}(\hat{x}_0)\right| + \left|\frac{\partial f}{\partial y}(x_0, \xi_y) - \frac{\partial f}{\partial y}(\hat{x}_0)\right|.
\end{aligned}
$$

Si ahora notamos que, de acuerdo con las expresiones 4.24 y 4.25 de las derivadas parciales de f, éstas son continuas en cualquier punto $\hat{x}_0 \neq \hat{0}$, y que los puntos (ξ_x, y) y (x_0, ξ_y) tienden al punto $(x_0, y_0) = \hat{x}_0$ si el punto $\hat{x} = (x, y)$ tiende al punto $(x_0, y_0) = \hat{x}_0$, tenemos que

$$
\begin{aligned}
0 &\leq \lim_{\hat{x} \to \hat{x}_0} \frac{|f(\hat{x}) - (Df(\hat{x}_0)(\hat{x} - \hat{x}_0) + f(\hat{x}_0))|}{\|\hat{x} - \hat{x}_0\|} \\
&\leq \lim_{\hat{x} \to \hat{x}_0} \left(\left|\frac{\partial f}{\partial x}(\xi_x, y) - \frac{\partial f}{\partial x}(\hat{x}_0)\right| + \left|\frac{\partial f}{\partial y}(x_0, \xi_y) - \frac{\partial f}{\partial y}(\hat{x}_0)\right|\right) \\
&= 0 + 0 \\
&= 0,
\end{aligned}
$$

de donde concluimos que

$$
\lim_{\hat{x} \to \hat{x}_0} \frac{|f(\hat{x}) - (Df(\hat{x}_0)(\hat{x} - \hat{x}_0) + f(\hat{x}_0))|}{\|\hat{x} - \hat{x}_0\|} = 0,
$$

o equivalentemente, que

$$
\lim_{\hat{x} \to \hat{x}_0} \frac{f(\hat{x}) - (Df(\hat{x}_0)(\hat{x} - \hat{x}_0) + f(\hat{x}_0))}{\|\hat{x} - \hat{x}_0\|} = 0,
$$

es decir, que f es derivable en $\hat{x}_0$.

Ojalá y el lector esté de acuerdo en que el arduo trabajo que nos tomamos para probar que la función del ejemplo 4.24 es derivable para cualquier $\hat{x}_0 \in \mathbb{R}^2$ distinto del $\hat{0}$ será redituable. En efecto, nótese que más allá de la forma explícita de f, lo importante en la argumentación anterior fueron dos cosas:

1. que las derivadas parciales de f existieron para todos los puntos de $\mathbb{R}^2$, y
2. que fueron continuas en $\hat{x}_0$.

Con respecto al punto 1, el lector estará de acuerdo (dado el caracter "local" de la derivada), en que en el desarrollo anterior hubiera bastado con que las derivadas parciales existieran en una vecindad del punto $\hat{x}_0$. Pues bien, lo que probaremos a continuación es que estas dos condiciones (con la modificación propuesta a la primera) son suficientes para poder asegurar que una función es derivable en un punto $\hat{x}_0$, lo que dejaremos formulado en la siguiente

Proposición 4.25 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$ y $r > 0$ tal que $B_r(\hat{x}_0) \subset U$. Si $\frac{\partial f}{\partial x_i}(\hat{x})$ existe para cada $\hat{x} \in B_r(\hat{x}_0)$ y $\frac{\partial f}{\partial x_i}$ es continua en $\hat{x}_0$ (para cada $i \in \{1, \ldots, n\}$), entonces f es derivable en $\hat{x}_0$.*

Demostración. Sea $\hat{x} \in B_r(\hat{x}_0)$, $\hat{x} \neq \hat{x}_0$. Si $\hat{x}_0 = \left(x_1^{(0)}, \ldots, x_n^{(0)}\right)$ y $\hat{x} = (x_1, \ldots, x_n)$, hacemos

$$
\hat{x}_i = \left(x_1, \ldots, x_i, x_{i+1}^{(0)}, \ldots, x_n^{(0)}\right)
$$

para cada $i \in \{1, \ldots, n-1\}$ y $\hat{x}_n = \hat{x}$. Por el inciso (a) del problema 12 del capítulo 1 sabemos que $\hat{x}_i \in B_r(\hat{x}_0)$ para cada $i \in \{1, \ldots, n-1\}$ y como $B_r(\hat{x}_0)$ es un conjunto convexo, entonces los segmentos $[\hat{x}_{i-1}, \hat{x}_i] \subset B_r(\hat{x}_0)$ para cada $i \in \{1, \ldots, n\}$.

Por otra parte, nótese que

$$\begin{aligned}
f(\hat{x}) - f(\hat{x}_0) &= (f(\hat{x}) - f(\hat{x}_{n-1})) + (f(\hat{x}_{n-1}) - f(\hat{x}_{n-2})) + \cdots + (f(\hat{x}_2) - f(\hat{x}_1)) + (f(\hat{x}_1) - f(\hat{x}_0)) \\
&= \sum_{i=1}^{n} (f(\hat{x}_i) - f(\hat{x}_{i-1})).
\end{aligned}$$

Ahora, dado que

$$\hat{x}_i - \hat{x}_{i-1} = \left(x_i - x_i^{(0)}\right)\hat{e}_i,$$

por la proposición 4.13 sabemos que existe $\hat{\xi}_i \in [\hat{x}_{i-1}, \hat{x}_i]$ tal que

$$f(\hat{x}_i) - f(\hat{x}_{i-1}) = \left(x_i - x_i^{(0)}\right)\frac{\partial f}{\partial x_i}\left(\hat{\xi}_i\right)$$

para cada $i \in \{1, \ldots, n\}$, de modo que

$$\begin{aligned}
f(\hat{x}) - f(\hat{x}_0) &= \sum_{i=1}^{n} (f(\hat{x}_i) - f(\hat{x}_{i-1})) \\
&= \sum_{i=1}^{n} \left(x_i - x_i^{(0)}\right)\frac{\partial f}{\partial x_i}\left(\hat{\xi}_i\right) \\
&= \left(\frac{\partial f}{\partial x_1}\left(\hat{\xi}_1\right), \frac{\partial f}{\partial x_2}\left(\hat{\xi}_2\right), \ldots, \frac{\partial f}{\partial x_2}\left(\hat{\xi}_n\right)\right) \cdot (\hat{x} - \hat{x}_0).
\end{aligned}$$

Por lo tanto

$$\begin{aligned}
&|f(\hat{x}) - (Df(\hat{x}_0)(\hat{x} - \hat{x}_0) + f(\hat{x}_0))| \\
&\quad = |f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)(\hat{x} - \hat{x}_0)| \\
&\quad = \left|\left(\frac{\partial f}{\partial x_1}\left(\hat{\xi}_1\right), \frac{\partial f}{\partial x_2}\left(\hat{\xi}_2\right), \ldots, \frac{\partial f}{\partial x_2}\left(\hat{\xi}_n\right)\right) \cdot (\hat{x} - \hat{x}_0) - \nabla f(\hat{x}_0) \cdot (\hat{x} - \hat{x}_0)\right| \\
&\quad = \left|\left(\left(\frac{\partial f}{\partial x_1}\left(\hat{\xi}_1\right), \frac{\partial f}{\partial x_2}\left(\hat{\xi}_2\right), \ldots, \frac{\partial f}{\partial x_2}\left(\hat{\xi}_n\right)\right) - \nabla f(\hat{x}_0)\right) \cdot (\hat{x} - \hat{x}_0)\right| \\
&\quad \leq \left\|\left(\left(\frac{\partial f}{\partial x_1}\left(\hat{\xi}_1\right), \frac{\partial f}{\partial x_2}\left(\hat{\xi}_2\right), \ldots, \frac{\partial f}{\partial x_2}\left(\hat{\xi}_n\right)\right) - \nabla f(\hat{x}_0)\right)\right\| \|(\hat{x} - \hat{x}_0)\|,
\end{aligned}$$

de modo que

$$\begin{aligned}
0 &\leq \frac{|f(\hat{x}) - (Df(\hat{x}_0)(\hat{x} - \hat{x}_0) + f(\hat{x}_0))|}{\|\hat{x} - \hat{x}_0\|} \\
&\leq \left\|\left(\left(\frac{\partial f}{\partial x_1}\left(\hat{\xi}_1\right), \frac{\partial f}{\partial x_2}\left(\hat{\xi}_2\right), \ldots, \frac{\partial f}{\partial x_2}\left(\hat{\xi}_n\right)\right) - \nabla f(\hat{x}_0)\right)\right\|.
\end{aligned}$$

Si ahora recordamos que por el inciso (b) del problema 12 del capítulo 1 se tiene que

$$\left\|\hat{\xi}_i - \hat{x}_0\right\| \leq \|\hat{x} - \hat{x}_0\|,$$

entonces $\hat{\xi}_i \to \hat{x}_0$ si $\hat{x} \to \hat{x}_0$, de modo que, como $\frac{\partial f}{\partial x_i}$ es continua en $\hat{x}_0$ (para cada $i \in \{1, \ldots, n\}$), tenemos que

$$\begin{aligned}
0 &\leq \lim_{\hat{x} \to \hat{x}_0} \frac{|f(\hat{x}) - (Df(\hat{x}_0)(\hat{x} - \hat{x}_0) + f(\hat{x}_0))|}{\|\hat{x} - \hat{x}_0\|} \\
&\leq \lim_{\hat{x} \to \hat{x}_0} \left\|\left(\left(\frac{\partial f}{\partial x_1}\left(\hat{\xi}_1\right), \frac{\partial f}{\partial x_2}\left(\hat{\xi}_2\right), \ldots, \frac{\partial f}{\partial x_2}\left(\hat{\xi}_n\right)\right) - \nabla f(\hat{x}_0)\right)\right\|
\end{aligned}$$

$$
\begin{aligned}
&= \|\nabla f(\hat{x}_0) - \nabla f(\hat{x}_0)\| \\
&= 0.
\end{aligned}
$$

Es decir,

$$
\lim_{\hat{x} \to \hat{x}_0} \frac{|f(\hat{x}) - (Df(\hat{x}_0)(\hat{x} - \hat{x}_0) + f(\hat{x}_0))|}{\|\hat{x} - \hat{x}_0\|} = 0,
$$

lo cual prueba que f es derivable en $\hat{x}_0$.

■

Es importante hacer notar que la condición de la proposición anterior sólo es una condición suficiente para la derivabilidad de f en $\hat{x}_0$; es decir, si bien es cierto que la existencia de la derivada de f en $\hat{x}_0$ garantiza la existencia de todas sus derivadas direccionales en este punto (incluyendo sus derivadas parciales), este hecho ni siquiera garantiza que existan todas las derivadas parciales de f en todos los puntos de alguna vecindad de $\hat{x}_0$.

Esta última afirmación queda ilustrada en el siguiente

Ejemplo 4.26 *Sea $f : \mathbb{R}^2 \to \mathbb{R}$ definida como*

$$
f(x, y) = |xy| .
$$

Nótese que

$$
\begin{aligned}
\frac{\partial f}{\partial x}(0,0) &= \lim_{h \to 0} \frac{f(h,0) - f(0,0)}{h} \\
&= \lim_{h \to 0} \frac{|h \cdot 0| - |0 \cdot 0|}{h} \\
&= 0
\end{aligned}
$$

y que

$$
\begin{aligned}
\frac{\partial f}{\partial y}(0,0) &= \lim_{h \to 0} \frac{f(0,h) - f(0,0)}{h} \\
&= \lim_{h \to 0} \frac{|0h| - |0 \cdot 0|}{h} \\
&= 0.
\end{aligned}
$$

Por lo tanto, tomando a $Df(0,0)$ como la constante 0, tenemos que

$$
\begin{aligned}
0 &\leq \lim_{(x,y) \to (0,0)} \frac{f(x,y) - (Df(0,0)((x,y) - (0,0)) + f(0,0))}{\|(x,y) - (0,0)\|} \\
&= \lim_{(x,y) \to (0,0)} \frac{|xy|}{\|(x,y)\|} \\
&\leq \lim_{(x,y) \to (0,0)} \frac{\|(x,y)\|^2}{\|(x,y)\|} \\
&= \lim_{(x,y) \to (0,0)} \|(x,y)\| \\
&= 0,
\end{aligned}
$$

de modo que

$$
\lim_{(x,y) \to (0,0)} \frac{f(x,y) - (Df(0,0)((x,y) - (0,0)) + f(0,0))}{\|(x,y) - (0,0)\|} = 0,
$$

lo que prueba que f es derivable en el $(0,0)$.

Por otra parte, nótese que tomando cualquier punto de la forma $(x_0, 0)$*, con* $x_0 \neq 0$*, de acuerdo con la definición de derivada parcial, se tiene que*

$$\begin{aligned}\frac{\partial f}{\partial y}(x_0,0) &= \lim_{h\to 0}\frac{f(x_0,h)-f(x_0,0)}{h}\\ &= \lim_{h\to 0}\frac{|x_0h|-0}{h}\\ &= |x_0|\lim_{h\to 0}\frac{|h|}{h}.\end{aligned}$$

Como este último límite no existe, entonces $\frac{\partial f}{\partial y}(x_0,0)$ *no existe.*

Análogamente, para cualquier punto de la forma $(0, y_0)$*, con* $y_0 \neq 0$*, se tiene que*

$$\begin{aligned}\frac{\partial f}{\partial x}(0,y_0) &= \lim_{h\to 0}\frac{f(h,y_0)-f(0,y_0)}{h}\\ &= \lim_{h\to 0}\frac{|hy_0|-0}{h}\\ &= |y_0|\lim_{h\to 0}\frac{|h|}{h},\end{aligned}$$

de modo que $\frac{\partial f}{\partial x}(0,y_0)$ *tampoco existe.*

Por tanto, en cualquier vecindad del $(0,0)$ *se tiene que existen puntos para los cuales* $\frac{\partial f}{\partial x}$ *y* $\frac{\partial f}{\partial y}$ *no existen, que es lo que deseábamos mostrar (ver figura 4.7).*

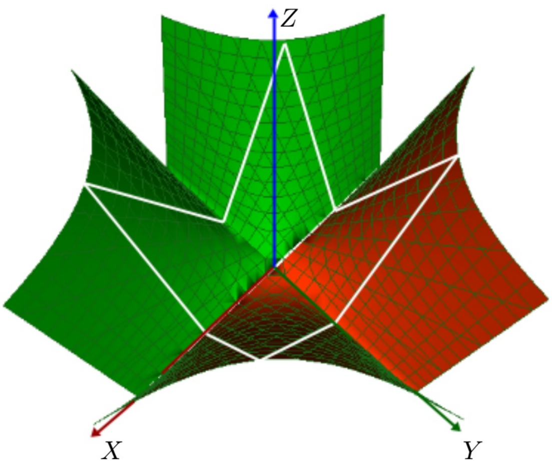


Figura 4.7: Gráfica de la función $f(x,y) = |xy|$ (del ejemplo 4.26) cuyas derivadas parciales no existen en los puntos de la forma $(x,0)$, con $x \neq 0$ y $(0,y)$, con $y \neq 0$.

Aun cuando el ejemplo anterior muestra sin lugar a dudas que la proposición 4.25 sólo nos proporciona una condición suficiente para la derivabilidad de una función en un punto, nos tomaremos el trabajo de dar otro ejemplo en el que las derivadas parciales, a diferencia del ejemplo anterior, sí existen en una vecindad del punto en cuestión, pero no son continuas en el punto en el que la función sí es derivable.

Ejemplo 4.27 *Sea* $f : \mathbb{R}^2 \to \mathbb{R}$ *definida como*

$$f(x,y) = \begin{cases} (x^2+y^2)\operatorname{sen}\left(\frac{1}{\sqrt{x^2+y^2}}\right) & si\ (x,y) \neq (0,0)\\[2ex] 0 & si\ (x,y) = (0,0)\end{cases}$$

En este caso tenemos que

$$\begin{aligned}\frac{\partial f}{\partial x}(0,0) &= \lim_{h\to 0}\frac{f(h,0)-f(0,0)}{h}\\ &= \lim_{h\to 0} h\,\mathrm{sen}\left(\frac{1}{|h|}\right)\\ &= 0\end{aligned}$$

y

$$\begin{aligned}\frac{\partial f}{\partial y}(0,0) &= \lim_{h\to 0}\frac{f(0,h)-f(0,0)}{h}\\ &= \lim_{h\to 0} h\,\mathrm{sen}\left(\frac{1}{|h|}\right)\\ &= 0.\end{aligned}$$

Si $(x,y)\neq(0,0)$, con base en la proposición 4.11, obtenemos que

$$\frac{\partial f}{\partial x}(x,y) = \begin{cases} 2x\,\mathrm{sen}\left(\frac{1}{\sqrt{x^2+y^2}}\right) - \frac{x}{\sqrt{x^2+y^2}}\cos\left(\frac{1}{\sqrt{x^2+y^2}}\right) & \text{si } (x,y)\neq(0,0)\\ 0 & \text{si } (x,y)=(0,0)\end{cases}$$

y

$$\frac{\partial f}{\partial y}(x,y) = \begin{cases} 2y\,\mathrm{sen}\left(\frac{1}{\sqrt{x^2+y^2}}\right) - \frac{y}{\sqrt{x^2+y^2}}\cos\left(\frac{1}{\sqrt{x^2+y^2}}\right) & \text{si } (x,y)\neq(0,0)\\ 0 & \text{si } (x,y)=(0,0)\end{cases}$$

Si ahora tomamos puntos de la forma $(t,0)$, con $t>0$, se tiene que

$$\frac{\partial f}{\partial x}(t,0) = 2t\,\mathrm{sen}\left(\frac{1}{t}\right) - \cos\left(\frac{1}{t}\right)$$

y como

$$\lim_{t\to 0}\left(2t\,\mathrm{sen}\left(\frac{1}{t}\right) - \cos\left(\frac{1}{t}\right)\right)$$

no existe, concluimos que

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

tampoco existe, de modo que $\frac{\partial f}{\partial x}$ no es continua en el $(0,0)$. Análogamente, evaluando en puntos de la forma $(0,t)$, con $t>0$, concluimos que $\frac{\partial f}{\partial y}$ no es continua en el $(0,0)$.

Por otra parte, se tiene que

$$\begin{aligned}\lim_{(x,y)\to(0,0)}\frac{f(x,y)-(Df(0,0)((x,y)-(0,0))+f(0,0))}{\|(x,y)-(0,0)\|} &= \lim_{(x,y)\to(0,0)}\frac{(x^2+y^2)\,\mathrm{sen}\left(\frac{1}{\sqrt{x^2+y^2}}\right)}{\|(x,y)\|}\\ &= \lim_{(x,y)\to(0,0)}\|(x,y)\|\,\mathrm{sen}\left(\frac{1}{\sqrt{x^2+y^2}}\right)\\ &= 0,\end{aligned}$$

lo que demuestra que f sí es derivable en el $(0,0)$.

Con el ejemplo anterior concluimos el análisis de las condiciones necesarias y suficientes relacionadas con la derivabilidad de una función. Lo siguiente que haremos será mostrar la relación entre este concepto y la aritmética de las funciones, la cual dejaremos plasmada en la siguiente

Proposición 4.28 *Sean* $f, g : U \subset \mathbb{R}^n \to \mathbb{R}$ *y* $\hat{x}_0 \in U$. *Si* f *y* g *son derivables en* $\hat{x}_0$, *se satisfacen las siguientes afirmaciones:*

1. $f + g$ *es derivable en* $\hat{x}_0$ *y además*
$$D(f+g)(\hat{x}_0) = Df(\hat{x}_0) + Dg(\hat{x}_0)$$

2. *si* $\alpha \in \mathbb{R}$, *entonces* αf *es derivable en* $\hat{x}_0$ *y además*
$$D(\alpha f)(\hat{x}_0) = \alpha Df(\hat{x}_0)$$

3. fg *es derivable en* $\hat{x}_0$ *y además*
$$D(fg)(\hat{x}_0) = f(\hat{x}_0)Dg(\hat{x}_0) + g(\hat{x}_0)Df(\hat{x}_0)$$

4. *si* $g(\hat{x}_0) \neq 0$, *entonces* f/g *es derivable en* $\hat{x}_0$ *y además*
$$D(f/g)(\hat{x}_0) = \frac{1}{g^2(\hat{x}_0)}\left(g(\hat{x}_0)Df(\hat{x}_0) - f(\hat{x}_0)Dg(\hat{x}_0)\right)$$

Demostración. A manera de ejemplo, probaremos el inciso 3 y el resto quedarán como problemas para el lector. Una vez dicho lo anterior, nótese que

$$\begin{aligned}
(fg)(\hat{x}) - (fg)(\hat{x}_0) - &\left[f(\hat{x}_0)Dg(\hat{x}_0) + g(\hat{x}_0)Df(\hat{x}_0)\right](\hat{x} - \hat{x}_0) \\
&= f(\hat{x})\left[g(\hat{x}) - g(\hat{x}_0)\right] + g(\hat{x}_0)\left[f(\hat{x}) - f(\hat{x}_0)\right] - \left[f(\hat{x}_0)Dg(\hat{x}_0) + g(\hat{x}_0)Df(\hat{x}_0)\right](\hat{x} - \hat{x}_0) \\
&= f(\hat{x})\left[g(\hat{x}) - g(\hat{x}_0) - Dg(\hat{x}_0)(\hat{x} - \hat{x}_0)\right] + \left[f(\hat{x}) - f(\hat{x}_0)\right]Dg(\hat{x}_0)(\hat{x} - \hat{x}_0) \\
&+ g(\hat{x}_0)\left[f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)(\hat{x} - \hat{x}_0)\right],
\end{aligned}$$

de tal forma que

$$\begin{aligned}
&\frac{(fg)(\hat{x}) - (fg)(\hat{x}_0) - (f(\hat{x}_0)Dg(\hat{x}_0) + g(\hat{x}_0)Df(\hat{x}_0))(\hat{x} - \hat{x}_0)}{\|\hat{x} - \hat{x}_0\|} \\
&\qquad = f(\hat{x})\frac{g(\hat{x}) - g(\hat{x}_0) - Dg(\hat{x}_0)(\hat{x} - \hat{x}_0)}{\|\hat{x} - \hat{x}_0\|} + \left[f(\hat{x}) - f(\hat{x}_0)\right]\frac{Dg(\hat{x}_0)(\hat{x} - \hat{x}_0)}{\|\hat{x} - \hat{x}_0\|} \\
&\qquad + g(\hat{x}_0)\frac{f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)(\hat{x} - \hat{x}_0)}{\|\hat{x} - \hat{x}_0\|}.
\end{aligned}$$

Para el primer y tercer sumando que aparecen en el lado derecho de esta última identidad, dado que f y g son derivables en $\hat{x}_0$ (y por lo tanto continuas en ese mismo punto), podemos concluir que

$$\begin{aligned}
\lim_{\hat{x} \to \hat{x}_0} f(\hat{x})\frac{g(\hat{x}) - g(\hat{x}_0) - Dg(\hat{x}_0)(\hat{x} - \hat{x}_0)}{\|\hat{x} - \hat{x}_0\|} &= \left(\lim_{\hat{x} \to \hat{x}_0} f(\hat{x})\right)\left(\lim_{\hat{x} \to \hat{x}_0} \frac{g(\hat{x}) - g(\hat{x}_0) - Dg(\hat{x}_0)(\hat{x} - \hat{x}_0)}{\|\hat{x} - \hat{x}_0\|}\right) \\
&= f(\hat{x}_0)(0) \\
&= 0
\end{aligned}$$

y

$$\begin{aligned}
\lim_{\hat{x} \to \hat{x}_0} g(\hat{x}_0)\frac{f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)(\hat{x} - \hat{x}_0)}{\|\hat{x} - \hat{x}_0\|} &= g(\hat{x}_0)\lim_{\hat{x} \to \hat{x}_0} \frac{f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)(\hat{x} - \hat{x}_0)}{\|\hat{x} - \hat{x}_0\|} \\
&= g(\hat{x}_0)(0) \\
&= 0.
\end{aligned}$$

Por otra parte, por el inciso (c) del problema 11 del capítulo 2, sabemos que existe $M \geq 0$ tal que

$$\left|Df(\hat{x}_0)(\hat{x} - \hat{x}_0)\right| \leq M\,\|\hat{x} - \hat{x}_0\|$$

para toda $\hat{x} \in \mathbb{R}^n$, y como f es continua en $\hat{x}_0$, concluimos que

$$
\begin{aligned}
0 &\leq \lim_{\hat{x} \to \hat{x}_0} \left| [f(\hat{x}) - f(\hat{x}_0)] \frac{Dg(\hat{x}_0)\,(\hat{x} - \hat{x}_0)}{\|\hat{x} - \hat{x}_0\|} \right| \\
&\leq M \lim_{\hat{x} \to \hat{x}_0} |f(\hat{x}) - f(\hat{x}_0)| \\
&= M(0) \\
&= 0
\end{aligned}
$$

y por lo tanto que

$$
\lim_{\hat{x} \to \hat{x}_0} [f(\hat{x}) - f(\hat{x}_0)] \frac{Dg(\hat{x}_0)\,(\hat{x} - \hat{x}_0)}{\|\hat{x} - \hat{x}_0\|} = 0.
$$

Sumando estos límites, concluimos que

$$
\lim_{\hat{x} \to \hat{x}_0} \frac{(fg)(\hat{x}) - (fg)(\hat{x}_0) - (f(\hat{x}_0)Dg(\hat{x}_0) + g(\hat{x}_0)Df(\hat{x}_0))\,(\hat{x} - \hat{x}_0)}{\|\hat{x} - \hat{x}_0\|} = 0,
$$

de modo que la función fg es derivable en $\hat{x}_0$ y además que

$$
D(fg)(\hat{x}_0) = f(\hat{x}_0)Dg(\hat{x}_0) + g(\hat{x}_0)Df(\hat{x}_0).
$$

■

Para terminar con las propiedades de la derivada relacionadas con las operaciones entre funciones, formularemos (en dos proposiciones separadas) aquellas que nos hablan de la composición de funciones.

Serán dos proposiciones diferentes puesto que existen dos formas de componer a una función de $\mathbb{R}^n$ en $\mathbb{R}$ con funciones para las cuales también tenemos una definición de derivada: por la izquierda, con una función de $\mathbb{R}$ en $\mathbb{R}$, o por la derecha, con una función de $\mathbb{R}$ en $\mathbb{R}^n$. Ambas proposiciones serán un caso particular de un resultado más general que veremos más adelante (y que se le conoce con el nombre de "regla de la cadena").

Proposición 4.29 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$, $(a, b) \subset \mathbb{R}$ tal que $f(U) \subset (a, b)$ y $g : (a, b) \subset \mathbb{R} \to \mathbb{R}$. Si f es derivable en $\hat{x}_0$ y g es derivable en $f(\hat{x}_0)$, entonces $g \circ f$ es derivable en $\hat{x}_0$ y además*

$$
D(g \circ f)(\hat{x}_0) = g'(f(\hat{x}_0))Df(\hat{x}_0)
$$

Demostración. Recordemos que, de acuerdo con la definición 4.14, lo que tenemos que demostrar es que

$$
\lim_{\hat{x} \to \hat{x}_0} \frac{(g \circ f)(\hat{x}) - (g \circ f)(\hat{x}_0) - g'(f(\hat{x}_0))Df(\hat{x}_0)(\hat{x} - \hat{x}_0)}{\|\hat{x} - \hat{x}_0\|} = 0,
$$

razón por la cual empezaremos por buscar una expresión "adecuadamente equivalente" al numerador del cociente anterior.

Para ello, definamos

$$
k_{\hat{x}} = f(\hat{x}) - f(\hat{x}_0)
$$

y observemos que, como f es derivable en $\hat{x}_0$, entonces f es continua en $\hat{x}_0$ (proposición 4.22), de donde se tiene que

$$
\begin{aligned}
\lim_{\hat{x} \to \hat{x}_0} k_{\hat{x}} &= \lim_{\hat{x} \to \hat{x}_0} (f(\hat{x}) - f(\hat{x}_0)) \\
&= 0.
\end{aligned}
$$

En el caso en que $k_{\hat{x}} \neq 0$, tenemos que

$$
\begin{aligned}
(g \circ f)(\hat{x}) - (g \circ f)(\hat{x}_0) &= g(f(\hat{x})) - g(f(\hat{x}_0)) \\
&= \frac{g(f(\hat{x}_0) + k_{\hat{x}}) - g(f(\hat{x}_0))}{k_{\hat{x}}} k_{\hat{x}}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{g(f(\hat{x}_0)+k_{\hat{x}})-g(f(\hat{x}_0))}{k_{\hat{x}}}(f(\hat{x})-f(\hat{x}_0))\\
&= \frac{g(f(\hat{x}_0)+k_{\hat{x}})-g(f(\hat{x}_0))}{k_{\hat{x}}}(f(\hat{x})-f(\hat{x}_0)-Df(\hat{x}_0)(\hat{x}-\hat{x}_0)+Df(\hat{x}_0)(\hat{x}-\hat{x}_0))\\
&= \frac{g(f(\hat{x}_0)+k_{\hat{x}})-g(f(\hat{x}_0))}{k_{\hat{x}}}(f(\hat{x})-f(\hat{x}_0)-Df(\hat{x}_0)(\hat{x}-\hat{x}_0))\\
&+ \frac{g(f(\hat{x}_0)+k_{\hat{x}})-g(f(\hat{x}_0))}{k_{\hat{x}}}Df(\hat{x}_0)(\hat{x}-\hat{x}_0),
\end{aligned}
$$

de tal manera que

$$
\begin{aligned}
&(g\circ f)(\hat{x})-(g\circ f)(\hat{x}_0)-g'(f(\hat{x}_0))Df(\hat{x}_0)(\hat{x}-\hat{x}_0)\\
&\qquad= \frac{g(f(\hat{x}_0)+k_{\hat{x}})-g(f(\hat{x}_0))}{k_{\hat{x}}}(f(\hat{x})-f(\hat{x}_0)-Df(\hat{x}_0)(\hat{x}-\hat{x}_0))\\
&\qquad+\left(\frac{g(f(\hat{x}_0)+k_{\hat{x}})-g(f(\hat{x}_0))}{k_{\hat{x}}}-g'(f(\hat{x}_0))\right)Df(\hat{x}_0)(\hat{x}-\hat{x}_0).
\end{aligned}
$$

Por lo tanto

$$
\begin{aligned}
&\frac{(g\circ f)(\hat{x})-(g\circ f)(\hat{x}_0)-g'(f(\hat{x}_0))Df(\hat{x}_0)(\hat{x}-\hat{x}_0)}{\|\hat{x}-\hat{x}_0\|}\\
&\qquad= \frac{g(f(\hat{x}_0)+k_{\hat{x}})-g(f(\hat{x}_0))}{k_{\hat{x}}}\left(\frac{f(\hat{x})-f(\hat{x}_0)-Df(\hat{x}_0)(\hat{x}-\hat{x}_0)}{\|\hat{x}-\hat{x}_0\|}\right)\\
&\qquad+\left(\frac{g(f(\hat{x}_0)+k_{\hat{x}})-g(f(\hat{x}_0))}{k_{\hat{x}}}-g'(f(\hat{x}_0))\right)Df(\hat{x}_0)\left(\frac{\hat{x}-\hat{x}_0}{\|\hat{x}-\hat{x}_0\|}\right).
\end{aligned}
\tag{4.26}
$$

Como seguramente el lector ya habrá notado, cada uno de los sumandos del lado derecho de esta última identidad tienden a 0 cuando $\hat{x}\to\hat{x}_0$. En efecto, los factores del primer sumando tienen límite; el primero tiende a $g'(f(\hat{x}_0))$ (puesto que g es derivable en $f(\hat{x}_0)$) y el segundo tiende a 0 (puesto que f es derivable en $\hat{x}_0$).

En cuanto al segundo sumando, su primer factor tiende a 0 (nuevamente porque g es derivable en $f(\hat{x}_0)$) mientras que para el segundo factor, por el inciso (c) del problema 11 del capítulo 2, sabemos que existe $M\geq 0$ tal que

$$|Df(\hat{x}_0)\,(\hat{x}-\hat{x}_0)|\leq M\,\|\hat{x}-\hat{x}_0\|\,,$$

de tal modo que

$$\left|Df(\hat{x}_0)\left(\frac{\hat{x}-\hat{x}_0}{\|\hat{x}-\hat{x}_0\|}\right)\right|\leq M$$

y por lo tanto todo el sumando tiende a 0 cuando $\hat{x}\to\hat{x}_0$.

Toda la argumentación anterior es correcta bajo el supuesto de que $k_{\hat{x}}=f(\hat{x})-f(\hat{x}_0)\neq 0$; lo que ahora mostraremos es que para considerar el caso en que $k_{\hat{x}}=0$ será necesario introducir una función auxiliar[1] $\varphi: U\subset\mathbb{R}^n\to\mathbb{R}$ definida de la siguiente forma

$$
\varphi(\hat{x})=\begin{cases}\frac{g(f(\hat{x}))-g(f(\hat{x}_0))}{f(\hat{x})-f(\hat{x}_0)} & \text{si } f(\hat{x})-f(\hat{x}_0)\neq 0\\[2ex] g'(f(\hat{x}_0)) & \text{si } f(\hat{x})-f(\hat{x}_0)=0\end{cases}
$$

Lo primero que haremos notar es que la igualdad 4.26 se escribe en términos de la función φ como

$$
\begin{aligned}
\frac{(g\circ f)(\hat{x})-(g\circ f)(\hat{x}_0)-g'(f(\hat{x}_0))Df(\hat{x}_0)(\hat{x}-\hat{x}_0)}{\|\hat{x}-\hat{x}_0\|}&=\varphi(\hat{x})\left(\frac{f(\hat{x})-f(\hat{x}_0)-Df(\hat{x}_0)(\hat{x}-\hat{x}_0)}{\|\hat{x}-\hat{x}_0\|}\right)\\
&+(\varphi(\hat{x})-g'(f(\hat{x}_0)))\,Df(\hat{x}_0)\left(\frac{\hat{x}-\hat{x}_0}{\|\hat{x}-\hat{x}_0\|}\right)
\end{aligned}
$$

[1]Esto es lo que siempre se hace en las pruebas de las diferentes variantes de la regla de la cadena, como seguramente el lector ya habrá notado a estas alturas.

y que ahora esta identidad se cumple para toda $\hat{x} \in U$, $\hat{x} \neq \hat{x}_0$. Si ahora recordamos que f es continua en $\hat{x}_0$, dado que g es derivable en $f(\hat{x}_0)$, aplicando el resultado del problema 49 del capítulo 2, sabemos que φ es continua en $\hat{x}_0$ y por tanto

$$\begin{aligned}
&\lim_{\hat{x}\to\hat{x}_0} \frac{(g\circ f)(\hat{x}) - (g\circ f)(\hat{x}_0) - g'(f(\hat{x}_0))Df(\hat{x}_0)(\hat{x}-\hat{x}_0)}{\|\hat{x}-\hat{x}_0\|}\\
&= \lim_{\hat{x}\to\hat{x}_0} \left[\varphi(\hat{x})\left(\frac{f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)(\hat{x}-\hat{x}_0)}{\|\hat{x}-\hat{x}_0\|}\right) + \left(\varphi(\hat{x}) - g'(f(\hat{x}_0))\right) Df(\hat{x}_0)\left(\frac{\hat{x}-\hat{x}_0}{\|\hat{x}-\hat{x}_0\|}\right)\right]\\
&= \lim_{\hat{x}\to\hat{x}_0} \varphi(\hat{x})\left(\frac{f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)(\hat{x}-\hat{x}_0)}{\|\hat{x}-\hat{x}_0\|}\right) + \lim_{\hat{x}\to\hat{x}_0} \left(\varphi(\hat{x}) - g'(f(\hat{x}_0))\right) Df(\hat{x}_0)\left(\frac{\hat{x}-\hat{x}_0}{\|\hat{x}-\hat{x}_0\|}\right)\\
&= \varphi(\hat{x}_0)0 + \lim_{\hat{x}\to\hat{x}_0} \left(\varphi(\hat{x}) - g'(f(\hat{x}_0))\right) Df(\hat{x}_0)\left(\frac{\hat{x}-\hat{x}_0}{\|\hat{x}-\hat{x}_0\|}\right)\\
&= \lim_{\hat{x}\to\hat{x}_0} \left(\varphi(\hat{x}) - g'(f(\hat{x}_0))\right) Df(\hat{x}_0)\left(\frac{\hat{x}-\hat{x}_0}{\|\hat{x}-\hat{x}_0\|}\right)\\
&= 0.
\end{aligned}$$

La última identidad se satisface dado que

$$\begin{aligned}
\lim_{\hat{x}\to\hat{x}_0} \left(\varphi(\hat{x}) - g'(f(\hat{x}_0))\right) &= \lim_{\hat{x}\to\hat{x}_0} \varphi(\hat{x}) - g'(f(\hat{x}_0))\\
&= \varphi(\hat{x}_0) - g'(f(\hat{x}_0))\\
&= g'(f(\hat{x}_0)) - g'(f(\hat{x}_0))\\
&= 0
\end{aligned}$$

y que, como mencionamos párrafos arriba, existe $M \geq 0$ tal que

$$\left| Df(\hat{x}_0)\left(\frac{\hat{x}-\hat{x}_0}{\|\hat{x}-\hat{x}_0\|}\right)\right| \leq M$$

para toda $\hat{x} \in \mathbb{R}^n$, $\hat{x} \neq \hat{x}_0$. ■

Como afirmamos anteriormente, también podemos componer (por la derecha) a una función f de $\mathbb{R}^n$ en $\mathbb{R}$, con una función γ de $\mathbb{R}$ en $\mathbb{R}^n$. Para esta composición también podemos establecer una regla de la cadena: bajo la hipótesis de que ambas funciones son derivables (en los puntos adecuados), entonces dicha composición también es derivable, y además nos da una forma de cómo calcular esta derivada.

En la formulación (y prueba) de este resultado representaremos a la derivada de f a través de su vector gradiente.

Proposición 4.30 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$, $\gamma : (a,b) \subset \mathbb{R} \to \mathbb{R}^n$ tal que $\gamma((a,b)) \subset U$, y $t_0 \in (a,b)$ tal que $\gamma(t_0) = \hat{x}_0$. Si f es derivable en $\hat{x}_0$ y γ es derivable en t_0, entonces $f\circ\gamma$ es derivable en t_0 y además*

$$(f\circ\gamma)'(t_0) = \nabla f(\hat{x}_0)\cdot\gamma'(t_0).$$

Demostración. Dado que la composición $f\circ\gamma$ es una función de $\mathbb{R}$ en $\mathbb{R}$, de acuerdo con la definición de la derivada de este tipo de funciones, debemos de mostrar que

$$\lim_{t\to t_0} \frac{(f\circ\gamma)(t) - (f\circ\gamma)(t_0)}{t-t_0}$$

existe.

De esta forma, lo primero que haremos será escribir de manera más "adecuada" a este cociente.

Nótese que, si $\gamma(t) - \gamma(t_0) \neq \hat{0}$, entonces

$$\begin{aligned}
&\frac{(f \circ \gamma)(t) - (f \circ \gamma)(t_0)}{t - t_0} \\
&= \frac{\|\gamma(t) - \gamma(t_0)\|}{t - t_0} \left(\frac{f(\gamma(t)) - f(\gamma(t_0)) - \nabla f(\gamma(t_0)) \cdot (\gamma(t) - \gamma(t_0)) + \nabla f(\gamma(t_0)) \cdot (\gamma(t) - \gamma(t_0))}{\|\gamma(t) - \gamma(t_0)\|} \right) \\
&= \frac{\|\gamma(t) - \gamma(t_0)\|}{t - t_0} \left(\frac{f(\gamma(t)) - f(\gamma(t_0)) - \nabla f(\gamma(t_0)) \cdot (\gamma(t) - \gamma(t_0))}{\|\gamma(t) - \gamma(t_0)\|} \right) + \nabla f(\gamma(t_0)) \cdot \left(\frac{\gamma(t) - \gamma(t_0)}{t - t_0} \right).
\end{aligned}$$

Ahora, dado que se satisfacen los siguientes tres hechos: uno, que $\gamma(t) \to \gamma(t_0)$ si $t \to t_0$ (puesto que γ es continua en t_0 ya que es derivable en t_0); dos, que la expresión

$$\frac{\|\gamma(t) - \gamma(t_0)\|}{t - t_0},$$

está acotada en una vecindad de t_0 pues, nuevamente porque γ es derivable en t_0, se tiene que

$$\begin{aligned}
\lim_{t \to t_0} \left| \frac{\|\gamma(t) - \gamma(t_0)\|}{t - t_0} \right| &= \lim_{t \to t_0} \left\| \frac{\gamma(t) - \gamma(t_0)}{t - t_0} \right\| \\
&= \|\gamma'(t_0)\|
\end{aligned}$$

y tres, que f es derivable en $\hat{x}_0 = \gamma(t_0)$, concluimos que

$$\lim_{t \to t_0} \frac{\|\gamma(t) - \gamma(t_0)\|}{t - t_0} \left(\frac{f(\gamma(t)) - f(\gamma(t_0)) - \nabla f(\gamma(t_0)) \cdot (\gamma(t) - \gamma(t_0))}{\|\gamma(t) - \gamma(t_0)\|} \right) = 0.$$

Finalmente, por el inciso 3 de la proposición 2.33, se tiene que

$$\begin{aligned}
\lim_{t \to t_0} \nabla f(\gamma(t_0)) \cdot \left(\frac{\gamma(t) - \gamma(t_0)}{t - t_0} \right) &= \lim_{t \to t_0} \nabla f(\gamma(t_0)) \cdot \lim_{t \to t_0} \frac{\gamma(t) - \gamma(t_0)}{t - t_0} \\
&= \nabla f(\gamma(t_0)) \cdot \gamma'(t_0),
\end{aligned}$$

con lo cual obtendríamos el resultado deseado.

Sin embargo (y como el lector ya habrá notado), el argumento anterior sólo es correcto si $\gamma(t) - \gamma(t_0) \neq \hat{0}$. Para que el razonamiento anterior sea válido incluso en el caso en el que $\gamma(t) - \gamma(t_0) = \hat{0}$, nuevamente recurriremos a una función auxiliar que resuelve el problema.

Sea $\varphi : (a, b) \subset \mathbb{R} \to \mathbb{R}$ definida como

$$\varphi(t) = \begin{cases} \frac{f(\gamma(t)) - f(\gamma(t_0)) - \nabla f(\gamma(t_0)) \cdot (\gamma(t) - \gamma(t_0))}{\|\gamma(t) - \gamma(t_0)\|} & \text{si } \gamma(t) - \gamma(t_0) \neq \hat{0} \\ 0 & \text{si } \gamma(t) - \gamma(t_0) = \hat{0} \end{cases}$$

Es fácil verificar que

$$\frac{(f \circ \gamma)(t) - (f \circ \gamma)(t_0)}{t - t_0} = \varphi(t) \frac{\|\gamma(t) - \gamma(t_0)\|}{t - t_0} + \nabla f(\gamma(t_0)) \cdot \left(\frac{\gamma(t) - \gamma(t_0)}{t - t_0} \right)$$

para toda $t \in (a, b)$, $t \neq t_0$. Por otra parte, como el lector probará en el problema 17, φ es continua en t_0 de tal forma que

$$\lim_{t \to t_0} \varphi(t) \frac{\|\gamma(t) - \gamma(t_0)\|}{t - t_0} = 0,$$

de donde

$$\begin{aligned}
\lim_{t \to t_0} \frac{(f \circ \gamma)(t) - (f \circ \gamma)(t_0)}{t - t_0} &= \lim_{t \to t_0} \left(\varphi(t) \frac{\|\gamma(t) - \gamma(t_0)\|}{t - t_0} + \nabla f(\gamma(t_0)) \cdot \left(\frac{\gamma(t) - \gamma(t_0)}{t - t_0} \right) \right) \\
&= \nabla f(\gamma(t_0)) \cdot \gamma'(t_0),
\end{aligned}$$

que es lo que deseábamos demostrar. ■

Esta última proposición, además de proporcionarnos una excelente regla de derivación, tiene consecuencias de caracter práctico muy importantes cuya discusión ocupará un espacio grande.

Consideremos una función f de $\mathbb{R}^n$ en $\mathbb{R}$ y un conjunto de nivel $N_c(f) := \{\hat{x} \in U \subset \mathbb{R}^n \mid f(\hat{x}) = c\}$ de ésta. Un problema que resulta de interés es el de encontrar un vector que esté en la dirección "normal" al conjunto de nivel $N_c(f)$, en un punto específico de dicho conjunto. Lo que haremos a continuación será mostrar que si $\hat{x}_0$ pertenece al conjunto de nivel $N_c(f) := \{\hat{x} \in U \subset \mathbb{R}^n \mid f(\hat{x}) = c\}$, entonces $\nabla f(\hat{x}_0)$ es "normal" a $N_c(f)$ en $\hat{x}_0$ en el sentido de que se satisface lo siguiente: para cualquier $\gamma : (a,b) \subset \mathbb{R} \to \mathbb{R}^n$ con las características de que $\gamma(t) \in N_c(f)$ para toda $t \in (a,b)$, γ es derivable en $t_0 \in (a,b)$ y $\gamma(t_0) = \hat{x}_0$, se cumple que

$$\nabla f(\gamma(t_0)) \cdot \gamma'(t_0) = 0.$$

Esta afirmación la dejaremos formulada como un corolario de la proposición 4.30 y su demostración se deja al lector.

Corolario 4.31 *Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}$ derivable en $\hat{x}_0 \in U$ tal que $\nabla f(\hat{x}_0) \neq \hat{0}$. Si $\hat{x}_0 \in N_c(f)$ (el conjunto de nivel c de f), entonces $\nabla f(\hat{x}_0)$ es "normal" a $N_c(f)$ en $\hat{x}_0$. Es decir que para cualquier $\gamma : (a,b) \subset \mathbb{R} \to \mathbb{R}^n$ tal que: $\gamma(t) \in N_c(f)$ para toda $t \in (a,b)$, γ es derivable en $t_0 \in (a,b)$ y $\gamma(t_0) = \hat{x}_0$, se tiene que*

$$\nabla f(\gamma(t_0)) \cdot \gamma'(t_0) = 0.$$

Como se ilustró en el capítulo 2, una buena cantidad de objetos geométricos muy conocidos (sobre todo en $\mathbb{R}^2$ y en $\mathbb{R}^3$) se pueden ver como conjuntos de nivel de una función de $\mathbb{R}^n$ en $\mathbb{R}$. Lo interesante del corolario anterior es que nos proporciona los elementos suficientes para definir (de forma muy sencilla) el concepto de recta tangente y plano tangente para el caso particular de conjuntos de nivel en $\mathbb{R}^2$ y en $\mathbb{R}^3$, respectivamente.

Definición 4.32 *Sea $f : U \subset \mathbb{R}^2 \to \mathbb{R}$ derivable en $\hat{x}_0 = (x_0, y_0) \in U$ tal que $\nabla f(\hat{x}_0) \neq \hat{0}$. Decimos que la recta (en $\mathbb{R}^2$) determinada por la ecuación cartesiana*

$$\frac{\partial f}{\partial x}(\hat{x}_0)\,(x - x_0) + \frac{\partial f}{\partial y}(\hat{x}_0)\,(y - y_0) = 0$$

es la recta tangente en $\hat{x}_0$ del conjunto (o curva) de nivel $N_{f(\hat{x}_0)}(f)$.

Definición 4.33 *Sea $f : U \subset \mathbb{R}^3 \to \mathbb{R}$ derivable en $\hat{x}_0 = (x_0, y_0, z_0) \in U$ tal que $\nabla f(\hat{x}_0) \neq \hat{0}$. Decimos que el plano (en $\mathbb{R}^3$) determinado por la ecuación cartesiana*

$$\frac{\partial f}{\partial x}(\hat{x}_0)\,(x - x_0) + \frac{\partial f}{\partial y}(\hat{x}_0)\,(y - y_0) + \frac{\partial f}{\partial z}(\hat{x}_0)\,(z - z_0) = 0$$

es el plano tangente en $\hat{x}_0$ del conjunto de nivel $N_{f(\hat{x}_0)}(f)$.

A continuación daremos un par de ejemplos en los que se muestra cómo se obtienen la recta y el plano tangente definidos anteriormente.

Ejemplo 4.34 *Consideremos:*

1. *la curva determinada por la ecuación*

$$(x^2 + y^2 - 2x)^2 = 4(x^2 + y^2),$$

   la cual corresponde a la ya conocida cardioide.

   Lo que ahora queremos hacer notar es que la cardioide es el conjunto de nivel 0 *de la función*

$$f(x,y) = (x^2 + y^2 - 2x)^2 - 4(x^2 + y^2),$$

   que resultará muy conveniente si se quieren calcular rectas tangentes a este conjunto.

Para esta función se tiene que

$$\frac{\partial f}{\partial x}(x,y) = 2(x^2+y^2-2x)\,(2x-2) - 8x$$

y

$$\frac{\partial f}{\partial y}(x,y) = 2(x^2+y^2-2x)\,(2y) - 8y,$$

de modo que

$$\begin{aligned}\nabla f(x,y) &= 4\left((x^2+y^2-2x)\,(x-1) - 2x, y(x^2+y^2-2x-2)\right)\\ &= 4\left(\left((x-1)^2+y^2-1\right)(x-1) - 2\,(x-1) - 2, y\left((x-1)^2+y^2-3\right)\right)\\ &= 4\left((x-1)\left((x-1)^2+y^2-3\right) - 2, y\left((x-1)^2+y^2-3\right)\right).\end{aligned}$$

De esta forma, si $\nabla f(x_0,y_0) \neq (0,0)$*, este vector es normal a la curva (en el punto* (x_0,y_0)*) y la recta dada por la ecuación*

$$\nabla f(x_0,y_0)\cdot(x-x_0, y-y_0) = 0,$$

o equivalentemente

$$\frac{\partial f}{\partial x}(x_0,y_0)(x-x_0) + \frac{\partial f}{\partial y}(x_0,y_0)(y-y_0) = 0,$$

es la ecuación de la recta tangente a la curva (en el punto (x_0,y_0)*).*

Si en particular tomamos los puntos en que la cardioide intersecta al eje Y*, es decir los puntos* $(0,2)$ *y* $(0,-2)$*, se tiene que*

$$\nabla f(0,2) = (-16,16) \qquad y \qquad \nabla f(0,-2) = (-16,-16).$$

Por lo tanto, la ecuación de las respectivas rectas tangentes están dadas por

$$-16x + 16(y-2) = 0 \qquad o \qquad y = x+2$$

y

$$-16x - 16(y+2) = 0 \qquad o \qquad y = -x-2,$$

que son las que ya habíamos obtenido en el capítulo 3.

2. *Sea* $S \subset \mathbb{R}^3$ *el conjunto definido como*

$$S = \left\{(x,y,z)\in\mathbb{R}^3 \mid \frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2} = 1\right\},$$

el cual corresponde a un elipsoide con centro en el $(0,0,0)$ *(ver figura 4.8).*

Como en el caso de la cardioide, este conjunto tampoco se puede obtener (completo) como la gráfica de una función de $\mathbb{R}^2$ *en* $\mathbb{R}$.

Aun cuando algunas partes de este conjunto sí se pueden ver como una de estas gráficas, en términos de los cálculos que hay que realizar (usando la ecuación 4.16), resulta mucho más fácil observar que se puede obtener como el conjunto de nivel 1 *de la función*

$$f(x,y,z) = \frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}.$$

Como

$$\nabla f(x,y,z) = 2\left(\frac{x}{a^2}, \frac{y}{b^2}, \frac{z}{c^2}\right),$$

si (x_0,y_0,z_0) *es cualquier punto del elipsoide, entonces* $\nabla f(x_0,y_0,z_0) \neq (0,0,0)$ *y por lo tanto la ecuación del plano tangente en este punto estará dada por*

$$\nabla f(x_0,y_0,z_0)\cdot(x-x_0,y-y_0,z-z_0) = 0 = \left(\frac{x_0}{a^2}, \frac{y_0}{b^2}, \frac{z_0}{c^2}\right)\cdot(x-x_0,y-y_0,z-z_0).$$

Al simplificarla, se reduce a la ecuación

$$\left(\frac{x_0}{a^2}\right)x + \left(\frac{y_0}{b^2}\right)y + \left(\frac{z_0}{c^2}\right)z = 1.$$

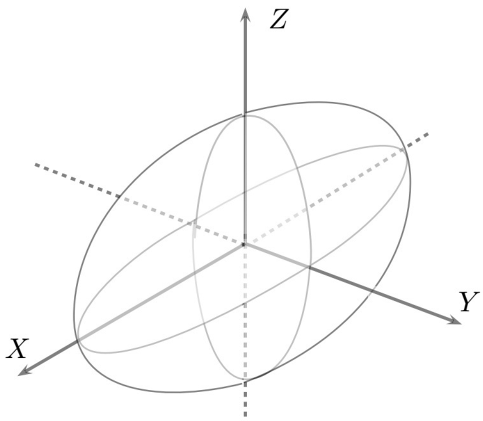


Figura 4.8: El elipsoide determinado por la ecuación $\frac{x^2}{8} + \frac{y^2}{3} + \frac{z^2}{5} = 1$, análogo al del ejemplo 4.34.

Como el lector habrá notado, en el ejemplo anterior se hace énfasis en que los conjuntos involucrados no se pueden obtener como la gráfica de una función de $\mathbb{R}$ en $\mathbb{R}$, para el primer inciso, o de una función de $\mathbb{R}^2$ en $\mathbb{R}$, para el segundo. La razón de ello es que el concepto de recta tangente y plano tangente ya se tenían definidos para este tipo de objetos, pero dada la incapacidad de verlos de esta forma, se quiso resaltar la "conveniencia" de poderlos obtener como un conjunto de nivel de cierta función, para así poder aplicar las definiciones 4.32 y 4.33.

Pero si bien es cierto que no todo conjunto de nivel en $\mathbb{R}^2$ se puede ver como la gráfica de una función de $\mathbb{R}$ en $\mathbb{R}$, ni todo conjunto de nivel en $\mathbb{R}^3$ se puede ver como la gráfica de una función de $\mathbb{R}^2$ en $\mathbb{R}$, lo recíproco sí se cumple (problema 10 del capítulo 2). Pues bien, en los problemas 10 y 11 de este capítulo pedimos al lector que pruebe que los conceptos de recta y plano tangente dados en las definiciones 4.32 y 4.33, coinciden con los conceptos correspondientes para los de gráfica de una función.

Más adelante, en el capítulo 5, probaremos un importante resultado (el Teorema de la Función Implícita) del cual podremos deducir, bajo ciertas hipótesis, que un conjunto de nivel en $\mathbb{R}^2$ (o en $\mathbb{R}^3$) sí se puede obtener, cuando menos "localmente", como la gráfica de una función de $\mathbb{R}$ en $\mathbb{R}$ (de $\mathbb{R}^2$ en $\mathbb{R}$). Apoyados en este teorema, se podrá probar entonces que la recta y plano tangente definidos en 4.32 y 4.33 también se pueden obtener como la recta (o plano) tangente a la gráfica de cierta función de $\mathbb{R}$ en $\mathbb{R}$ (o de $\mathbb{R}^2$ en $\mathbb{R}$), respectivamente.

Otra interpretación del resultado de la proposición 4.30 que tendrá un uso muy importante es el siguiente: si $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$, $\gamma : (a, b) \subset \mathbb{R} \to \mathbb{R}^n$, con $\gamma((a, b)) \subset U$, y $t_0 \in (a, b)$ son tales que $\gamma(t_0) = \hat{x}_0$, es un hecho que la composición $(f \circ \gamma)(t) = f(\gamma(t))$ nos da los valores de f sobre los puntos de la curva recorrida por la función γ.

Con base en lo anterior, es fácil convencerse de que la derivada de esta composición en t_0 se puede interpretar como la razón de cambio de f en $\hat{x}_0$ cuando nos aproximamos a este punto por la curva descrita por γ. Aquí es importante destacar que resulta "natural" que esta razón de cambio se vea afectada por la velocidad con la que γ se aproxima al punto $\hat{x}_0$, lo que queda evidenciado en la fórmula

$$(f \circ \gamma)'(t_0) = \nabla f(\gamma(t_0)) \cdot \gamma'(t_0) \tag{4.27}$$

De esta manera, si lo que deseamos es interpretar a $(f \circ \gamma)'(t_0)$ como la razón de cambio de f en $\hat{x}_0$ (cuando nos acercamos a este punto por la curva descrita por la función γ), es importante que γ sea tal que $\gamma'(t_0) \neq \hat{0}$. De hecho, si suponemos que γ es una parametrización por longitud de arco, es decir que $\|\gamma'(t)\| = 1$ para cada $t \in (a, b)$, para que no haya problema con la velocidad (o para ser más exactos, con la rapidez) con la que nos aproximamos a $\hat{x}_0$, la derivada de $f \circ \gamma$ en t_0 no es más que la derivada direccional de f en $\hat{x}_0$, en la dirección del vector $\gamma'(t_0)$ (que ahora sería de norma 1).

En efecto, si ahora usamos la notación $Df(\hat{x}_0)$ para referirnos a la derivada de f en $\hat{x}_0$, usando la proposición 4.16, se tiene que

$$\begin{aligned}(f \circ \gamma)'(t_0) &= \nabla f(\gamma(t_0)) \cdot \gamma'(t_0)\\ &= Df(\hat{x}_0)(\gamma'(t_0))\\ &= D_{\gamma'(t_0)} f(\hat{x}_0).\end{aligned}$$

Más aún, si sólo se tiene que $\gamma'(t_0) \neq \hat{0}$ (aunque este vector no sea de norma 1), la identidad 4.27 se puede escribir como

$$\begin{aligned}(f \circ \gamma)'(t_0) &= \nabla f(\gamma(t_0)) \cdot \gamma'(t_0)\\ &= \|\gamma'(t_0)\| \left(\nabla f(\gamma(t_0)) \cdot \left(\frac{\gamma'(t_0)}{\|\gamma'(t_0)\|}\right)\right)\\ &= \|\gamma'(t_0)\|\, Df(\hat{x}_0)\left(\frac{\gamma'(t_0)}{\|\gamma'(t_0)\|}\right)\\ &= \|\gamma'(t_0)\|\, D_{\frac{\gamma'(t_0)}{\|\gamma'(t_0)\|}} f(\hat{x}_0). \qquad (4.28)\end{aligned}$$

Que la razón de cambio de una función f sobre la curva descrita por una función γ en un punto $\hat{x}_0$ sea igual a la derivada direccional de f en $\hat{x}_0$, en la dirección de $\gamma'(t_0)$ (normalizado) multiplicada por la rapidez con que γ se aproxima a $\hat{x}_0$, no hace más que confirmar que el concepto de derivada es un concepto “local” (simplemente recuerde el lector que, cerca del punto $\hat{x}_0 = \gamma(t_0)$, γ y su recta tangente se “parecen” mucho).

Ilustremos la discusión anterior con el siguiente

Ejemplo 4.35 *Sean* $f : \mathbb{R}^2 \setminus \{(0,0)\} \to \mathbb{R}$ *definida como*

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

y $\gamma : \mathbb{R} \to \mathbb{R}^2$ *definida como* $\gamma(t) = (r\cos(t), r\,\mathrm{sen}(t))$.

Nótese que

$$\begin{aligned}(f \circ \gamma)(t) &= f(\gamma(t))\\ &= \frac{(r\cos(t))\,(r\,\mathrm{sen}(t))}{(r\cos(t))^2 + (r\,\mathrm{sen}(t))^2}\\ &= \frac{\cos(t)\,\mathrm{sen}(t)}{\cos^2(t) + \mathrm{sen}^2(t)}\\ &= \cos(t)\,\mathrm{sen}(t),\end{aligned}$$

de modo que

$$\begin{aligned}(f \circ \gamma)'(t) &= \cos(t)\,\mathrm{sen}'(t) + \cos'(t)\,\mathrm{sen}(t)\\ &= \cos^2(t) - \mathrm{sen}^2(t).\end{aligned}$$

Por otra parte, se tiene que

$$\begin{aligned}\nabla f(x,y) &= \left(\frac{\partial f}{\partial x}(x,y), \frac{\partial f}{\partial y}(x,y)\right)\\ &= \left(\frac{y\left(x^2+y^2\right) - 2x\,(xy)}{\left(x^2+y^2\right)^2}, \frac{x\left(x^2+y^2\right) - 2y\,(xy)}{\left(x^2+y^2\right)^2}\right)\\ &= \left(\frac{y\left(y^2-x^2\right)}{\left(x^2+y^2\right)^2}, \frac{x\left(x^2-y^2\right)}{\left(x^2+y^2\right)^2}\right)\end{aligned}$$

y

$$\gamma'(t) = (-r\,\mathrm{sen}(t), r\cos(t)).$$

Por tanto, de acuerdo con la identidad 4.27 se debe tener que

$$\begin{aligned}
(f\circ\gamma)'(t) &= \nabla f(\gamma(t))\cdot\gamma'(t)\\
&= \left(\frac{(r\,\mathrm{sen}(t))\left((r\,\mathrm{sen}(t))^2-(r\cos(t))^2\right)}{\left((r\cos(t))^2+(r\,\mathrm{sen}(t))^2\right)^2}, \frac{(r\cos(t))\left((r\cos(t))^2-(r\,\mathrm{sen}(t))^2\right)}{\left((r\cos(t))^2+(r\,\mathrm{sen}(t))^2\right)^2}\right)\cdot\gamma'(t)\\
&= \frac{1}{r}\left(\mathrm{sen}(t)\left(\mathrm{sen}^2(t)-\cos^2(t)\right), \cos(t)\left(\cos^2(t)-\mathrm{sen}^2(t)\right)\right)\cdot(-r\,\mathrm{sen}(t), r\cos(t))\\
&= \left(\cos^2(t)-\mathrm{sen}^2(t)\right)\left[(-\,\mathrm{sen}(t),\cos(t))\cdot(-\,\mathrm{sen}(t),\cos(t))\right]\\
&= \cos^2(t)-\mathrm{sen}^2(t),
\end{aligned}$$

lo cual coincide con el primer cálculo que hicimos.

La identidad

$$(f\circ\gamma)'(t_0) = \|\gamma'(t_0)\|\, D_{\frac{\gamma'(t_0)}{\|\gamma'(t_0)\|}} f(\hat{x}_0)$$

(bajo el supuesto de que $\gamma'(t_0)\neq\hat{0}$) tendrá implicaciones prácticas muy importantes, sobre todo en lo relacionado con el cálculo de la derivada de una función f cuando ésta esté descrita en términos de coordenadas diferentes a las cartesianas, como las que mencionamos en el capítulo 1. Justo esto es lo que nos disponemos a desarrollar en la siguiente sección.

4.3.3 La derivada en otras coordenadas

En el capítulo 1 introdujimos, para el caso particular de los puntos o vectores del plano o del espacio, otros sistemas de coordenadas diferentes de las coordenadas cartesianas. En ese capítulo ya describimos las características más importantes de esas otras coordenadas, de tal forma que aquí las vamos a usar con toda libertad.

Nuestro objetivo en esta sección es encontrar una representación de la derivada de una función que está dada en términos de algunas de estas coordenadas. Para ello, es importante tener presentes los siguientes dos hechos:

1. dado que la derivada es una función lineal, para tener una representación de ella es suficiente conocer sus valores en los elementos de una base ortonormal de $\mathbb{R}^2$ o $\mathbb{R}^3$ (dependiendo de dónde esté contenido el dominio de la función con la que estemos tratando), y

2. si nuestra función está expresada en términos de algunas de estas coordenadas (polares, cilíndricas o esféricas), los únicos cálculos que estaremos en condiciones de hacer, consistirán en "derivar" estas expresiones con respecto a las variables en las que estén escritas.

Empezaremos por analizar el segundo punto, y para ello supondremos que tenemos una función $f : U \subset \mathbb{R}^2 \to \mathbb{R}$ que está expresada en términos de las coordenadas polares (ρ,θ) de cada punto $\hat{x}\in U$.

Derivar la expresión que define a la función f con respecto a una de sus variables, por ejemplo ρ, significará que en dicha expresión sólo consideraremos como variable a ρ, y a la variable θ la trataremos como una constante. En términos más elementales, es decir, recordando que todo proceso de derivación con respecto de una variable consiste en el cálculo de un límite (independientemente de que para obtener estas derivadas hagamos uso de algunas de las reglas de derivación que ya conocemos), lo que estaremos haciendo es obtener el siguiente límite:

$$\lim_{h\to 0}\frac{f(\rho+h,\theta)-f(\rho,\theta)}{h}. \tag{4.29}$$

Sin duda el cociente que aparece en el límite anterior representa una razón de cambio de la función f, y lo que analizaremos con más cuidado es sobre qué puntos del dominio de f se está calculando esta razón de cambio.

Para tener una idea más precisa de qué tipo de puntos son aquellos cuyas coordenadas polares son de la forma $(\rho + h, \theta)$, recurriremos a las fórmulas de conversión entre coordenadas polares y coordenadas cartesianas que dedujimos en el capítulo 1. De esta forma, si llamamos $\hat{x} \in U$ al punto cuyas coordenadas polares son (ρ, θ), tenemos que las coordenadas cartesianas de $\hat{x}$ están dadas por la pareja $(\rho \cos(\theta), \rho \operatorname{sen}(\theta))$, y si llamamos $\hat{x}_h \in U$ al punto cuyas coordenadas polares son $(\rho + h, \theta)$, sus coordenadas cartesianas estarán dadas por $((\rho + h)\cos(\theta), (\rho + h)\operatorname{sen}(\theta))$.

Nótese entonces que

$$\begin{aligned}\hat{x}_h &= ((\rho + h)\cos(\theta), (\rho + h)\operatorname{sen}(\theta))\\ &= (\rho\cos(\theta), \rho\operatorname{sen}(\theta)) + h(\cos(\theta), \operatorname{sen}(\theta))\\ &= \hat{x} + h(\cos(\theta), \operatorname{sen}(\theta)),\end{aligned}$$

lo que significa que $\hat{x}_h$ siempre es un punto que está en la recta que pasa por $\hat{x}$, en la dirección del vector $(\cos(\theta), \operatorname{sen}(\theta))$ (justo la que hace un ángulo θ con el eje polar) (ver figura 4.9).

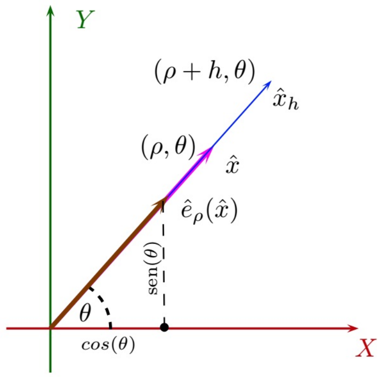


Figura 4.9: Si un vector $\hat{x}$ tiene coordenadas polares (ρ, θ), entonces el vector $\hat{x}_h$ de coordenadas polares $(\rho+h, \theta)$ está en la misma dirección que $\hat{x}$ (si ρ y $\rho+h$ tienen el mismo signo), y que el vector unitario $\hat{e}_\rho(\hat{x})$ de coordenadas cartesianas $(cos(\theta), \operatorname{sen}(\theta))$.

En virtud de lo anterior, si denotamos por $\hat{e}_\rho(\hat{x})$ al vector de coordenadas (cartesianas) $(\cos(\theta), \operatorname{sen}(\theta))$, tenemos que el límite 4.29 se reescribe como

$$\lim_{h\to 0} \frac{f(\rho + h, \theta) - f(\rho, \theta)}{h} = \lim_{h\to 0} \frac{f(\hat{x} + h\hat{e}_\rho(\hat{x})) - f(\hat{x})}{h},$$

de tal forma que, por la definición 4.2 y la proposición 4.16, se tiene que

$$\begin{aligned}\lim_{h\to 0} \frac{f(\rho + h, \theta) - f(\rho, \theta)}{h} &= \lim_{h\to 0} \frac{f(\hat{x} + h\hat{e}_\rho(\hat{x})) - f(\hat{x})}{h}\\ &= D_{\hat{e}_\rho(\hat{x})} f(\hat{x})\\ &= Df(\hat{x})(\hat{e}_\rho(\hat{x})),\end{aligned}$$

o lo que es lo mismo, que

$$Df(\hat{x})(\hat{e}_\rho(\hat{x})) = \lim_{h\to 0} \frac{f(\rho + h, \theta) - f(\rho, \theta)}{h}. \tag{4.30}$$

Esta última identidad es muy afortunada, ya que nos dice que el proceso de derivación con respecto de la variable ρ que describimos al inicio de esta discusión, coincide con ser la evaluación de la derivada de f en $\hat{x}$ $(Df(\hat{x}))$ en el vector unitario $\hat{e}_\rho(\hat{x})$. Esto es sin duda de mucha ayuda para lograr nuestro objetivo (el de obtener una expresión de $Df(\hat{x})$, como lo mencionamos al inicio de esta sección).

Seguramente el lector estará de acuerdo en que ahora lo que hay que hacer es la otra derivada, es decir, derivar la expresión que define a f con respecto de la variable θ (tratando a la variable ρ como constante), y esperar que esta otra derivada también coincida con ser la evaluación de la derivada de f en $\hat{x}$ $(Df(\hat{x}))$

en algún otro vector unitario, que junto con el vector $\hat{e}_\rho(\hat{x})$, forme una base ortonormal de $\mathbb{R}^2$ (¡soñar no cuesta nada!). Aun cuando este sueño no se cumplirá completo, ¡lograremos despertar con una solución al problema que nos planteamos resolver en esta sección! ¡Manos a la obra!

Derivar la expresión que define a f con respecto de la variable θ (tratando a la variable ρ como constante), nuevamente se traduce en calcular el siguiente límite:

$$\lim_{h\to 0} \frac{f(\rho, \theta + h) - f(\rho, \theta)}{h}. \tag{4.31}$$

Como en el caso anterior, el cociente que aparece en este límite también representa una razón de cambio de la función f, y lo que haremos otra vez, será analizar sobre qué puntos del dominio de f se está calculando esta razón de cambio, para lo cual recurriremos nuevamente a las fórmulas de conversión entre coordenadas polares y coordenadas cartesianas.

Si, como hicimos antes, llamamos $\hat{x}$ al punto cuyas coordenadas polares son (ρ, θ), y $\hat{x}_h$ ahora al de coordenadas polares $(\rho, \theta+h)$, tendremos que las correspondientes coordenadas cartesianas serán $(\rho\cos(\theta), \rho\,\mathrm{sen}(\theta))$ y $(\rho\cos(\theta+h), \rho\,\mathrm{sen}(\theta+h))$, respectivamente. Con base en estas coordenadas cartesianas, es fácil notar ahora que $\hat{x}$ y $\hat{x}_h$ son puntos que pertenecen a la misma circunferencia (la de radio ρ con centro en el origen), y que si h es un número "cercano" a 0, entonces $\hat{x}_h$ es un punto "cercano" a $\hat{x}$ (ver figura 4.10).

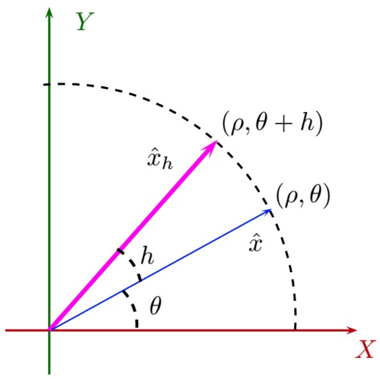


Figura 4.10: Si un vector $\hat{x}$ tiene coordenadas polares (ρ, θ), entonces el vector $\hat{x}_h$ de coordenadas polares $(\rho, \theta+h)$ está en la misma circunferencia (la de radio ρ) que $\hat{x}$.

De esta forma, la derivada con respecto de la variable θ de la función f (que se obtiene al calcular el límite 4.31) se puede interpretar como el cálculo de la razón de cambio de la función f cuando nos aproximamos a $\hat{x}$ por puntos de la misma circunferencia que pasa por este punto.

Por lo tanto, si parametrizamos esta circunferencia por la función $\gamma : \mathbb{R} \to \mathbb{R}^2$ definida como

$$\gamma(h) = (\rho\cos(\theta + h), \rho\,\mathrm{sen}(\theta + h)),$$

el límite 4.31 también se puede reescribir como

$$\begin{aligned}\lim_{h\to 0} \frac{f(\rho, \theta + h) - f(\rho, \theta)}{h} &= \lim_{h\to 0} \frac{f(\gamma(h)) - f(\gamma(0))}{h} \\ &= (f \circ \gamma)'(0).\end{aligned}$$

Como seguramente el lector recordará, la proposición 4.30 nos asegura que

$$\begin{aligned}(f \circ \gamma)'(0) &= \nabla f(\gamma(0)) \cdot \gamma'(0) \\ &= \nabla f(\hat{x}) \cdot (-\rho\,\mathrm{sen}(\theta), \rho\cos(\theta)) \\ &= \rho\nabla f(\hat{x}) \cdot (-\,\mathrm{sen}(\theta), \cos(\theta)) \\ &= \rho Df(\hat{x})(-\,\mathrm{sen}(\theta), \cos(\theta)),\end{aligned}$$

de tal forma que si ahora denotamos por $\hat{e}_\theta(\hat{x})$ al vector de coordenadas cartesianas $(-\operatorname{sen}(\theta), \cos(\theta))$ (¡que es de norma 1 y perpendicular al vector $\hat{e}_\rho(\hat{x})$!), tenemos que

$$\lim_{h\to 0} \frac{f(\rho, \theta + h) - f(\rho, \theta)}{h} = \rho Df(\hat{x})(\hat{e}_\theta(\hat{x})).$$

Aun cuando esta última identidad no nos asegura que la derivada con respecto de la variable θ de la función f sea exactamente igual a la derivada de f en $\hat{x}$ ($Df(\hat{x})$) evaluada en algún vector unitario (como sucedió en el caso anterior), si $\rho \neq 0$, concluimos que

$$Df(\hat{x})(\hat{e}_\theta(\hat{x})) = \frac{1}{\rho} \lim_{h\to 0} \frac{f(\rho, \theta + h) - f(\rho, \theta)}{h}. \tag{4.32}$$

Las identidades 4.30 y 4.32 nos proporcionan la solución al problema que nos planteamos al inicio de esta sección (para el caso de las coordenadas polares): cómo encontrar una representación de la derivada de una función que está dada en términos de este tipo de coordenadas.

Es fundamental que el lector tenga claro la forma en que quedó resuelto este problema. Como mencionamos en el punto número 1 al inicio de esta sección, dado que la derivada de una función es una función lineal, ésta queda completamente determinada si se conocen sus valores en los elementos de una base ortonormal (en este caso de $\mathbb{R}^2$). Y las identidades 4.30 y 4.32 justo nos proporcionan estos valores para la base ortonormal formada por los vectores $\hat{e}_\rho(\hat{x})$ y $\hat{e}_\theta(\hat{x})$, que sin duda forman una de estas bases.

Lo que es importante destacar es que estos vectores dependen del punto en el que estemos "calculando" la derivada de f (lo que por cierto, explica la notación que usamos para representarlos); si cambiamos de punto, cambiamos de base en la cual estamos representando a la función lineal $Df(\hat{x})$ (ver figura 4.11).

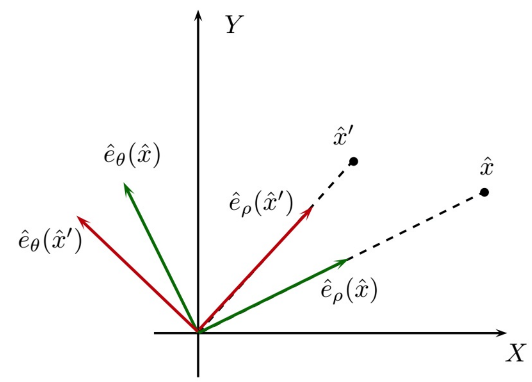


Figura 4.11: Cuando una función f está dada en términos de coordenadas polares, la derivada de f se expresa en una base ortonormal que depende del punto en donde se evalue, $\{\hat{e}_\rho(\hat{x}), \hat{e}_\theta(\hat{x})\}$ para el punto $\hat{x}$, y $\{\hat{e}_\rho(\hat{x}'), \hat{e}_\theta(\hat{x}')\}$ para el punto $\hat{x}'$.

El otro aspecto importante de la solución que encontramos a nuestro problema está en la forma en que están dados los valores de la derivada de f en $\hat{x}$ ($Df(\hat{x})$) en los vectores $\hat{e}_\rho(\hat{x})$ y $\hat{e}_\theta(\hat{x})$: calculando lo único que podíamos calcular (como lo mencionamos en el punto 2 al inicio de esta sección), la derivada con respecto a las variables ρ y θ de la expresión que define a f, y que se obtienen por medio de los límites que aparecen en las identidades 4.30 y 4.32.

Estas "derivadas parciales", aunque análogas a las que definimos en 4.9 no son del mismo tipo en virtud de que no siempre resultan ser una derivada direccional. Por esta razón, lo siguiente que haremos será definir este otro tipo de derivadas parciales.

Definición 4.36 *Sea $f : U \subset \mathbb{R}^2 \to \mathbb{R}$ una función que está expresada en términos de las coordenadas polares (ρ, θ) de cada punto $\hat{x} \in U$. Si $\hat{x}_0 \in U$ tiene coordenadas polares (ρ_0, θ_0), definimos:*

1. *la derivada parcial de f con respecto de ρ en $\hat{x}_0$, que denotamos por $\frac{\partial f}{\partial \rho}(\hat{x}_0)$, como*

$$\frac{\partial f}{\partial \rho}(\hat{x}_0) := \lim_{h\to 0} \frac{f(\rho_0 + h, \theta_0) - f(\rho_0, \theta_0)}{h}$$

2. la derivada parcial de f con respecto de θ en $\hat{x}_0$, que denotamos por $\frac{\partial f}{\partial \theta}(\hat{x}_0)$, como

$$\frac{\partial f}{\partial \theta}(\hat{x}_0) := \lim_{h \to 0} \frac{f(\rho_0, \theta_0 + h) - f(\rho_0, \theta_0)}{h}.$$

Con base en estas definiciones podemos resumir lo obtenido hasta ahora de la siguiente manera: si $f : U \subset \mathbb{R}^2 \to \mathbb{R}$ es una función que está expresada en términos de las coordenadas polares (ρ, θ) de cada punto $\hat{x} \in U$, y $\hat{x}_0 \in U$, $\hat{x}_0 \neq (0,0)$ tiene coordenadas polares (ρ_0, θ_0), $\rho_0 \neq 0$, la derivada de f en $\hat{x}_0$ $(Df(\hat{x}_0))$ está representada en la base ortonormal formada por los vectores

$$\begin{aligned} \hat{e}_\rho(\hat{x}_0) &:= (\cos(\theta_0), \operatorname{sen}(\theta_0)) \\ \hat{e}_\theta(\hat{x}_0) &:= (-\operatorname{sen}(\theta_0), \cos(\theta_0)), \end{aligned}$$

por la matriz (de 1×2)

$$\begin{bmatrix} \frac{\partial f}{\partial \rho}(\hat{x}_0) & \frac{1}{\rho_0}\frac{\partial f}{\partial \theta}(\hat{x}_0) \end{bmatrix}$$

o por el vector

$$\left(\frac{\partial f}{\partial \rho}(\hat{x}_0), \frac{1}{\rho_0}\frac{\partial f}{\partial \theta}(\hat{x}_0) \right).$$

Esto significa que

$$\nabla f(\hat{x}_0) = \frac{\partial f}{\partial \rho}(\hat{x}_0)\hat{e}_\rho(\hat{x}_0) + \frac{1}{\rho_0}\frac{\partial f}{\partial \theta}(\hat{x}_0)\hat{e}_\theta(\hat{x}_0),$$

es decir, que $\frac{\partial f}{\partial \rho}(\hat{x}_0)$ y $\frac{1}{\rho_0}\frac{\partial f}{\partial \theta}(\hat{x}_0)$ son las coordenadas del gradiente de f en $\hat{x}_0$ $(\nabla f(\hat{x}_0))$ en la base $\{\hat{e}_\rho(\hat{x}_0), \hat{e}_\theta(\hat{x}_0)\}$.

Por otra parte, de acuerdo con las identidades 4.19 y 4.18, como

$$\begin{aligned} \hat{e}_\rho(\hat{x}_0) &= (\cos(\theta_0), \operatorname{sen}(\theta_0)) \\ &= \cos(\theta_0)\hat{e}_1 + \operatorname{sen}(\theta_0)\hat{e}_2 \end{aligned}$$

y

$$\begin{aligned} \hat{e}_\theta(\hat{x}_0) &= (-\operatorname{sen}(\theta_0), \cos(\theta_0)) \\ &= -\operatorname{sen}(\theta_0)\hat{e}_1 + \cos(\theta_0)\hat{e}_2, \end{aligned}$$

se tiene que

$$\begin{bmatrix} \frac{\partial f}{\partial x}(\hat{x}_0) & \frac{\partial f}{\partial y}(\hat{x}_0) \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial \rho}(\hat{x}_0) & \frac{1}{\rho_0}\frac{\partial f}{\partial \theta}(\hat{x}_0) \end{bmatrix} \begin{bmatrix} \cos(\theta_0) & \operatorname{sen}(\theta_0) \\ -\operatorname{sen}(\theta_0) & \cos(\theta_0) \end{bmatrix} \tag{4.33}$$

y

$$\begin{bmatrix} \frac{\partial f}{\partial \rho}(\hat{x}_0) & \frac{1}{\rho_0}\frac{\partial f}{\partial \theta}(\hat{x}_0) \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x}(\hat{x}_0) & \frac{\partial f}{\partial y}(\hat{x}_0) \end{bmatrix} \begin{bmatrix} \cos(\theta_0) & -\operatorname{sen}(\theta_0) \\ \operatorname{sen}(\theta_0) & \cos(\theta_0) \end{bmatrix}. \tag{4.34}$$

De estas últimas identidades de matrices se concluye que

$$\begin{aligned} \frac{\partial f}{\partial x}(\hat{x}_0) &= \cos(\theta_0)\frac{\partial f}{\partial \rho}(\hat{x}_0) - \frac{1}{\rho_0}\operatorname{sen}(\theta_0)\frac{\partial f}{\partial \theta}(\hat{x}_0) \\ \frac{\partial f}{\partial y}(\hat{x}_0) &= \operatorname{sen}(\theta_0)\frac{\partial f}{\partial \rho}(\hat{x}_0) + \frac{1}{\rho_0}\cos(\theta_0)\frac{\partial f}{\partial \theta}(\hat{x}_0) \end{aligned}$$

y

$$\begin{aligned} \frac{\partial f}{\partial \rho}(\hat{x}_0) &= \cos(\theta_0)\frac{\partial f}{\partial x}(\hat{x}_0) + \operatorname{sen}(\theta_0)\frac{\partial f}{\partial y}(\hat{x}_0) \\ \frac{\partial f}{\partial \theta}(\hat{x}_0) &= -\rho_0\operatorname{sen}(\theta_0)\frac{\partial f}{\partial x}(\hat{x}_0) + \rho_0\cos(\theta_0)\frac{\partial f}{\partial y}(\hat{x}_0), \end{aligned} \tag{4.35}$$

en donde recuerde que (ρ_0, θ_0) son las coordenadas polares de $\hat{x}_0$.

Con relación a la condición que impusimos de que $\hat{x}_0 \neq (0,0)$, como el lector ya sabrá, las coordenadas polares del origen son cualquier pareja de la forma $(0, \theta)$, en donde θ puede ser cualquier valor. Este hecho imposibilita la elección de la base ortonormal $\{\hat{e}_\rho(\hat{x}_0), \hat{e}_\theta(\hat{x}_0)\}$, lo que a su vez (en algunos casos) se ve reflejado en la imposibilidad de calcular el lado derecho de la identidad 4.32. Esto es lo que nos hace imponer la condición de que $\hat{x}_0 \neq (0,0)$.

El lector seguramente estará de acuerdo en que lo siguiente que habrá que hacer es un ejemplo que ilustre todo lo desarrollado en esta sección. Y eso es justo lo que haremos.

Ejemplo 4.37 *Considere la función $f : \mathbb{R}^2 \to \mathbb{R}$, que está expresada en términos de las coordenadas polares (ρ, θ) de cada punto $\hat{x} \in \mathbb{R}^2$ por*

$$f(\rho, \theta) = (\rho^2 - 2\rho\cos(\theta))^2 - 4\rho^2.$$

En este caso, tenemos que

$$\frac{\partial f}{\partial \rho}(\rho, \theta) = 2(\rho^2 - 2\rho\cos(\theta))\,(2\rho - 2\cos(\theta)) - 8\rho$$

y

$$\frac{\partial f}{\partial \theta}(\rho, \theta) = 2(\rho^2 - 2\rho\cos(\theta))\,(2\rho\operatorname{sen}(\theta)),$$

de tal forma que la derivada de f en un punto $\hat{x} \neq \hat{0}$ cuyas coordenadas polares sean (ρ, θ), es la función lineal $Df(\hat{x})$ cuya matriz asociada (de 1×2) en la base ortonormal $\{\hat{e}_\rho(\hat{x}), \hat{e}_\theta(\hat{x})\}$ es

$$\begin{bmatrix} \frac{\partial f}{\partial \rho}(\rho,\theta) & \frac{1}{\rho}\frac{\partial f}{\partial \theta}(\rho,\theta) \end{bmatrix} = \begin{bmatrix} 2(\rho^2 - 2\rho\cos(\theta))\,(2\rho - 2\cos(\theta)) - 8\rho & 4\operatorname{sen}(\theta)(\rho^2 - 2\rho\cos(\theta)) \end{bmatrix},$$

o la determinada por el vector que tiene coordenadas (también en la base $\{\hat{e}_\rho(\hat{x}), \hat{e}_\theta(\hat{x})\}$)

$$\left(2(\rho^2 - 2\rho\cos(\theta))\,(2\rho - 2\cos(\theta)) - 8\rho, 4\operatorname{sen}(\theta)(\rho^2 - 2\rho\cos(\theta))\right).$$

Es decir, que

$$\nabla f(\rho, \theta) = \left(2(\rho^2 - 2\rho\cos(\theta))\,(2\rho - 2\cos(\theta)) - 8\rho, 4\operatorname{sen}(\theta)(\rho^2 - 2\rho\cos(\theta))\right).$$

Si en particular tomamos el punto $\hat{x}_0$, cuyas coordenadas polares son $(2, \pi/2)$, tenemos que las coordenadas del gradiente de f en $\hat{x}_0$ en la base $\{\hat{e}_\rho(\hat{x}_0), \hat{e}_\theta(\hat{x}_0)\}$ están dadas por

$$\nabla f(\hat{x}_0) = (16, 16).$$

Ahora, observe el lector que la función f de este ejemplo es la misma función que dimos en el inciso 1 del ejemplo 4.34 (basta con que recurra a las fórmulas de cambio de coordenadas para que compruebe este hecho). Nótese también, que el punto $\hat{x}_0$ que tiene coordenadas polares $(2, \pi/2)$, tiene coordenadas cartesianas $(0, 2)$ (en la base canónica $\{\hat{e}_1, \hat{e}_2\}$).

De esta forma, como

$$\begin{aligned} \hat{e}_\rho(\hat{x}_0) &= (\cos(\pi/2), \operatorname{sen}(\pi/2)) \\ &= (0, 1) \\ &= 0\hat{e}_1 + 1\hat{e}_2 \end{aligned}$$

y

$$\begin{aligned} \hat{e}_\theta(\hat{x}_0) &= (-\operatorname{sen}(\pi/2), \cos(\pi/2)) \\ &= (-1, 0) \\ &= -1\hat{e}_1 + 0\hat{e}_2, \end{aligned}$$

de acuerdo con la identidad 4.19, se debe tener que

$$\begin{bmatrix} \frac{\partial f}{\partial x}(\hat{x}_0) & \frac{\partial f}{\partial y}(\hat{x}_0) \end{bmatrix} = \begin{bmatrix} 16 & 16 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} -16 & 16 \end{bmatrix}.$$

Lo anterior es equivalente a que las coordenadas de $\nabla f(\hat{x}_0)$, en la base canónica $\{\hat{e}_1, \hat{e}_2\}$, deben estar dadas por

$$\begin{aligned}\nabla f(\hat{x}_0) &= \left(\frac{\partial f}{\partial x}(\hat{x}_0), \frac{\partial f}{\partial y}(\hat{x}_0)\right)\\ &= (-16, 16)\,,\end{aligned}$$

lo cual coincide con lo obtenido en el ejemplo mencionado.

Aun cuando hicimos el cálculo de la derivada sólo para funciones de $\mathbb{R}^2$ en $\mathbb{R}$ expresadas en términos de coordenadas polares, para funciones de $\mathbb{R}^3$ en $\mathbb{R}$ expresadas en términos de coordenadas cilíndricas o esféricas, el análisis es completamente análogo al anterior y lo dejaremos como un par de problemas para el lector.

4.4 Derivadas direccionales de orden superior

Si para una función $f : U \subset \mathbb{R}^n \to \mathbb{R}$ y un vector unitario $\hat{u} \in \mathbb{R}^n$ fijo, se tiene que $D_{\hat{u}}f(\hat{x})$ existe para todo $\hat{x} \in U$, entonces $D_{\hat{u}}f$ es nuevamente una función de $U \subset \mathbb{R}^n$ en $\mathbb{R}$, es decir, $D_{\hat{u}}f : U \subset \mathbb{R}^n \to \mathbb{R}$. De esta forma, podemos trabajar todos los conceptos de derivación que hasta ahora hemos definido para esta nueva función. En particular, podemos ahora preguntarnos por la existencia de la derivada direccional de $D_{\hat{u}}f$ en un punto $\hat{x} \in U$, en la dirección de cualquier otro vector unitario $\hat{v} \in \mathbb{R}^n$. De acuerdo con la definición 4.2, dicha derivada se denotaría por $D_{\hat{v}}(D_{\hat{u}}f)(\hat{x})$ y estaría dada por

$$D_{\hat{v}}(D_{\hat{u}}f)(\hat{x}) := \lim_{h\to 0} \frac{D_{\hat{u}}f(\hat{x} + h\hat{v}) - D_{\hat{u}}f(\hat{x})}{h}.$$

Más aún, si ahora para $\hat{u}$ y $\hat{v}$ fijos, se tiene que $D_{\hat{v}}(D_{\hat{u}}f)(\hat{x})$ existe para toda $\hat{x} \in U$, entonces $D_{\hat{v}}(D_{\hat{u}}f)$ será otra vez una función de $U \subset \mathbb{R}^n$ en $\mathbb{R}$ y por tanto tiene sentido tomar otro vector unitario $\hat{w} \in \mathbb{R}^n$ y preguntarnos por la existencia de la derivada direccional de $D_{\hat{v}}(D_{\hat{u}}f)$ en un punto $\hat{x} \in U$, en la dirección del vector $\hat{w}$. Es decir, si existe

$$D_{\hat{w}}(D_{\hat{v}}(D_{\hat{u}}f))(\hat{x}) := \lim_{h\to 0} \frac{D_{\hat{v}}(D_{\hat{u}}f)(\hat{x} + h\hat{w}) - D_{\hat{v}}(D_{\hat{u}}f)(\hat{x})}{h}.$$

Antes de seguir por este camino, tomemos un respiro y notemos lo siguiente: seguramente a estas alturas, tomando en cuenta resultados como los de la proposición 4.25, el lector ya intuye la importancia de las derivadas direccionales de una función f en la dirección de los vectores de una base ortonormal $\{\hat{e}_1, \dots, \hat{e}_n\}$, es decir, de las derivadas parciales.

Por esta razón, y aun cuando las ideas planteadas en los párrafos anteriores se pueden seguir trabajando en ese contexto más general, nos concentraremos en el caso particular de las derivadas parciales, ideas que dejaremos expresadas en la siguiente

Definición 4.38 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$, $r > 0$ tal que $B_r(\hat{x}_0) \subset U$, y $x_1, \dots, x_n$ las coordenadas determinadas por una base ortonormal $\{\hat{e}_1, \dots, \hat{e}_n\}$ de $\mathbb{R}^n$. Si $\frac{\partial f}{\partial x_i}(\hat{x})$ existe para toda $\hat{x} \in B_r(\hat{x}_0)$, definimos la segunda derivada parcial de f con respecto a x_i y respecto a x_j en $\hat{x}_0$, que denotaremos por $\frac{\partial^2 f}{\partial x_j \partial x_i}(\hat{x}_0)$, como la derivada parcial de la función $\frac{\partial f}{\partial x_i}$ con respecto de x_j en $\hat{x}_0$. Es decir,*

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(\hat{x}_0) := \lim_{h\to 0} \frac{\frac{\partial f}{\partial x_i}(\hat{x}_0 + h\hat{e}_j) - \frac{\partial f}{\partial x_i}(\hat{x}_0)}{h}.$$

Si $j = i$, denotamos la derivada anterior por $\frac{\partial^2 f}{\partial x_i^2}(\hat{x}_0)$ y decimos que es la segunda derivada parcial de f con respecto de x_i, en $\hat{x}_0$.

Como en el caso de la (primera) derivada parcial, desde un punto de vista práctico, el cálculo de las segundas derivadas parciales se realiza siguiendo el mismo esquema: en cada paso, derivar sólo con respecto a una variable, suponiendo que las demás son constantes. Como es de esperarse, lo que ahora sigue es un ejemplo que ilustre este concepto.

Ejemplo 4.39 *Sea $f : \mathbb{R}^2 \to \mathbb{R}$ dada por la expresión*

$$f(x,y) = x\cos(xy) + y\,\mathrm{sen}(xy).$$

Lo que haremos será calcular todas las posibles segundas derivadas parciales de f (que son cuatro) para cualquier $\hat{x} = (x,y) \in \mathbb{R}^2$.
Para ello, notemos primero que

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

y que

$$\frac{\partial f}{\partial y}(x,y) = -x^2\,\mathrm{sen}(xy) + xy\cos(xy) + \mathrm{sen}(xy).$$

Por tanto, se tiene que

$$\begin{aligned}\frac{\partial^2 f}{\partial x^2}(x,y) &= \frac{\partial\left(\frac{\partial f}{\partial x}\right)}{\partial x}(x,y)\\ &= -xy^2\cos(xy) - y\,\mathrm{sen}(xy) - y\,\mathrm{sen}(xy) - y^3\,\mathrm{sen}(xy)\\ &= -xy^2\cos(xy) - 2y\,\mathrm{sen}(xy) - y^3\,\mathrm{sen}(xy)\end{aligned}$$

y que

$$\begin{aligned}\frac{\partial^2 f}{\partial y\partial x}(x,y) &= \frac{\partial\left(\frac{\partial f}{\partial x}\right)}{\partial y}(x,y)\\ &= -x^2y\cos(xy) - x\,\mathrm{sen}(xy) - x\,\mathrm{sen}(xy) - xy^2\,\mathrm{sen}(xy) + 2y\cos(xy)\\ &= -x^2y\cos(xy) - 2x\,\mathrm{sen}(xy) - xy^2\,\mathrm{sen}(xy) + 2y\cos(xy).\end{aligned}$$

Finalmente,

$$\begin{aligned}\frac{\partial^2 f}{\partial x\partial y}(x,y) &= \frac{\partial\left(\frac{\partial f}{\partial y}\right)}{\partial x}(x,y)\\ &= -x^2y\cos(xy) - 2x\,\mathrm{sen}(xy) - xy^2\,\mathrm{sen}(xy) + y\cos(xy) + y\cos(xy)\\ &= -x^2y\cos(xy) - 2x\,\mathrm{sen}(xy) - xy^2\,\mathrm{sen}(xy) + 2y\cos(xy)\end{aligned}$$

y

$$\begin{aligned}\frac{\partial^2 f}{\partial y^2}(x,y) &= \frac{\partial\left(\frac{\partial f}{\partial y}\right)}{\partial y}(x,y)\\ &= -x^3 cos(xy) - x^2y\,\mathrm{sen}(xy) + x\cos(xy) + x\cos(xy)\\ &= -x^3(xy) - x^2y\,\mathrm{sen}(xy) + 2x\cos(xy).\end{aligned}$$

Del ejemplo anterior es muy importante hacer dos observaciones con respecto al cálculo de las segundas derivadas parciales en las que aparecen ambas variables (x y y), y que se les conoce con el nombre de *derivadas parciales mixtas* (o *cruzadas*).

Una, que se refiere al orden en que se toman estas variables para calcular las correspondientes derivadas, y que es de derecha a izquierda, según aparezcan en el denominador correspondiente; y dos, que en este ejemplo, estas segundas derivadas parciales resultaron ser iguales.

De esta segunda observación, sin duda surge la pregunta de si dicha igualdad siempre se cumple. La respuesta es que no. En el problema 28 el lector encontrará un ejemplo de que esta identidad no siempre sucede. Por nuestra parte, lo siguiente que haremos será justo enunciar un resultado que nos proporciona las condiciones suficientes para que dicha igualdad sí se cumpla.

Un punto importante en la prueba de este resultado tiene que ver con un hecho muy sencillo, relacionado con los valores de una función sobre cualesquiera cuatro puntos de su dominio (que los podemos pensar como cuatro vértices de un cuadrilátero), y el cálculo de ciertas diferencias tomadas a partir de estos valores.

Nótese que, si $\hat{x}_1, \hat{x}_2, \hat{x}_3$ y $\hat{x}_4$ son puntos del dominio de una función f, entonces se tiene que

$$(f(\hat{x}_3) - f(\hat{x}_2)) - (f(\hat{x}_4) - f(\hat{x}_1)) = (f(\hat{x}_3) - f(\hat{x}_4)) - (f(\hat{x}_2) - f(\hat{x}_1)).$$

Teorema 4.40 (de las parciales cruzadas) *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $i, j \in \{1, \ldots, n\}$, $\hat{x}_0 \in U$ y $r > 0$ tal que $B_r(\hat{x}_0) \subset U$. Si las segundas derivadas parciales $\frac{\partial^2 f}{\partial x_j \partial x_i}$ y $\frac{\partial^2 f}{\partial x_i \partial x_j}$ existen para cada $\hat{x} \in B_r(\hat{x}_0)$ y son continuas en $\hat{x}_0$, entonces*

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(\hat{x}_0) = \frac{\partial^2 f}{\partial x_i \partial x_j}(\hat{x}_0).$$

Demostración. Sea $B_r(\hat{0}) = \{(h_1, h_2) \in \mathbb{R}^2 \mid \|(h_1, h_2)\| < r\}$. Nótese que si $(h_1, h_2) \in B_r(\hat{0})$, entonces $\hat{x}_0 + h_1\hat{e}_i + h_2\hat{e}_j$, $\hat{x}_0 + h_1\hat{e}_i$ y $\hat{x}_0 + h_2\hat{e}_j$ pertenecen a $B_r(\hat{x}_0) \subset \mathbb{R}^n$, de modo que podemos definir $H : B_r(\hat{0}) \subset \mathbb{R}^2 \to \mathbb{R}$ como

$$H(h_1, h_2) = (f(\hat{x}_0 + h_1\hat{e}_i + h_2\hat{e}_j) - f(\hat{x}_0 + h_1\hat{e}_i)) - (f(\hat{x}_0 + h_2\hat{e}_j) - f(\hat{x}_0)). \tag{4.36}$$

Ahora, dado $(h_1, h_2) \in B_r(\hat{0})$, definimos

$$g : \left\{x \in \mathbb{R} \mid |x| < \sqrt{r^2 - h_2^2}\right\} = \left(-\sqrt{r^2 - h_2^2}, \sqrt{r^2 - h_2^2}\right) \subset \mathbb{R} \to \mathbb{R}$$

como

$$g(x) = f(\hat{x}_0 + x\hat{e}_i + h_2\hat{e}_j) - f(\hat{x}_0 + x\hat{e}_i).$$

Ya que $h_1 \in \left(-\sqrt{r^2 - h_2^2}, \sqrt{r^2 - h_2^2}\right)$, se tiene que

$$\begin{aligned} H(h_1, h_2) &= (f(\hat{x}_0 + h_1\hat{e}_i + h_2\hat{e}_j) - f(\hat{x}_0 + h_1\hat{e}_i)) - (f(\hat{x}_0 + h_2\hat{e}_j) - f(\hat{x}_0)) \\ &= g(h_1) - g(0). \end{aligned}$$

Por otra parte, observe que por el problema 2 de este capítulo, g es derivable en su dominio y

$$g'(x) = \frac{\partial f}{\partial x_i}(\hat{x}_0 + x\hat{e}_i + h_2\hat{e}_j) - \frac{\partial f}{\partial x_i}(\hat{x}_0 + x\hat{e}_i),$$

de modo que por el Teorema de Valor Medio (para funciones de $\mathbb{R}$ en $\mathbb{R}$), existe

$$\xi \in \left(-\sqrt{r^2 - h_2^2}, \sqrt{r^2 - h_2^2}\right)$$

tal que $0 < |\xi| < |h_1|$ y

$$\begin{aligned} H(h_1, h_2) &= g(h_1) - g(0) \\ &= g'(\xi)h_1 \\ &= \left(\frac{\partial f}{\partial x_i}(\hat{x}_0 + \xi\hat{e}_i + h_2\hat{e}_j) - \frac{\partial f}{\partial x_i}(\hat{x}_0 + \xi\hat{e}_i)\right) h_1. \end{aligned}$$

Finalmente, como $[\hat{x}_0 + \xi\hat{e}_i, \hat{x}_0 + \xi\hat{e}_i + h_2\hat{e}_j] \subset B_r(\hat{x}_0)$ y $(\hat{x}_0 + \xi\hat{e}_i + h_2\hat{e}_j) - (\hat{x}_0 + \xi\hat{e}_i) = h_2\hat{e}_j$, por la proposición 4.13 (aplicada a la función $\frac{\partial f}{\partial x_i}$) se tiene que existe $\eta \in (0, |h_2|)$ tal que

$$\begin{aligned} H(h_1, h_2) &= \left(\frac{\partial f}{\partial x_i}(\hat{x}_0 + \xi\hat{e}_i + h_2\hat{e}_j) - \frac{\partial f}{\partial x_i}(\hat{x}_0 + \xi\hat{e}_i)\right) h_1 \\ &= \frac{\partial\left(\frac{\partial f}{\partial x_i}\right)}{\partial x_j}(\hat{x}_0 + \xi\hat{e}_i + \eta\hat{e}_j)h_1h_2 \end{aligned}$$

$$= \frac{\partial^2 f}{\partial x_j \partial x_i}(\hat{x}_0 + \xi\hat{e}_i + \eta\hat{e}_j)h_1h_2.$$

Por lo tanto, si $h_1h_2 \neq 0$, por la continuidad de $\frac{\partial^2 f}{\partial x_j \partial x_i}$ en $\hat{x}_0$, y del hecho de que $(\xi, \eta) \to (0,0)$ si $(h_1, h_2) \to (0,0)$, concluimos que

$$\begin{aligned}
\lim_{(h_1,h_2)\to(0,0)} \frac{H(h_1,h_2)}{h_1h_2} &= \lim_{(\xi,\eta)\to(0,0)} \frac{\partial^2 f}{\partial x_j \partial x_i}(\hat{x}_0 + \xi\hat{e}_i + \eta\hat{e}_j) \\
&= \frac{\partial^2 f}{\partial x_j \partial x_i}(\hat{x}_0).
\end{aligned}$$

Observe ahora que para cada $(h_1, h_2) \in B_r(\hat{0})$, la función H definida en 4.36 también se puede escribir como

$$H(h_1,h_2) = (f(\hat{x}_0 + h_1\hat{e}_i + h_2\hat{e}_j) - f(\hat{x}_0 + h_2\hat{e}_j)) - (f(\hat{x}_0 + h_1\hat{e}_i) - f(\hat{x}_0))$$

de tal forma que, dado $(h_1, h_2) \in B_r(\hat{0})$, ahora definimos

$$\tilde{g} : \left\{ x \in \mathbb{R} \mid |x| < \sqrt{r^2 - h_1^2} \right\} = \left(-\sqrt{r^2 - h_1^2}, \sqrt{r^2 - h_1^2} \right) \subset \mathbb{R} \to \mathbb{R}$$

como

$$\tilde{g}(x) = f(\hat{x}_0 + h_1\hat{e}_i + x\hat{e}_j) - f(\hat{x}_0 + x\hat{e}_j).$$

Ya que $h_2 \in \left(-\sqrt{r^2 - h_1^2}, \sqrt{r^2 - h_1^2} \right)$, se tiene que

$$H(h_1,h_2) = \tilde{g}(h_2) - \tilde{g}(0).$$

Procediendo con la función $\tilde{g}$ como lo hicimos anteriormente con la función g, se tiene que existe $\xi' \in \left(-\sqrt{r^2 - h_1^2}, \sqrt{r^2 - h_1^2} \right)$ tal que $0 < |\xi'| < |h_2|$ y

$$\begin{aligned}
H(h_1,h_2) &= \tilde{g}(h_2) - \tilde{g}(0) \\
&= \tilde{g}'(\xi')h_2 \\
&= \left(\frac{\partial f}{\partial x_j}(\hat{x}_0 + h_1\hat{e}_i + \xi'\hat{e}_j) - \frac{\partial f}{\partial x_j}(\hat{x}_0 + \xi'\hat{e}_j) \right) h_2.
\end{aligned}$$

Como $[\hat{x}_0 + \xi'\hat{e}_j, \hat{x}_0 + h_1\hat{e}_i + \xi'\hat{e}_j] \subset B_r(\hat{x}_0)$, con $(\hat{x}_0 + h_1\hat{e}_i + \xi'\hat{e}_j) - (\hat{x}_0 + \xi'\hat{e}_j) = h_1\hat{e}_i$, nuevamente por la proposición 4.13 (aplicada a la función $\frac{\partial f}{\partial x_j}$) se tiene que existe $\eta' \in (0, |h_1|)$ tal que

$$\begin{aligned}
H(h_1,h_2) &= \left(\frac{\partial f}{\partial x_j}(\hat{x}_0 + h_1\hat{e}_i + \xi'\hat{e}_j) - \frac{\partial f}{\partial x_j}(\hat{x}_0 + \xi'\hat{e}_j) \right) h_2 \\
&= \frac{\partial \left(\frac{\partial f}{\partial x_j} \right)}{\partial x_i}(\hat{x}_0 + \eta'\hat{e}_i + \xi'\hat{e}_j)h_1h_2 \\
&= \frac{\partial^2 f}{\partial x_i \partial x_j}(\hat{x}_0 + \eta'\hat{e}_i + \xi'\hat{e}_j)h_1h_2.
\end{aligned}$$

De esta forma, si $h_1h_2 \neq 0$, por la continuidad de $\frac{\partial^2 f}{\partial x_i \partial x_j}$ en $\hat{x}_0$ y que $(\xi', \eta') \to (0,0)$ si $(h_1, h_2) \to (0,0)$, concluimos que

$$\begin{aligned}
\lim_{(h_1,h_2)\to(0,0)} \frac{H(h_1,h_2)}{h_1h_2} &= \lim_{(\xi',\eta')\to(0,0)} \frac{\partial^2 f}{\partial x_i \partial x_j}(\hat{x}_0 + \eta'\hat{e}_i + \xi'\hat{e}_j) \\
&= \frac{\partial^2 f}{\partial x_i \partial x_j}(\hat{x}_0).
\end{aligned}$$

Por tanto, como el límite de cualquier función, si existe, es único, tenemos que

$$\begin{aligned}\frac{\partial^2 f}{\partial x_j \partial x_i}(\hat{x}_0) &= \lim_{(h_1,h_2)\to(0,0)} \frac{H(h_1,h_2)}{h_1 h_2} \\ &= \frac{\partial^2 f}{\partial x_i \partial x_j}(\hat{x}_0),\end{aligned}$$

que es la identidad que deseabamos probar. ■

Como mencionamos al inicio de esta sección, podemos calcular sucesivamente tantas derivadas direccionales (en la misma o en diferentes direcciones) como nos lo permita la función con la que estemos trabajando.

Dado que en términos prácticos es suficiente hacer esto para los vectores de una base ortonormal de $\mathbb{R}^n$, lo siguiente que haremos será "extender" la definición 4.38 para cualquier $k \in \mathbb{N}$.

Definición 4.41 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$, $r > 0$ tal que $B_r(\hat{x}_0) \subset U$, $k \in \mathbb{N}$ y $x_1, \ldots, x_n$ las coordenadas determinadas por una base ortonormal $\{\hat{e}_1, \ldots, \hat{e}_n\}$ de $\mathbb{R}^n$. Si $i_1, \ldots, i_k, i_{k+1} \in \{1, \ldots, n\}$ son tales que $\frac{\partial^k f}{\partial x_{i_k} \cdots \partial x_{i_1}}(\hat{x})$ existe para toda $\hat{x} \in B_r(\hat{x}_0)$, definimos la $k+1$ derivada parcial de f con respecto a $x_{i_1}, \ldots, x_{i_k}, x_{i_{k+1}}$, en $\hat{x}_0$, que denotaremos por $\frac{\partial^{k+1} f}{\partial x_{i_{k+1}} \partial x_{i_k} \cdots \partial x_{i_1}}(\hat{x}_0)$, como la derivada parcial con respecto de $x_{i_{k+1}}$ de la función $\frac{\partial^k f}{\partial x_{i_k} \cdots \partial x_{i_1}}$ en $\hat{x}_0$. Es decir,*

$$\frac{\partial^{k+1} f}{\partial x_{i_{k+1}} \partial x_{i_k} \cdots \partial x_{i_1}}(\hat{x}_0) := \lim_{h\to 0} \frac{\frac{\partial^k f}{\partial x_{i_k} \cdots \partial x_{i_1}}(\hat{x}_0 + h\hat{e}_{i_{k+1}}) - \frac{\partial^k f}{\partial x_{i_k} \cdots \partial x_{i_1}}(\hat{x}_0)}{h}.$$

Si $i_1 = i_2 = \cdots = i_k = i_{k+1} = i$, la derivada anterior la denotamos por $\frac{\partial^{k+1} f}{\partial x_i^{k+1}}(\hat{x}_0)$ y decimos que es la $k+1$ derivada parcial de f con respecto de x_i, en $\hat{x}_0$.

A propósito de la notación usada en esta definición, y en concordancia con la última parte, en general, si en la expresión $\partial x_{i_k} \cdots \partial x_{i_1}$ existen m índices consecutivos iguales, esta parte de la expresión la sustituiremos por ∂x_i^m, en donde x_i es la variable que se repite m veces consecutivas.

Por ejemplo, la derivada parcial de orden 6

$$\frac{\partial^6 f}{\partial x_3 \partial x_2 \partial x_2 \partial x_3 \partial x_1 \partial x_1}(\hat{x}_0)$$

se podrá escribir como

$$\frac{\partial^6 f}{\partial x_3 \partial x_2^2 \partial x_3 \partial x_1^2}(\hat{x}_0)$$

Dada $k \in \mathbb{N}$, como seguramente el lector recordará de sus cursos de álgebra, existen n^k formas de elegir ordenaciones (con repetición) con elementos del conjunto $\{1, \ldots, n\}$, de tal forma que existen n^k derivadas parciales de orden k para una función f de $\mathbb{R}^n$ en $\mathbb{R}$. Cuando todas estas derivadas parciales existen en los puntos de un conjunto (abierto) sobre el cual esté definida la función f, y además son continuas en este mismo conjunto, diremos que f *es una función de clase C^k*. Este concepto jugará un papel muy importante más adelante, y lo dejaremos plasmado en la siguiente

Definición 4.42 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$ y $k \in \mathbb{N}$. Decimos que f es una función de clase C^k en U si existen todas las derivadas parciales de orden k de f en cada punto de U, y además estas derivadas parciales son continuas en cada punto de U.*

El concepto anterior está muy relacionado con las hipótesis de la proposición 4.25 y el teorema 4.40, razón por la cual enunciaremos el siguiente par de proposiciones y dejaremos su prueba al lector.

Proposición 4.43 *Si $f : U \subset \mathbb{R}^n \to \mathbb{R}$ es de clase C^1 en U, entonces f es derivable para toda $\hat{x} \in U$.*

Proposición 4.44 *Si $f : U \subset \mathbb{R}^n \to \mathbb{R}$ es de clase C^2 en U, entonces*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\hat{x}) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\hat{x})$$

para toda $\hat{x} \in U$ y para todas $i, j \in \{1, \ldots, n\}$.

Cuando una función es de clase C^k, muchas de sus derivadas parciales de orden k coinciden. En efecto, si en dos derivadas parciales de orden k de una función f se deriva con respecto a las mismas variables, y además se deriva (con respecto de cada una de ellas y sin importar el orden) el mismo número de veces, entonces estas derivadas parciales son iguales. Enunciar con toda precisión (¡y probar!) la afirmación anterior no está dentro de los objetivos de este texto, y sólo baste decir que en el fondo su prueba se basa en el teorema 4.40.

Antes de iniciar con otro tema, es importante enfatizar qué conceptos hemos definido (¡y cuáles no hemos definido!) en esta sección.

Como lo dice su nombre, aquí sólo definimos el concepto de derivadas direccionales de orden superior de una función f, lo que no significa que hayamos definido el concepto de derivada (global) de orden superior de f.

Como hemos venido insistiendo, la derivada de una función f es una función que a cada $\hat{x}$ del dominio de f le asocia una función lineal de $\mathbb{R}^n$ en $\mathbb{R}$ que denotamos por $Df(\hat{x})$. Siguiendo la idea de que la segunda derivada de una función f debería de ser la derivada de la derivada de f, para definir este concepto tendríamos que definir lo que significa derivar una función cuyo dominio está contenido en $\mathbb{R}^n$ y que tiene como contradominio al conjunto de las funciones lineales de $\mathbb{R}^n$ en $\mathbb{R}$.

Seguramente el lector estará de acuerdo en que abordar el problema que se planteó en el párrafo presedente, escapa a los objetivos de este texto. No obstante lo anterior, con los conceptos desarrollados en esta sección nos es suficiente para abordar un tema muy importante relacionado con la "aproximación" de una función cerca de un punto, tema que abordaremos en la siguiente sección.

4.5 Aproximación polinomial

Como el lector recordará, el concepto de derivada (global) de una función f de $\mathbb{R}^n$ en $\mathbb{R}$ en un punto $\hat{x}_0$, está íntimamente relacionado con el problema de aproximar a dicha función alrededor (o cerca) de $\hat{x}_0$ por medio de una función lineal. En términos más concretos, la derivabilidad de f en $\hat{x}_0$ se reduce al hecho de que la función

$$P(\hat{x}) = Df(\hat{x}_0)(\hat{x} - \hat{x}_0) + f(\hat{x}_0)$$

es la única función "de este tipo" que tiene la propiedad de que

$$\lim_{\hat{x} \to \hat{x}_0} \frac{f(\hat{x}) - P(\hat{x})}{\|\hat{x} - \hat{x}_0\|} = 0. \tag{4.37}$$

Si ahora recordamos que

$$Df(\hat{x}_0)(\hat{x} - \hat{x}_0) = \nabla f(\hat{x}_0) \cdot (\hat{x} - \hat{x}_0),$$

tendremos que en un sistema de coordenadas $x_1, \ldots, x_n$, en donde $\hat{x}_0 = \left(x_1^{(0)}, \ldots, x_n^{(0)}\right)$, la función P toma la forma

$$\begin{aligned}
P(\hat{x}) &= P(x_1, \ldots, x_n) \\
&= \nabla f(\hat{x}_0) \cdot (\hat{x} - \hat{x}_0) + f(\hat{x}_0) \\
&= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\hat{x}_0) \left(x_i - x_i^{(0)}\right) + f(\hat{x}_0), \qquad (4.38)
\end{aligned}$$

de donde concluimos que P es una función *polinomial* de grado a lo más 1 de las coodenadas (o variables) $x_1, \ldots, x_n$, con la particularidad de que $P(\hat{x}_0) = f(\hat{x}_0)$.

Visto de esta forma, con base en la proposición 4.25 podemos asegurar que si las derivadas parciales $\frac{\partial f}{\partial x_i}$ existen en una vecindad del punto $\hat{x}_0$ y son continuas en éste (lo que se satisface sobradamente si f es de clase C^1 en una vecindad de $\hat{x}_0$), entonces el polinomio de grado a lo más 1 en las variables $x_1, \ldots, x_n$ dado por 4.38 es el único polinomio que satisface la condición 4.37.

Como seguramente el lector ya se estará imaginando, la siguiente pregunta que nos haremos será la siguiente: ¿qué propiedades deberá tener una función f para que podamos asegurar que existe una función polinomial de grado a lo más 2 en las variables $x_1, \ldots, x_n$, con la particularidad de que valga lo mismo que f en $\hat{x}_0$, tal que *se parezca mucho a f alrededor de $\hat{x}_0$*?

Para abordar este problema, empezaremos analizando el caso de una función definida en $\mathbb{R}^2$. Supongamos entonces que $f : U \subset \mathbb{R}^2 \to \mathbb{R}$ y que $\hat{x}_0 = (x_0, y_0) \in U$.

Primero notemos que una función $P_2(x, y)$ polinomial en las variables x y y de grado a lo más 2, es de la forma

$$P_2(x, y) = Ax^2 + By^2 + Cxy + Dx + Ey + F.$$

Si se desea que esta función tome el valor $f(\hat{x}_0)$ cuando $(x, y) = (x_0, y_0)$, entonces deberá ser de la forma

$$P_2(x, y) = A(x - x_0)^2 + B(y - y_0)^2 + C(x - x_0)(y - y_0) + D(x - x_0) + E(y - y_0) + f(\hat{x}_0), \tag{4.39}$$

por lo que nuestro problema se "reduce" a encontrar los coeficientes A, B, C, D y E.

Ahora, si el criterio para decir que P_2 se parece mucho a f alrededor de $\hat{x}_0$ es que satisfaga la condición 4.37, dado que

$$(x - x_0)^2, (y - y_0)^2, |(x - x_0)(y - y_0)| \leq \|\hat{x} - \hat{x}_0\|^2,$$

entonces

$$\lim_{(x,y)\to(x_0,y_0)} \frac{(x - x_0)^2}{\|\hat{x} - \hat{x}_0\|} = \lim_{(x,y)\to(x_0,y_0)} \frac{(y - y_0)^2}{\|\hat{x} - \hat{x}_0\|} = \lim_{(x,y)\to(x_0,y_0)} \frac{(x - x_0)(y - y_0)}{\|\hat{x} - \hat{x}_0\|} = 0.$$

Por lo tanto,

$$\begin{aligned}
0 &= \lim_{\hat{x}\to\hat{x}_0} \frac{f(\hat{x}) - P_2(\hat{x})}{\|\hat{x} - \hat{x}_0\|} \\
&= \lim_{\hat{x}\to\hat{x}_0} \frac{f(\hat{x}) - \left(A(x - x_0)^2 + B(y - y_0)^2 + C(x - x_0)(y - y_0) + D(x - x_0) + E(y - y_0) + f(\hat{x}_0)\right)}{\|\hat{x} - \hat{x}_0\|} \\
&= \lim_{\hat{x}\to\hat{x}_0} \frac{f(\hat{x}) - (D(x - x_0) + E(y - y_0) + f(\hat{x}_0))}{\|\hat{x} - \hat{x}_0\|},
\end{aligned}$$

lo cual es equivalente a que f sea derivable en el punto $\hat{x}_0$. De esta forma, se deberá tener que

$$D = \frac{\partial f}{\partial x}(\hat{x}_0) \qquad \text{y} \qquad E = \frac{\partial f}{\partial y}(\hat{x}_0).$$

En conclusión, si f es derivable en $\hat{x}_0$, entonces la función polinomial

$$P_2(x, y) = A(x - x_0)^2 + B(y - y_0)^2 + C(x - x_0)(y - y_0) + \frac{\partial f}{\partial x}(\hat{x}_0)(x - x_0) + \frac{\partial f}{\partial y}(\hat{x}_0)(y - y_0) + f(\hat{x}_0) \tag{4.40}$$

cumple la condición 4.37 para cualesquiera A, B y C números reales.

Tal vez esta última parte no sea del todo una buena noticia, pues nos asegura que hay una gran cantidad de funciones polinomiales de grado 2 que se "parecen" mucho a f alrededor de $\hat{x}_0$, lo que sin duda nos deja con la pregunta de si de entre todas ellas no habrá alguna que sea "la mejor" de todas.

Lo que haremos a continuación será mostrar que sí existe un criterio con base en el cual podemos elegir, de entre todas las funciones polinomiales dadas por 4.40, la que "se parece más" a f alrededor de $\hat{x}_0$. Para ello, basta con hacer notar que una forma de interpretar la multicitada condición 4.37 es que el número $f(\hat{x}) - P(\hat{x})$ se hace 0 mucho más "rápido" que el número $\|\hat{x} - \hat{x}_0\|$.

Con base en esta interpretación, y dado que, si $\|\hat{x} - \hat{x}_0\| \leq 1$, se tiene que $\|\hat{x} - \hat{x}_0\|^2 \leq \|\hat{x} - \hat{x}_0\|$, entonces pedir que

$$\lim_{\hat{x}\to\hat{x}_0} \frac{f(\hat{x}) - P(\hat{x})}{\|\hat{x} - \hat{x}_0\|^2} = 0 \tag{4.41}$$

sin duda es una condición “más fuerte” que la condición 4.37 (de hecho, por el problema 37 del capítulo 2, sabemos que si se cumple la condición anterior, entonces se cumple la condición 4.37 (¡pero no lo recíproco!)). Por esta razón, si suponemos que una función polinomial de grado a lo más 2 satisface la condición 4.41, podemos dar por hecho que ya es de la forma 4.40.

Supongamos entonces que P_2 es una función polinomial como en 4.40 que satisface la condición 4.41. Es decir, supongamos que

$$\begin{aligned}
&\lim_{\hat{x}\to\hat{x}_0} \frac{f(\hat{x}) - P_2(\hat{x})}{\|\hat{x}-\hat{x}_0\|^2} \\
&= \lim_{\hat{x}\to\hat{x}_0} \frac{f(\hat{x}) - \left(A(x-x_0)^2 + B(y-y_0)^2 + C(x-x_0)(y-y_0) + \frac{\partial f}{\partial x}(\hat{x}_0)(x-x_0) + \frac{\partial f}{\partial y}(\hat{x}_0)(y-y_0) + f(\hat{x}_0)\right)}{\|\hat{x}-\hat{x}_0\|^2} \\
&= 0.
\end{aligned}$$

Si nos aproximamos a (x_0, y_0) por puntos de la forma $(x, y) = (x_0 + h, y_0)$, se deberá tener que

$$\begin{aligned}
0 &= \lim_{h\to 0} \frac{f(x_0+h,y_0) - \left(Ah^2 + \frac{\partial f}{\partial x}(x_0,y_0)h + f(x_0,y_0)\right)}{h^2} \\
&= \lim_{h\to 0} \left[\frac{f(x_0+h,y_0) - \frac{\partial f}{\partial x}(x_0,y_0)h - f(x_0,y_0)}{h^2} - A\right],
\end{aligned}$$

de modo que al cociente

$$\frac{f(x_0+h,y_0) - \frac{\partial f}{\partial x}(x_0,y_0)h - f(x_0,y_0)}{h^2}$$

lo podemos tratar como un cociente de dos funciones que dependen de la variable (real) h.

Ahora, si suponemos que f y sus derivadas parciales son funciones derivables en todos los puntos de una vecindad del punto $\hat{x}_0$ (lo cual se puede garantizar suponiendo que f es de clase C^2 en una vecindad de este punto), aplicando (las veces que haga falta) la regla de L'Hospital para este tipo de funciones (y por supuesto la regla de la cadena probada en la proposición 4.30), tenemos que

$$\begin{aligned}
\lim_{h\to 0} \frac{f(x_0+h,y_0) - \frac{\partial f}{\partial x}(x_0,y_0)h - f(x_0,y_0)}{h^2} &= \lim_{h\to 0} \frac{\frac{\partial f}{\partial x}(x_0+h,y_0) - \frac{\partial f}{\partial x}(x_0,y_0)}{2h} \\
&= \frac{1}{2}\lim_{h\to 0} \frac{\frac{\partial f}{\partial x}(x_0+h,y_0) - \frac{\partial f}{\partial x}(x_0,y_0)}{h} \\
&= \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(x_0,y_0),
\end{aligned}$$

de donde concluimos que

$$A = \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(x_0,y_0).$$

Aproximándonos a (x_0, y_0) por puntos de la forma $(x, y) = (x_0, y_0 + h)$, por un procedimiento análogo se concluye que

$$B = \frac{1}{2}\frac{\partial^2 f}{\partial y^2}(x_0,y_0).$$

Finalmente, para deducir el valor del coeficiente C, nos aproximaremos a (x_0, y_0) por puntos de la forma $(x, y) = (x_0 + h, y_0 + h)$.

En este caso, usando nuevamente la regla de L'Hospital en la segunda y tercera identidad, se deberá tener que

$$0 = \lim_{h\to 0} \frac{f(x_0+h,y_0+h) - \left(\frac{1}{2}\frac{\partial^2 f}{\partial x^2}(\hat{x}_0)h^2 + \frac{1}{2}\frac{\partial^2 f}{\partial y^2}(\hat{x}_0)h^2 + Ch^2 + \frac{\partial f}{\partial x}(\hat{x}_0)h + \frac{\partial f}{\partial y}(\hat{x}_0)h + f(\hat{x}_0)\right)}{h^2}$$

$$= \lim_{h\to 0} \frac{\frac{\partial f}{\partial x}(x_0+h,y_0+h)+\frac{\partial f}{\partial y}(x_0+h,y_0+h)-\left(\frac{\partial^2 f}{\partial x^2}(\hat{x}_0)h+\frac{\partial^2 f}{\partial y^2}(\hat{x}_0)h+2Ch+\frac{\partial f}{\partial x}(\hat{x}_0)+\frac{\partial f}{\partial y}(\hat{x}_0)\right)}{2h}$$

$$= \lim_{h\to 0}\frac{1}{2}\left[\frac{\partial^2 f}{\partial x^2}(x_0+h,y_0+h)+\frac{\partial^2 f}{\partial y\partial x}(x_0+h,y_0+h)+\frac{\partial^2 f}{\partial x\partial y}(x_0+h,y_0+h)+\frac{\partial^2 f}{\partial y^2}(x_0+h,y_0+h)\right.$$
$$\left.-\left(\frac{\partial^2 f}{\partial x^2}(\hat{x}_0)+\frac{\partial^2 f}{\partial y^2}(\hat{x}_0)+2C\right)\right]$$
$$= \frac{1}{2}\left(\frac{\partial^2 f}{\partial y\partial x}(\hat{x}_0)+\frac{\partial^2 f}{\partial x\partial y}(\hat{x}_0)-2C\right),$$

de donde obtenemos que

$$C=\frac{1}{2}\left(\frac{\partial^2 f}{\partial y\partial x}(\hat{x}_0)+\frac{\partial^2 f}{\partial x\partial y}(\hat{x}_0)\right).$$

De lo anterior concluimos que, si existen las segundas derivadas parciales de f (en al menos una vecindad del punto $\hat{x}_0$ y en este punto son continuas), y P_2 es una función polinomial de grado menor o igual a 2 que satisface la condición 4.41, entonces necesariamente P_2 está dado por

$$P_2(x,y)=\frac{1}{2}\left(\frac{\partial^2 f}{\partial x^2}(\hat{x}_0)(x-x_0)^2+\frac{\partial^2 f}{\partial y^2}(y-y_0)^2+\left(\frac{\partial^2 f}{\partial y\partial x}(\hat{x}_0)+\frac{\partial^2 f}{\partial x\partial y}(\hat{x}_0)\right)(x-x_0)(y-y_0)\right) \quad (4.42)$$
$$+\frac{\partial f}{\partial x}(\hat{x}_0)(x-x_0)+\frac{\partial f}{\partial y}(\hat{x}_0)(y-y_0)+f(\hat{x}_0).$$

Es importante enfatizar que el resultado anterior se obtuvo bajo el supuesto de que f es de clase C^2 en una vecindad del punto $\hat{x}_0$, y de que existe una función polinomial de grado menor o igual a 2 que satisface la condición 4.41. La buena noticia es que, bajo las mismas hipótesis sobre f, lo recíproco también es cierto, es decir la función polinomial dada por 4.42 cumple la condición 4.41.

Seguramente a esta alturas el lector ya estará empezando a vislumbrar el resultado más general hacia el cual estamos dirigiendo nuestros pasos. Todo parece indicar que, si $f: U\subset\mathbb{R}^n\to\mathbb{R}$ y $\hat{x}_0\in U$ son tales que f es de clase C^N en una vecindad del punto $\hat{x}_0$ para alguna $N\in\mathbb{N}$, entonces la función polinomial de grado a lo más N dada por

$$P_N(\hat{x})$$
$$= f(\hat{x}_0)+\sum_{i=1}^{n}\frac{\partial f}{\partial x_i}(\hat{x}_0)\left(x_i-x_i^{(0)}\right)+\cdots+\frac{1}{N!}\sum_{i_1,\ldots,i_N=1}^{n}\frac{\partial^N f}{\partial x_{i_1}\cdots\partial x_{i_N}}(\hat{x}_0)\left(x_{i_1}-x_{i_1}^{(0)}\right)\cdots\left(x_{i_N}-x_{i_N}^{(0)}\right)$$

en donde $\hat{x}=(x_1,\ldots,x_n)$ y $\hat{x}_0=\left(x_1^{(0)},\ldots,x_n^{(0)}\right)$, debe ser tal que

$$\lim_{\hat{x}\to\hat{x}_0}\frac{f(\hat{x})-P_N(\hat{x})}{\|\hat{x}-\hat{x}_0\|^N}=0$$

y además es la única con esta propiedad.

Este resultado es conocido como el *Teorema de Taylor*, y aquí lo formularemos un poco diferente a como lo acabamos de describir en el párrafo anterior. También es importante mencionar que la prueba de este teorema se basa en el Teorema de Taylor, en su versión para funciones de $\mathbb{R}$ en $\mathbb{R}$.

Antes de formular y probar el teorema de Taylor, daremos las siguientes definiciones.

Definición 4.45 *Sean $f: U\subset\mathbb{R}^n\to\mathbb{R}$, $\hat{x}_0\in U$, $r>0$ y $N\in\mathbb{N}$ tales que f es de clase C^N en $B_r(\hat{x}_0)\subset U$.*

1. *Definimos el polinomio de Taylor de grado N de la función f en $\hat{x}_0$, que denotamos por $P_{N,f,\hat{x}_0}$, como*

$$P_{N,f,\hat{x}_0}(\hat{x})$$
$$:= f(\hat{x}_0)+\sum_{i=1}^{n}\frac{\partial f}{\partial x_i}(\hat{x}_0)\left(x_i-x_i^{(0)}\right)+\cdots+\frac{1}{N!}\sum_{i_1,\ldots,i_N=1}^{n}\frac{\partial^N f}{\partial x_{i_N}\cdots\partial x_{i_1}}(\hat{x}_0)\left(x_{i_1}-x_{i_1}^{(0)}\right)\cdots\left(x_{i_N}-x_{i_N}^{(0)}\right)$$

(en donde $\hat{x}=(x_1,\ldots,x_n)\in\mathbb{R}^n$ y $\hat{x}_0=\left(x_1^{(0)},\ldots,x_n^{(0)}\right)$).

2. *Definimos el residuo de orden N de f en $\hat{x}_0$, que denotamos por $R_{N,f,\hat{x}_0}$, como*

$$R_{N,f,\hat{x}_0}(\hat{x}) := f(\hat{x}) - P_{N,f,\hat{x}_0}(\hat{x})$$

para $\hat{x} \in U$.

Teorema 4.46 (de Taylor) *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$, $r > 0$ y $N \in \mathbb{N}$ tales que f es de clase C^{N+1} en $B_r(\hat{x}_0) \subset U$. El polinomio $P_{N,f,\hat{x}_0}$ y su residuo $R_{N,f,\hat{x}_0}$ satisfacen que:*

1. *si $\hat{x} \in B_r(\hat{x}_0)$ existe $\hat{\xi} \in [\hat{x}_0, \hat{x}]$ tal que*

$$\begin{aligned}&R_{N,f,\hat{x}_0}(\hat{x})\\ &= \frac{1}{(N+1)!} \sum_{i_1,\ldots,i_N,i_{N+1}=1}^{n} \frac{\partial^{N+1} f}{\partial x_{i_{N+1}} \partial x_{i_N} \cdots \partial x_{i_1}} \left(\hat{\xi}\right) \left(x_{i_1} - x_{i_1}^{(0)}\right) \cdots \left(x_{i_N} - x_{i_N}^{(0)}\right) \left(x_{i_{N+1}} - x_{i_{N+1}}^{(0)}\right)\end{aligned}$$

2.

$$\lim_{\hat{x} \to \hat{x}_0} \frac{R_{N,f,\hat{x}_0}(\hat{x})}{\left\|\hat{x} - \hat{x}_0\right\|^N} = \lim_{\hat{x} \to \hat{x}_0} \frac{f(\hat{x}) - P_{N,f,\hat{x}_0}(\hat{x})}{\left\|\hat{x} - \hat{x}_0\right\|^N} = 0$$

3. *$P_{N,f,\hat{x}_0}$ es el único polinomio de grado menor o igual a N que satisface la condición del inciso* 2.

Demostración. Para la prueba del inciso (1), tomamos $\hat{x} \in B_r(\hat{x}_0)$, $\hat{x} \neq \hat{x}_0$ (el caso $\hat{x} = \hat{x}_0$ es inmediato tomando $\hat{\xi} = \hat{x}_0$), y definimos $g : (-r/\left\|\hat{x} - \hat{x}_0\right\|, r/\left\|\hat{x} - \hat{x}_0\right\|) \subset \mathbb{R} \to \mathbb{R}$ como $g(t) = f(\hat{x}_0 + t(\hat{x} - \hat{x}_0))$.

Por el problema 34 (tomando $\hat{u} = \hat{x} - \hat{x}_0$) sabemos que g es de clase C^{N+1} en su dominio de tal forma que, por el Teorema de Taylor para funciones de $\mathbb{R}$ en $\mathbb{R}$, tenemos que

$$\begin{aligned}&R_{N,g,0}(1)\\ &= \frac{1}{(N+1)!} g^{(N+1)}(\xi)(1-0)\\ &= \frac{1}{(N+1)!} g^{(N+1)}(\xi)\\ &= \frac{1}{(N+1)!} \sum_{i_1,\ldots,i_N,i_{N+1}=1}^{n} \frac{\partial^{N+1} f}{\partial x_{i_{N+1}} \partial x_{i_N} \cdots \partial x_{i_1}} \left(\hat{\xi}\right) \left(x_{i_1} - x_{i_1}^{(0)}\right) \cdots \left(x_{i_N} - x_{i_N}^{(0)}\right) \left(x_{i_{N+1}} - x_{i_{N+1}}^{(0)}\right),\end{aligned}$$

en donde $\hat{\xi} = \hat{x}_0 + \xi(\hat{x} - \hat{x}_0)$ para alguna $\xi \in (0,1)$, de modo que $\hat{\xi} \in [\hat{x}_0, \hat{x}]$.

Por otra parte, como

$$\begin{aligned}R_{N,g,0}(1) &= g(1) - P_{N,g,0}(1)\\ &= f(\hat{x}) - \sum_{i=0}^{N} \frac{g^{(i)}(0)}{i!}(1-0)^i\\ &= f(\hat{x}) - (g(0) + g'(0) + \frac{1}{2!}g''(0) + \cdots + \frac{1}{N!}g^{(N)}(0)\\ &= f(\hat{x}) - \left(f(\hat{x}_0) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\hat{x}_0)\left(x_i - x_i^{(0)}\right) + \cdots + \right.\\ &\left. + \frac{1}{N!} \sum_{i_1,\ldots,i_N=1}^{n} \frac{\partial^N f}{\partial x_{i_N} \cdots \partial x_{i_1}}(\hat{x}_0)\left(x_{i_1} - x_{i_1}^{(0)}\right) \cdots \left(x_{i_N} - x_{i_N}^{(0)}\right)\right)\\ &= f(\hat{x}) - P_{N,f,\hat{x}_0}(\hat{x})\\ &= R_{N,f,\hat{x}_0}(\hat{x}),\end{aligned}$$

concluimos que

$$R_{N,f,\hat{x}_0}(\hat{x}) = \frac{1}{(N+1)!} \sum_{i_1,\ldots,i_N,i_{N+1}=1}^{n} \frac{\partial^{N+1} f}{\partial x_{i_{N+1}} \partial x_{i_N} \cdots \partial x_{i_1}} \left(\hat{\xi}\right) \left(x_{i_1} - x_{i_1}^{(0)}\right) \cdots \left(x_{i_N} - x_{i_N}^{(0)}\right) \left(x_{i_{N+1}} - x_{i_{N+1}}^{(0)}\right)$$

para alguna $\hat{\xi} \in [\hat{x}_0, \hat{x}]$.

Ahora, dado que

$$\left|x_i - x_i^{(0)}\right| \leq \|\hat{x} - \hat{x}_0\|$$

para cada $i \in \{1, \ldots, n\}$, entonces

$$\begin{aligned}
&\left|R_{N,f,\hat{x}_0}(\hat{x})\right| \\
&= \left| \frac{1}{(N+1)!} \sum_{i_1,\ldots,i_N,i_{N+1}=1}^{n} \frac{\partial^{N+1} f}{\partial x_{i_{N+1}} \partial x_{i_N} \cdots \partial x_{i_1}} \left(\hat{\xi}\right) \left(x_{i_1} - x_{i_1}^{(0)}\right) \cdots \left(x_{i_N} - x_{i_N}^{(0)}\right) \left(x_{i_{N+1}} - x_{i_{N+1}}^{(0)}\right) \right| \\
&\leq \frac{1}{(N+1)!} \sum_{i_1,\ldots,i_N,i_{N+1}=1}^{n} \left| \frac{\partial^{N+1} f}{\partial x_{i_{N+1}} \partial x_{i_N} \cdots \partial x_{i_1}} \left(\hat{\xi}\right) \right| \left|x_{i_1} - x_{i_1}^{(0)}\right| \cdots \left|x_{i_N} - x_{i_N}^{(0)}\right| \left|x_{i_{N+1}} - x_{i_{N+1}}^{(0)}\right| \\
&\leq \frac{1}{(N+1)!} \sum_{i_1,\ldots,i_N,i_{N+1}=1}^{n} \left| \frac{\partial^{N+1} f}{\partial x_{i_{N+1}} \partial x_{i_N} \cdots \partial x_{i_1}} \left(\hat{\xi}\right) \right| \|\hat{x} - \hat{x}_0\| \cdots \|\hat{x} - \hat{x}_0\| \|\hat{x} - \hat{x}_0\| \\
&= \frac{\|\hat{x} - \hat{x}_0\|^{N+1}}{(N+1)!} \sum_{i_1,\ldots,i_N,i_{N+1}=1}^{n} \left| \frac{\partial^{N+1} f}{\partial x_{i_{N+1}} \partial x_{i_N} \cdots \partial x_{i_1}} \left(\hat{\xi}\right) \right|,
\end{aligned}$$

de tal forma que

$$\frac{\left|R_{N,f,\hat{x}_0}(\hat{x})\right|}{\|\hat{x} - \hat{x}_0\|^N} \leq \|\hat{x} - \hat{x}_0\| \left(\frac{1}{(N+1)!} \sum_{i_1,\ldots,i_N,i_{N+1}=1}^{n} \left| \frac{\partial^{N+1} f}{\partial x_{i_{N+1}} \partial x_{i_N} \cdots \partial x_{i_1}} \left(\hat{\xi}\right) \right| \right).$$

Por tanto, dado que f es de clase C^{N+1} en $B_r(\hat{x}_0)$, se tiene que

$$\lim_{\hat{x} \to \hat{x}_0} \frac{\partial^{N+1} f}{\partial x_{i_{N+1}} \partial x_{i_N} \cdots \partial x_{i_1}} \left(\hat{\xi}\right) = \frac{\partial^{N+1} f}{\partial x_{i_{N+1}} \partial x_{i_N} \cdots \partial x_{i_1}} (\hat{x}_0)$$

para todas $i_1, , i_N, i_{N+1} \in \{1, \ldots, n\}$. Así concluimos que

$$\lim_{\hat{x} \to \hat{x}_0} \frac{R_{N,f,\hat{x}_0}(\hat{x})}{\|\hat{x} - \hat{x}_0\|^N} = 0,$$

lo cual prueba el inciso (2).

Finalmente, para la prueba del inciso (3), supongamos que P_N es otro polinomio de grado menor o igual a N para el cual se satisface que

$$\lim_{\hat{x} \to \hat{x}_0} \frac{f(\hat{x}) - P_N(\hat{x})}{\|\hat{x} - \hat{x}_0\|^N} = 0.$$

Dado que

$$\begin{aligned}
\lim_{\hat{x} \to \hat{x}_0} \frac{f(\hat{x}) - P_{N,f,\hat{x}_0}(\hat{x})}{\|\hat{x} - \hat{x}_0\|^N} &= \lim_{\hat{x} \to \hat{x}_0} \frac{R_{N,f,\hat{x}_0}(\hat{x})}{\|\hat{x} - \hat{x}_0\|^N} \\
&= 0,
\end{aligned}$$

restando estos dos últimos límites tenemos que

$$\lim_{\hat{x} \to \hat{x}_0} \frac{P_N(\hat{x}) - P_{N,f,\hat{x}_0}(\hat{x})}{\|\hat{x} - \hat{x}_0\|^N} = 0,$$

de tal forma que, si hacemos $P(\hat{x}) = P_N(\hat{x}) - P_{N,f,\hat{x}_0}(\hat{x})$, se tiene que P es un polinomio de grado menor o igual a N con la propiedad de que

$$\lim_{\hat{x}\to\hat{x}_0} \frac{P(\hat{x})}{\|\hat{x}-\hat{x}_0\|^N} = 0.$$

El siguiente paso será demostrar que un polinomio con estas características tiene que ser la función constante 0. Para ello, tomemos $\hat{x} \in \mathbb{R}^n$, fija. Dado que la identidad $P(\hat{x}_0) = 0$ se sigue inmediatamente del límite anterior y de la continuidad de P, supondremos que $\hat{x} \neq \hat{x}_0$.

Definimos $p(t) = P(\hat{x}_0 + t(\hat{x}-\hat{x}_0))$ para $t \in \mathbb{R}$; nótese que p será un polinomio en la variable t de grado a lo más N, con la propiedad de que

$$\begin{aligned}\lim_{t\to 0}\left|\frac{p(t)}{t^N}\right| &= \|\hat{x}-\hat{x}_0\|^N \lim_{t\to 0} \frac{|P(\hat{x}_0 + t(\hat{x}-\hat{x}_0))|}{\|(\hat{x}_0 + t(\hat{x}-\hat{x}_0)) - \hat{x}_0\|^N} \\ &= 0,\end{aligned}$$

o equivalentemente, que

$$\lim_{t\to 0} \frac{p(t)}{t^N} = 0.$$

Por lo tanto, como para el caso de una variable ya sabemos que un polinomio con estas características tiene que ser la constante cero, concluimos que

$$\begin{aligned}0 &= p(1) \\ &= P(\hat{x}) \\ &= P_N(\hat{x}) - P_{N,f,\hat{x}_0}(\hat{x}).\end{aligned}$$

Es decir, que $P_N(\hat{x}) = P_{N,f,\hat{x}_0}(\hat{x})$ para toda $\hat{x} \in \mathbb{R}^n$ que es lo que queríamos demostrar. ■

Los polinomios de Taylor son una herramienta muy importante en el análisis del comportamiento de una función alrededor de un punto, y nos serán muy útiles cuando en la próxima sección abordemos el tema de los valores máximos y mínimos de una función. Por ahora, veremos con un ejemplo cómo se pueden usar para el cálculo de límites.

Ejemplo 4.47 *Considere la función*

$$f(x,y) = \frac{\ln(1+x^2) + \ln(1+y^2)}{x^2+y^2}$$

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

Nuestro objetivo es calcular

$$\lim_{(x,y)\to(0,0)} f(x,y).$$

Para ello, bastará con calcular el polinomio de Taylor de grado 2 *en el* $(0,0)$ *de la función*

$$g(x,y) = \ln(1+x^2) + \ln(1+y^2),$$

dado que el denominador de f *es* $x^2+y^2 = \|(x,y)\|^2$.

Para la función g *se tiene que*

$$\frac{\partial g}{\partial x}(x,y) = \frac{2x}{1+x^2} \qquad y \qquad \frac{\partial g}{\partial y}(x,y) = \frac{2y}{1+y^2},$$

de modo que

$$\frac{\partial^2 g}{\partial x^2}(x,y) = \frac{2}{1+x^2} - \frac{4x^2}{(1+x^2)^2} \qquad y \qquad \frac{\partial^2 g}{\partial y \partial x}(x,y) = 0.$$

Análogamente,

$$\frac{\partial^2 g}{\partial x \partial y}(x,y) = 0 \qquad y \qquad \frac{\partial^2 g}{\partial y^2}(x,y) = \frac{2}{1+y^2} - \frac{4y^2}{(1+y^2)^2}.$$

De esta forma, tenemos que

$$\begin{aligned} P_{2,g,\hat{0}}(x,y) &= g(\hat{0}) + \frac{\partial g}{\partial x}(\hat{0})x + \frac{\partial g}{\partial y}(\hat{0})y + \frac{1}{2}\left(\frac{\partial^2 g}{\partial x^2}(\hat{0})x^2 + \frac{\partial^2 g}{\partial y \partial x}(\hat{0})xy + \frac{\partial^2 g}{\partial x \partial y}(\hat{0})yx + \frac{\partial^2 g}{\partial y^2}(\hat{0})y^2\right) \\ &= x^2 + y^2 \end{aligned}$$

y, por lo tanto, que

$$\begin{aligned} \lim_{(x,y)\to(0,0)} f(x,y) &= \lim_{(x,y)\to(0,0)} \frac{\ln(1+x^2) + \ln(1+y^2)}{x^2+y^2} \\ &= \lim_{(x,y)\to(0,0)} \frac{g(x,y)}{x^2+y^2} \\ &= \lim_{(x,y)\to(0,0)} \frac{P_{2,g,\hat{0}}(x,y) + R_{2,g,\hat{0}}(x,y)}{x^2+y^2} \\ &= \lim_{(x,y)\to(0,0)} \frac{x^2+y^2+R_{2,g,\hat{0}}(x,y)}{x^2+y^2} \\ &= \lim_{(x,y)\to(0,0)} \left(1 + \frac{R_{2,g,\hat{0}}(x,y)}{\|(x,y)\|^2}\right) \\ &= 1+0 \\ &= 1. \end{aligned}$$

4.6 Máximos y mínimos

Una de las aplicaciones más comunes de la derivada es la localización de los puntos en los que una función alcanza sus valores máximo y mínimo. De acuerdo con el corolario 2.52 del capítulo 2, si una función $f : A \subset \mathbb{R}^n \to \mathbb{R}$ es continua en un subconjunto B de su dominio A, y este subconjunto es cerrado y acotado (es decir compacto), sólo en este caso podemos estar seguros de que existen un par de puntos en B en los cuales la función f alcanza su valor máximo y su valor mínimo, en donde además estos valores son máximo y mínimo sólo con respecto a los valores de f sobre los elementos del subconjunto B.

Puesto que todo conjunto $B \subset \mathbb{R}^n$ que sea cerrado se puede "descomponer" en la unión de su interior y su frontera ($B = \text{Fr}(B) \cup \text{int}(B)$), que son dos conjuntos ajenos, y dado que $\text{int}(B)$ siempre es un conjunto abierto y $\text{Fr}(B)$ siempre es un conjunto cerrado (y "flaco", ya que por el problema 16 del capítulo 1 se tiene que $\text{int}(\text{Fr}(B)) = \varnothing$), vamos a "reducir" nuestro análisis a estos dos casos: cuando los puntos en los que una función f alcanza su valor máximo o mínimo (o alguno de ellos) pertenecen al interior de B, y cuando estos mismos puntos (o alguno de ellos) pertenecen a la frontera de B.

De acuerdo con el enfoque anterior, lo que haremos será analizar en general qué características específicas satisfacen los puntos en los que una función f alcanza su valor máximo y su valor mínimo, dependiendo de si éstos pertenecen a un conjunto abierto (como el $\text{int}(B)$) o a un conjunto cerrado con interior vacío (como la $\text{Fr}(B)$, cuando B es cerrado).

Definición 4.48 *Sean $f : A \subset \mathbb{R}^n \to \mathbb{R}$, $B \subset A$ y $\hat{x}_0 \in B$. Decimos que:*

1. *f alcanza (o tiene) un valor máximo (mínimo) en $\hat{x}_0$ sobre B si $f(\hat{x}) \leq f(\hat{x}_0)$ ($f(\hat{x}_0) \leq f(\hat{x})$) para todo $\hat{x} \in B$.*

2. *f alcanza (o tiene) un valor máximo local (mínimo local) en $\hat{x}_0$ sobre B si existe $r > 0$ tal que $f(\hat{x}) \leq f(\hat{x}_0)$ ($f(\hat{x}_0) \leq f(\hat{x})$) para todo $\hat{x} \in B \cap B_r(\hat{x}_0)$.*

Con respecto a las definiciones anteriores es importante hacer dos observaciones. Una es que, necesariamente, si en un punto $\hat{x}_0$ una función f alcanza un valor máximo (mínimo) sobre un conjunto B, entonces también alcanza (o tiene) un valor máximo local (mínimo local) sobre B, pero no recíprocamente. Es decir, si f alcanza en $\hat{x}_0$ un valor máximo local (mínimo local) sobre B, este no es necesariamente máximo (mínimo) sobre todo B (ver figura 4.12). La segunda observación es que, si B es un conjunto abierto, entonces f alcanza en $\hat{x}_0$ un valor máximo local (mínimo local) sobre B si y sólo si existe $r > 0$ tal que $f(\hat{x}) \leq f(\hat{x}_0)$ ($f(\hat{x}_0) \leq f(\hat{x})$) para todo $\hat{x} \in B_r(\hat{x}_0) \subset B$.

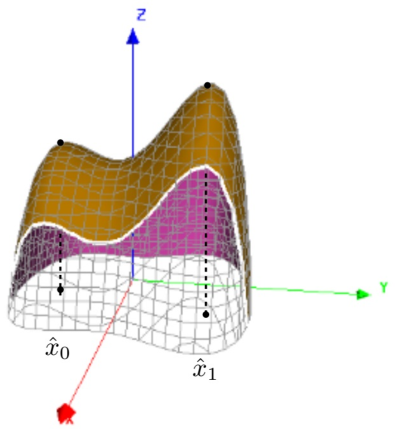


Figura 4.12: En el punto $\hat{x}_0$ la función f tiene un máximo local pero no global, y en el punto $\hat{x}_1$ tiene un máximo local que también es global.

El primer resultado importante que vamos a probar en esta sección tiene que ver con los puntos en los que una función f alcanza un valor máximo (o mínimo) local sobre un conjunto abierto. Si en un punto $\hat{x}_0$, además de satisfacerse la propiedad anterior, f también es derivable, entonces la derivada en $\hat{x}_0$ ($Df(\hat{x}_0)$) tiene que ser la función lineal constante 0 (como seguramente el lector ya está observando, este resultado generaliza lo que sucede para las funciones de $\mathbb{R}$ en $\mathbb{R}$).

Proposición 4.49 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$ (U abierto) y $\hat{x}_0 \in U$. Si f alcanza en $\hat{x}_0$ un valor máximo local (mínimo local) sobre U y f es derivable en $\hat{x}_0$, entonces $Df(\hat{x}_0)$ es la constante cero ($Df(\hat{x}_0) \equiv 0$).*

Demostración. Dado que la derivada de f en $\hat{x}_0$ está representada (en un sistema coordenado ortonormal) por la matriz de $1 \times n$ cuya $i-$ésima entrada está dada por $\frac{\partial f}{\partial x_i}(\hat{x}_0)$, bastará con demostrar que cada uno de éstos números es 0.

Suponiendo que f alcanza en $\hat{x}_0$ un valor máximo local sobre U (el caso del mínimo local se prueba de manera análoga), sea $r > 0$ tal que $f(\hat{x}) \leq f(\hat{x}_0)$ para todo $\hat{x} \in B_r(\hat{x}_0) \subset U$. Por tanto, si $|h| < r$ tenemos que $f(\hat{x}_0 + h\hat{e}_i) \leq f(\hat{x}_0)$, de tal forma que

$$f(\hat{x}_0 + h\hat{e}_i) - f(\hat{x}_0) \leq 0.$$

Ahora, si tomamos h tal que $0 < h < r$, entonces

$$\frac{f(\hat{x}_0 + h\hat{e}_i) - f(\hat{x}_0)}{h} \leq 0$$

y por lo tanto

$$\begin{aligned}\frac{\partial f}{\partial x_i}(\hat{x}_0) &= \lim_{h \to 0} \frac{f(\hat{x}_0 + h\hat{e}_i) - f(\hat{x}_0)}{h} \\ &\leq 0.\end{aligned}$$

Por otra parte, si tomamos h tal que $-r < h < 0$, entonces

$$0 \leq \frac{f(\hat{x}_0 + h\hat{e}_i) - f(\hat{x}_0)}{h},$$

de donde

$$\begin{aligned} 0 &\leq \lim_{h \to 0} \frac{f(\hat{x}_0 + h\hat{e}_i) - f(\hat{x}_0)}{h} \\ &= \frac{\partial f}{\partial x_i}(\hat{x}_0). \end{aligned}$$

Entonces se tiene que

$$0 \leq \frac{\partial f}{\partial x_i}(\hat{x}_0) \leq 0$$

y por lo tanto

$$\frac{\partial f}{\partial x_i}(\hat{x}_0) = 0.$$

■

Como seguramente el lector ya habrá notado, si bien la proposición anterior es importante, ésta sólo nos proporciona una consecuencia (o condición) necesaria del hecho de que una función alcance en $\hat{x}_0$ un valor máximo (o mínimo) local. Como suele suceder con frecuencia, esta condición no es suficiente. En efecto, el hecho de que la derivada de una función f en un punto $\hat{x}_0$ sea la constante 0, no es suficiente para que podamos asegurar que f alcance en $\hat{x}_0$ un valor máximo (o mínimo) local.

Antes de dar un ejemplo que ilustra este hecho, mencionaremos que a pesar de todo, aquellos puntos en que la derivada de una función es la constante 0 resultan ser muy importantes y reciben un nombre especial: *puntos críticos*, lo cual dejamos plasmado en la siguiente

Definición 4.50 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$ y $\hat{x}_0 \in U$. Si f es derivable en $\hat{x}_0$ y la derivada de f en $\hat{x}_0$ ($Df(\hat{x}_0)$) es la constante 0, decimos que $\hat{x}_0$ es un punto crítico de f.*

Con base en la definición anterior, lo que mostraremos en el siguiente ejemplo es que los puntos críticos de una función f no tienen por qué ser necesariamente puntos en los que f alcance un valor máximo (o mínimo) local.

Ejemplo 4.51 *Sea $f : \mathbb{R}^2 \to \mathbb{R}$ dada por la expresión*

$$f(x, y) = x^2 - y^2.$$

Dado que

$$\frac{\partial f}{\partial x}(x, y) = 2x \qquad y \qquad \frac{\partial f}{\partial y}(x, y) = -2y,$$

se tiene que el único punto crítico de f es el $(0, 0)$.

Sin embargo, para cualquier $r > 0$ se tiene que: si $0 < |x| < r$, entonces $(x, 0) \in B_r((0, 0))$ y

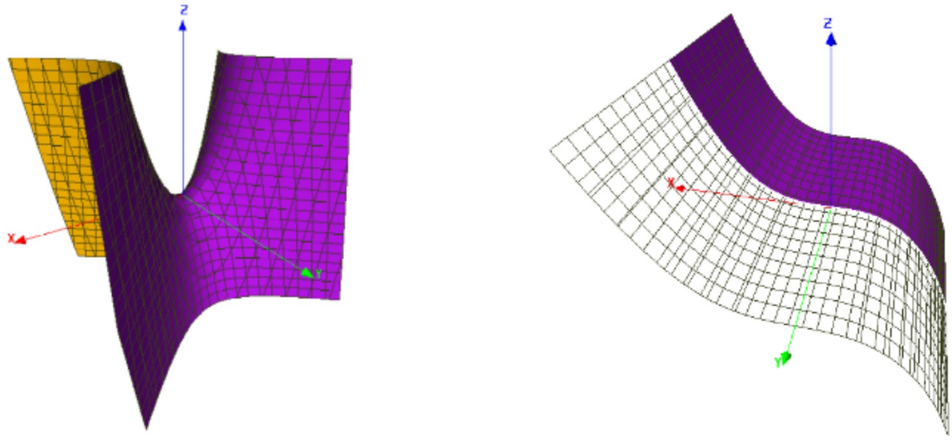
$$\begin{aligned} f(x, 0) &= x^2 \\ &> 0 \\ &= f(0, 0). \end{aligned}$$

Por otra parte, si $0 < |y| < r$, entonces $(0, y) \in B_r((0, 0))$ y

$$\begin{aligned} f(0, y) &= -y^2 \\ &< 0 \\ &= f(0, 0), \end{aligned}$$

lo que demuestra que f no tiene un máximo local ni un mínimo local en el $(0, 0)$ (ver figura 4.13 (a)).

Aquellos puntos críticos en los que una función no tiene un máximo local o un mínimo local, justo por el aspecto geométrico del ejemplo anterior en una vecindad del $(0,0)$ (el de una silla de montar), reciben el nombre de *punto silla*, aunque no todos los puntos silla tienen este aspecto. Por ejemplo para la función $f(x,y) = x^3$, todos sus puntos críticos (los puntos de la forma $(0,y)$) son puntos silla, en el sentido de que en ellos la función f no tiene un máximo local o un mínimo local. Sin embargo, geométricamente la función f no tiene el aspecto de una silla de montar (ver figura 4.13 (b)).



(a) (b)

Figura 4.13: La función $f(x,y) = x^2 - y^2$ (figura (a)) tiene un punto silla en el punto $(0,0)$, y la función $f(x,y) = x^3$ (figura (b)) tiene puntos silla en los puntos de la forma $(0,y)$.

El ejemplo anterior, y el hecho de que la proposición 4.49 es válida para máximos o mínimos locales nos plantea el problema de contar con un criterio que nos permita "clasificar" a los puntos críticos de una función.

Lo que haremos a continuación será desarrollar un criterio análogo al criterio de la segunda derivada para la clasificación de puntos críticos de funciones de $\mathbb{R}$ en $\mathbb{R}$, pero para ello será necesario introducir el concepto (que tomaremos "prestado" nuevamente del ¡Álgebra Lineal!) de: *forma cuadrática*. Más adelante haremos una breve revisión de este concepto.

De acuerdo con el problema 37 (que es una consecuencia del teorema de Taylor), si $f : U \subset \mathbb{R}^n \to \mathbb{R}$ es una función de clase C^2 en U, $\hat{x}_0 \in U$ y $r > 0$ son tales que $B_r(\hat{x}_0) \subset U$, sabemos que si $\hat{h} = (h_1, \ldots, h_n) \in \mathbb{R}^n$ es tal que $\left\|\hat{h}\right\| < r$, entonces

$$\begin{aligned} f(\hat{x}_0 + \hat{h}) &= P_{2,f,\hat{x}_0}(\hat{x}_0 + \hat{h}) + R_{2,f,\hat{x}_0}(\hat{x}_0 + \hat{h}) \\ &= f(\hat{x}_0) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\hat{x}_0) h_i + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}(\hat{x}_0) h_i h_j + R_{2,f,\hat{x}_0}(\hat{x}_0 + \hat{h}), \end{aligned}$$

en donde

$$\lim_{\hat{h} \to \hat{0}} \frac{R_{2,f,\hat{x}_0}(\hat{x}_0 + \hat{h})}{\left\|\hat{h}\right\|^2} = 0.$$

De este modo, si $\hat{x}_0$ es un punto crítico de f, se tiene que

$$f(\hat{x}_0 + \hat{h}) - f(\hat{x}_0) = \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}(\hat{x}_0) h_i h_j + R_{2,f,\hat{x}_0}(\hat{x}_0 + \hat{h}).$$

De esta última identidad, nos detendremos a estudiar con detalle a la expresión

$$\sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}(\hat{x}_0) h_i h_j \tag{4.43}$$

y lo primero que haremos notar es que se puede escribir en forma matricial, de la siguiente manera

$$\sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}(\hat{x}_0) h_i h_j = \left[\begin{array}{ccc} h_1 & \cdots & h_n \end{array}\right] A \left[\begin{array}{c} h_1 \\ \vdots \\ h_n \end{array}\right],$$

en donde la matriz A es

$$A = \left[\begin{array}{ccc} \frac{\partial^2 f}{\partial x_1^2}(\hat{x}_0) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1}(\hat{x}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(\hat{x}_0) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\hat{x}_0) \end{array}\right]. \tag{4.44}$$

La ventaja de escribir a la expresión 4.43 en forma matricial es que nos permite relacionarla con un concepto que ya mencionamos: las formas cuadráticas.

Las formas cuadráticas en $\mathbb{R}^n$ son (además de la función constante 0, que es la única forma cuadrática que es constante), las funciones polinomiales de las coordenadas $x_1, \ldots, x_n$ que tienen la particularidad de que todos los monomios que la forman son expresiones de grado 2. Por ejemplo, en el caso de $\mathbb{R}^2$, cualquier forma cuadrática en las variables x, y se escribe como

$$ax^2 + 2bxy + cy^2 \tag{4.45}$$

la cual, como en el caso de la expresión 4.43, se puede escribir en forma matricial de la siguiente manera:

$$\begin{aligned} ax^2 + 2bxy + cy^2 &= \left[\begin{array}{cc} x & y \end{array}\right] \left[\begin{array}{cc} a & b \\ b & c \end{array}\right] \left[\begin{array}{cc} x & y \end{array}\right]^t \\ &= \left[\begin{array}{cc} x & y \end{array}\right] \left[\begin{array}{cc} a & b \\ b & c \end{array}\right] \left[\begin{array}{c} x \\ y \end{array}\right]. \end{aligned}$$

En general, toda forma cuadrática en $\mathbb{R}^n$ se identifica con una matriz A de $n \times n$ con entradas reales ($A \in M_{2\times 2}(\mathbb{R})$) la cual es simétrica, es decir, que es igual a su transpuesta ($A = A^t$). Y recíprocamente, si A es una de estas matrices, la función $F : \mathbb{R}^n \to \mathbb{R}$ definida como

$$\begin{aligned} F(x_1, \ldots, x_n) &= \left[\begin{array}{ccc} x_1 & \cdots & x_n \end{array}\right] A \left[\begin{array}{ccc} x_1 & \cdots & x_n \end{array}\right]^t \\ &= \left[\begin{array}{ccc} x_1 & \cdots & x_n \end{array}\right] A \left[\begin{array}{c} x_1 \\ \vdots \\ x_n \end{array}\right] \end{aligned}$$

claramente define una forma cuadrática en $\mathbb{R}^n$.

A la matriz A dada por 4.44 se le conoce como la *matriz hessiana*[2] de f en $\hat{x}_0$, que de aquí en adelante denotaremos por $Hf(\hat{x}_0)$, y a la forma cuadrática asociada a ésta le llamaremos *la hessiana de* f *en* $\hat{x}_0$. La hessiana de f en $\hat{x}_0$ la denotaremos por $Hf_{\hat{x}_0}$ de tal forma que si $\hat{x} = (x_1, \ldots, x_n)$, entonces

$$\begin{aligned} Hf_{\hat{x}_0}(\hat{x}) &= Hf_{\hat{x}_0}(x_1, \ldots, x_n) \\ &= \left[\begin{array}{ccc} x_1 & \cdots & x_n \end{array}\right] Hf(\hat{x}_0) \left[\begin{array}{ccc} x_1 & \cdots & x_n \end{array}\right]^t \\ &= \left[\begin{array}{ccc} x_1 & \cdots & x_n \end{array}\right] Hf(\hat{x}_0) \left[\begin{array}{c} x_1 \\ \vdots \\ x_n \end{array}\right]. \end{aligned}$$

Hay una clasificación de las formas cuadráticas que en particular resultará muy importante para el problema de la identificación de los puntos críticos de una función. Se dice que la forma cuadrática $F : \mathbb{R}^n \to \mathbb{R}$ es *semipositiva* (análogamente *seminegativa*) si $F(\hat{x}) \geq 0$ ($F(\hat{x}) \leq 0$) para toda $\hat{x} \in \mathbb{R}^n$ y si F satisface la condición de que $F(\hat{x}) = 0$ si y sólo si $\hat{x} = \hat{0}$, entonces decimos que F es *no degenerada*.

[2]Llamada así en honor de Ludwig Otto Hesse (22 abril 1811 - 4 agosto 1874) quien fue un matemático alemán. Hesse nació en K'onigsberg, Prussia, y murió en Munich, Bavaria. Trabajó en la teoría de invariantes.

Lo interesante de esta clasificación de las formas cuadráticas es que, con base en ella, podremos establecer criterios que nos permitan determinar si los puntos críticos de una función son puntos en los que ésta alcanza un valor máximo o mínimo, locales en ambos casos.

En efecto, lo siguiente que haremos será probar un resultado en el que, dependiendo de qué tipo de forma cuadrática resulte ser la hessiana de f en un punto $\hat{x}_0$, podremos asegurar que f tiene un cierto tipo de valor extremo (máximo o mínimo local). Para ello, será necesario probar antes un sencillo resultado que formularemos en el siguiente

Lema 4.52 *Si $F : \mathbb{R}^n \to \mathbb{R}$ es una forma cuadrática semipositiva (seminegativa) y no degenerada asociada a una matriz A (es decir que $F(\hat{x}) = \hat{x}A\hat{x}^t$), entonces existe $M > 0$ ($m < 0$) tal que*

$$F(\hat{h}) \geq M \left\|\hat{h}\right\|^2 \qquad \left(F(\hat{h}) \leq m \left\|\hat{h}\right\|^2\right)$$

para toda $\hat{h} \in \mathbb{R}^n$.

Demostración. Dado que $S^{n-1} = \{\hat{x} \in \mathbb{R}^n \mid \|\hat{x}\| = 1\}$ es un conjunto cerrado y acotado (compacto) y toda forma cuadrática es una función continua, por el corolario 2.52 del capítulo 2 sabemos que existe un valor mínino para F (si F es semipositiva), que llamaremos $M > 0$ (un valor máximo para F, si F es seminegativa, que llamaremos $m < 0$), tal que

$$F(\hat{x}) \geq M \qquad (F(\hat{x}) \leq m)$$

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

Por tanto, dado que las desigualdades que deseamos probar se satisfacen si $\hat{h} = \hat{0}$, si tomamos $\hat{h} \neq \hat{0}$ y hacemos $\hat{x} = \hat{h}/\left\|\hat{h}\right\|$, entonces se tiene que $\hat{x} \in S^{n-1}$ y por lo tanto

$$\begin{aligned} M &\leq F(\hat{x}) \\ &= \hat{x}A\hat{x}^t \\ &= \left(\frac{\hat{h}}{\left\|\hat{h}\right\|}\right) A \left(\frac{\hat{h}}{\left\|\hat{h}\right\|}\right)^t \\ &= \frac{1}{\left\|\hat{h}\right\|^2}\left(\hat{h}A\hat{h}^t\right) \\ &= \frac{1}{\left\|\hat{h}\right\|^2} F(\hat{h}). \end{aligned}$$

Es decir, se tiene que

$$F(\hat{h}) \geq M \left\|\hat{h}\right\|^2$$

para toda $\hat{h} \in \mathbb{R}^n$ (análogamente se prueba que $F(\hat{h}) \leq m \left\|\hat{h}\right\|^2$ para toda $\hat{h} \in \mathbb{R}^n$). ■

Una vez hecho todo lo anterior, estamos en condiciones de formular un resultado que nos permitirá determinar si un punto crítico $\hat{x}_0$ de una función f es un punto en donde ésta alcanza un máximo local o un mínimo local. Sólo probaremos uno de estos casos y como es de imaginar, el otro caso quedará como un problema para el lector.

Proposición 4.53 *Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}$ de clase C^2 en U, y $\hat{x}_0 \in U$ un punto crítico de f. Se satisface lo siguiente:*

1. *si la hessiana de f en $\hat{x}_0$ es una forma cuadrática semipositiva y no degenerada, entonces f tiene un mínimo local en $\hat{x}_0$, y*

2. *si la hessiana de f en $\hat{x}_0$ es una forma cuadrática seminegativa y no degenerada, entonces f tiene un máximo local en $\hat{x}_0$.*

Demostración. Sólo probaremos el inciso 1. Sea $r_0 > 0$ tal que $B_{r_0}(\hat{x}_0) \subset U$. Como ya habíamos mencionado anteriormente, por el problema 37 sabemos que, como $f : U \subset \mathbb{R}^n \to \mathbb{R}$ es una función de clase C^2 en U y $\hat{x}_0 \in U$ es un punto crítico de f, si $\hat{h} = (h_1, \dots, h_n) \in \mathbb{R}^2$ es tal que $\left\|\hat{h}\right\| < r_0$, entonces

$$\begin{aligned} f(\hat{x}_0 + \hat{h}) &= P_{2,f,\hat{x}_0}(\hat{x}_0 + \hat{h}) + R_{2,f,\hat{x}_0}(\hat{x}_0 + \hat{h}) \\ &= f(\hat{x}_0) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\hat{x}_0) h_i + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}(\hat{x}_0) h_i h_j + R_{2,f,\hat{x}_0}(\hat{x}_0 + \hat{h}) \\ &= f(\hat{x}_0) + \frac{1}{2} H f_{\hat{x}_0}(\hat{h}) + R_{2,f,\hat{x}_0}(\hat{x}_0 + \hat{h}), \end{aligned}$$

en donde

$$\lim_{\hat{h} \to \hat{0}} \frac{R_{2,f,\hat{x}_0}(\hat{x}_0 + \hat{h})}{\left\|\hat{h}\right\|^2} = 0. \tag{4.46}$$

Ahora, dado que $H f_{\hat{x}_0}$ es semipositiva, por el lema 4.52 sabemos que existe $M > 0$ tal que

$$H f_{\hat{x}_0}(\hat{h}) \geq M \left\|\hat{h}\right\|^2$$

para toda $\hat{h} \in \mathbb{R}^n$. Por tanto, por 4.46 sabemos que existe $\delta > 0$ tal que si $\left\|\hat{h}\right\| < \delta$, entonces

$$\frac{\left|R_{2,f,\hat{x}_0}(\hat{x}_0 + \hat{h})\right|}{\left\|\hat{h}\right\|^2} < \frac{M}{2}$$

y en particular, que

$$0 < \frac{M}{2} + \frac{R_{2,f,\hat{x}_0}(\hat{x}_0 + \hat{h})}{\left\|\hat{h}\right\|^2}.$$

Por lo tanto, si $r = \min\{\delta, r_0\}$ y $0 < \left\|\hat{h}\right\| < r$, tendremos que

$$\begin{aligned} \frac{f(\hat{x}_0 + \hat{h}) - f(\hat{x}_0)}{\left\|\hat{h}\right\|^2} &= \frac{\frac{1}{2} H f_{\hat{x}_0}(\hat{h}) + R_{2,f,\hat{x}_0}(\hat{x}_0 + \hat{h})}{\left\|\hat{h}\right\|^2} \\ &= \frac{1}{2} \frac{H f_{\hat{x}_0}(\hat{h})}{\left\|\hat{h}\right\|^2} + \frac{R_{2,f,\hat{x}_0}(\hat{x}_0 + \hat{h})}{\left\|\hat{h}\right\|^2} \\ &\geq \frac{M}{2} + \frac{R_{2,f,\hat{x}_0}(\hat{x}_0 + \hat{h})}{\left\|\hat{h}\right\|^2} \\ &> 0, \end{aligned}$$

de donde concluimos que

$$f(\hat{x}_0) \leq f(\hat{x}_0 + \hat{h})$$

si $\left\|\hat{h}\right\| < r$, lo que demuestra que f tiene un mínimo local en $\hat{x}_0$. ■

Como el lector habrá notado, el resultado anterior es una especie de "equivalente" (en $\mathbb{R}^n$) al "criterio de la segunda derivada" para la clasificación de puntos críticos de funciones de $\mathbb{R}$ en $\mathbb{R}$, y como sucede en ese caso, su "recíproco" no se cumple, lo que ilustramos en el siguiente

Ejemplo 4.54 *Sea $f : \mathbb{R}^2 \to \mathbb{R}$ definida como*

$$f(x, y) = x^4 + y^4$$

para cada $(x, y) \in \mathbb{R}^2$. Como es fácil de verificar (ver figura 4.14), el $(0, 0)$ es el punto en el que f alcanza su mínimo global (y por tanto también local) y dado que

$$\frac{\partial^2 f}{\partial x^2}(0,0) = \frac{\partial^2 f}{\partial y^2}(0,0) = \frac{\partial^2 f}{\partial y \partial x}(0,0) = 0$$

se tiene que la hessiana de f en el $(0, 0)$ es la función constante 0, la cual es la más degenerada (¡en el sentido matemático de la palabra!) de todas las formas cuadráticas.

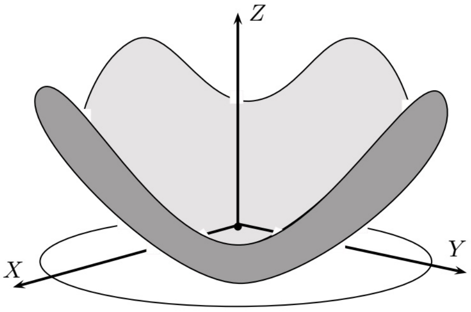


Figura 4.14: La función $f(x, y) = x^4 + y^4$ tiene un mínimo global (y por tanto también local) en el punto $(0, 0)$, y su forma cuadrática asociada en ese punto (su hessiana) es degenerada.

A pesar del ejemplo anterior, para la proposición 4.53 se puede establecer una proposición "casi-recíproca". Como seguramente el lector recordará, para el caso de funciones de $\mathbb{R}$ en $\mathbb{R}$ se sabe que si f tiene un máximo (mínimo) local en un punto x_0, y $f''(x_0)$ existe, entonces se debe cumplir que $f''(x_0) \leq 0$ ($f''(x_0) \geq 0$) (resultado que por cierto nos será de mucha utilidad en la prueba de nuestra siguiente proposición).

Para una función f de $\mathbb{R}^n$ en $\mathbb{R}$ se "extiende" el resultado anterior en el sentido de que, si f tiene un máximo (o un mínimo) local en un punto $\hat{x}_0$, entonces la hessiana de f en $\hat{x}_0$ será seminegativa (o semipositiva). Es decir, lo único que no podremos asegurar es que la hessiana es no degenerada (así como en el caso real no podemos asugurar que $f''(x_0)$ no sea 0).

Proposición 4.55 *Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}$ de clase C^2 en U, y $\hat{x}_0 \in U$ un punto crítico de f. Se satisface lo siguiente:*

1. *si f tiene un mínimo local en $\hat{x}_0$, entonces la hessiana de f en $\hat{x}_0$ es una forma cuadrática semipositiva, y*

2. *si f tiene un máximo local en $\hat{x}_0$, entonces la hessiana de f en $\hat{x}_0$ es una forma cuadrática seminegativa.*

Demostración. Sólo probaremos el inciso 2. Sean, $r_0 > 0$ tal que $B_{r_0}(\hat{x}_0) \subset U$, y $\hat{u} = (u_1, \ldots, u_n) \in \mathbb{R}^n$ tal que $\|\hat{u}\| = 1$. Definimos $\gamma : (-r_0, r_0) \subset \mathbb{R} \to \mathbb{R}^n$ como $\gamma(t) = \hat{x}_0 + t\hat{u}$ y $g : (-r_0, r_0) \subset \mathbb{R} \to \mathbb{R}$ como $g = f \circ \gamma$.

Nótese que, por la proposición 4.30 sabemos que g es derivable en su dominio y que

$$\begin{aligned} g'(t) &= \nabla f(\gamma(t)) \cdot \gamma'(t) \\ &= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\gamma(t)) u_i. \end{aligned}$$

Dado que f es de clase C^2 en U, por la misma proposición 4.30 (aplicada a cada miembro de la suma anterior), se tiene

$$\begin{aligned} g''(t) &= \sum_{i=1}^{n}\left(\sum_{j=1}^{n}\frac{\partial^2 f}{\partial x_j \partial x_i}(\gamma(t))u_j\right)u_i \\ &= \sum_{i,j=1}^{n}\frac{\partial^2 f}{\partial x_j \partial x_i}(\gamma(t))u_i u_j, \end{aligned}$$

de tal forma que

$$\begin{aligned} g''(0) &= \sum_{i,j=1}^{n}\frac{\partial^2 f}{\partial x_j \partial x_i}(\gamma(0))u_i u_j \\ &= \sum_{i,j=1}^{n}\frac{\partial^2 f}{\partial x_j \partial x_i}(\hat{x}_0)u_i u_j \\ &= Hf_{\hat{x}_0}(\hat{u}). \end{aligned}$$

Por otra parte, es sencillo demostrar (problema 44) que, si f tiene un máximo local en $\hat{x}_0$, entonces g tiene un máximo local en $t = 0$, de tal forma que, por el resultado para funciones de $\mathbb{R}$ en $\mathbb{R}$ que mencionamos previamente a esta proposición, se debe tener que

$$\begin{aligned} Hf_{\hat{x}_0}(\hat{u}) &= g''(0) \\ &\leq 0. \end{aligned}$$

Por lo tanto, si $\hat{x} \in \mathbb{R}^n$, $\hat{x} \neq \hat{0}$, y hacemos $\hat{u} = \hat{x}/\left\|\hat{x}\right\|$, se tendrá que

$$\begin{aligned} Hf_{\hat{x}_0}(\hat{u}) &= Hf_{\hat{x}_0}\left(\frac{\hat{x}}{\left\|\hat{x}\right\|}\right) \\ &= \frac{1}{\left\|\hat{x}\right\|^2}Hf_{\hat{x}_0}(\hat{x}) \\ &\leq 0, \end{aligned}$$

de donde $Hf_{\hat{x}_0}(\hat{x}) \leq 0$ para toda $\hat{x} \in \mathbb{R}^n$, lo que prueba que la hessiana de f en $\hat{x}_0$ es seminegativa. ■

Aun cuando todo lo desarrollado hasta aquí parece funcionar muy bien, no podemos dejar de mencionar que el problema de determinar si una función f tiene un máximo o un mínimo local (o ninguno de estos dos) en un punto crítico $\hat{x}_0$, se "trasladadó" al problema de determinar si la hessiana de f en este punto $\hat{x}_0$ es una función cuadrática semipositiva o seminegativa (o ninguna de estas dos).

En la siguiente sección mencionaremos (sin probar) algunas condiciones necesarias y suficientes para que una forma cuadrática F en $\mathbb{R}^n$ (en términos de su matriz asociada) sea semipositiva o seminegativa, y no degenerada. Como ya mencionamos, estas condiciones necesarias y suficientes son un tema importante de Álgebra Lineal que desafortunadamente no podemos incluir en este texto (pero que el lector interesado puede consultar en [2]) para saber más de él).

Sin embargo, para que el amable lector no se sienta muy desalentado por esta ausencia, mostraremos y probaremos cuáles son estas condiciones para el caso particular de las formas cuadráticas en $\mathbb{R}^2$, las que dejaremos plasmadas en la siguiente

Proposición 4.56 *Sea $F : \mathbb{R}^2 \to \mathbb{R}$ la forma cuadrática asociada a la matriz*

$$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

y que está dada por la expresión

$$F(x, y) = \begin{bmatrix} x & y \end{bmatrix} A \begin{bmatrix} x & y \end{bmatrix}^t$$

$$\begin{aligned} &= \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\ &= ax^2 + 2bxy + cy^2. \end{aligned}$$

F es semipositiva (seminegativa) y no degenerada si y sólo si $a > 0$ ($a < 0$) y $ac - b^2 = \det(A) > 0$.

Demostración. Supongamos que F es semipositiva (el caso en que sea seminegativa se prueba de manera análoga) y no degenerada. Por esta última propiedad, dado que

$$a = F(1, 0) > 0,$$

concluimos que $a > 0$. Por tanto, podemos reescribir a F como

$$\begin{aligned} F(x, y) &= ax^2 + 2bxy + cy^2 \\ &= a\left(x + \frac{b}{a}y\right)^2 + \left(c - \frac{b^2}{a}\right)y^2 \\ &= a\left(x + \frac{b}{a}y\right)^2 + \left(\frac{ac - b^2}{a}\right)y^2, \end{aligned} \tag{4.47}$$

de tal forma que

$$\begin{aligned} 0 &< F(b/a, -1) \\ &= \frac{ac - b^2}{a} \end{aligned}$$

y por lo tanto $ac - b^2 > 0$.

Supongamos ahora que $a > 0$ y $ac - b^2 > 0$. De la identidad 4.47 concluimos que $F(x, y) \geq 0$ para toda $(x, y) \in \mathbb{R}^2$, es decir, F es semipositiva.

Finalmente, si (x, y) es tal que $F(x, y) = 0$, se debe tener que

$$\left(\frac{ac - b^2}{a}\right)y^2 = 0$$

y

$$\left(x + \frac{b}{a}y\right)^2 = 0,$$

de tal forma que de la primera identidad concluimos que $y = 0$, y sustituyendo esto último en la segunda identidad, concluimos que $x = 0$. Esto prueba que F es no degenerada. ■

Como una consecuencia inmediata de la proposición anterior y de la proposición 4.53, podemos formular el siguiente resultado, que establece criterios muy concretos para determinar si un punto crítico de una función de $\mathbb{R}^2$ en $\mathbb{R}$ es un máximo o mínimo local.

Proposición 4.57 *Sea $f : U \subset \mathbb{R}^2 \to \mathbb{R}$ de clase C^2 en U, y $\hat{x}_0 \in U$ un punto crítico de f.*

1. *Si*
$$\frac{\partial^2 f}{\partial x^2}(\hat{x}_0) > 0 \qquad y \qquad \frac{\partial^2 f}{\partial x^2}(\hat{x}_0)\frac{\partial^2 f}{\partial y^2}(\hat{x}_0) - \left(\frac{\partial^2 f}{\partial y \partial x}(\hat{x}_0)\right)^2 > 0,$$
entonces f tiene un mínimo local en $\hat{x}_0$, y

2. *si*
$$\frac{\partial^2 f}{\partial x^2}(\hat{x}_0) < 0 \qquad y \qquad \frac{\partial^2 f}{\partial x^2}(\hat{x}_0)\frac{\partial^2 f}{\partial y^2}(\hat{x}_0) - \left(\frac{\partial^2 f}{\partial y \partial x}(\hat{x}_0)\right)^2 > 0,$$
entonces f tiene un máximo local en $\hat{x}_0$.

Antes de hacer unos breves comentarios adicionales acerca de las formas cuadráticas, daremos algunos ejemplos que ilustren el trabajo realizado hasta ahora sobre el tema de máximos y mínimos.

Ejemplo 4.58

1. *Sea* $f : \mathbb{R}^2 \to \mathbb{R}$ *dada por*

$$f(x, y) = (x + 1)^2 + y^2$$

y $B = \left\{(x, y) \in \mathbb{R}^2 \mid \|(x, y)\| \leq 2\right\}$.

De acuerdo con lo realizado hasta este momento, si los puntos en los que f *alcanza un valor máximo o mínimo local (o global) sobre* B*, se encuentran en el interior de* B*, entonces deben ser puntos críticos de* f*, de modo que lo primero que se tendrá que hacer será localizar este tipo de puntos.*

Para lograr esto, hay que resolver las ecuaciones

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

que en este caso son

$$2(x + 1) = 0 \qquad y \qquad 2y = 0,$$

de donde obtenemos que $(-1, 0)$ *es el único punto crítico de* f *(en todo* $\mathbb{R}^2$*).*

Por otra parte, dado que $f(x, y) \geq 0$ *para toda* $(x, y) \in \mathbb{R}^2$*, que* $f(-1, 0) = 0$ *y* $(-1, 0) \in B$*, sin necesidad de mayores cálculos, podemos concluir que* f *alcanza su valor mínimo sobre* B *(de hecho, sobre todo* $\mathbb{R}^2$*) en el punto* $(-1, 0)$ *y que dicho valor mínimo es* 0.

En cuanto al valor máximo de f *sobre* B*, dado que* f *no tiene más puntos críticos, este valor máximo se debe de alcanzar en algún punto de la frontera de* B*, es decir, sobre el conjunto*

$$\mathrm{Fr}(B) = \left\{(x, y) \in \mathbb{R}^2 \mid \|(x, y)\| = 2\right\}.$$

Dado que para este tipo de conjuntos (cerrados con interior vacío) no hemos desarrollado ninguna herramienta que nos permita resolver este problema, para este caso particular recurriremos al siguiente hecho: la $\mathrm{Fr}(B)$ *se puede obtener como la imagen de una función de* $\mathbb{R}$ *en* $\mathbb{R}^2$*. En efecto, nótese que la función* $\gamma : [0, 2\pi] \subset \mathbb{R} \to \mathbb{R}^2$ *dada por*

$$\gamma(t) = (2\cos(t), 2\,\mathrm{sen}(t))$$

es una parametrización del conjunto $\mathrm{Fr}(B)$*, es decir,* $\gamma([0, 2\pi]) = \mathrm{Fr}(B)$.

Aprovechando este hecho, si definimos la función $g : [0, 2\pi] \subset \mathbb{R} \to \mathbb{R}$ *como*

$$\begin{aligned} g(t) &= (f \circ \gamma)(t) \\ &= f(\gamma(t)) \\ &= (2\cos(t) + 1)^2 + 4\,\mathrm{sen}^2(t) \\ &= 4\cos(t) + 5 \end{aligned}$$

y encontramos los puntos del dominio de g *en los que esta función alcanza sus valores máximo y mínimo, podremos entonces localizar los valores máximo y mínimo de* f *sobre la* $\mathrm{Fr}(B)$.

Como ya se sabe, hay que proceder de manera análoga a lo que estamos haciendo con f*; es decir, hay que localizar los puntos críticos de* g *en el intervalo abierto* $(0, 2\pi)$*, determinar si en estos puntos críticos* g *alcanza un máximo o mínimo local, y comparar el valor de* g *en estos puntos con los valores de* g *en los extremos del intervalo* $[0, 2\pi]$ *(que son la frontera de* $[0, 2\pi]$*).*

De esta forma, lo primero que hay que localizar son los valores de $t \in (0, 2\pi)$ *para los cuales se cumple que* $g'(t) = -4\,\mathrm{sen}(t) = 0$ *y que claramente sólo sucede para* $t = \pi$*. Dado que* $g(t) \geq 1$ *para toda* $t \in [0, 2\pi]$ *y que* $g(\pi) = 1$*, concluimos que* g *alcanza su valor mínimo en* $t = \pi$ *y que este valor mínimo es* 1.

Finalmente, dado que $g(0) = g(2\pi) = 9$*, concluimos que el valor máximo de* g *sobre el intervalo* $[0, 2\pi]$ *es* 9*, y que este valor lo alcanza en* $t = 0$ *y* $t = 2\pi$ *(es decir, en la frontera de su dominio). De todo lo anterior concluimos las siguiente dos cosas: una, que el valor mínimo de* f *sobre* $\mathrm{Fr}(B)$ *es* 1 *y que este valor lo alcanza en el punto* $\gamma(\pi) = (-2, 0)$*; dos, que su valor máximo sobre la* $\mathrm{Fr}(B)$ *lo alcanza en el punto* $\gamma(0) = \gamma(2\pi) = (2, 0)$ *y que este valor máximo es* 9*.*

Por lo tanto, comparando los valores de f *en los puntos críticos que están en el interior de* B *(*$f(-1, 0) = 0$*), junto con los valores máximo y mínimo de* f *sobre la* $\mathrm{Fr}(B)$ *(*$f(-2, 0) = 1$ *y* $f(2, 0) = 9$*), concluimos que los valores máximo y mínimo de* f *sobre* B *son* 0 *y* 9*, y que estos valores se alcanzan en los puntos* $(-1, 0)$ *y* $(2, 0)$*, respectivamente (ver figura 4.15).*

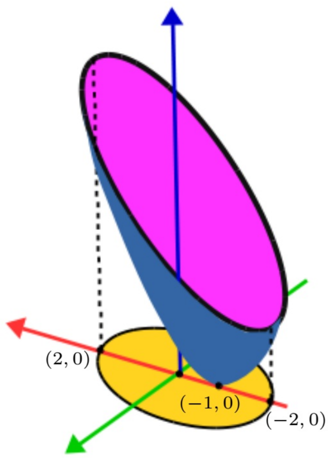


Figura 4.15: Los valores máximo y mínimo de la función $f(x, y) = (x + 1)^2 + y^2$ sobre el conjunto $B = \{(x, y) \in \mathbb{R}^2 \mid \|(x, y)\| \le 2\}$ se alcanzan en los puntos $(2, 0)$ y $(-1, 0)$, respectivamente.

2. *Sea* $f : \mathbb{R}^2 \to \mathbb{R}$ *dada por*
$$f(x, y) = x(x - 1)^2 + y^2$$
y $B = \{(x, y) \in \mathbb{R}^2 \mid \|(x, y)\| \le 1\}$*.*

Nuevamente, para determinar los puntos en los que f *alcanza un valor máximo o mínimo local (o global) sobre* B*, procederemos como en el inciso anterior. Primero encontraremos los puntos críticos de* f*, para lo cual hay que resolver las ecuaciones*
$$\begin{aligned}\frac{\partial f}{\partial x}(x, y) &= (x - 1)^2 + 2x(x - 1)\\ &= (x - 1)(3x - 1)\\ &= 0\end{aligned}$$
y
$$\begin{aligned}\frac{\partial f}{\partial y}(x, y) &= 2y\\ &= 0\end{aligned}$$
y cuyas soluciones son las parejas $(1, 0)$ *y* $(1/3, 0)$*, en donde este último punto es el único que queda dentro del interior de* B*.*

Ahora será necesario calcular las segundas derivadas parciales de f*, las cuales resultan ser:*
$$\frac{\partial^2 f}{\partial x^2}(x, y) = 6x - 4$$

$$\begin{aligned}
\frac{\partial^2 f}{\partial y \partial x}(x,y) &= 0 \\
\frac{\partial^2 f}{\partial y^2}(x,y) &= 2,
\end{aligned}$$

de modo que para el punto crítico $(1/3,0)$ *se tiene que*

$$\begin{aligned}
\frac{\partial^2 f}{\partial x^2}(1/3,0) &= -2 \\
\frac{\partial^2 f}{\partial y \partial x}(1/3,0) &= 0 \\
\frac{\partial^2 f}{\partial y^2}(1/3,0) &= 2.
\end{aligned}$$

Es decir,

$$\begin{aligned}
\frac{\partial^2 f}{\partial x^2}(1/3,0)\frac{\partial^2 f}{\partial y^2}(1/3,0) - \left(\frac{\partial^2 f}{\partial y \partial x}(1/3,0)\right)^2 &= -4 \\
&< 0
\end{aligned}$$

y por lo tanto, por el problema 51 (o el problema 50), concluimos que f *tiene un punto silla en el punto* $(1/3,0)$.

Como en el inciso anterior, nos resta localizar los puntos en los que f *alcanza sus valores máximo y mínimo sobre la frontera del conjunto* B*, que ahora es el conjunto*

$$\mathrm{Fr}(B) = \left\{(x,y) \in \mathbb{R}^2 \mid \|(x,y)\| = 1\right\}$$

y que, como en el inciso anterior, podemos parametrizar por la función $\gamma : [0,2\pi] \subset \mathbb{R} \to \mathbb{R}^2$ *dada por*

$$\gamma(t) = (\cos(t), \mathrm{sen}(t)).$$

Como en el caso anterior, definimos $g : [0,2\pi] \subset \mathbb{R} \to \mathbb{R}$ *como*

$$\begin{aligned}
g(t) &= (f \circ \gamma)(t) \\
&= f(\gamma(t)) \\
&= \cos(t)(\cos(t)-1)^2 + \mathrm{sen}^2(t)
\end{aligned}$$

de tal forma que

$$\begin{aligned}
g'(t) &= \nabla f(\gamma(t)) \cdot \gamma'(t) \\
&= (\cos(t)-1)(3\cos(t)-1)(-\mathrm{sen}(t)) + 2\,\mathrm{sen}(t)\cos(t) \\
&= \mathrm{sen}(t)\left(2\cos(t) - (\cos(t)-1)(3\cos(t)-1)\right) \\
&= \mathrm{sen}(t)\left(-3\cos^2(t) + 6\cos(t) - 1\right).
\end{aligned}$$

Dado que la raíces de la ecuación cuadrática $-3x^2+6x-1=0$ *son* $1-\sqrt{6}/3$ *y* $1+\sqrt{6}/3$*, concluimos que los puntos críticos de* g *en el intervalo abierto* $(0,2\pi)$ *son* π, t_0*, y* $2\pi - t_0$*, en donde* $t_0 \in (0,\pi)$ *es tal que* $\cos(t_0) = 1-\sqrt{6}/3$*. Por lo tanto, como* $g(0)=0$*,* $g(t_0) = 4\sqrt{6}/9 = g(2\pi - t_0)$*,* $g(\pi) = -4$ *y* $g(2\pi)=0$*, concluimos que* f *alcanza su valor máximo sobre la* $\mathrm{Fr}(B)$ *en los puntos*

$$\gamma(t_0) = \left(1-\sqrt{6}/3, \sqrt{2(\sqrt{6}-1)/3}\right), \gamma(2\pi - t_0) = \left(1-\sqrt{6}/3, -\sqrt{2(\sqrt{6}-1)/3}\right)$$

y este máximo es $4\sqrt{6}/9$*, mientras que su valor mínimo lo alcanza en el punto* $\gamma(\pi) = (-1,0)$ *y este valor es* -4.

En resumen, comparando los valores de f en los puntos

$$(1,0), \left(1-\sqrt{6}/3, \sqrt{2(\sqrt{6}-1)/3}\right), \left(1-\sqrt{6}/3, -\sqrt{2(\sqrt{6}-1)/3}\right), (-1,0)$$

(el punto crítico $(1/3,0)$ no lo consideramos, puesto que ahí la función f tiene un punto silla), concluimos que el valor mínimo de f sobre el conjunto B es -4 y lo alcanza en el punto $(-1,0)$, mientras que su valor máximo es $4\sqrt{6}/9$ y lo alcanza en los puntos

$$\left(1-\sqrt{6}/3, \sqrt{2(\sqrt{6}-1)/3}\right) \qquad y \qquad \left(1-\sqrt{6}/3, -\sqrt{2(\sqrt{6}-1)/3}\right).$$

4.6.1 Breve comentario sobre formas cuadráticas

En esta breve subsección, justificaremos "intuitivamente" el resultado más conocido relacionado con las formas cuadráticas. En este resultado se establecen (para el caso general en $\mathbb{R}^n$) condiciones necesarias y suficientes para que una forma cuadrática F asociada a una matriz A sea semipositiva (o seminegativa) y no degenerada.

Sin lugar a dudas, las formas cuadráticas del tipo

$$\tilde{F}(x_1,\ldots,x_n) = \lambda_1 x_1^2 + \cdots + \lambda_n x_n^2 \tag{4.48}$$

son las más simples de analizar, pues es muy fácil comprobar que F es una forma cuadrática semipositiva (seminegativa) y no degenerada si y sólo si $\lambda_i > 0$ ($\lambda_i < 0$) para cada $i \in \{1,\ldots,n\}$ (nótese que $\tilde{F}(\hat{e}_i) = \lambda_i$).

También es fácil comprobar que

$$\tilde{F}(x_1,\ldots,x_n) = \left[\begin{array}{ccc} x_1 & \cdots & x_n \end{array}\right] \left[\begin{array}{ccc} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{array}\right] \left[\begin{array}{ccc} x_1 & \cdots & x_n \end{array}\right]^t,$$

de tal manera que la matriz asociada a $\tilde{F}$ es la matriz diagonal

$$D = \left[\begin{array}{ccc} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{array}\right].$$

Un hecho interesante de este tipo de matrices es que si definimos, para cada $k \in \{1,\ldots,n\}$,

$$D_k = \left[\begin{array}{ccc} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_k \end{array}\right],$$

entonces

$$\det(D_k) = \lambda_1 \cdots \lambda_k$$

de tal forma que ahora podemos reformular la afirmación anterior de la siguiente manera: $\tilde{F}$ es una forma cuadrática semipositiva (seminegativa) y no degenerada si y sólo si $\det(D_k) > 0$ ($\det(D_k) < 0$ si k es impar y $\det(D_k) > 0$ si k es par) para cada $k \in \{1,\ldots,n\}$.

La buena noticia es que este criterio (por medio de los determinantes de las submatrices D_k) que se cumple para el caso particular de $\tilde{F}$, se sigue cumpliendo en el caso general.

Es decir, si $F : \mathbb{R}^n \to \mathbb{R}$ es una forma cuadrática que tiene asociada la matriz simétrica

$$A = \left[\begin{array}{ccc} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{array}\right],$$

se prueba que: F es semipositiva (seminegativa) y no degenerada si y sólo si $\det(A_k) > 0$ ($\det(A_k) < 0$ si k es impar y $\det(A_k) > 0$ si k es par) para cada $k \in \{1, \ldots, n\}$, en donde

$$A_k = \begin{bmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{kk} \end{bmatrix}.$$

Como el lector puede comprobar fácilmente, en la proposición 4.56 se prueba esta afirmación para el caso $n = 2$.

Para reforzar un poco más las ideas anteriores, resulta relevante mencionar un resultado muy importante del Álgebra Lineal (Teorema 6.20 de [2]): si $A \in M_{n\times n}(\mathbb{R})$ es una matriz simétrica, entonces existe $B \in M_{n\times n}(\mathbb{R})$ una matriz ortonormal tal que BAB^t es una matriz diagonal. Es decir, que

$$BAB^t = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}. \tag{4.49}$$

Como seguramente el lector recordará, las matrices ortonormales son justo las matrices que se obtienen al (y que se usan para) realizar un cambio entre bases ortonormales de $\mathbb{R}^n$. De esta forma, con base en este resultado, podemos afirmar que: si $F : \mathbb{R}^n \to \mathbb{R}$ es una forma cuadrática que tiene asociada la matriz simétrica A (en una base ortonormal $\{\hat{e}_1, \ldots, \hat{e}_n\}$), es decir que

$$\begin{aligned} F(\hat{x}) &= F(x_1, \ldots, x_n) \\ &= \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} A \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^t, \end{aligned}$$

entonces existe otra base ortonormal $\{\hat{e}'_1, \ldots, \hat{e}'_n\}$ tal que la matriz asociada a F en esta nueva base es una matriz diagonal.

En efecto, observe que si

$$B = \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nn} \end{bmatrix}$$

es la matriz ortonormal tal que BAB^t es una matriz diagonal, entonces tomando

$$\hat{e}'_i = b_{i1}\hat{e}_1 + \cdots + b_{in}\hat{e}_n$$

para $i \in \{1, \ldots, n\}$, obtenemos que $\{\hat{e}'_1, \ldots, \hat{e}'_n\}$ es una base ortonormal de $\mathbb{R}^n$. De acuerdo con la identidad 4.1, si $\hat{x} \in \mathbb{R}^n$ tiene coordenadas $(x'_1, \ldots, x'_n)$ en esta base, entonces sus coordenadas $(x_1, \ldots, x_n)$ en la base $\{\hat{e}_1, \ldots, \hat{e}_n\}$ están dadas por la identidad

$$\begin{aligned} \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} &= \begin{bmatrix} x'_1 & \cdots & x'_n \end{bmatrix} \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nn} \end{bmatrix} \\ &= \begin{bmatrix} x'_1 & \cdots & x'_n \end{bmatrix} B. \end{aligned}$$

De esta manera, la forma cuadrática F expresada en términos de las nuevas coordenadas $x'_1, \ldots, x'_n$ se escribe como

$$\begin{aligned} F(\hat{x}) &= F(x_1, \ldots, x_n) \\ &= \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} A \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^t \\ &= \left(\begin{bmatrix} x'_1 & \cdots & x'_n \end{bmatrix} B\right) A \left(\begin{bmatrix} x'_1 & \cdots & x'_n \end{bmatrix} B\right)^t \\ &= \begin{bmatrix} x'_1 & \cdots & x'_n \end{bmatrix} \left(BAB^t\right) \begin{bmatrix} x'_1 & \cdots & x'_n \end{bmatrix}^t \end{aligned}$$

$$
\begin{aligned}
&= \begin{bmatrix} x_1' & \cdots & x_n' \end{bmatrix} \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_k \end{bmatrix} \begin{bmatrix} x_1' & \cdots & x_n' \end{bmatrix}^t \\
&= \lambda_1 \left(x_1'\right)^2 + \cdots + \lambda_n \left(x_n'\right)^2 .
\end{aligned}
$$

Es decir, en la base $\{\hat{e}_1', \ldots, \hat{e}_n'\}$ la función cuadrática F tiene la forma 4.48.

Sin duda que la parte medular de estas últimas ideas está en cómo, dada la matriz A, se encuentra la matriz ortonormal B y los escalares $\lambda_1, \ldots, \lambda_n$ que satisfagan la identidad 4.49. Pero este problema es todo un tema de un curso de Álgebra Lineal.

4.6.2 Máximos y mínimos sobre restricciones

Como se muestra en el ejemplo 4.58, para localizar los puntos en donde una función f alcanza sus valores extremos, no basta con encontrar sus puntos críticos (incluso puede suceder que estos valores extremos no se alcancen en este tipo de puntos, como sucede en el inciso 2 de este mismo ejemplo).

En la mayoría de los casos, la localización de los valores extremos (globales) de una función de $\mathbb{R}^n$ en $\mathbb{R}$, requerirá la localización de valores extremos sobre conjuntos "más pequeños", cuya característica principal será que son "flacos" (es decir, de interior vacío). En los incisos del ejemplo 4.58 estos conjuntos fueron

$$
\left\{(x,y) \in \mathbb{R}^2 \mid \|(x,y)\| = 2\right\} \qquad \text{y} \qquad \left\{(x,y) \in \mathbb{R}^2 \mid \|(x,y)\| = 1\right\},
$$

que además de poderse parametrizar por una función de $\mathbb{R}$ en $\mathbb{R}^2$ (propiedad que fue fundamental para encontrar los valores extremos sobre estos conjuntos), también tienen la propiedad de ser los conjuntos de nivel de alguna función de $\mathbb{R}^2$ en $\mathbb{R}$. De hecho, ambos conjuntos son conjuntos de nivel de la función $g(x,y) = x^2 + y^2$ para $c = 4$ y $c = 1$, es decir $N_4(g)$ y $N_1(g)$, respectivamente.

Lo interesante de la observación anterior es que, si ahora nos fijamos en los conjuntos de nivel de la función $f(x,y) = (x+1)^2 + y^2$ (la función del inciso 1 del mencionado ejemplo) para $c = 1$ y $c = 9$ (que son los valores extremos que alcanzó f sobre $N_4(g)$, y que denotamos por $N_1(f)$ y $N_9(f)$, respectivamente), notaremos que estos conjuntos de nivel se intersectan "tangencialmente" con el conjunto de nivel $N_4(g)$ en los puntos $(-2,0)$ y $(2,0)$, que son justo los puntos en donde f alcanza estos valores extremos (ver figura 4.16).

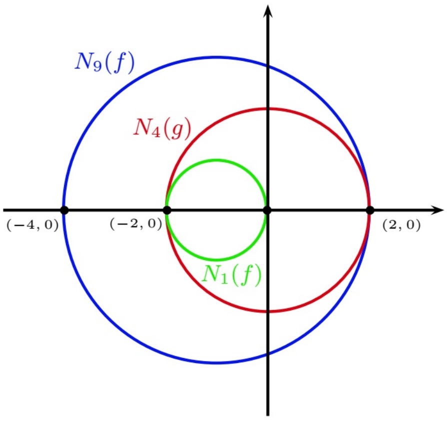


Figura 4.16: Los conjuntos de nivel $N_9(f)$ y $N_1(f)$ (de la función $f(x,y) = (x+1)^2 + y^2$) intersectan tangencialmente al conjunto de nivel $N_4(g)$ (de la función $g(x,y) = x^2 + y^2$) en los puntos $(2,0)$ y $(-2,0)$, respectivamente.

Lo importante de este hecho geométrico es que éste se puede "traducir" en una identidad (es decir, se puede "escribir" en forma "analítica"). En efecto, si ahora recordamos que el gradiente de una función en un punto $\hat{x}_0$ siempre es normal al conjunto de nivel que contiene a este punto, concluimos entonces que los vectores $\nabla f(-2,0)$ y $\nabla g(-2,0)$ deben ser paralelos (es decir, uno debe ser múltiplo del otro), y lo mismo debe suceder para los vectores $\nabla f(2,0)$ y $\nabla g(2,0)$, lo que es muy fácil de verificar, ya que

$$\nabla f(x,y) = (2(x+1), 2y) \qquad \text{y} \qquad \nabla g(x,y) = (2x, 2y)$$

para todo $(x,y) \in \mathbb{R}^2$, de modo que

$$\nabla f(2,0) = (6,0) \qquad \text{y} \qquad \nabla g(2,0) = (4,0),$$

y

$$\nabla f(-2,0) = (-2,0) \qquad \text{y} \qquad \nabla g(-2,0) = (-4,0).$$

Esto comprueba que

$$\nabla f(2,0) = \frac{3}{2}\nabla g(2,0) \qquad \text{y} \qquad \nabla f(-2,0) = \frac{1}{2}\nabla g(-2,0).$$

La discusión anterior parece sugerir lo siguiente: si tenemos una función f para la cual queremos calcular sus valores extremos sobre un conjunto de nivel de una función g (digamos $N_c(g)$), y localizar los puntos de este conjunto en los cuales alcanza estos valores extremos, es suficiente con encontrar los puntos $\hat{x} \in N_c(g)$ en los cuales se satisface que $\nabla f(\hat{x})$ y $\nabla g(\hat{x})$ son paralelos.

Aprovechando que ya sabemos cuáles son estos puntos para las funciones f y g del inciso 2 del ejemplo 4.58, sigamos el procedimiento descrito en el párrafo anterior y veamos si llegamos a las mismas soluciones. ¡Manos a la obra!

Como $f(x,y) = x(x-1)^2 + y^2$ y $g(x,y) = x^2 + y^2$, se tiene que

$$\begin{aligned}\nabla f(x,y) &= ((x-1)^2 + 2x(x-1), 2y)\\ &= ((x-1)(3x-1), 2y)\end{aligned}$$

y

$$\nabla g(x,y) = (2x, 2y),$$

de modo que nuestro problema es localizar las parejas $(x,y) \in N_1(g)$ para las cuales se satisfaga que

$$\nabla f(x,y) = \lambda \nabla g(x,y)$$

para alguna $\lambda \in \mathbb{R}$.

Lo anterior se traduce en encontrar las parejas $(x,y) \in \mathbb{R}^2$ y los números $\lambda \in \mathbb{R}$ que se satisfagan las siguientes ecuaciones:

$$\begin{aligned}\frac{\partial f}{\partial x}(x,y) &= \lambda \frac{\partial g}{\partial x}(x,y)\\ \frac{\partial f}{\partial y}(x,y) &= \lambda \frac{\partial g}{\partial y}(x,y)\\ g(x,y) &= c\end{aligned}$$

que en nuestro caso se traducen en el sistema de ecuaciones

$$\begin{aligned}(x-1)(3x-1) &= \lambda 2x \qquad (1)\\ 2y &= \lambda 2y \qquad (2)\\ x^2 + y^2 &= 1 \qquad (3)\end{aligned}$$

que en principio se debería poder resolver, pues es un sistema de tres ecuaciones con tres incógnitas (x, y y λ).

Empecemos por observar que, si alguna solución de las ecuaciones anteriores fuera tal que $y \neq 0$, entonces de la ecuación (2) deducimos que $\lambda = 1$, de modo que la ecuación (1) se convierte en la ecuación cuadrática (en la variable x)

$$3x^2 - 6x + 1 = 0,$$

cuyas soluciones son

$$\begin{aligned} x_1 &= \frac{6 + \sqrt{6^2 - 4(3)(1)}}{2(3)} \\ &= 1 + \frac{\sqrt{6}}{3} \end{aligned}$$

y

$$\begin{aligned} x_2 &= \frac{6 - \sqrt{6^2 - 4(3)(1)}}{2(3)} \\ &= 1 - \frac{\sqrt{6}}{3}. \end{aligned}$$

Si sustituimos estas soluciones en la ecuación (3), concluimos que sólo podemos tomar el segundo valor, obteniendo las siguientes soluciones a nuestro sistema original

$$(x, y) = \left(1 - \frac{\sqrt{6}}{3}, \sqrt{\frac{2(\sqrt{6} - 1)}{3}}\right) \qquad \text{con} \qquad \lambda = 1$$

y

$$(x, y) = \left(1 - \frac{\sqrt{6}}{3}, -\sqrt{\frac{2(\sqrt{6} - 1)}{3}}\right) \qquad \text{con} \qquad \lambda = 1.$$

Si ahora suponemos que tenemos una solución de nuestro sistema de ecuaciones con $y = 0$, de la ecuación (3) de este sistema concluimos que $x = 1$ o $x = -1$. Esto, partiendo de la ecuación (1), nos conduce a las soluciones

$$(x, y) = (1, 0) \qquad \text{con} \qquad \lambda = 0$$

y

$$(x, y) = (-1, 0) \qquad \text{con} \qquad \lambda = 4.$$

Seguramente el lector ya se percató de que con este procedimiento obtuvimos los mismos puntos que obtuvimos en el inciso 2 del ejemplo 4.58. La buena noticia es que el procedimiento anterior sigue siendo válido en cualquier $\mathbb{R}^n$. Es decir, aseguramos que: si f es una función derivable sobre un conjunto de nivel $N_c(g)$ de una función derivable g, y $\hat{x}_0 \in N_c(g)$ es un punto en el que la función f tiene un máximo o mínimo local (sobre $N_c(g)$), entonces se debe cumplir que $\nabla f(\hat{x}_0)$ y $\nabla g(\hat{x}_0)$ son vectores paralelos, es decir, que existe $\lambda \in \mathbb{R}$ tal que

$$\nabla f(\hat{x}_0) = \lambda \nabla g(\hat{x}_0).$$

La importancia de la afirmación anterior es que, si deseamos localizar los puntos en los que una función f alcanza sus valores extremos (incluso locales) sobre un conjunto de nivel $N_c(g)$ de una función derivable g, basta con encontrar los puntos de este conjunto para los cuales se satisface que los gradientes de f y de g son paralelos. Ilustraremos la afirmación anterior con un problema geométrico (en $\mathbb{R}^3$) que sin duda el lector conoce (y sabe la respuesta): dado un plano P cuya ecuación está dada por

$$Ax + By + Cz + D = 0$$

(con $A^2 + B^2 + C^2 > 0$) y un punto $\hat{x}_0 = (x_0, y_0, z_0) \in \mathbb{R}^3$, calcular la distancia del punto $\hat{x}_0$ al plano P.

Como sabemos, la distancia del punto $\hat{x}_0$ al plano P se obtiene como la mínima distancia entre el punto $\hat{x}_0$ y los puntos del plano P. Es decir, podemos plantear nuestro problema de la siguiente manera: consideramos la función d que nos da la distancia entre el punto $\hat{x}_0$ y cualquier otro punto $\hat{x} = (x, y, z) \in \mathbb{R}^3$, es decir

$$d(x, y, z) = \sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2}.$$

Nos planteamos el problema de localizar el mínimo valor de la función d sobre el conjunto $N_0(g)$, en donde g la definimos como

$$g(x, y, z) = Ax + By + Cz + D.$$

Para simplificar los cálculos, dado que el punto en $N_0(g)$ en el que se minimiza la distacia con $\hat{x}_0$ es el mismo en el que se minimiza la distacia al cuadrado, trabajaremos con la función

$$\begin{aligned} f(x,y,z) &= d^2(x,y,z) \\ &= (x-x_0)^2 + (y-y_0)^2 + (z-z_0)^2 \end{aligned}$$

(esta es una simplificación que el lector debería de tener siempre muy presente).

Dado que

$$\nabla f(x,y,z) = (2(x-x_0), 2(y-y_0), 2(z-z_0))$$

y

$$\nabla g(x,y,z) = (A,B,C)$$

nuestro problema se reduce a resolver el sistema de ecuaciones

$$\begin{aligned} 2(x-x_0) &= \lambda A \\ 2(y-y_0) &= \lambda B \\ 2(z-z_0) &= \lambda C \\ Ax+By+Cz+D &= 0. \end{aligned}$$

De las primeras tres ecuaciones concluimos que

$$\begin{aligned} x &= \frac{1}{2}\lambda A + x_0 \\ y &= \frac{1}{2}\lambda B + y_0 \\ z &= \frac{1}{2}\lambda C + z_0. \end{aligned}$$

Si sustituimos estos valores en la cuarta ecuación, obtenemos que

$$A\left(\frac{1}{2}\lambda A + x_0\right) + B\left(\frac{1}{2}\lambda B + y_0\right) + C\left(\frac{1}{2}\lambda C + z_0\right) + D = 0,$$

es decir, que

$$\lambda = -\frac{2\left(Ax_0+By_0+Cz_0+D\right)}{A^2+B^2+C^2}. \tag{4.50}$$

Por lo tanto, sustituyendo λ en los valores despejados de x, y y z, el punto $\hat{x}_p$ dado por la terna

$$\begin{aligned} \hat{x}_p &= \left(x_0 - A\frac{Ax_0+By_0+Cz_0+D}{A^2+B^2+C^2}, y_0 - B\frac{Ax_0+By_0+Cz_0+D}{A^2+B^2+C^2}, z_0 - C\frac{Ax_0+By_0+Cz_0+D}{A^2+B^2+C^2}\right) \\ &= \hat{x}_0 - \frac{Ax_0+By_0+Cz_0+D}{A^2+B^2+C^2}(A,B,C) \end{aligned}$$

es el único punto del plano P para el cual se satisface que

$$\nabla f(\hat{x}_p) = \lambda \nabla g(\hat{x}_p),$$

en donde λ está dada por 4.50 (nótese que, si $\hat{x}_0 \in P$, entonces $\hat{x}_p = \hat{x}_0$ ¡como era de esperarse!), de modo que la distacia del punto $\hat{x}_0$ al plano P estará dada por

$$\begin{aligned} d(\hat{x}_0, P) &= \sqrt{f(\hat{x}_p)} \\ &= \|\hat{x}_p - \hat{x}_0\| \\ &= \left\| -\frac{Ax_0+By_0+Cz_0+D}{A^2+B^2+C^2}(A,B,C) \right\| \end{aligned}$$

$$\begin{aligned} &= \frac{|Ax_0 + By_0 + Cz_0 + D|}{\|(A,B,C)\|} \\ &= \frac{|Ax_0 + By_0 + Cz_0 + D|}{\sqrt{A^2 + B^2 + C^2}}, \end{aligned}$$

lo que sin duda el lector recordará muy bien de sus cursos de Geometría Analítica.

El ejemplo anterior no sólo refuerza nuestra sospecha de que, si una función f alcanza un valor extremo sobre un conjunto de nivel de otra función g en un punto $\hat{x}_0$ de este conjunto, entonces los vectores gradiente de f y g en $\hat{x}_0$ deben ser paralelos, sino que nos permite plantear un problema que nos conducirá a formular un resultado más general que el descrito en el párrafo anterior.

El problema ahora es el siguiente:

1. dada una recta l determinada por la intersección de los planos P_1 y P_2 cuya ecuación está dada por

$$\begin{aligned} A_1x + B_1y + C_1z + D_1 = 0 \qquad\qquad \text{y} \\ A_2x + B_2y + C_2z + D_2 = 0 \end{aligned}$$

respectivamente, (con (A_1, B_1, C_1) y (A_2, B_2, C_2) vectores no paralelos (en particular diferentes de $\hat{0}$) para que sí determinen una recta), y

2. dado un punto $\hat{x}_0 = (x_0, y_0, z_0) \in \mathbb{R}^3$,

calcular la distacia del punto $\hat{x}_0$ a la recta l.

Como en el ejemplo anterior, usaremos la función que nos calcula la distacia al cuadrado entre el punto $\hat{x}_0$ y un punto de la recta l. Nuestra tarea será localizar el punto de l en el que esta función alcanza su valor mínimo (con respecto a los puntos de la recta l). Es decir, tomaremos la función

$$\begin{aligned} f(x,y,z) &= d^2(x,y,z) \\ &= (x-x_0)^2 + (y-y_0)^2 + (z-z_0)^2 \end{aligned}$$

y nuestro objetivo será localizar un punto $\hat{x}_l$ del conjunto

$$S = \left\{(x,y,z) \in \mathbb{R}^3 \mid g_1(x,y,z) = 0, g_2(x,y,z) = 0\right\}$$

tal que f alcance su valor mínimo (con respecto a los puntos de S) en $\hat{x}_l$, en donde g_1 y g_2 están dadas por

$$\begin{aligned} g_1(x,y,z) &= A_1x + B_1y + C_1z + D_1 \\ g_2(x,y,z) &= A_2x + B_2y + C_2z + D_2. \end{aligned}$$

Desde un punto de vista geométrico, si $\hat{x}_l$ es el punto de la recta l que está más cerca del punto $\hat{x}_0$, esto significa que $\hat{x}_l$ es el único punto de esta recta que pertenece al conjunto de nivel $N_c(f)$ de f, en donde

$$\begin{aligned} c &= f(\hat{x}_l) \\ &= \|\hat{x}_l - \hat{x}_0\|^2 \end{aligned}$$

y $N_c(f)$ es la esfera de radio $\sqrt{c}$ con centro en $\hat{x}_0$. Dicho de otra forma, la recta l es "tangente" a la esfera $N_c(f)$ en el punto $\hat{x}_l$. Nótese ahora que, si tomamos el vector que va de $\hat{x}_0$ a $\hat{x}_l$, que llamaremos $\hat{v}$, este debe ser perpendicular a la recta l, y si por otra parte recordamos que los vectores (A_1, B_1, C_1) y (A_2, B_2, C_2) son normales a los planos P_1 y P_2 (que determinan a la recta l y que por lo tanto la contienen), tendremos que dichos vectores también deberán ser perpendiculares a l (ver figura 4.17).

Del razonamiento anterior concluimos que, como los vectores $\hat{v}, (A_1, B_1, C_1)$ y (A_2, B_2, C_2) son perpendiculares a la misma recta l, y los dos últimos son linealmente independientes, se debe cumplir que el vector $\hat{v}$ se puede escribir como una combinación lineal de los vectores (A_1, B_1, C_1) y (A_2, B_2, C_2).

Finalmente, si ahora observamos:

1. que el vector (A_1, B_1, C_1) coincide con ser el vector gradiente de g_1 evaluado en cualquier punto del plano P_1 (en particular en $\hat{x}_l$),

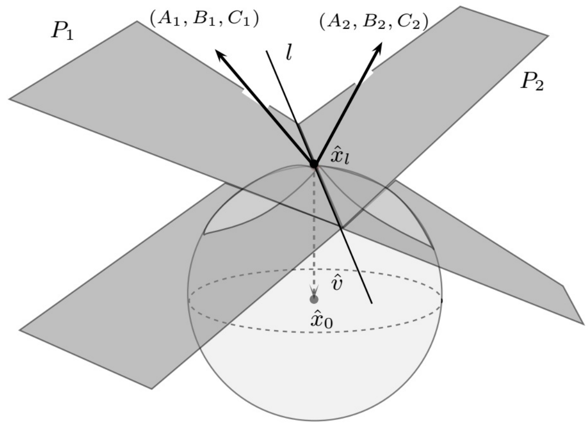


Figura 4.17: Dado que la recta l, que es la intersección de los planos P_1 y P_2, es tangente en el punto $\hat{x}_l$ a la esfera con centro en $\hat{x}_0$, los vectores $\hat{v} = \hat{x}_l - \hat{x}_0$, (A_1, B_1, C_1) y (A_2, B_2, C_2) (los vectores normales a P_1 y P_2, respectivamente) son perpendiculares a l, y por lo mismo pertenecen a un mismo plano.

2. que lo mismo sucede con el vector (A_2, B_2, C_2) y la función g_2, y
3. que los vectores $\nabla f(\hat{x}_l)$ y $\hat{v}$ son paralelos, pues ambos son normales a la esfera $N_c(f)$,

concluimos que el vector $\nabla f(\hat{x}_l)$ se debe poder expresar como una combinación lineal de los vectores $\nabla g_1(\hat{x}_l)$ y $\nabla g_2(\hat{x}_l)$. Es decir, deben existir $\lambda_1, \lambda_2 \in \mathbb{R}$ tales que

$$\nabla f(\hat{x}_l) = \lambda_1 \nabla g_1(\hat{x}_l) + \lambda_2 \nabla g_2(\hat{x}_l).$$

Para aprovechar la discusión anterior, formularemos un resultado mas general que seguramente el lector ya está intuyendo: sea $S \subset \mathbb{R}^3$ la intersección de los conjuntos de nivel (que sin perdida de generalidad supondremos que son los conjuntos de nivel 0) de dos funciones derivables g_1 y g_2, es decir que

$$S = \left\{(x, y, z) \in \mathbb{R}^3 \mid g_1(x, y, z) = 0, g_2(x, y, z) = 0\right\}.$$

Si $\hat{x}_0 \in S$ es un punto tal que:

1. los vectores $\nabla g_1(\hat{x}_0)$ y $\nabla g_2(\hat{x}_0)$ son linealmente independientes, y
2. una función f (derivable) alcanza un valor extremo (incluso local) en $\hat{x}_0$ sobre (o con respecto a) S,

entonces se debe cumplir que el vector $\nabla f(\hat{x}_0)$ se puede expresar como una combinación lineal de los vectores $\nabla g_1(\hat{x}_0)$ y $\nabla g_2(\hat{x}_0)$, es decir, deben existir $\lambda_1, \lambda_2 \in \mathbb{R}$ tales que

$$\nabla f(\hat{x}_0) = \lambda_1 \nabla g_1(\hat{x}_0) + \lambda_2 \nabla g_2(\hat{x}_0).$$

Si el lector ya está covencido de que el resultado anterior debe ser cierto, seguramente estará de acuerdo en que su siguiente generalización (para funciones de $\mathbb{R}^n$ en $\mathbb{R}$) también debe ser cierta: si $g_1, \ldots, g_m : U \subset \mathbb{R}^n \to \mathbb{R}$ son m funciones de clase C^1 en U, con $m < n$, $S \subset \mathbb{R}^n$ es el conjunto dado por

$$S = \{\hat{x} \in U \subset \mathbb{R}^n \mid g_1(\hat{x}) = 0, \ldots, g_m(\hat{x}) = 0\}$$

y $\hat{x}_0 \in S$ es un punto tal que:

1. los vectores $\nabla g_1(\hat{x}_0), \ldots, \nabla g_m(\hat{x}_0)$ son linealmente independientes, y
2. una función $f : U \subset \mathbb{R}^n \to \mathbb{R}$ de clase C^1 en U alcanza un valor extremo (incluso local) en $\hat{x}_0$ sobre (o con respecto a) S,

entonces se debe cumplir que el vector $\nabla f(\hat{x}_0)$ se puede expresar como una combinación lineal de los vectores $\nabla g_1(\hat{x}_0), \ldots, \nabla g_m(\hat{x}_0)$. Es decir, deben existir $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ tales que

$$\nabla f(\hat{x}_0) = \lambda_1 \nabla g_1(\hat{x}_0) + \cdots + \lambda_m \nabla g_m(\hat{x}_0).$$

Todo el trabajo que nos hemos tomado a lo largo de esta sección ha sido con el único objetivo de llegar al resultado anterior, el cual es conocido como el *Teorema de los multiplicadores de Lagrange*[3]. Antes de formularlo de manera más formal, conviene hacer algunas observaciones, empezando por su nombre: el término "multiplicadores" se refiere a los escalares $\lambda_1, \ldots, \lambda_m$. Al conjunto S se le conoce con el nombre de "*conjunto de restricciones*" y la condición de que m sea menor a n tiene que ver con los siguientes dos hechos: uno, que sólo si m es menor o igual a n se puede cumplir que m vectores en $\mathbb{R}^n$ sean linealmente independientes, y dos, si $m = n$ lo más probable es que el conjunto S sea vacío, o contega a lo más un número finito de puntos. Finalmente, la condición de que las funciones $g_1, \ldots, g_m$ y f sean de clase C^1 en U está directamente relacionada con las "herramientas" que usaremos en su prueba, que por cierto, daremos hasta el siguiente capítulo, pues mucha de esta "herramienta" será desarrollada hasta entonces.

La versión que probaremos en este texto del teorema de los multiplicadores de Lagrange es la siguiente.

Teorema 4.59 (de los multiplicadores de Lagrange) *Sean:*

1. $g_1, \ldots, g_m : U \subset \mathbb{R}^n \to \mathbb{R}$ *funciones de clase* C^1 *en* U, *con* $m < n$,
2. $S \subset \mathbb{R}^n$ *el conjunto dado por*

$$S = \{\hat{x} \in U \subset \mathbb{R}^n \mid g_1(\hat{x}) = 0, \ldots, g_m(\hat{x}) = 0\}$$

3. $\hat{x}_0 \in S$ *tal que* $\nabla g_1(\hat{x}_0), \ldots, \nabla g_m(\hat{x}_0)$ *son linealmente independientes.*

Si $f : U \subset \mathbb{R}^n \to \mathbb{R}$ *es una función de clase* C^1 *en* U *tal que* f *tiene un máximo o mínimo (local) en* $\hat{x}_0$ *sobre* S, *entonces existen* $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ *tales que*

$$\nabla f(\hat{x}_0) = \lambda_1 \nabla g_1(\hat{x}_0) + \cdots + \lambda_m \nabla g_m(\hat{x}_0).$$

Dado que la prueba de este teorema tendrá que esperar hasta el siguiente capítulo, por ahora sólo haremos una importante observación con relación a la tercera hipótesis de su enunciado. Esta observación se relaciona con el hecho de que un mismo conjunto de restricciones S se puede obtener usando diferentes colecciones de funciones $g_1, \ldots, g_m$, y no en todos los casos estas colecciones satisfacen la condición del inciso 3, la cual es fundamental para la validez del teorema. En el ejemplo siguiente ilustramos esta situación.

Ejemplo 4.60 *Sea* $S = \left\{(t, 0, 0) \in \mathbb{R}^3 \mid t \in \mathbb{R}\right\}$ *y*

$$\begin{aligned} f(x, y, z) &= \|(x, y, z) - (0, -1, -1)\|^2 \\ &= x^2 + (y+1)^2 + (z+1)^2 \end{aligned}$$

(el cuadrado de la distancia entre el punto (x, y, z) *y el punto* $(0, -1, -1)$*).*

Seguramente el lector estará de acuerdo en que $\hat{0} = (0, 0, 0)$ *es el punto en el cual la función* f *alcanza su mínimo valor (global) sobre el conjunto* S *y que en este punto se tiene que*

$$\nabla f(\hat{0}) = (0, 2, 2).$$

Primero observemos que, si $g_1(x, y, z) = z$ *y* $g_2(x, y, z) = z + y^2$, *entonces el conjunto* S *coincide con ser*

$$S = \left\{(x, y, z) \in \mathbb{R}^3 \mid g_1(x, y, z) = 0, g_2(x, y, z) = 0\right\}$$

[3]Joseph-Louis Lagrange, bautizado como Giuseppe Lodovico Lagrangia, también llamado Giuseppe Luigi Lagrangia o Lagrange (Turín, 25 de enero de 1736 - París, 10 de abril de 1813), fue un físico, matemático y astrónomo italiano que después vivió en Prusia y Francia. Lagrange trabajó para Federico II de Prusia, en Berlín, durante veinte años. Lagrange demostró el teorema del valor medio, desarrolló la mecánica Lagrangiana y tuvo una importante contribución en astronomía. (fuente: Wikipedia).

y si $g_3(x,y,z) = y$ *también se tiene que*

$$S = \left\{(x,y,z) \in \mathbb{R}^3 \mid g_1(x,y,z) = 0, g_3(x,y,z) = 0\right\}.$$

En el primer caso se tiene que $\nabla g_1(\hat{0}) = (0,0,1) = \nabla g_2(\hat{0})$*, es decir, no son linealmente independientes, y el vector* $\nabla f(\hat{0}) = (0,2,2)$ *no se puede escribir como combinación lineal de ellos (ver figura 4.18). En el segundo caso se tiene que* $\nabla g_1(\hat{0}) = (0,0,1)$ *y* $\nabla g_3(\hat{0}) = (0,1,0)$*, los cuales sí son linealmente independientes y se verifica que*

$$\nabla f(\hat{0}) = 2\nabla g_1(\hat{0}) + 2\nabla g_3(\hat{0}),$$

que es lo que asegura el teorema de los multiplicadores de Lagrange (ver figura 4.19).

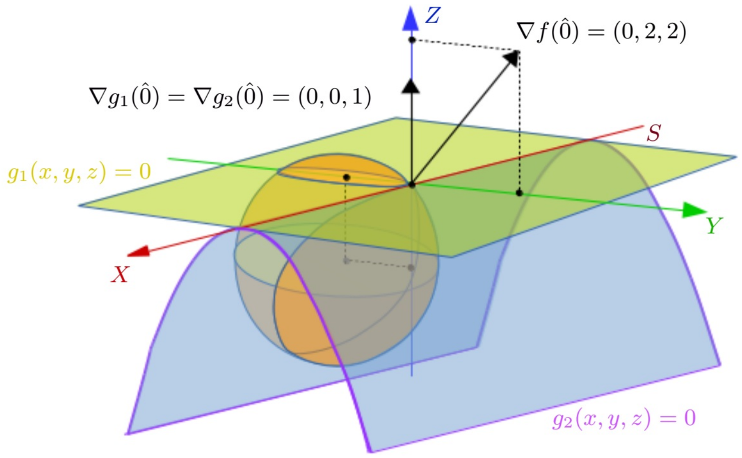


Figura 4.18: En el ejemplo 4.60, los vectores $\nabla g_1(\hat{0}) = \nabla g_2(\hat{0}) = (0,0,1)$ no son linealmente independientes y el vector $\nabla f(\hat{0}) = (0,2,2)$ no se puede escribir como combinación lineal de ellos.

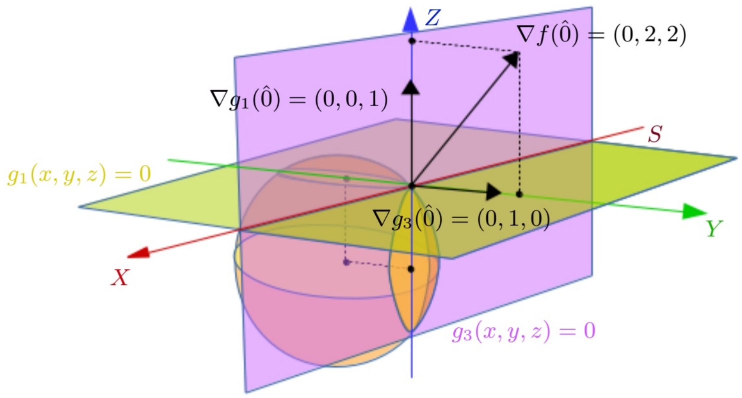


Figura 4.19: En el ejemplo 4.60, los vectores $\nabla g_1(\hat{0}) = (0,0,1)$ y $= \nabla g_3(\hat{0}) = (0,1,0)$ sí son linealmente independientes y el vector $\nabla f(\hat{0}) = (0,2,2)$ sí se puede escribir como combinación lineal de ellos, como asegura el teorema de los multiplicadores de Lagrange.

A continuación mostraremos cómo se usa el teorema 4.59 resolviendo el problema con el cual lo motivamos: encontrar la distancia de un punto $\hat{x}_0 = (x_0, y_0, z_0) \in \mathbb{R}^3$ a una recta l, la cual está determinada como la intersección de dos planos.

Desafortunadamente, resolver el caso general planteado inicialmente implica realizar largas y tediosas manipulaciones algebraicas, razón por la cual nos limitaremos a resolver el problema para un punto y una recta específicos.

Ejemplo 4.61 *Tomemos el punto* $(1,1,1) \in \mathbb{R}^3$ *y la recta determinada por la intersección de los planos*

$$\begin{aligned} x + y - 1 &= 0 \\ y - z + 1 &= 0, \end{aligned}$$

y calculemos el valor mínimo de la función

$$f(x,y,z) = (x-1)^2 + (y-1)^2 + (z-1)^2$$

sobre el conjunto $S \subset \mathbb{R}^3$ *dado por*

$$S = \left\{(x,y,z) \in \mathbb{R}^3 \mid g_1(x,y,z) = 0, g_2(x,y,z) = 0\right\},$$

en donde $g_1(x,y,z) = x + y - 1$ *y* $g_2(x,y,z) = y - z + 1$.

De acuerdo con el teorema de los multiplicadores de Lagrange, si $\hat{x}_0 = (x_0, y_0, z_0) \in S$ *es el punto en el que* f *alcanza su valor mínimo (sobre* S*), deben existir* $\lambda_1, \lambda_2 \in \mathbb{R}$ *tales que*

$$\nabla f(\hat{x}_0) = \lambda_1 \nabla g_1(\hat{x}_0) + \lambda_2 \nabla g_2(\hat{x}_0).$$

Esto nos lleva a resolver el sistema de ecuaciones

$$\begin{aligned} \frac{\partial f}{\partial x}(\hat{x}_0) &= \lambda_1 \frac{\partial g_1}{\partial x}(\hat{x}_0) + \lambda_2 \frac{\partial g_2}{\partial x}(\hat{x}_0) \\ \frac{\partial f}{\partial y}(\hat{x}_0) &= \lambda_1 \frac{\partial g_1}{\partial y}(\hat{x}_0) + \lambda_2 \frac{\partial g_2}{\partial y}(\hat{x}_0) \\ \frac{\partial f}{\partial z}(\hat{x}_0) &= \lambda_1 \frac{\partial g_1}{\partial z}(\hat{x}_0) + \lambda_2 \frac{\partial g_2}{\partial z}(\hat{x}_0) \\ g_1(\hat{x}_0) &= 0 \\ g_2(\hat{x}_0) &= 0 \end{aligned}$$

que en el caso particular que estamos analizando, se traduce en

$$\begin{aligned} 2(x_0 - 1) &= \lambda_1 \\ 2(y_0 - 1) &= \lambda_1 + \lambda_2 \\ 2(z_0 - 1) &= -\lambda_2 \\ x_0 + y_0 - 1 &= 0 \\ y_0 - z_0 + 1 &= 0. \end{aligned}$$

Para resolver este sistema, de las primeras tres ecuaciones despejamos a x_0, y_0 *y* z_0 *en términos de* λ_1 *y* λ_2 *y los sustituimos en las últimas dos ecuaciones, con lo que obtenemos el siguiente sistema en las incógnitas* λ_1 *y* λ_2*:*

$$\begin{aligned} 2\lambda_1 + \lambda_2 &= -2 \\ \lambda_1 + 2\lambda_2 &= -2, \end{aligned}$$

que tiene como solución $\lambda_1 = -2/3 = \lambda_2$. *Sustituyendo estos valores en las primeras tres ecuaciones del sistema original, concluimos que*

$$(x_0, y_0, z_0) = \frac{1}{3}(2,1,4),$$

de manera que la distancia que se estaba buscando es

$$\begin{aligned} \sqrt{f\left(\frac{1}{3}(2,1,-4)\right)} &= \sqrt{\left(\frac{2}{3} - 1\right)^2 + \left(\frac{1}{3} - 1\right)^2 + \left(\frac{4}{3} - 1\right)^2} \\ &= \frac{\sqrt{6}}{3}. \end{aligned}$$

Si el lector resolvió problemas de este estilo en sus cursos de Geometría Analítica, seguramente notó que en general los cálculos que acabamos de hacer son más sencillos, además de que se obtuvo explícitamente cuál es el punto de la recta que está más cercano al punto $\hat{x}_0$, lo que no se suele hacer por los métodos desarrollados en esos cursos.

Finalmente, daremos otro ejemplo de cómo usar el Teorema de los Multiplicadores de Lagrange para demostrar algunas desigualdades importantes. Aun cuando en el capítulo 1 dimos una prueba de la desigualdad de Cauchy-Schwarz, en el siguiente ejemplo daremos otra prueba usando este teorema.

Ejemplo 4.62 *Como sabemos, la desigualdad de Cauchy-Schwarz establece que*

$$|\hat{x} \cdot \hat{y}| \leq \|\hat{x}\| \, \|\hat{y}\|$$

para todos $\hat{x}, \hat{y} \in \mathbb{R}^n$.

Dado que, si $\hat{x} = \hat{0}$ o $\hat{y} = \hat{0}$ es claro que esta desigualdad se satisface, la probaremos para el caso en que $\hat{x} \neq \hat{0}$ y $\hat{y} \neq \hat{0}$. Bajo este supuesto podemos dividir la desigualdad anterior por $\|\hat{x}\| \, \|\hat{y}\|$ y nuestro problema se reduce a probar que

$$\left| \frac{\hat{x}}{\|\hat{x}\|} \cdot \frac{\hat{y}}{\|\hat{y}\|} \right| \leq 1.$$

Como

$$\left\| \frac{\hat{x}}{\|\hat{x}\|} \right\| = \frac{1}{\|\hat{x}\|} \|\hat{x}\| = 1 = \frac{1}{\|\hat{y}\|} \|\hat{y}\| = \left\| \frac{\hat{y}}{\|\hat{y}\|} \right\|$$

lo que debemos probar es que

$$|\hat{x} \cdot \hat{y}| \leq 1$$

para todos $\hat{x}, \hat{y} \in \mathbb{R}^n$ tales que $\|\hat{x}\| = 1 = \|\hat{y}\|$.

Con base en lo anterior, definimos $f, g_1, g_2 : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ como

$$\begin{aligned} f(x_1, \ldots, x_n, y_1, \ldots, y_n) &= x_1 y_1 + \cdots + x_n y_n \\ g_1(x_1, \ldots, x_n, y_1, \ldots, y_n) &= x_1^2 + \cdots + x_n^2 - 1 \\ g_2(x_1, \ldots, x_n, y_1, \ldots, y_n) &= y_1^2 + \cdots + y_n^2 - 1 \end{aligned}$$

y

$$S = \{(x_1, \ldots, x_n, y_1, \ldots, y_n) \in \mathbb{R}^n \times \mathbb{R}^n \mid g_i(x_1, \ldots, x_n, y_1, \ldots, y_n) = 0, i = 1, 2\}$$

de tal forma que ahora nuestro objetivo será mostrar, apoyados en el Teorema de los multiplicadores de Lagrange, que los valores mínimo y máximo de f sobre el conjunto S son 1 y -1, respectivamente, los cuales deben de existir puesto que f es continua en $\mathbb{R}^n$ y S es un conjunto cerrado y acotado (es decir, compacto).

Si $(\hat{x}_0, \hat{y}_0) = \left(x_1^{(0)}, \ldots, x_n^{(0)}, y_1^{(0)}, \ldots, y_n^{(0)}\right) \in S$ es un punto en donde f alcanza uno de estos valores extremos, por dicho teorema sabemos que existen $\lambda_1, \lambda_2 \in \mathbb{R}$ tales que

$$\nabla f(\hat{x}_0, \hat{y}_0) = \lambda_1 \nabla g_1(\hat{x}_0, \hat{y}_0) + \lambda_2 \nabla g_2 f(\hat{x}_0, \hat{y}_0),$$

es decir, que

$$\begin{aligned} y_i^{(0)} &= 2\lambda_1 x_i^{(0)} \\ x_i^{(0)} &= 2\lambda_2 y_i^{(0)} \end{aligned} \tag{4.51}$$

para cada $i \in \{1, \ldots, n\}$.

Si a la primera de estas identidades la multiplicamos por $x_i^{(0)}$, sumamos sobre el índice i, y usamos el hecho de que

$$\begin{aligned} \left(x_1^{(0)}\right)^2 + \cdots + \left(x_n^{(0)}\right)^2 - 1 &= g_1\left(x_1^{(0)}, \ldots, x_n^{(0)}, y_1^{(0)}, \ldots, y_n^{(0)}\right) \\ &= 0, \end{aligned}$$

obtenemos que

$$\begin{aligned} x_1^{(0)}y_1^{(0)} + \cdots + x_n^{(0)}y_n^{(0)} &= 2\lambda_1 \left(\left(x_1^{(0)}\right)^2 + \cdots + \left(x_n^{(0)}\right)^2 \right) \\ &= 2\lambda_1. \end{aligned}$$

Análogamente, si a la segunda identidad de 4.51 la multiplicamos por $y_i^{(0)}$*, sumamos sobre el índice* i*, y ahora usamos el hecho de que*

$$\begin{aligned} \left(y_1^{(0)}\right)^2 + \cdots + \left(y_n^{(0)}\right)^2 - 1 &= g_2\left(x_1^{(0)}, \ldots, x_n^{(0)}, y_1^{(0)}, \ldots, y_n^{(0)}\right) \\ &= 0, \end{aligned}$$

obtenemos que

$$\begin{aligned} x_1^{(0)}y_1^{(0)} + \cdots + x_n^{(0)}y_n^{(0)} &= 2\lambda_2 \left(\left(y_1^{(0)}\right)^2 + \cdots + \left(y_n^{(0)}\right)^2 \right) \\ &= 2\lambda_2. \end{aligned}$$

Como resultado de estos dos procedimientos, concluimos que $\lambda_1 = \lambda_2$*, de tal forma que si llamamos* $\lambda = \lambda_1 = \lambda_2$*, y sustituimos este valor en las identidades 4.51, obtenemos que*

$$\begin{aligned} y_i^{(0)} &= 2\lambda x_i^{(0)} = 2\lambda\left(2\lambda y_i^{(0)}\right) = 4\lambda^2 y_i^{(0)} \\ x_i^{(0)} &= 2\lambda y_i^{(0)} = 2\lambda\left(2\lambda x_i^{(0)}\right) = 4\lambda^2 x_i^{(0)} \end{aligned}$$

para cada $i \in \{1, \ldots, n\}$*.*

Dado que $y_i^{(0)}$ *o* $x_i^{(0)}$ *debe ser distinto de* 0 *para alguna* $i \in \{1, \ldots, n\}$*, se tiene que*

$$4\lambda^2 = 1,$$

y por lo tanto que

$$\lambda = \pm\frac{1}{2}.$$

Si sustituimos este valor de λ *en cualquiera de las identidades 4.51, se tiene que*

$$y_i^{(0)} = \pm x_i^{(0)}$$

o

$$x_i^{(0)} = \pm y_i^{(0)}.$$

Nótese que en cualquiera de estos dos casos, obtenemos que

$$\begin{aligned} f\left(x_1^{(0)}, \ldots, x_n^{(0)}, y_1^{(0)}, \ldots, y_n^{(0)}\right) &= f\left(x_1^{(0)}, \ldots, x_n^{(0)}, \pm x_1^{(0)}, \ldots, \pm x_n^{(0)}\right) \\ &= x_1^{(0)}\left(\pm x_1^{(0)}\right) + \cdots + x_n^{(0)}\left(\pm x_n^{(0)}\right) \\ &= \pm\left(\left(x_1^{(0)}\right)^2 + \cdots + \left(x_n^{(0)}\right)^2 \right) \\ &= \pm 1 \end{aligned}$$

o

$$\begin{aligned} f\left(x_1^{(0)}, \ldots, x_n^{(0)}, y_1^{(0)}, \ldots, y_n^{(0)}\right) &= f\left(\pm y_1^{(0)}, \ldots, \pm y_n^{(0)}, y_1^{(0)}, \ldots, y_n^{(0)}\right) \\ &= y_1^{(0)}\left(\pm y_1^{(0)}\right) + \cdots + y_n^{(0)}\left(\pm y_n^{(0)}\right) \end{aligned}$$

$$= \pm\left(\left(y_1^{(0)}\right)^2 + \cdots + \left(y_n^{(0)}\right)^2\right)$$
$$= \pm 1,$$

de donde concluimos que el valor mínimo de f sobre el conjunto S es -1, y el máximo es 1. Es decir que

$$|\hat{x} \cdot \hat{y}| = |f(\hat{x}, \hat{y})| \leq 1$$

para todos $\hat{x}, \hat{y} \in \mathbb{R}^n$ tales que $\|\hat{x}\| = 1 = \|\hat{y}\|$, que es lo que deseábamos demostrar.

4.7 Problemas

1. Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$ y $\hat{u} \in \mathbb{R}^n$ tal que $\|\hat{u}\| = 1$. Pruebe que, si la derivada direccional $D_{\hat{u}}f(\hat{x}_0)$ existe, entonces la derivada direccional $D_{-\hat{u}}f(\hat{x}_0)$ también existe y además
$$D_{-\hat{u}}f(\hat{x}_0) = -D_{\hat{u}}f(\hat{x}_0).$$

2. Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$ y $\hat{u} \in \mathbb{R}^n$ tal que $\|\hat{u}\| = 1$. Si $r > 0$ es tal que $B_r(\hat{x}_0) \subset U$, definimos $g : (-r, r) \subset \mathbb{R} \to \mathbb{R}$ como $g(x) = f(\hat{x}_0 + x\hat{u})$. Pruebe que, si $D_{\hat{u}}f(\hat{x})$ existe para toda $\hat{x} \in B_r(\hat{x}_0)$, entonces g es derivable para toda $x \in (-r, r)$ y además
$$g'(x) = D_{\hat{u}}f(\hat{x}_0 + x\hat{u}).$$

3. Pruebe las proposiciones 4.4 y 4.5.

4. Sea $f : \mathbb{R}^2 \to \mathbb{R}$ tal que $f(x, y) = f(y, x)$ para todo $(x, y) \in \mathbb{R}^2$. Pruebe que $\frac{\partial f}{\partial x}(a, b) = \frac{\partial f}{\partial y}(b, a)$ para todo $(a, b) \in \mathbb{R}^2$.

5. Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}$ tal que U es un conjunto abierto e $i \in \{1, \ldots, n\}$ fija. Pruebe que, si para todo par de puntos $\hat{x}, \hat{y} \in U$ tales que $\hat{x} - \hat{y} = \lambda\hat{e}_i$ para alguna $\lambda \in \mathbb{R}$ se tiene que $f(\hat{x}) = f(\hat{y})$ (es decir, que f no depende de la variable x_i), entonces $\frac{\partial f}{\partial x_i}$ existe para toda $\hat{x} \in U$ y además
$$\frac{\partial f}{\partial x_i}(\hat{x}) = 0.$$
¿Se cumple lo recíproco? ¿Es necesaria otra hipótesis sobre el conjunto U? Pruebe sus respuestas.

6. Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}$ tal que f es constante sobre el conjunto abierto U. Pruebe que f es derivable para toda $\hat{x} \in U$ y que además $Df(\hat{x})$ es la función lineal constante 0 ($Df(\hat{x}) \equiv 0$) para toda $\hat{x} \in U$. ¿Se cumple lo recíproco? ¿Es necesaria otra hipótesis sobre el conjunto U? Pruebe sus respuestas.

7. Si $L : \mathbb{R}^n \to \mathbb{R}$ es una función lineal, pruebe que L es derivable para toda $\hat{x} \in \mathbb{R}^n$ y que $DL(\hat{x}) = L$.

8. Dé una estimación para las siguientes cantidades. Justifique su respuesta.

   $i)(0.99e^{0.02})^8$ $\quad$ $ii)(0.99)^3 + (2.01)^3 - 6(0.99)(2.01)$ $\quad$ $iii)((4.01)^2 + (3.98)^2 + (2.02)^2)^{1/2}$

9. El capitán Nemo está en problemas cerca de la parte soleada de Mercurio. La temperatura del casco de su nave, cuando se localiza en el punto (x, y, z), está dada por la función $T(x, y, z) = e^{-x-2y^2-3z^2}$, donde x, y y z están medidos en metros. Si la nave se encuentra en el punto $(1, 1, 1)$:

   (a) ¿en qué dirección debe mover la nave para que la temperatura decrezca más rápidamente?

   (b) si la nave viaja a e^8 metros por segundo, ¿con qué rapidez decrecerá la temperatura si se mueve en la dirección del inciso anterior?

   (c) desafortunadamente, el metal del casco se fracturará si este se enfría a una razón mayor de $\sqrt{14e^2}$ grados por segundo. Describa el conjunto de posibles direcciones en las que se puede mover la nave para que la temperatura decrezca a una razón menor que ésta.

10. Sea $f : I \subset \mathbb{R} \to \mathbb{R}$ derivable en $x_0 \in I$, con I un intervalo abierto. Definimos $h : I \times \mathbb{R} \subset \mathbb{R}^2 \to \mathbb{R}$ como $h(x, y) = f(x) - y$. Pruebe:

    (a) que h es derivable en el punto $(x_0, y_0) = (x_0, f(x_0)) \in I \times \mathbb{R}$

    (b) que $G_f = N_0(h)$ y que la recta tangente al conjunto de nivel $N_0(h)$ en el punto $(x_0, f(x_0))$, calculada de acuerdo a la definición 4.32, coincide con ser la recta tangente a la gráfica de f en el punto $(x_0, f(x_0))$, calculada de acuerdo a como lo hacía en su primer curso de cálculo.

11. Sea $f : U \subset \mathbb{R}^2 \to \mathbb{R}$ derivable en $\hat{x}_0 = (x_0, y_0) \in U$. Definimos $h : U \times \mathbb{R} \subset \mathbb{R}^3 \to \mathbb{R}$ como $h(x, y, z) = f(x, y) - z$. Pruebe:

    (a) que h es derivable en el punto $(x_0, y_0, z_0) = (x_0, y_0, f(x_0, y_0)) \in U \times \mathbb{R}$

    (b) que $G_f = N_0(h)$ y que el plano tangente al conjunto de nivel $N_0(h)$ en el punto (x_0, y_0, z_0), calculada de acuerdo a la definición 4.33, coincide con ser el plano tangente a la gráfica de f en el punto (x_0, y_0, z_0), calculado de acuerdo con la definición 4.18.

12. De un ejemplo de una función $f : \mathbb{R}^2 \to \mathbb{R}$ que sólo sea derivable en el $(0, 0)$.

13. Muestre que las derivadas parciales de la función definida en el ejemplo 4.24 no son continuas en el $(0, 0)$ (*sugerencia*: hay dos formas de hacer esto, una directa, mostrando que el límite de ambas derivadas en el $(0, 0)$ no existe, y una indirecta, usando los resultados probados en el texto).

14. Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}$ tal que $\frac{\partial f}{\partial x_i}$ existe y es continua en cada punto $\hat{x} \in U$, para $i \in \{1, \dots, n\}$. Si $x_1', \dots, x_n'$ es otro sistema de coordenadas (inducido por otra base ortonormal $\{\hat{e}_1', \dots, \hat{e}_n'\}$), pruebe que $\frac{\partial f}{\partial x_i'}$ existe y es continua en cada punto $\hat{x} \in U$, para $i \in \{1, \dots, n\}$.

15. Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$ y $r > 0$ tal que $B_r(\hat{x}_0) \subset U$. Pruebe que, si para toda $i \in \{1, \dots, n\}$ se tiene que $\frac{\partial f}{\partial x_i}$ existe para cada $\hat{x} \in B_r(\hat{x}_0)$ y $\frac{\partial f}{\partial x_i}$ está acotada en $B_r(\hat{x}_0)$, entonces f es continua en $\hat{x}_0$.

16. Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$ y $r > 0$ tal que $B_r(\hat{x}_0) \subset U$. Pruebe que, si para toda $i \in \{1, \dots, n-1\}$ se tiene que $\frac{\partial f}{\partial x_i}$ existe para cada $\hat{x} \in B_r(\hat{x}_0)$ y es continua en $\hat{x}_0$, y además $\frac{\partial f}{\partial x_n}$ existe en $\hat{x}_0$, entonces f es derivable en $\hat{x}_0$.

17. Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$, $\gamma : (a, b) \subset \mathbb{R} \to \mathbb{R}^n$ tal que $\gamma(a, b) \subset U$, y $t_0 \in (a, b)$ tal que $\gamma(t_0) = \hat{x}_0$. Definimos $\varphi : (a, b) \subset \mathbb{R} \to \mathbb{R}$ como

$$\varphi(t) = \begin{cases} \frac{f(\gamma(t)) - f(\gamma(t_0)) - \nabla f(\gamma(t_0)) \cdot (\gamma(t) - \gamma(t_0))}{\|\gamma(t) - \gamma(t_0)\|} & \text{si } \gamma(t) - \gamma(t_0) \neq \hat{0} \\ 0 & \text{si } \gamma(t) - \gamma(t_0) = \hat{0} \end{cases}$$

    Pruebe que, si f es derivable en $\hat{x}_0$ y γ es continua en t_0, entonces φ es continua en t_0.

18. Pruebe la proposición 4.28.

19. Pruebe el corolario 4.31.

20. Sea $f : \mathbb{R}^2 \to \mathbb{R}$ continua. Use la regla de la cadena para demostrar que:

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

    (Defina $F(u, v) = \int_0^u f(v, y)dy$ y dé por hecho que $\frac{\partial F}{\partial v}(u, v) = \int_0^u \frac{\partial f}{\partial x}(v, y)dy$).

21. Sea $f : A \subset \mathbb{R}^n \to \mathbb{R}$. Se dice que f es homogénea de grado $p \in \mathbb{N}$ si $f(\lambda \hat{x}) = \lambda^p f(\hat{x})$ para toda $\lambda \in \mathbb{R}$ y toda $\hat{x} \in A$ tales que $\lambda \hat{x} \in A$. Pruebe que, si f es derivable en $\hat{x}_0 \in A$, entonces $\hat{x}_0 \cdot \nabla f(\hat{x}_0) = p f(\hat{x}_0)$ (este resultado es conocido como el teorema de Euler para funciones homogéneas).

22. Calcula el plano tangente a las superficies determinadas por las siguientes ecuaciones, en el punto indicado:

    (a) $x^2 + 2y^2 + 3xz = 10$ en $(1, 2, 1/3)$

    (b) $xe^z + yz = 1$ en $(1, 1, 0)$

    (c) $z = \cos(x)\,\mathrm{sen}(y)$ en $(0, \pi/2, 1)$.

23. Sean $f, g : \mathbb{R}^3 \to \mathbb{R}$. Supóngase que f es derivable y que $\nabla f(\hat{x}) = g(\hat{x})\hat{x}$ para toda $\hat{x} \in \mathbb{R}^3$. Pruebe que cada esfera en $\mathbb{R}^3$ con centro en el origen está contenida en un conjunto de nivel de f, es decir, que f es constante sobre esferas (*sugerencia*: use el problema 10 del capítulo 3).

24. Sea $f : U \subset \mathbb{R}^3 \to \mathbb{R}$ una función que está expresada en términos de las coordenadas cilíndricas (ρ, θ, z) de cada punto $\hat{x} \in U$. Si $\hat{x}_0 \in U$ tiene coordenadas cilíndricas (ρ_0, θ_0, z_0)

    (a) defina $\frac{\partial f}{\partial \rho}$, $\frac{\partial f}{\partial \theta}$ y $\frac{\partial f}{\partial z}$ en el punto $\hat{x}_0$. Interprete el significado geométrico de estas derivadas

    (b) relacione a cada una de las derivadas $\frac{\partial f}{\partial \rho}$, $\frac{\partial f}{\partial \theta}$ y $\frac{\partial f}{\partial z}$ en el punto $\hat{x}_0$ con alguna derivada direccional en el mismo punto

    (c) para cada punto $\hat{x} \in U$ defina una base ortonormal de $\mathbb{R}^3$ en la cual se pueda expresar la derivada de f en $\hat{x}$ $(Df(\hat{x}))$ en términos de las derivadas definidas en el inciso anterior

    (d) exprese a cada una de las derivadas $\frac{\partial f}{\partial \rho}$, $\frac{\partial f}{\partial \theta}$ y $\frac{\partial f}{\partial z}$ en el punto $\hat{x}_0$ en términos de las derivadas $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$ y $\frac{\partial f}{\partial z}$ en el punto $\hat{x}_0$, y viceversa.

25. Repita el problema anterior, suponiendo ahora que f está expresada en términos de coordenadas esféricas (ρ, θ, φ).

26. Sean, $f : U \subset \mathbb{R}^n \to \mathbb{R}$ de clase C^2 en U, y $\hat{u}, \hat{v} \in \mathbb{R}^n$ vectores unitarios. Pruebe que $D_{\hat{v}}(D_{\hat{u}}f)$ y $D_{\hat{u}}(D_{\hat{v}}f)$ existen y
$$D_{\hat{v}}(D_{\hat{u}}f)(\hat{x}) = D_{\hat{u}}(D_{\hat{v}}f)(\hat{x})$$
para toda $\hat{x} \in U$.

27. (a) Defina las derivadas parciales de orden dos en coordenadas polares (ver definición 4.36), en coordenadas cilíndricas (ver problema 24), y en coordenadas esféricas (ver problema 25).

    (b) Pruebe que el teorema de las derivadas parciales cruzadas también se cumple para estas derivadas parciales de orden dos.

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

    (a) calcule $\frac{\partial f}{\partial x}(x, y)$ y $\frac{\partial f}{\partial y}(x, y)$ para $(x, y) \neq (0, 0)$

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

    (c) pruebe que $\frac{\partial^2 f}{\partial y \partial x}(0, 0) = -1$ y que $\frac{\partial^2 f}{\partial x \partial y}(0, 0) = 1$

    (d) ¿este ejemplo contradice el resultado de la proposición 4.40? Justifique su respuesta.

29. Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $i, j \in \{1, \ldots, n\}$, $\hat{x}_0 \in U$ y $r > 0$ tal que $B_r(\hat{x}_0) \subset U$. Pruebe que, si $\frac{\partial^2 f}{\partial x_j \partial x_i}(\hat{x})$ existe para toda $\hat{x} \in B_r(\hat{x}_0)$ y es continua en $\hat{x}_0$, y $\frac{\partial f}{\partial x_j}(\hat{x})$ existe para toda $\hat{x} \in B_r(\hat{x}_0)$, entonces $\frac{\partial^2 f}{\partial x_i \partial x_j}(\hat{x}_0)$ existe y además
$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\hat{x}_0) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\hat{x}_0).$$
(*Sugerencia*: aplique el problema 41 del capítulo 2 a la función $H(h_1, h_2)/h_1 h_2$ definida en la prueba del teorema 4.40).

30. Sean $F_1, \ldots, F_n : U \subset \mathbb{R}^n \to \mathbb{R}$ tal que F_i es de clase C^1 en U para $i = 1, \ldots, n$, y supóngase que existe $f : U \subset \mathbb{R}^n \to \mathbb{R}$ tal que $\frac{\partial f}{\partial x_i}(\hat{x}) = F_i(\hat{x})$ para toda $\hat{x} \in U$ $(i = 1, \ldots, n)$. Pruebe que $\frac{\partial F_i}{\partial x_j}(\hat{x}) = \frac{\partial F_j}{\partial x_i}(\hat{x})$ para toda $\hat{x} \in U$ $(i, j = 1, \ldots, n)$.

31. Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}$ de clase C^{k+1} en U. Pruebe que cualquier derivada parcial de orden k de f es (como función de $\mathbb{R}^n$ en $\mathbb{R}$) derivable en cualquier punto $\hat{x} \in U$.

32. Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}$. Pruebe que, si f es de clase C^{k+1} en U, entonces f es de clase C^k en U (para toda $k \in \mathbb{N}$).

33. Sea $f : \mathbb{R}^n \to \mathbb{R}$ de clase C^2 en $\mathbb{R}^n$ y $g : \mathbb{R} \to \mathbb{R}$ dos veces derivable. Pruebe que $\frac{\partial^2 (g \circ f)}{\partial x_j \partial x_i}(\hat{x})$ existe para toda $\hat{x} \in \mathbb{R}^n$ $(i, j = 1, \ldots, n)$ y dé una expresión para estas derivadas parciales en términos de f y g.

34. Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$ y $r > 0$ tales que f es de clase C^k en $B_r(\hat{x}_0) \subset U$. Si $\hat{u} = (u_1, \ldots, u_n)$ es tal que $0 < \|\hat{u}\| < r$, y $\gamma : (-r/\|\hat{u}\|, r/\|\hat{u}\|) \subset \mathbb{R} \to \mathbb{R}^n$ está dada por $\gamma(t) = \hat{x}_0 + t\hat{u}$, pruebe que la función
$$f \circ \gamma : (-r/\|\hat{u}\|, r/\|\hat{u}\|) \subset \mathbb{R} \to \mathbb{R}$$
es de clase C^k en $(-r/\|\hat{u}\|, r/\|\hat{u}\|)$ y que además
$$(f \circ \gamma)^{(m)}(t) = \sum_{i_1, \ldots, i_m = 1}^{n} \frac{\partial^m f}{\partial x_{i_m} \cdots \partial x_{i_1}}(\gamma(t))\, u_{i_1} \cdots u_{i_m}$$
para cada $t \in (-r/\|\hat{u}\|, r/\|\hat{u}\|)$ y cada $m \in \{1, \ldots, k\}$.

35. Sean $f, g : \mathbb{R} \to \mathbb{R}$ dos veces derivables.

    (a) Definimos $h : \mathbb{R}^2 \to \mathbb{R}$ como $h(x,t) = f(x - at) + g(x + at)$ con $a \in \mathbb{R}$ una constante. Pruebe que:
    $$a^2 \frac{\partial^2 h}{\partial x^2}(x,t) = \frac{\partial^2 h}{\partial t^2}(x,t) \text{ para toda } (x,t) \in \mathbb{R}^2.$$

    (b) Definimos $h : \mathbb{R}^2 \to \mathbb{R}$ como $h(x,y) = xf(x+y) + yg(x+y)$. Pruebe que:
    $$\frac{\partial^2 h}{\partial x^2}(x,y) + \frac{\partial^2 h}{\partial y^2}(x,y) = 2\frac{\partial^2 h}{\partial y \partial x}(x,y) \text{ para toda } (x,y) \in \mathbb{R}^2.$$

    (c) Definimos $h : \mathbb{R}^2 \to \mathbb{R}$ como $h(x,y) = f(x^2 + y^2)$. Pruebe que:
    $$y^2 \frac{\partial^2 h}{\partial x^2}(x,y) - x^2 \frac{\partial^2 h}{\partial y^2}(x,y) + x\frac{\partial h}{\partial x}(x,y) - y\frac{\partial h}{\partial y}(x,y) = 0 \text{ para toda } (x,y) \in \mathbb{R}^2.$$

36. Sea $f : U \subset \mathbb{R}^2 \to \mathbb{R}$ y $\hat{x}_0 \in U$.

    (a) Pruebe que, si f es derivable en $\hat{x}_0$, entonces la función polinomial
    $$P_2(x,y) = A(x-x_0)^2 + B(y-y_0)^2 + C(x-x_0)(y-y_0) + \frac{\partial f}{\partial x}(\hat{x}_0)(x-x_0) + \frac{\partial f}{\partial y}(\hat{x}_0)(y-y_0) + f(\hat{x}_0)$$
    satisface que
    $$\lim_{\hat{x} \to \hat{x}_0} \frac{f(\hat{x}) - P_2(\hat{x})}{\|\hat{x} - \hat{x}_0\|} = 0$$
    para cualesquiera A, B y C números reales.

    (b) Muestre con un ejemplo que el inciso anterior es falso si sólo suponemos que $\frac{\partial f}{\partial x}(\hat{x}_0)$ y $\frac{\partial f}{\partial y}(\hat{x}_0)$ existen.

37. Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}$ de clase C^N en U y $\hat{x}_0 \in U$. Pruebe que el polinomio de Taylor $P_{N,f,\hat{x}_0}$ satisface que:
$$\lim_{\hat{x}\to\hat{x}_0} \frac{f(\hat{x}) - P_{N,f,\hat{x}_0}(\hat{x})}{\|\hat{x} - \hat{x}_0\|^N} = 0$$
(*sugerencia*: use la expresión del residuo vista en la demostración del teorema de Taylor para el caso $N - 1$).

38. Use el polinomio de Taylor de grado dos de la función $f(x, y) = \cos(x + y)$ para calcular el siguiente límite:
$$\lim_{(x,y)\to(0,0)} \frac{1 - \cos(x + y)}{(x^2 + y^2)^{1/2}}.$$
Pruebe su respuesta.

39. Sea $f(x, y, z) = Ax^2 + By^2 + Cz^2 + Dxy + Eyz + Fxz + G$. Pruebe que el residuo del polinomio de Taylor de orden dos de esta función siempre vale cero (en cualquier punto). Con base en lo anterior:
    (a) escriba el polinomio $x^2 + y^2 + z^2$ en potencias de $x - 1, y - 1$ y $z - 3$
    (b) escriba el polinomio $x^2 - 2y^2 + 4z^2 + xy - yz + xz + 1$ en potencias de $x - 1, y + 1$ y z.

40. Sea $p(x_1, \ldots, x_n)$ una función polinomial de las variables $x_1, \ldots, x_n$. Decimos que p es homogénea de orden $l \in \mathbb{N}$, si
$$p(tx_1, \ldots, tx_n) = t^l p(x_1, \ldots, x_n)$$
para toda $t \in \mathbb{R}$ y todo $(x_1, \ldots, x_n) \in \mathbb{R}^n$. Pruebe que:
    (a) si $f : \mathbb{R} \to \mathbb{R}$ es tal que $f^{(k)}(0)$ existe y $f \circ p$ es de clase C^{kl} en $\mathbb{R}^n$, entonces
$$P_{kl,f\circ p,\hat{0}} = P_{k,f,0} \circ p$$
    (b) para $s \in \{1, \ldots, kl\}$, $s \neq il$, con $i \in \{1, \ldots, k\}$, se tiene que
$$\frac{\partial^s (f \circ p)}{\partial x_n^{i_n} \cdots \partial x_1^{i_1}}(\hat{0}) = 0$$
    en donde $0 \leq i_1, \ldots, i_n \in \mathbb{N}$ son tales que $i_1 + \cdots + i_n = s$ (si $i_j = 0$, esto significa que en la expresión anterior no aparece la derivada parcial con respecto a la variable x_j).

41. Usando el polinomio de Taylor de grado adecuado de la función e^t, y sin hacer demasiados cálculos, encuentre todas las derivadas parciales de orden tres, en el $(0, 0)$, de la función $f(x, y) = (x^2+y^2)e^{x^2+y^2}$.

42. Por un procedimiento análogo al del problema anterior, encuentre el valor de todas las derivadas parciales de orden tres, en el $(0, 0)$, de la función $f(x, y) = (e^{xy} - 1)\operatorname{sen}(x + y)$.

43. Sean $x_1, \ldots, x_n \in \mathbb{R}$ y $N \in \mathbb{N}$. Use el teorema de Taylor para probar que
$$(x_1 + \cdots + x_n)^N = \sum_{\substack{k_1+\cdots+k_n=N \\ 0\leq k_i}} \begin{pmatrix} N \\ k_1 \cdots k_n \end{pmatrix} x_1^{k_1} \cdots x_n^{k_n}$$
donde
$$\begin{pmatrix} N \\ k_1 \cdots k_n \end{pmatrix} = \frac{N!}{k_1! \cdots k_n!}$$
(este número es conocido con el nombre de *coeficiente multinomial*).

44. Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $H : V \subset \mathbb{R}^k \to \mathbb{R}^n$, U y V abiertos, $\hat{x}_0 \in U$ y $\hat{y}_0 \in V$ tales que $H(\hat{y}_0) = \hat{x}_0$. Pruebe que, si f alcanza un valor máximo (mínimo) local en $\hat{x}_0$ y H es continua en $\hat{y}_0$, entonces $f \circ H$ alcanza un valor máximo (mínimo) local en $\hat{y}_0$. ¿Este resultado sigue siendo cierto si H no es continua en $\hat{y}_0$? Pruebe su respuesta.

45. Sea $n \in \mathbb{N}$, $n \geq 2$, y $f(x,y) = ax^n + by^n$, con $ab \neq 0$. Muestre que el $(0,0)$ es el único punto crítico de f y determine su tipo en términos de a, b y n.

46. Sea $f(x,y,z) = ax^2 + by^2 + cz^2$, $abc \neq 0$. Muestre que el $(0,0,0)$ es el único punto crítico de f y determine su tipo en términos de a, b y c.

47. Pruebe el inciso 2 de la proposición 4.53.

48. Pruebe el inciso 1 de la proposición 4.55.

49. Sea f una función de clase C^2 en una vecindad de $\hat{x}_0 \in \mathbb{R}^n$. Pruebe que, si f alcanza un mínimo local en $\hat{x}_0$, entonces
$$\frac{\partial^2 f}{\partial x_i^2}(\hat{x}_0) \geq 0$$
para cada $i \in \{1, \ldots, n\}$ (*sugerencia*: use la proposición 4.55).

50. Sea $\hat{x}_0 = (x_1, \ldots, x_n)$ un punto crítico de una función f de clase C^2 en una vecindad de $\hat{x}_0 \in \mathbb{R}^n$, tal que
$$\frac{\partial^2 f}{\partial x_i^2}(\hat{x}_0)\frac{\partial^2 f}{\partial x_j^2}(\hat{x}_0) < 0$$
para algunas $i, j \in \{1, \ldots, n\}$, con $i \neq j$. Pruebe que $\hat{x}_0$ es un punto silla de f.

51. Sea $\hat{x}_0 = (x_0, y_0)$ un punto crítico de una función f de clase C^2 en una vecindad de $\hat{x}_0 \in \mathbb{R}^2$, tal que
$$\frac{\partial^2 f}{\partial x^2}(\hat{x}_0)\frac{\partial^2 f}{\partial y^2}(\hat{x}_0) - \left(\frac{\partial^2 f}{\partial y \partial x}(\hat{x}_0)\right)^2 < 0.$$
Pruebe que $\hat{x}_0$ es un punto silla de f (*sugerencia*: observe que la cantidad anterior está relacionada con el discriminante del polinomio de grado 2 que se obtiene en la variable m al evaluar $Hf_{\hat{x}_0}(1, m)$, para $m \in \mathbb{R}$).

52. Sea $f : \mathbb{R}^2 \to \mathbb{R}$ definida como $f(x,y) = (y - 3x^2)(y - x^2)$. Pruebe que:
    (a) el origen es un punto crítico de f
    (b) si $a, b \in \mathbb{R}$ y $g(t) = (at, bt)$ con $t \in \mathbb{R}$, entonces $f \circ g$ tiene un mínimo local en $t = 0$
    (c) el origen no es un mínimo local de f.

53. Sea $f(x,y) = -(x^2 - 1)^2 - (x^2 y - x - 1)^2$. Pruebe que:
    (a) f sólo tiene dos puntos críticos
    (b) ambos puntos críticos son máximos locales
    (c) ¿se puede presentar una situación análoga para funciones de $\mathbb{R}$ en $\mathbb{R}$? Pruebe su respuesta.

54. Sea $f(x,y) = 3xe^y - x^3 - e^{3y}$. Pruebe que:
    (a) f sólo tiene un punto crítico
    (b) el punto crítico es un máximo local
    (c) f no tiene un máximo global
    (d) ¿se puede presentar una situación análoga para funciones de $\mathbb{R}$ en $\mathbb{R}$? Pruebe su respuesta.

55. Encuentre el máximo y el mínimo de la función $f(x,y) = xy - y + x - 1$ sobre el conjunto
$$A = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 2\}.$$

56. Sean $(x_1, y_1), \ldots, (x_n, y_n)$ puntos en $\mathbb{R}^2$ con $x_1 < x_2 < \ldots < x_n$. Pruebe que:

(a) la función

$$d(m,b) = \sum_{i=1}^{n} (y_i - (mx_i + b))^2$$

alcanza un valor mínimo en $\mathbb{R}^2$. Encuentre los valores m_0 y b_0 para los cuales se alcanza este valor mínimo (la recta $y = m_0x + b_0$ es la que “mejor” aproxima a los puntos $(x_1, y_1), \ldots, (x_n, y_n)$ y al método que se usa para calcular m_0 y b_0 se le conoce como el **método de los cuadrados mínimos**).

(b) si m_0 y b_0 son los valores del inciso anterior, pruebe que

$$\sum_{i=1}^{n} (y_i - (m_0x_i + b_0)) = 0$$

(c) si $n = 2$ (es decir, sólo se toman dos puntos), entonces $y = m_0x + b_0$ es la recta que pasa por dichos puntos.

57. Encuentre los extremos de f relativos a S, donde:

(a) $f(x, y) = x^2 - y^2$ y $S = \{(x, \cos(x)) \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$

(b) $f(x, y, z) = x^2 + y^2 + z^2$ y $S = \{(x, y, z) \in \mathbb{R}^3 \mid z \geq 2 + x^2 + y^2\}$.

58. Escriba el número 120 como suma de tres números, de modo que la suma de sus productos, tomados de dos en dos, sea máxima.

59. Una compañía planea fabricar cajas rectangulares cerradas con un volumen de 8 litros. El material para la base y la tapa cuesta el doble que el que se usa para los lados. Encuentre las dimensiones para las cuales el costo es mínimo.

60. Una ventana tiene la forma de un triángulo isósceles montado sobre un rectángulo. Si la base del rectángulo mide x, su altura mide y, los ángulos de la base del triángulo miden θ, y el perímetro de la ventana mide 4 metros, encuentre los valores de x, y y θ que hacen que el área de la ventana sea máxima.

61. Tres alelos (formas mutantes de genes) **A**, **B** y **O** determinan los cuatro tipos sanguíneos: **A** (**AA** o **AO**), **B** (**BB** o **BO**), **O** (**OO**) y **AB**. La ley de Hardy-Weinberg establece que la proporción P de individuos de una población que llevan dos alelos diferentes es $P = 2pq + 2pr + 2qr$, donde p, q y r son las proporciones de los alelos **A**, **B** y **O** que se presentan en dicha población, respectivamente. Use el hecho de que $p + q + r = 1$ para demostrar que $P \leq 2/3$.

62. Sea $P \in S = N_1(f) \subset \mathbb{R}^3$ con f de clase C^1 en $\mathbb{R}^3$. Supóngase que P es un punto donde se maximiza la distancia del origen a S. Pruebe que el vector que sale del origen y termina en P es perpendicular a S.

63. Sea A una matriz de 3×3, simétrica y diferente de la matriz cero. Definimos $f(x, y, z) = \frac{1}{2}(A(x, y, z)^t) \cdot (x, y, z)$ y $S = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1\}$. Pruebe que:

(a) si $\hat{x}_0 \in S$ es un punto en donde f alcanza su valor máximo (o mínimo) sobre S, entonces $\hat{x}_0$ es un vector propio de A, es decir, que existe λ tal que $A\hat{x}_0 = \lambda\hat{x}_0$

(b) existe $\hat{x}_0 \in S$ tal que $\hat{x}_0$ es un vector propio de A correspondiente a un valor propio distinto de cero, es decir, que existe $\lambda \neq 0$ tal que $A\hat{x}_0 = \lambda\hat{x}_0$.

64. Sean $a_1, \ldots, a_k$ números positivos. Use multiplicadores de Lagrange para probar que:

$$(a_1 \cdots a_k)^{1/k} \leq \frac{a_1 + \cdots + a_k}{k}.$$

65. Sean $a_i, b_i \geq 0$ para $i = 1, \ldots, n$, y $p, q > 1$ tales que $\frac{1}{p} + \frac{1}{q} = 1$.

(a) Use multiplicadores de Lagrange para demostrar que

$$a_1b_1 + \cdots + a_nb_n \leq \left(a_1^p + \cdots + a_n^p\right)^{1/p} \left(b_1^q + \cdots + b_n^q\right)^{1/q}$$

(esta desigualdad es conocida como *desigualdad de Hölder*).

(b) Use la desigualdad anterior para probar que

$$\left((a_1 + b_1)^p + \cdots + (a_n + b_n)^p\right)^{1/p} \leq \left(a_1^p + \cdots + a_n^p\right)^{1/p} + \left(b_1^p + \cdots + b_n^p\right)^{1/p}$$

(esta desigualdad es conocida como *desigualdad de Minkowski*).