Capítulo 5

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

En este capítulo introduciremos el concepto de derivada para funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$, que entre otras cosas, tendrá que coincidir con los conceptos que hemos desarrollado en los capítulos anteriores cuando $n = 1$ o $m = 1$. Además de definir el concepto de derivada para este tipo de funciones, formularemos y probaremos sus propiedades más importantes, algunas de las cuales generalizan a las correspondientes propiedades vistas en los capítulos 3 y 4. Por la razón anterior, mucho del material que veremos ahora estará basado en las ideas y conceptos de los capítulos anteriores (sobre todo del capítulo 4).

Para finalizar este capítulo (¡y este texto!) formularemos y probaremos dos teoremas, que sin duda son de los más importantes del Cálculo Diferencial de varias Variables: el Teorema de la Función Implícita y el Teorema de la Función Inversa.

5.1 La derivada

Como era de esperarse, la definición de derivada de una función $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ en un punto $\hat{x}_0 \in U$ está basada en la misma idea que se usó para las funciones de $\mathbb{R}^n$ en $\mathbb{R}$: la existencia de "la mejor aproximación lineal a f alrededor del punto $\hat{x}_0$".

Para empezar, notemos que, como en el capítulo anterior, si $L : \mathbb{R}^n \to \mathbb{R}^m$ es una función lineal de $\mathbb{R}^n$ en $\mathbb{R}^m$, entonces $L(\hat{x} - \hat{x}_0) + f(\hat{x}_0)$ es una función afín de $\mathbb{R}^n$ en $\mathbb{R}^m$ cuyo valor en $\hat{x}_0$ coincide con el valor de f en $\hat{x}_0$. Ahora, como el lector estará de acuerdo, decir que una función afín de éstas "se parece mucho a f alrededor del punto $\hat{x}_0$", significará que la diferencia de estas funciones, evaluada en puntos $\hat{x}$ "cercanos" a $\hat{x}_0$, tiende a $\hat{0}$ más rápido de lo que la diferencia $\hat{x} - \hat{x}_0$ también tiende a $\hat{0}$.

De manera más precisa, la discusión anterior queda plasmada en la siguiente

Definición 5.1 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ y $\hat{x}_0 \in U$. Decimos que f es derivable en $\hat{x}_0$ si existe $L : \mathbb{R}^n \to \mathbb{R}^m$ una función lineal tal que*

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

o lo que es equivalente (problema 33 del capítulo 2), si

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

Lo que ahora procede es dar un ejemplo de cómo calcular la derivada que acabamos de definir, pero que ésta no sea de una función del tipo de las que ya vimos en el capítulo 3 o en el capítulo 4, es decir, la derivada de una función de $\mathbb{R}^n$ en $\mathbb{R}^m$ en donde $n > 1$ y $m > 1$.

Observación 5.2 *Antes de ello, y en beneficio del propio ejemplo, conviene hacer notar que la definición anterior es en efecto una generalización de las definiciones de derivada dadas en los capítulos antes mencionados.*

Que la definición 4.14 del capítulo 4 es un caso particular de la definición 5.1, es un hecho que el lector puede verificar con sólo leer su enunciado. Para verificar la equivalencia entre la definición 5.1 y la definición

3.2 dada en el capítulo 3, el lector tendrá que recurrir al problema 9 de este capítulo, del cual se desprende fácilmente lo que aquí afirmamos.

Una vez dicho lo anterior, procedemos a dar el siguiente

Ejemplo 5.3 *Sea $f : \mathbb{R}^2 \to \mathbb{R}^3$ definida como*

$$f(x,y) = (x^2\cos(y), x^2\operatorname{sen}(y), x^2).$$

Dado que la definición 5.1, como ya mencionamos, es una generalización de los conceptos de derivada que vimos en los capítulos 3 y 4, por la proposición 3.8 del primero de estos dos capítulos, debemos sospechar que la derivabilidad (¡y la derivada!) de f está determinada por la derivabilidad (¡y la derivada!) de cada una de sus funciones coordenadas, funciones que son del tipo de las que vimos en el capítulo 4.

En virtud de lo anterior, todo parece indicar que para obtener la derivada de f en un punto $(x_0, y_0) \in \mathbb{R}^2$, será necesario calcular la derivada de sus funciones coordenadas $f_1(x,y) = x^2\cos(y)$, $f_2(x,y) = x^2\operatorname{sen}(y)$ y $f_3(x,y) = x^2$ en este punto.

Ahora, por los resultados obtenidos en el capítulo 4, sabemos que las funciones lineales L_1,L_2 y L_3 de $\mathbb{R}^2$ en $\mathbb{R}$ asociadas a las matrices de 1×2

$$\begin{bmatrix} \frac{\partial f_1}{\partial x}(x_0,y_0) & \frac{\partial f_1}{\partial y}(x_0,y_0) \end{bmatrix} = \begin{bmatrix} 2x_0\cos(y_0) & -x_0^2\operatorname{sen}(y_0) \end{bmatrix},$$
$$\begin{bmatrix} \frac{\partial f_2}{\partial x}(x_0,y_0) & \frac{\partial f_2}{\partial y}(x_0,y_0) \end{bmatrix} = \begin{bmatrix} 2x_0\operatorname{sen}(y_0) & x_0^2\cos(y_0) \end{bmatrix}$$

y

$$\begin{bmatrix} \frac{\partial f_3}{\partial x}(x_0,y_0) & \frac{\partial f_3}{\partial y}(x_0,y_0) \end{bmatrix} = \begin{bmatrix} 2x_0 & 0 \end{bmatrix}$$

son las funciones lineales que más se le "parecen" a las respectivas funciones coordenadas alrededor del punto (x_0, y_0). Por esta razón, la función lineal L de $\mathbb{R}^2$ en $\mathbb{R}^3$ que tiene como funciones coordenadas a las funciones L_1,L_2 y L_3, y cuya matriz asociada (de 3×2) está dada por

$$\begin{bmatrix} \frac{\partial f_1}{\partial x}(x_0,y_0) & \frac{\partial f_1}{\partial y}(x_0,y_0) \\ \frac{\partial f_2}{\partial x}(x_0,y_0) & \frac{\partial f_2}{\partial y}(x_0,y_0) \\ \frac{\partial f_3}{\partial x}(x_0,y_0) & \frac{\partial f_3}{\partial y}(x_0,y_0) \end{bmatrix} = \begin{bmatrix} 2x_0\cos(y_0) & -x_0^2\operatorname{sen}(y_0) \\ 2x_0\operatorname{sen}(y_0) & x_0^2\cos(y_0) \\ 2x_0 & 0 \end{bmatrix}$$

sin duda es la mejor candidata para satisfacer la condición de la definición 5.1.

En efecto, de la definición 4.14 del capítulo 4 sabemos que para cada $i \in \{1,2,3\}$ se satisface que

$$\lim_{(x,y)\to(x_0,y_0)} \frac{f_i(x,y) - (L_i(x-x_0, y-y_0) + f_i(x_0,y_0))}{\|(x-x_0, y-y_0)\|} = 0.$$

Por la proposición 2.30 del capítulo 2 podemos concluir que

$$\lim_{(x,y)\to(x_0,y_0)} \frac{f(x,y) - (L(x-x_0, y-y_0) + f(x_0,y_0))}{\|(x-x_0, y-y_0)\|} = \hat{0},$$

de modo que f es derivable en el punto (x_0, y_0).

Tomaremos el ejemplo anterior como punto de partida para obtener algunas conclusiones sobre las funciones de $\mathbb{R}^2$ en $\mathbb{R}^3$. Empezaremos por aquellas que son de carácter geométrico.

5.1.1 Elementos básicos acerca de superficies

Como seguramente el lector recordará, en el capítulo 2 mencionamos que desde un punto de vista geométrico, lo importante de las funciones de $\mathbb{R}^2$ en $\mathbb{R}^3$ es su imagen. Así mismo, recordará que el concepto de derivada para funciones de $\mathbb{R}$ en $\mathbb{R}^3$ nos fue útil para definir la recta tangente, en un punto, de un conjunto que se puede ver como la imagen de una función de este tipo.

Ahora veremos que si un conjunto $S \subset \mathbb{R}^3$ se puede obtener como la imagen de una función, que en esta subsección denotaremos como $\sigma : U \subset \mathbb{R}^2 \to \mathbb{R}^3$, la cual es derivable en un punto $\hat{x}_0 \in U$, y esta derivada cumple con otra condición que precisaremos más adelante, entonces podremos definir lo que llamaremos el plano tangente a S en el punto $\sigma(\hat{x}_0)$.

Sea pues $S \subset \mathbb{R}^3$ un conjunto que se puede ver como la imagen de una función

$$\sigma = (\sigma_1, \sigma_2, \sigma_3) : U \subset \mathbb{R}^2 \to \mathbb{R}^3,$$

la cual es derivable en un punto $\hat{x}_0 = (x_0, y_0) \in U$. Dado que estamos suponiendo que U es un conjunto abierto, entonces existe $r > 0$ tal que $B_r(\hat{x}_0) \subset U$. De esta forma, podemos definir las funciones

$$\gamma_x, \gamma_y : (-r, r) \subset \mathbb{R} \to \mathbb{R}^3$$

como

$$\begin{aligned}\gamma_x(t) &= \sigma(x_0 + t, y_0)\\ &= (\sigma_1(x_0 + t, y_0), \sigma_2(x_0 + t, y_0), \sigma_3(x_0 + t, y_0))\end{aligned}$$

y

$$\begin{aligned}\gamma_y(t) &= \sigma(x_0, y_0 + t)\\ &= (\sigma_1(x_0, y_0 + t), \sigma_2(x_0, y_0 + t), \sigma_3(x_0, y_0 + t)).\end{aligned}$$

Afirmamos que estas funciones son derivables en $t = 0$.

En efecto, de acuerdo con la definición 3.2 del capítulo 3, se tiene que

$$\begin{aligned}\gamma_x'(0) &= \lim_{t\to 0} \frac{\gamma_x(t) - \gamma_x(0)}{t - 0}\\ &= \lim_{t\to 0} \frac{(\sigma_1(x_0 + t, y_0) - \sigma_1(x_0, y_0), \sigma_2(x_0 + t, y_0) - \sigma_2(x_0, y_0), \sigma_3(x_0 + t, y_0) - \sigma_3(x_0, y_0))}{t}\\ &= \lim_{t\to 0} \left(\frac{\sigma_1(x_0 + t, y_0) - \sigma_1(x_0, y_0)}{t}, \frac{\sigma_2(x_0 + t, y_0) - \sigma_2(x_0, y_0)}{t}, \frac{\sigma_3(x_0 + t, y_0) - \sigma_3(x_0, y_0)}{t} \right)\\ &= \left(\frac{\partial \sigma_1}{\partial x}(x_0, y_0), \frac{\partial \sigma_2}{\partial x}(x_0, y_0), \frac{\partial \sigma_3}{\partial x}(x_0, y_0) \right)\end{aligned}$$

y

$$\begin{aligned}\gamma_y'(0) &= \lim_{t\to 0} \frac{\gamma_y(t) - \gamma_y(0)}{t - 0}\\ &= \lim_{t\to 0} \frac{(\sigma_1(x_0, y_0 + t) - \sigma_1(x_0, y_0), \sigma_2(x_0, y_0 + t) - \sigma_2(x_0, y_0), \sigma_3(x_0, y_0 + t) - \sigma_3(x_0, y_0))}{t}\\ &= \lim_{t\to 0} \left(\frac{\sigma_1(x_0, y_0 + t) - \sigma_1(x_0, y_0)}{t}, \frac{\sigma_2(x_0, y_0 + t) - \sigma_2(x_0, y_0)}{t}, \frac{\sigma_3(x_0, y_0 + t) - \sigma_3(x_0, y_0)}{t} \right)\\ &= \left(\frac{\partial \sigma_1}{\partial y}(x_0, y_0), \frac{\partial \sigma_2}{\partial y}(x_0, y_0), \frac{\partial \sigma_3}{\partial y}(x_0, y_0) \right).\end{aligned}$$

Por lo anterior, si se satisface que

$$\gamma_x'(0) = \left(\frac{\partial \sigma_1}{\partial x}(x_0, y_0), \frac{\partial \sigma_2}{\partial x}(x_0, y_0), \frac{\partial \sigma_3}{\partial x}(x_0, y_0) \right) \neq \hat{0}$$

y

$$\gamma_y'(0) = \left(\frac{\partial \sigma_1}{\partial y}(x_0, y_0), \frac{\partial \sigma_2}{\partial y}(x_0, y_0), \frac{\partial \sigma_3}{\partial y}(x_0, y_0) \right) \neq \hat{0},$$

entonces $\gamma_x((-r, r))$ y $\gamma_y((-r, r))$ son dos curvas contenidas en S que tienen recta tengente en el punto $\sigma(\hat{x}_0)$.

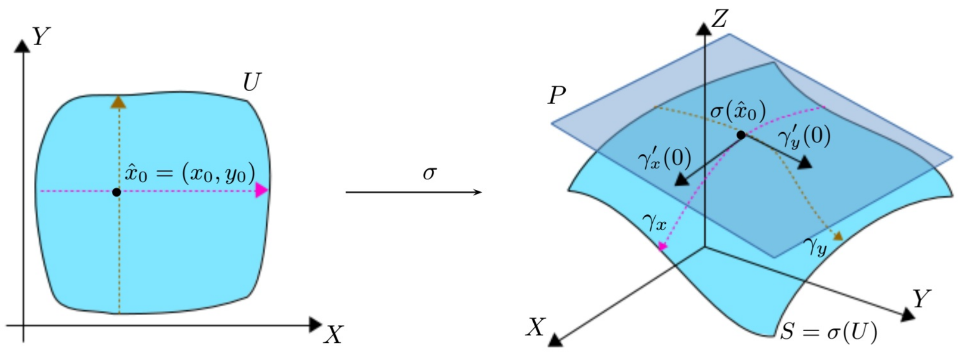

Figura 5.1: Si los vectores $\gamma_x'(0)$ y $\gamma_y'(0)$ son linealmente independientes, entonces el plano P generado por ellos, trasladado al punto $\sigma(\hat{x}_0)$, es tangente a la superficie $S = \sigma(U)$.

Ahora, si además los vectores $\gamma_x'(0)$ y $\gamma_y'(0)$ son linealmente independientes, el conjunto dado por

$$\left\{\sigma\left(\hat{x}_0\right) + t\gamma_x'(0) + s\gamma_y'(0) \in \mathbb{R}^3 \mid t, s \in \mathbb{R}\right\}$$

representa un plano que tiene todo el aspecto de ser tangente a S en $\sigma\left(\hat{x}_0\right)$ (ver figura 5.1).

Con base en la discusión anterior, definiremos los siguientes conceptos.

Definición 5.4 *Sea $S \subset \mathbb{R}^3$.*

1. *Decimos que S es una superficie, si existen $\sigma = (\sigma_1, \sigma_2, \sigma_3) : U \subset \mathbb{R}^2 \to \mathbb{R}^3$ derivable en cada punto de U, y $A \subset U$ tales que $\sigma(A) = S$. En este caso decimos que σ es una parametrización de S.*

2. *Si $\hat{x}_0 \in A$ y los vectores*
$$\frac{\partial \sigma}{\partial x}\left(\hat{x}_0\right) := \left(\frac{\partial \sigma_1}{\partial x}(\hat{x}_0), \frac{\partial \sigma_2}{\partial x}(\hat{x}_0), \frac{\partial \sigma_3}{\partial x}(\hat{x}_0)\right)$$
y
$$\frac{\partial \sigma}{\partial y}\left(\hat{x}_0\right) := \left(\frac{\partial \sigma_1}{\partial y}(\hat{x}_0), \frac{\partial \sigma_2}{\partial y}(\hat{x}_0), \frac{\partial \sigma_3}{\partial y}(\hat{x}_0)\right)$$
son linealmente independientes, decimos que el plano $P \subset \mathbb{R}^3$ definido paramétricamente como
$$P := \left\{\sigma\left(\hat{x}_0\right) + t\frac{\partial \sigma}{\partial x}\left(\hat{x}_0\right) + s\frac{\partial \sigma}{\partial y}\left(\hat{x}_0\right) \in \mathbb{R}^3 \mid t, s \in \mathbb{R}\right\}$$
es el plano tangente a S en el punto $\sigma\left(\hat{x}_0\right)$, y que S es suave en $\sigma\left(\hat{x}_0\right)$.

Con relación a esta nueva definición de plano tangente en un punto de una superficie, es importante hacer algunas observaciones. La primera de ellas es que, aún cuando una superficie $S \subset \mathbb{R}^3$ es la imagen de alguna función derivable, esto no significa que el conjunto S no pueda tener “picos”. Una prueba de este hecho nos lo proporciona la función

$$\sigma(x, y) = (x^2\cos(y), x^2\operatorname{sen}(y), x^2)$$

del ejemplo 5.3 la cual, ya sabemos, es derivable en cualquier punto $(x, y) \in \mathbb{R}^2$.

Como el lector podrá verificar fácilmente, el conjunto $S = \sigma(\mathbb{R}^2)$ es la parte superior del cono determinado por la ecuación cartesiana $u^2 + v^2 = w^2$ (figura 5.2), el cual claramente tiene un “pico” en el punto $\sigma(0, 0) = (0, 0, 0)$.

Una segunda observación es que no cualquier parametrización de una superficie $S \subset \mathbb{R}^3$ es útil para calcular su plano tangente en algún punto (del mismo modo que sucedió en el caso de las curvas).

El conjunto $S \subset \mathbb{R}^3$ y el par de parametrizaciones de éste que daremos en el siguiente ejemplo nos proporcionan una prueba de este hecho.

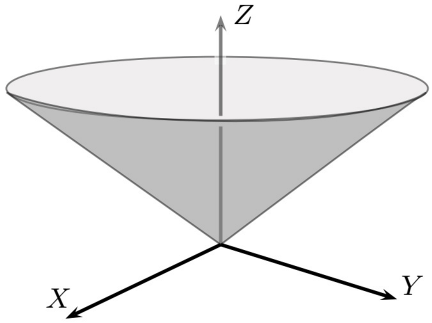

Figura 5.2: El cono determinado por la ecuación cartesiana $u^2 + v^2 = w^2$ tiene un "pico" en el punto $(0,0,0)$.

Ejemplo 5.5 *Sea $S \subset \mathbb{R}^3$ el siguiente conjunto:*

$$S = \left\{(u,v,w) \in \mathbb{R}^3 \mid w = u^2 + v^2\right\}.$$

La figura 5.3 muestra un esbozo de S, el cual corresponde a un paraboloide. A continuación daremos un par de parametrizaciones de este conjunto.

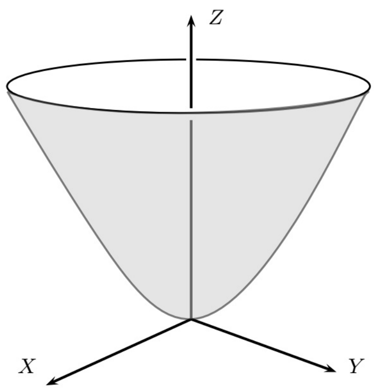

Figura 5.3: El paraboloide determinado por la ecuación cartesiana $u^2 + v^2 = w$ tiene como plano tangente en el punto $(0,0,0)$ al plano XY.

1. *Sea $\sigma : \mathbb{R}^2 \to \mathbb{R}^3$ definida como $\sigma(x,y) = \left(x, y, x^2 + y^2\right)$. Como el lector podrá verificar fácilmente, se tiene que $S = \sigma(\mathbb{R}^2)$. Por otra parte, procediendo como en el ejemplo 5.3, se concluye que σ es derivable para cualquier $(x,y) \in \mathbb{R}^2$ y además que*

$$\frac{\partial \sigma}{\partial x}(x,y) = (1,0,2x)$$

y

$$\frac{\partial \sigma}{\partial y}(x,y) = (0,1,2y).$$

Por lo tanto, para el punto $(0,0)$ se tiene que

$$\frac{\partial \sigma}{\partial x}(0,0) = (1,0,0)$$

y

$$\frac{\partial \sigma}{\partial y}(0,0) = (0,1,0),$$

los cuales son vectores linealmente independientes. De esta forma, el conjunto

$$\begin{aligned} P &= \left\{ \sigma(0,0) + t\frac{\partial\sigma}{\partial x}(0,0) + s\frac{\partial\sigma}{\partial y}(0,0) \in \mathbb{R}^3 \mid t,s \in \mathbb{R} \right\} \\ &= \left\{ (0,0,0) + t(1,0,0) + s(0,1,0) \in \mathbb{R}^3 \mid t,s \in \mathbb{R} \right\} \\ &= \left\{ (t,s,0) \in \mathbb{R}^3 \mid t,s \in \mathbb{R} \right\} \end{aligned}$$

sí resulta ser un plano (el plano XY), que sin duda es el plano tangente a S en el punto $\sigma(0,0) = (0,0,0)$.

2. *Sea ahora $\sigma : \mathbb{R}^2 \to \mathbb{R}^3$ definida como $\sigma(x,y) = \left(x\cos(y), x\,\mathrm{sen}(y), x^2\right)$. Nuevamente el lector podrá verificar fácilmente que $S = \sigma(\mathbb{R}^2)$. Por otra parte, para esta función también es sencillo mostrar que es derivable para cualquier $(x,y) \in \mathbb{R}^2$ y además, que*

$$\frac{\partial\sigma}{\partial x}(x,y) = (\cos(y), \mathrm{sen}(y), 2x)$$

y

$$\frac{\partial\sigma}{\partial y}(x,y) = (-x\,\mathrm{sen}(y), x\cos(y), 0)\,.$$

Por lo tanto, para el punto $(0,0)$ se tiene que

$$\frac{\partial\sigma}{\partial x}(0,0) = (1,0,0)$$

y

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

los cuales no son vectores linealmente independientes, de tal forma que el conjunto

$$\begin{aligned} P &= \left\{ \sigma(0,0) + t\frac{\partial\sigma}{\partial x}(0,0) + s\frac{\partial\sigma}{\partial y}(0,0) \in \mathbb{R}^3 \mid t,s \in \mathbb{R} \right\} \\ &= \left\{ (0,0,0) + t(1,0,0) + s(0,0,0) \in \mathbb{R}^3 \mid t,s \in \mathbb{R} \right\} \\ &= \left\{ (t,0,0) \in \mathbb{R}^3 \mid t,s \in \mathbb{R} \right\} \end{aligned}$$

no resulta ser un plano.

Concluimos nuestras observaciones acerca de la definición de plano tangente a una superficie recordando que en el capítulo 4 se definió este mismo concepto para dos objetos geométricos diferentes; la gráfica de una función de $\mathbb{R}^2$ en $\mathbb{R}$ (definición 4.18), y el conjunto de nivel de una función de $\mathbb{R}^3$ en $\mathbb{R}$ (definición 4.33). En el problema 11 de este mismo capítulo, el lector probó que la gráfica de una función de $\mathbb{R}^2$ en $\mathbb{R}$ siempre se puede obtener como el conjunto de nivel de una cierta función de $\mathbb{R}^3$ en $\mathbb{R}$, y que el plano tangente que se obtiene usando cualquiera de las definiciones antes mencionadas es el mismo. Algo análogo probará el lector en el problema 1 de este capítulo; la gráfica de una función de $\mathbb{R}^2$ en $\mathbb{R}$ siempre se puede obtener como la imagen de una cierta función $\mathbb{R}^2$ en $\mathbb{R}^3$, y el plano tangente que se obtiene usando la definición 5.4 es el mismo que se obtiene usando la definición 4.18.

Como mencionamos en el capítulo 4, en este capítulo probaremos el Teorema de la Función Implícita, del cual podremos deducir (bajo ciertas hipótesis) que todo conjunto de nivel de una función de $\mathbb{R}^3$ en $\mathbb{R}$ se puede obtener (al menos “por partes”) como la gráfica de una función de $\mathbb{R}^2$ en $\mathbb{R}$ y, por lo dicho en el párrafo anterior, entonces también como la imagen de una cierta función de $\mathbb{R}^2$ en $\mathbb{R}^3$.

En el problema 21 de este capítulo el lector probará la afirmación anterior, y además que el plano tangente calculado de acuerdo con la definición 5.4 es el mismo que se obtiene usando la definición 4.33.

5.2 Propiedades de la derivada

Seguramente el lector estará de acuerdo en que, para determinar la derivabilidad de la función del ejemplo 5.3, fue muy importante saber que sus funciones coordenadas eran derivables. Esto sin duda nos lleva a concluir que en general, para deducir la derivabilidad de una función de $\mathbb{R}^n$ en $\mathbb{R}^m$ en un punto, es suficiente con que sus funciones coordenadas lo sean en ese mismo punto.

La buena noticia es que la afirmación recíproca también es cierta, es decir, si una función de $\mathbb{R}^n$ en $\mathbb{R}^m$ es derivable en un punto, sus funciones coordenadas también deben ser derivables en ese punto. Este hecho es el primer criterio importante que veremos para determinar la derivabilidad (¡o no derivabilidad!) de una función de $\mathbb{R}^n$ en $\mathbb{R}^m$ y lo dejamos plasmado en la siguiente

Proposición 5.6 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$, $\hat{x}_0 \in U$ y $f_1, \dots, f_m : U \subset \mathbb{R}^n \to \mathbb{R}$ funciones coordenadas de f (en una base ortonormal $\{\hat{e}_1, \dots, \hat{e}_m\}$ de $\mathbb{R}^m$). La función f es derivable en $\hat{x}_0$ si y sólo si f_i es derivable en $\hat{x}_0$ para cada $i \in \{1, \dots, m\}$.*

Demostración. Sea $L : \mathbb{R}^n \to \mathbb{R}^m$ una función lineal y supongamos que $L_i : \mathbb{R}^n \to \mathbb{R}$ son las funciones coordenadas de L (que por lo tanto también son lineales) determinadas por la misma base ortonormal de $\mathbb{R}^m$ que determina a las funciones coordenadas de f.

Dado que la $i-$ésima coordenada del vector

$$\frac{1}{\|\hat{x} - \hat{x}_0\|}\left(f(\hat{x}) - (L(\hat{x} - \hat{x}_0) + f(\hat{x}_0))\right) = \frac{f(\hat{x}) - (L(\hat{x} - \hat{x}_0) + f(\hat{x}_0))}{\|\hat{x} - \hat{x}_0\|}$$

está dada por

$$\frac{f_i(\hat{x}) - (L_i(\hat{x} - \hat{x}_0) + f_i(\hat{x}_0))}{\|\hat{x} - \hat{x}_0\|}$$

para cada $i \in \{1, \dots, m\}$, por la proposición 2.30 (del capítulo 2), se sigue que

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

si y sólo si

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

de donde las afirmaciones de esta proposición se concluyen inmediatamente. ∎

Muchos resultados y propiedades de la derivada de funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$ serán una consecuencia inmediata de la proposición anterior (y de los correspondientes resultados y propiedades para funciones de $\mathbb{R}^n$ en $\mathbb{R}$), y para empezar, la usaremos para probar la unicidad de la derivada de las funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$. En efecto, como por la proposición 4.15 del capítulo 4 sabemos que la derivada de una función de $\mathbb{R}^n$ en $\mathbb{R}$ es única, por la proposición anterior concluimos que esto mismo sucede para las funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$. Por esta razón, de aquí en adelante hablaremos de *la* derivada de f en $\hat{x}_0$ y la denotamos por $Df(\hat{x}_0)$ (igual que en el caso de las funciones de $\mathbb{R}^n$ en $\mathbb{R}$); es decir, $Df(\hat{x}_0)$ designará a la función lineal que al trasladarla al punto $\hat{x}_0$, construyendo la función $Df(\hat{x}_0)(\hat{x} - \hat{x}_0) + f(\hat{x}_0)$, obtenemos la función lineal afín que más se le parece a f alrededor del punto $\hat{x}_0$. Por otra parte, como seguramente el lector recordará de su curso de Álgebra Lineal, toda función lineal $L : \mathbb{R}^n \to \mathbb{R}^m$ se puede representar por una matriz de $m \times n$ con entradas reales (que depende de las bases que se elijan para $\mathbb{R}^n$ y $\mathbb{R}^m$), y que se construye de la siguiente manera: Si $L : \mathbb{R}^n \to \mathbb{R}^m$ es una función lineal, $\{\hat{e}_1, \dots, \hat{e}_n\}$ y $\{\tilde{e}_1, \dots, \tilde{e}_m\}$ son bases ortonormales de $\mathbb{R}^n$ y $\mathbb{R}^m$, respectivamente, y $\hat{x} \in \mathbb{R}^n$ es tal que

$$\begin{aligned}\hat{x} &= (x_1, \dots, x_n)\\ &= x_1\hat{e}_1 + \dots + x_n\hat{e}_n,\end{aligned}$$

entonces

$$L(\hat{x}) = L\left(x_1\hat{e}_1 + \dots + x_n\hat{e}_n\right) = x_1 L\left(\hat{e}_1\right) + \dots + x_n L\left(\hat{e}_n\right)$$

De esta forma, si

$$L\left(\hat{e}_i\right) = a_{1i}\tilde{e}_1 + \cdots + a_{mi}\tilde{e}_m$$

para cada $i \in \{1, \ldots, n\}$, se tiene que

$$\begin{aligned}
L(\hat{x}) &= x_1 L\left(\hat{e}_1\right) + \cdots + x_n L\left(\hat{e}_n\right) \\
&= x_1\left(a_{11}\tilde{e}_1 + \cdots + a_{m1}\tilde{e}_m\right) + \\
&\quad \vdots \\
&\quad + x_n\left(a_{1n}\tilde{e}_1 + \cdots + a_{mn}\tilde{e}_m\right) \\
&= \left(x_1, \ldots, x_n\right) \cdot \left(a_{11}, \ldots, a_{1n}\right)\tilde{e}_1 + \\
&\quad \vdots \\
&\quad + \left(x_1, \ldots, x_n\right) \cdot \left(a_{m1}, \ldots, a_{mn}\right)\tilde{e}_m
\end{aligned}$$

de donde, usando matrices, obtenemos que

$$L(\hat{x}) = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

Por tanto, se tiene que la matriz

$$\beta = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

representa a la función L (en las bases $\{\hat{e}_1, \ldots, \hat{e}_n\}$ y $\{\tilde{e}_1, \ldots, \tilde{e}_m\}$).

Obsérvese que las matrices β_i de $1 \times n$ dadas por

$$\beta_i = \begin{bmatrix} a_{i1} & \cdots & a_{in} \end{bmatrix}$$

para cada $i \in \{1, \ldots, m\}$ representarán a sus correspondientes funciones coordenadas L_i, y recíprocamente, si la matriz β_i representa a la función lineal L_i, para cada $i \in \{1, \ldots, m\}$, entonces la matriz β representa a la función lineal L. De lo anterior, y de la identidad 4.17 del capítulo 4 tendremos que, si $f_1, \ldots, f_m$ son las funciones coordenadas de la función f (las cuales, reiteramos que dependen de la base de $\mathbb{R}^m$ que se elija), entonces la derivada de f en $\hat{x}_0$ ($Df(\hat{x}_0)$) estará representada por la matriz

$$\begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial f_1}{\partial x_n}(\hat{x}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial f_m}{\partial x_n}(\hat{x}_0) \end{bmatrix}, \tag{5.1}$$

a la que se le conococe con el nombre de *matriz jacobiana*[1], y no sin cierto abuso de notación (¡nuevamente!) escribiremos que

$$Df(\hat{x}_0) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial f_1}{\partial x_n}(\hat{x}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial f_m}{\partial x_n}(\hat{x}_0) \end{bmatrix}, \tag{5.2}$$

Sin duda una condición que es necesaria y suficiente resulta ser muy útil (como la que se da en la proposición 5.6), sin embargo, las que sólo son necesarias o sólo son suficientes, también lo son. Y aprovechando

[1]Llamada así en honor de Carl Gustav Jacob Jacobi (10 de diciembre de 1804 en Potsdam, Prusia, actual Alemania -18 de febrero de 1851 en Berlín). Autor muy prolífico, contribuyó en varios campos de la matemática, principalmente en el área de las funciones elípticas, el álgebra, la teoría de números y las ecuaciones diferenciales. También destacó en su labor pedagógica, por la que se le ha considerado el profesor más estimulante de su tiempo. (fuente: Wikipedia).

que conocemos dos de este tipo de propiedades para funciones de $\mathbb{R}^n$ en $\mathbb{R}$, enunciaremos sus equivalentes para funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$.

La primera de ellas es una condición (o consecuencia) necesaria de la derivabilidad de una función en un punto, y es una consecuencia inmediata de las proposiciones 5.6, 4.22 y 2.43.

Proposición 5.7 *Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$. Si f es derivable en $\hat{x}_0 \in U$, entonces f es continua en $\hat{x}_0$.*

La otra propiedad que es muy importante mencionar, y que es una condición suficiente para la derivabilidad de una función en un punto, es una consecuencia inmediata de las proposiciones 5.6 y 4.25.

Proposición 5.8 *Sean, $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$, $\hat{x}_0 \in U$ y $r > 0$ tal que $B_r(\hat{x}_0) \subset U$. Si $f_1, \ldots, f_m$ son funciones coordenadas de f tales que $\frac{\partial f_j}{\partial x_i}(\hat{x})$ existe para cada $\hat{x} \in B_r(\hat{x}_0)$ y $\frac{\partial f_j}{\partial x_i}$ es continua en $\hat{x}_0$, para cada $i \in \{1, \ldots, n\}$ y para cada $j \in \{1, \ldots, m\}$, entonces f es derivable en $\hat{x}_0$.*

Como vimos en el capítulo 4, la hipótesis de la proposición 4.25 (análoga a la anterior) dio lugar al concepto de función de clase C^k, concepto que jugó un papel muy importante en el tema de aproximación polinomial, y que ahora vamos a generalizar, apoyados justo en la proposición 5.6, a las funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$ en la siguiente

Definición 5.9 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$, $f_1, \ldots, f_m$ funciones coordenadas de f, y $k \in \mathbb{N}$. Decimos que f es una función de clase C^k en U si cada f_j es una función de clase C^k en U, para $j \in \{1, \ldots, m\}$. Es decir, si existen todas las derivadas parciales de orden k de f_j en cada punto de U, y además estas derivadas parciales son continuas en cada punto de U, para $j \in \{1, \ldots, m\}$.*

Con respecto a la definición anterior, hay que mencionar lo siguiente: por el problema 14 del capítulo 4 y la identidad 5.3 que probaremos en la siguiente subsección, el hecho de que una función sea de clase C^k en un conjunto (abierto) es independiente tanto del sistema coordenado que estemos usando para representar a cada punto $\hat{x} \in \mathbb{R}^n$, como del que estemos usando para representar a cada punto $\hat{y} \in \mathbb{R}^m$. Es decir, este concepto es independiente de las variables coordenadas $x_1, \ldots, x_n$ y de las funciones coordenadas $f_1, \ldots, f_m$.

Como mencionamos anteriormente, con base en el concepto de función de clase C^k (para $k = 1$) podemos establecer el siguiente resultado, cuya prueba es una consecuencia inmediata de la proposición 4.43 del capítulo 4 y de la proposición 5.6.

Proposición 5.10 *Si $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ es de clase C^1 en U, entonces f es derivable para toda $\hat{x} \in U$.*

Las propiedades de la derivada de funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$ relacionadas con la aritmética de éstas también se obtienen de manera inmediata a partir de la proposición 5.6 y de las correspondientes propiedades para funciones de $\mathbb{R}^n$ en $\mathbb{R}$. Por esta razón, no las probaremos y sólo las dejaremos formuladas en la siguiente

Proposición 5.11 *Sean $f, g : U \subset \mathbb{R}^n \to \mathbb{R}^m$, $\tilde{c} : U \subset \mathbb{R} \to \mathbb{R}^m$, $\hat{x}_0 \in U$ y $c \in \mathbb{R}$. Si f, g y $\tilde{c}$ son derivables en $\hat{x}_0$, entonces:*

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. *cf es derivable en $\hat{x}_0$ y además*
$$D(cf)(\hat{x}_0) = cDf(\hat{x}_0)$$

3. *en general, la función $\tilde{c}f$ es derivable en $\hat{x}_0$ y además*
$$D(\tilde{c}f)(\hat{x}_0) = \tilde{c}(\hat{x}_0)Df(\hat{x}_0) + \begin{bmatrix} f_1(\hat{x}_0) & \cdots & f_m(\hat{x}_0) \end{bmatrix}^t D\tilde{c}(\hat{x}_0)$$

4. *$f \cdot g$ es derivable en $\hat{x}_0$ y además*
$$D(f \cdot g)(\hat{x}_0) = \begin{bmatrix} f_1(\hat{x}_0) & \cdots & f_m(\hat{x}_0) \end{bmatrix} Dg(\hat{x}_0) + \begin{bmatrix} g_1(\hat{x}_0) & \cdots & g_m(\hat{x}_0) \end{bmatrix} Df(\hat{x}_0),$$
en donde $f_1, \ldots, f_m$ y $g_1, \ldots, g_m$ son funciones coordenadas de f y g, respectivamente.

5.2.1 Breve comentario sobre funciones coordenadas

Si el lector revisa con cuidado los conceptos de límite, continuidad, continuidad uniforme y derivabilidad de una función f de $\mathbb{R}^n$ en $\mathbb{R}^m$, todos ellos han sido definidos sin tener que recurrir a algún sistema coordenado, del dominio o del contradominio de la función. Los sistemas coordenados son necesarios e importantes para realizar cálculos específicos y por esta razón hemos probado resultados que expresan estos conceptos en términos de funciones coordenadas.

Así como en el capítulo anterior hicimos énfasis en que la variable $\hat{x} \in \mathbb{R}^n$ de nuestra función f se puede describir en términos de diferentes coordenadas, dependiendo de la base ortonormal de $\mathbb{R}^n$ en la que lo estemos representando, de igual manera las funciones coordenadas asociadas a f también dependen de la base ortonormal de $\mathbb{R}^m$ con la que se esté trabajando. En efecto, si $f_1, \ldots, f_m$ son las funciones coordenadas de f para una cierta base ortonormal $\{\tilde{e}_1, \ldots, \tilde{e}_m\}$ de $\mathbb{R}^m$ (lo que significa que $f_i(\hat{x})$ es la $j-$ésima coordenada del vector $f(\hat{x})$ en esta base), es decir que

$$f(\hat{x}) = f_1(\hat{x})\hat{e}_1 + \cdots + f_m(\hat{x})\hat{e}_m$$

y $\{\tilde{e}'_1, \ldots, \tilde{e}'_m\}$ es otra base ortonormal de $\mathbb{R}^m$, las funciones coordenadas de f en esta nueva base, que denotaremos por $\tilde{f}_1, \ldots, \tilde{f}_m$, se podrán obtener a partir de las funciones coordenadas $f_1, \ldots, f_m$ a través de una matriz ortonormal $\tilde{M} \in M_{m\times m}(\mathbb{R})$, matriz que se obtiene de la misma forma en que obtuvimos la matriz de cambio de coordenadas M de la identidad 4.2 del capítulo 4.

Esto es, si

$$\begin{aligned}\tilde{e}_j &= \left(b_1^{(j)}, \ldots, b_m^{(j)}\right) \\ &= b_1^{(j)}\tilde{e}'_1 + \cdots + b_m^{(j)}\tilde{e}'_m\end{aligned}$$

para cada $j \in \{1, \ldots, m\}$, entonces

$$\begin{aligned}f(\hat{x}) &= f_1(\hat{x})\tilde{e}_1 + \cdots + f_m(\hat{x})\tilde{e}_m \\ &= f_1(\hat{x})\left(b_1^{(1)}\tilde{e}'_1 + \cdots + b_m^{(1)}\tilde{e}'_m\right) + \\ &\vdots \\ &+ f_m(\hat{x})\left(b_1^{(m)}\tilde{e}'_1 + \cdots + b_m^{(m)}\tilde{e}'_m\right) \\ &= \left(b_1^{(1)}, \ldots, b_1^{(m)}\right) \cdot (f_1(\hat{x}), \ldots, f_m(\hat{x}))\tilde{e}'_1 + \\ &\vdots \\ &+ \left(b_m^{(1)}, \ldots, b_m^{(m)}\right) \cdot (f_1(\hat{x}), \ldots, f_m(\hat{x}))\tilde{e}'_m,\end{aligned}$$

de donde concluimos que la $j-$ésima coordenada del vector $f(\hat{x})$ en la base $\{\tilde{e}'_1, \ldots, \tilde{e}'_m\}$, que en el párrafo anterior dijimos que denotaríamos por $\tilde{f}_j(\hat{x})$, estará dada por

$$\tilde{f}_j(\hat{x}) = b_j^{(1)} f_1(\hat{x}) + \cdots + b_j^{(m)} f_m(\hat{x}) \tag{5.3}$$

para cada $j \in \{1, \ldots, m\}$.

Equivalentemente, pero escrito en forma matricial, se tiene que

$$\left[\begin{array}{ccc} \tilde{f}_1(\hat{x}) & \cdots & \tilde{f}_m(\hat{x}) \end{array}\right] = \left[\begin{array}{ccc} f_1(\hat{x}) & \cdots & f_m(\hat{x}) \end{array}\right] \left[\begin{array}{ccc} b_1^{(1)} & \cdots & b_m^{(1)} \\ \vdots & \ddots & \vdots \\ b_1^{(m)} & \cdots & b_m^{(m)} \end{array}\right],$$

en donde la matriz

$$\tilde{M} = \left[\begin{array}{ccc} b_1^{(1)} & \cdots & b_m^{(1)} \\ \vdots & \ddots & \vdots \\ b_1^{(m)} & \cdots & b_m^{(m)} \end{array}\right]$$

es una matriz ortonormal.

Lo relevante de la identidad 5.3 es que establece la forma sencilla en que se relacionan las diferentes funciones coordenadas que expresan (en diferentes bases ortonormales) a la misma función. Es decir, si una función f de $\mathbb{R}^n$ en $\mathbb{R}^m$ tiene funciones coordenadas $f_1, \ldots, f_m$ en una cierta base ortonormal $\{\tilde{e}_1, \ldots, \tilde{e}_m\}$ de $\mathbb{R}^m$, y tomamos otra base ortonormal $\{\tilde{e}'_1, \ldots, \tilde{e}'_m\}$ de este mismo conjunto, sus funciones coordenadas $\tilde{f}_1, \ldots, \tilde{f}_m$ en esta nueva base se obtienen como una combinación lineal de las primeras. Esta relación resulta de particular importancia para todos los conceptos que mencionamos al inicio de esta subsección, pues confirma el hecho de que todos ellos son independientes del sistema coordenado de $\mathbb{R}^m$ que se esté usando para representar a f.

Para el caso del concepto de derivada, la identidad 5.3 resulta particularmente importante, pues si tenemos una expresión para la derivada $Df(\hat{x}_0)$ en términos de unas funciones coordenadas $f_1, \ldots, f_m$, de esta identidad podemos deducir cuál será una expresión de esta derivada, pero ahora en términos de las funciones coordenadas $\tilde{f}_1, \ldots, \tilde{f}_m$. En efecto, si escribimos a $Df(\hat{x}_0)$ como $D\left(\tilde{f}_1, \ldots, \tilde{f}_m\right)(\hat{x}_0)$ para enfatizar que estamos escribiendo a f en términos de las funciones coordenadas $\tilde{f}_1, \ldots, \tilde{f}_m$, de acuerdo con la identidad 5.2 se debe tener que

$$\begin{aligned}
Df(\hat{x}_0) &= D\left(\tilde{f}_1, \ldots, \tilde{f}_m\right)(\hat{x}_0) \\
&= \begin{bmatrix} \frac{\partial \tilde{f}_1}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial \tilde{f}_1}{\partial x_n}(\hat{x}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial \tilde{f}_m}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial \tilde{f}_m}{\partial x_n}(\hat{x}_0) \end{bmatrix},
\end{aligned}$$

de tal forma que, de la identidad 5.3 y los incisos 1. y 2. de la proposición 4.11 del capítulo 4, concluimos que, para cada $i \in \{1, \ldots, n\}$ y cada $j \in \{1, \ldots, m\}$, se tiene que

$$\begin{aligned}
\frac{\partial \tilde{f}_j}{\partial x_i}(\hat{x}_0) &= b_j^{(1)} \frac{f_1}{\partial x_i}(\hat{x}_0) + \cdots + b_j^{(m)} \frac{f_m}{\partial x_i}(\hat{x}_0) \\
&= \left(b_j^{(1)}, \ldots, b_j^{(m)}\right) \cdot \left(\frac{f_1}{\partial x_i}(\hat{x}_0), \ldots, \frac{f_m}{\partial x_i}(\hat{x}_0)\right),
\end{aligned}$$

de donde obtenemos que

$$\begin{aligned}
D\left(\tilde{f}_1, \ldots, \tilde{f}_m\right)(\hat{x}_0) &= \begin{bmatrix} \frac{\partial \tilde{f}_1}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial \tilde{f}_1}{\partial x_n}(\hat{x}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial \tilde{f}_m}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial \tilde{f}_m}{\partial x_n}(\hat{x}_0) \end{bmatrix} \\
&= \begin{bmatrix} b_1^{(1)} & \cdots & b_1^{(m)} \\ \vdots & \ddots & \vdots \\ b_m^{(1)} & \cdots & b_m^{(m)} \end{bmatrix} \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial f_1}{\partial x_n}(\hat{x}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial f_m}{\partial x_n}(\hat{x}_0) \end{bmatrix} \\
&= \tilde{M}^t D\left(f_1, \ldots, f_m\right)(\hat{x}_0).
\end{aligned}$$

5.3 La regla de la cadena

Para terminar de revisar las propiedades más relevantes de la derivada de una función de $\mathbb{R}^n$ en $\mathbb{R}^m$, formularemos un resultado muy importante en el cual se establecen las condiciones para asegurar la derivabilidad de la composición de una función de $\mathbb{R}^n$ en $\mathbb{R}^m$ con una de $\mathbb{R}^m$ en $\mathbb{R}^k$, y la fórmula que nos permite calcular la derivada de esta composición. Esta será la versión más general de *la regla de la cadena* que daremos en este texto, y para probarla podríamos hacer uso de la proposición 5.6 de la siguiente manera: si $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ y $g : V \subset \mathbb{R}^m \to \mathbb{R}^k$, con $g_1, \ldots, g_m$ funciones coordenadas de g, son tales que $f(U) \subset V$, se tiene que la composición $g \circ f$ está bien definida y que las funciones $g_i \circ f$, para $i \in \{1, \ldots, m\}$, son funciones coordenadas de $g \circ f$.

De esta forma, y justo por la proposición 5.6, para probar la derivabilidad de $g \circ f$ bastaría con probar la derivabilidad de cada $g_i \circ f$. Sin embargo, la prueba de la regla de la cadena para la composición $g \circ f$, suponiendo que g es una función de $\mathbb{R}^m$ en $\mathbb{R}$, es casi igual a la prueba que se puede hacer suponiendo que g es una función de $\mathbb{R}^m$ en $\mathbb{R}^k$, en virtud de lo cual optaremos por esta última.

Proposición 5.12 (Regla de la cadena) *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$, $g : V \subset \mathbb{R}^m \to \mathbb{R}^k$ y $\hat{x}_0 \in U$ tales que $f(U) \subset V$. Si f es derivable en $\hat{x}_0$ y g es derivable en $\hat{y}_0 = f(\hat{x}_0)$, entonces $g \circ f$ es derivable en $\hat{x}_0$ y además se tiene que*

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

O equivalentemente, si pensamos a las derivadas $D(g \circ f)(\hat{x}_0)$, $Dg(f(\hat{x}_0))$ y $Df(\hat{x}_0)$ como matrices, escribimos que

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

Demostración. La idea detrás de la prueba de este resultado es completamente análoga a la seguida en la prueba de la proposicón 4.30 del capítulo 4, en virtud de lo cual haremos con menos detalles los pasos que seguiremos en este caso.

Como hicimos en esa proposición, será necesario introducir una función auxiliar

$$\varphi : U \subset \mathbb{R}^n \to \mathbb{R}^k$$

definida de la siguiente forma:

$$\varphi(\hat{x}) = \begin{cases} \frac{g(f(\hat{x})) - g(f(\hat{x}_0)) - Dg(f(\hat{x}_0))(f(\hat{x}) - f(\hat{x}_0))}{\|f(\hat{x}) - f(\hat{x}_0)\|} & \text{si } f(\hat{x}) - f(\hat{x}_0) \neq \hat{0} \\ \\ \hat{0} & \text{si } f(\hat{x}) - f(\hat{x}_0) = \hat{0} \end{cases}$$

Como el lector podrá verificar muy fácilmente, se tiene que

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

para toda $\hat{x} \in U$, $\hat{x} \neq \hat{x}_0$.

Ahora, del hecho de que f es derivable en $\hat{x}_0$, y por los problemas 5 y 4 de este capítulo, obtenemos las dos siguientes conclusiones: una, que la función φ es continua en $\hat{x}_0$, y dos, que la expresión

$$\frac{\|f(\hat{x}) - f(\hat{x}_0)\|}{\|\hat{x} - \hat{x}_0\|}$$

está acotada en una vecindad (agujerada) de $\hat{x}_0$, de tal forma que

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

Por otra parte, dado que f es derivable en $\hat{x}_0$, sabemos que

$$\frac{f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)\left(\hat{x} - \hat{x}_0\right)}{\|\hat{x} - \hat{x}_0\|} = \hat{0}$$

y como toda función lineal es continua, tenemos que

$$\lim_{\hat{x} \to \hat{x}_0} Dg(f(\hat{x}_0))\left(\frac{f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)\left(\hat{x} - \hat{x}_0\right)}{\|\hat{x} - \hat{x}_0\|}\right)$$

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

De esta forma,

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

lo que prueba que $g\circ f$ es derivable en $\hat{x}_0$ y que su derivada es $Dg(f(\hat{x}_0))\circ Df(\hat{x}_0)$. ■

5.3.1 Cambio de coordenadas y regla de la cadena

Desde el capítulo 1, a lo largo de todo este texto se ha venido insistiendo en ver a $\mathbb{R}^n$ como el "representante por excelencia" (o "prototipo") de los espacios vectoriales de dimensión n sobre los números reales. Por esta misma razón, se ha enfatizado la posibilidad de que los elementos de $\mathbb{R}^n$ se pueden describir por medio de diferentes sistemas coordenados (lo que no deja de ser un poco extraño, pues los elementos de $\mathbb{R}^n$ están definidos en términos de n-adas), incluyendo sistemas coordenados no cartesianos, como lo son los sistemas coordenados cilíndrico y esférico (en $\mathbb{R}^3$), y polar (en $\mathbb{R}^2$).

Con base en lo anterior, se ha puesto particular interés en mostrar que el concepto de derivada es independiente de estos sistemas coordenados, pero al mismo tiempo se ha visto cómo se expresa en éstos, así como la manera de "pasar" de una expresión a otra cuando "pasamos" de un sistema coordenado a otro (ejemplo de esto son las identidades 4.18, 4.19, 4.33 y 4.34). Lo que ahora deseamos mostrar es que estas mismas "identidades" se pueden obtener usando la regla de la cadena que acabamos de probar. Por ejemplo, supongamos que $\{\hat{e}_1,\ldots,\hat{e}_n\}$ y $\{\hat{e}'_1,\ldots,\hat{e}'_n\}$ son dos bases ortonormales de $\mathbb{R}^n$, y que $x_1,\ldots,x_n$ y $x'_1,\ldots,x'_n$ denotan las coordenadas de un mismo punto $\hat{x}\in\mathbb{R}^n$ en cada una de estas bases, respectivamente. Si

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

para cada $i\in\{1,\ldots,n\}$, en el capítulo 4 dedujimos que si $(x'_1,\ldots,x'_n)$ son las coordenadas de un punto $\hat{x}\in\mathbb{R}^n$ en la base $\{\hat{e}'_1,\ldots,\hat{e}'_n\}$, entonces las coordenadas $(x_1,\ldots,x_n)$ del mismo punto en la base $\{\hat{e}_1,\ldots,\hat{e}_n\}$ están dadas por

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

De esta forma, si definimos $g:\mathbb{R}^n\to\mathbb{R}^n$ como

$$\begin{aligned}
g(\hat{x}) &= g(x'_1,\ldots,x'_n)\\
&= \left(x'_1a_1^{(1)}+\cdots+x'_na_1^{(n)},\ldots,x'_1a_n^{(1)}+\cdots+x'_na_n^{(n)}\right),
\end{aligned}$$

se tendrá que g no es más que la función identidad, sólo que estamos expresando a los elementos de su dominio en la base $\{\hat{e}'_1,\ldots,\hat{e}'_n\}$, mientras que a los de su contradominio los estamos expresando en la base $\{\hat{e}_1,\ldots,\hat{e}_n\}$. Este tipo de funciones son conocidas como *funciones de cambio de coordenadas.*

Dado que g es la función identidad, sin duda g es derivable en todos los puntos de $\mathbb{R}^n$. Por otra parte, de acuerdo con la identidad 5.1, tenemos que la matriz jacobiana de g (es decir $Dg(\hat{x})$), expresada con respecto a las bases $\{\hat{e}_1, \ldots, \hat{e}_n\}$ y $\{\hat{e}'_1, \ldots, \hat{e}'_n\}$, está dada por

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

para toda $\hat{x} \in \mathbb{R}^n$, que coincide con la matriz transpuesta de la matriz de cambio de coordenadas que obtuvimos anteriormente.

Ahora, si $f : U \subset \mathbb{R}^n \to \mathbb{R}$ es una función derivable que está expresada en términos de las coordenadas $x_1, \ldots, x_n$, dado que g es la función identidad, que la composición $f \circ g$ está bien definida y que en sentido estricto $(f \circ g)(\hat{x}) = f(\hat{x})$ para toda $\hat{x} \in U$ (identidad que no es equivalente a escribir que $f(g(x'_1, \ldots, x'_n)) = f(x'_1, \ldots, x'_n)$, puesto que f no depende de las coordenadas $x'_1, \ldots, x'_n$), podemos decir que $f \circ g$ es la misma función f, pero expresada en términos de las coordenadas $x'_1, \ldots, x'_n$. Este hecho es lo que justifica que nos podamos tomar la libertad de escribir que

$$\frac{\partial\,(f \circ g)}{\partial x'_i}(\hat{x}) = \frac{\partial f}{\partial x'_i}(\hat{x}) \tag{5.4}$$

para cada $i \in \{1, \ldots, n\}$ y cada $\hat{x} \in U$, aún cuando f no dependa directamente de las coordenadas $x'_1, \ldots, x'_n$.

Por otra parte, de acuerdo con la identidad (matricial) de la regla de la cadena (proposición 5.12), se tiene que

$$\begin{aligned} D(f \circ g)(\hat{x}) &= Df(g(\hat{x}))Dg(\hat{x}) \\ &= Df(\hat{x})Dg(\hat{x}). \end{aligned}$$

Es decir, que

$$\begin{aligned} \begin{bmatrix} \frac{\partial f}{\partial x'_1}(\hat{x}) & \cdots & \frac{\partial f}{\partial x'_n}(\hat{x}) \end{bmatrix} &= \begin{bmatrix} \frac{\partial (f\circ g)}{\partial x'_1}(\hat{x}) & \cdots & \frac{\partial (f\circ g)}{\partial x'_n}(\hat{x}) \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial f}{\partial x_1}(\hat{x}) & \cdots & \frac{\partial f}{\partial x_n}(\hat{x}) \end{bmatrix} \begin{bmatrix} a_1^{(1)} & \cdots & a_1^{(n)} \\ \vdots & \ddots & \vdots \\ a_n^{(1)} & \cdots & a_n^{(n)} \end{bmatrix} \end{aligned} \tag{5.5}$$

o equivalentemente, que para cada $i \in \{1, \ldots, n\}$ se cumple que

$$\begin{aligned} \frac{\partial f}{\partial x'_i}(\hat{x}) &= a_1^{(i)} \frac{\partial f}{\partial x_1}(\hat{x}) + \cdots + a_n^{(i)} \frac{\partial f}{\partial x_n}(\hat{x}) \\ &= \sum_{j=1}^{n} a_j^{(i)} \frac{\partial f}{\partial x_j}(\hat{x}). \end{aligned} \tag{5.6}$$

Como seguramente el lector reconocerá, la identidad 5.5 es la misma que obtuvimos en el capítulo 4 (identidad 4.18).

Observación 5.13 *Es importante insistir en que las identidades 5.4, 5.5 y 5.6 contienen un cierto abuso de notación. En estricto sentido, no podemos hablar de las derivadas parciales de f con respecto a las variables $x'_1, \ldots, x'_n$ puesto que supusimos que esta función está expresada en términos (o sólo depende) de las variables $x_1, \ldots, x_n$. Si, por ejemplo, deseamos poner a la identidad 5.6 en términos de todas estas coordenadas, lo correcto es escribir que*

$$\begin{aligned} \frac{\partial\,(f \circ g)}{\partial x'_i}(x'_1, \ldots, x'_n) &= a_1^{(i)} \frac{\partial f}{\partial x_1}(g\,(x'_1, \ldots, x'_n)) + \cdots + a_n^{(i)} \frac{\partial f}{\partial x_n}(g\,(x'_1, \ldots, x'_n)) \\ &= a_1^{(i)} \left(\frac{\partial f}{\partial x_1} \circ g\right)(x'_1, \ldots, x'_n) + \cdots + a_n^{(i)} \left(\frac{\partial f}{\partial x_n} \circ g\right)(x'_1, \ldots, x'_n). \end{aligned}$$

Es muy importante tener presente lo anterior, sobre todo si se desea calcular derivadas parciales de orden superior de la función $f \circ g$.

Para concluir esta subsección, lo que ahora queremos mostrar es que se puede proceder de forma análoga al caso anterior y obtener identidades equivalentes, aún cuando los sistemas coordenados involucrados no sean cartesianos.

Por ejemplo, consideremos la función $g : \mathbb{R}^2 \to \mathbb{R}^2$ dada por $g(u,v) = (u\cos(v), u\,\mathrm{sen}(v))$, la que sin duda está definida en términos de las ecuaciones de cambio de coordenadas polares a coordenadas cartesianas (para puntos en $\mathbb{R}^2$) y que por esta misma razón podemos decir que g es la *función de cambio de coordenadas polares a coordenadas cartesianas*. De esta forma, y como en el primer caso que tratamos en esta subsección, si sólo vemos a g como función de $\mathbb{R}^2$ en $\mathbb{R}^2$, podemos afirmar que g no es más que la función identidad (es decir, si no usamos coordenadas de ningún tipo, podemos escribir que $g(\hat{x}) = \hat{x}$ para toda $\hat{x} \in \mathbb{R}^2$).

A diferencia del caso cartesiano, esta función g es tal que su derivada no es constante, además de que la matriz que la representa (tomando la base canónica de $\mathbb{R}^2$, tanto en el dominio como en el contradominio) no es una matriz ortonormal para toda pareja (u, v). Sin embargo, la aplicación de la regla de la cadena nos permitirá deducir algunas identidades que nos resultarán muy familiares.

En efecto, si nuevamente $f : \mathbb{R}^2 \to \mathbb{R}$ es una función derivable que está expresada en términos de las coordenadas cartesianas x y y de un punto $\hat{x} \in \mathbb{R}^2$, por la regla de la cadena sabemos que

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

o equivalentemente, que

$$\nabla(f \circ g)(\hat{x}) = \nabla f(g(\hat{x}))Dg(\hat{x}).$$

Si ahora, como hemos venido haciendo, escribimos que

$$\begin{aligned} Dg(\hat{x}) &= \begin{bmatrix} \frac{\partial}{\partial u}(u\cos(v)) & \frac{\partial}{\partial v}(u\cos(v)) \\ \frac{\partial}{\partial u}(u\,\mathrm{sen}(v)) & \frac{\partial}{\partial v}(u\,\mathrm{sen}(v)) \end{bmatrix} \\ &= \begin{bmatrix} \cos(v) & -u\,\mathrm{sen}(v) \\ \mathrm{sen}(v) & u\cos(v) \end{bmatrix} \end{aligned}$$

y

$$\begin{aligned} \nabla(f \circ g)(\hat{x}) &= \begin{bmatrix} \frac{\partial(f\circ g)}{\partial u}(\hat{x}) & \frac{\partial(f\circ g)}{\partial v}(\hat{x}) \end{bmatrix} \\ \nabla f(g(\hat{x})) &= \begin{bmatrix} \frac{\partial f}{\partial x}(g(\hat{x})) & \frac{\partial f}{\partial y}(g(\hat{x})) \end{bmatrix}, \end{aligned}$$

concluimos que

$$\begin{aligned} \begin{bmatrix} \frac{\partial(f\circ g)}{\partial u}(\hat{x}) & \frac{\partial(f\circ g)}{\partial v}(\hat{x}) \end{bmatrix} &= \nabla(f \circ g)(\hat{x}) \\ &= \nabla f(g(\hat{x}))Dg(\hat{x}) \\ &= \begin{bmatrix} \frac{\partial f}{\partial x}(g(\hat{x})) & \frac{\partial f}{\partial y}(g(\hat{x})) \end{bmatrix} \begin{bmatrix} \cos(v) & -u\,\mathrm{sen}(v) \\ \mathrm{sen}(v) & u\cos(v) \end{bmatrix} \end{aligned}$$

o equivalentemente que

$$\begin{aligned} \frac{\partial(f\circ g)}{\partial u}(\hat{x}) &= \cos(v)\frac{\partial f}{\partial x}(g(\hat{x})) + \mathrm{sen}(v)\frac{\partial f}{\partial y}(g(\hat{x})) \\ \frac{\partial(f\circ g)}{\partial v}(\hat{x}) &= -u\,\mathrm{sen}(v)\frac{\partial f}{\partial x}(g(\hat{x})) + u\cos(v)\frac{\partial f}{\partial y}(g(\hat{x})) \end{aligned}$$

Si como dijimos antes no usamos algún tipo de coordenadas para representar al punto $\hat{x}$, entonces podemos asumir que $g(\hat{x}) = \hat{x}$ y así tomarnos la libertad de escribir que $f \circ g = f$. Haciendo esto, las identidades anteriores tomarían la forma

$$\begin{aligned} \frac{\partial f}{\partial u}(\hat{x}) &= \cos(v)\frac{\partial f}{\partial x}(\hat{x}) + \mathrm{sen}(v)\frac{\partial f}{\partial y}(\hat{x}) \\ \frac{\partial f}{\partial v}(\hat{x}) &= -u\,\mathrm{sen}(v)\frac{\partial f}{\partial x}(\hat{x}) + u\cos(v)\frac{\partial f}{\partial y}(\hat{x}) \end{aligned}$$

que como el lector podrá verficar fácilmente (cambiando u por ρ_0 y v por θ_0) coinciden con las identidades 4.35 que dedujimos en el capítulo 4.

Aun corriendo el riesgo de parecer repetitivos, es importante insistir en que las identidades anteriores contienen un cierto abuso de notación. Si se quieren escribir en términos de las coordenadas "polares" (u, v) del punto $\hat{x}$, lo correcto será escribir que

$$\begin{aligned}\frac{\partial (f \circ g)}{\partial u}(u, v) &= \cos(v)\frac{\partial f}{\partial x}(g(u, v)) + \operatorname{sen}(v)\frac{\partial f}{\partial y}(g(u, v))\\ &= \cos(v)\left(\frac{\partial f}{\partial x} \circ g\right)(u, v) + \operatorname{sen}(v)\left(\frac{\partial f}{\partial y} \circ g\right)(u, v)\end{aligned}$$

y

$$\begin{aligned}\frac{\partial (f \circ g)}{\partial v}(u, v) &= -\rho\operatorname{sen}(v)\frac{\partial f}{\partial x}(g(u, v)) + \rho\cos(v)\frac{\partial f}{\partial y}(g(u, v))\\ &= -u\operatorname{sen}(v)\left(\frac{\partial f}{\partial x} \circ g\right)(u, v) + \rho\cos(v)\left(\frac{\partial f}{\partial y} \circ g\right)(u, v).\end{aligned}$$

Como ya se mencionó antes, estas identidades son las que hay que considerar si se desea calcular las derivadas parciales de orden superior (con respecto a las variables u o v) de la función $f \circ g$.

5.4 El teorema de la función implícita

Una vez que hemos desarrollado las herramientas básicas relacionadas con el concepto de derivada para funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$, estamos en condiciones de abordar uno de los teoremas más importantes del cálculo diferencial de varias variables: el Teorema de la Función Implícita.

5.4.1 El caso lineal

Con el fin de motivar y deducir el contenido de este teorema, empezaremos por analizar algunas propiedades y características de las soluciones de los sistemas de ecuaciones lineales en $\mathbb{R}^n$ que tienen la particularidad de tener menos ecuaciones que incógnitas, es decir, sistemas de la forma

$$\begin{aligned}a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n + b_1 &= 0\\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n + b_2 &= 0\\ &\vdots\\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n + b_m &= 0,\end{aligned} \tag{5.7}$$

en donde $m < n$.

Sin duda el caso más sencillo de este tipo de sistemas de ecuaciones lo tenemos en $\mathbb{R}^2$, y se trata de un sistema que consta de una sola ecuación de la forma

$$ax + by + c = 0.$$

Como seguramente recordará el lector, si $a^2 + b^2 > 0$ la ecuación anterior representa una recta en el plano, y si de obtener sus soluciones se trata, es fácil probar que, si $a \neq 0$, entonces éstas son de la forma

$$\left(\frac{-by - c}{a}, y\right)$$

para cualquier $y \in \mathbb{R}$, y si $b \neq 0$, entonces todas las soluciones las podemos escribir como las parejas

$$\left(x, \frac{-ax - c}{b}\right).$$

Es decir, si el coeficiente de la incógnita x es diferente de cero ($a \neq 0$), esta incógnita se puede poner en función de la otra incógnita y ($x = h(y) = (-by - c)/a$) y todas las soluciones de la ecuación son de la forma $(h(y), y)$, con $y \in \mathbb{R}$.

Sucede lo análogo si el coeficiente de la incógnita y es diferente de cero ($b \neq 0$); en este caso y se puede poner en función de x ($y = h(x) = (-ax - c)/b$) y todas las soluciones de la ecuación son de la forma $(x, h(x))$, con $x \in \mathbb{R}$.

Para sistemas de ecuaciones en $\mathbb{R}^3$, la situación empieza a ponerse un poco más interesante. El primer caso que se puede tener es el de una sóla ecuación de la forma

$$ax + by + cz + d = 0 \tag{5.8}$$

(la cual geométricamente representa a un plano si $a^2 + b^2 + c^2 > 0$), y la obtención de sus soluciones es muy similar al caso anterior. En efecto, nótese que si el coeficiente de la incógnita x es diferente de cero ($a \neq 0$), entonces podemos poner a x en función de las incógnitas y y z como

$$\begin{aligned} x &= h(y, z) \\ &= \frac{-(by + cz + d)}{a} \end{aligned}$$

y todas las ternas de la forma

$$(h(y, z), y, z) = \left(\frac{-(by + cz + d)}{a}, y, z \right),$$

con $(y, z) \in \mathbb{R}^2$, son las soluciones de la ecuación 5.8. Como el lector puede constatar fácilmente, se tiene una situación análoga si $b \neq 0$ o $c \neq 0$.

El caso que se empieza a poner aún más interesante es cuando tenemos un sistema de dos ecuaciones de la forma

$$\begin{aligned} a_{11}x + a_{12}y + a_{13}z + b_1 &= 0 \\ a_{21}x + a_{22}y + a_{23}z + b_2 &= 0. \end{aligned} \tag{5.9}$$

Lo primero que haremos, será recordar que en un sistema de este tipo se pueden presentar las siguientes situaciones:

1. que una de las ecuaciones sea "múltiplo" de la otra, es decir, que una de ellas se puede obtener de la otra multiplicando por un escalar, lo que significa que en realidad nuestro sistema sólo consta de una ecuación, caso que ya analizamos

2. que los planos que representan cada una de las ecuaciones, sean dos planos distintos pero paralelos, es decir, que los vectores normales a cada uno de ellos ((a_{11}, a_{12}, a_{13}) y (a_{21}, a_{22}, a_{23})), sean paralelos; en este caso no existen soluciones y no hay nada más que se pueda hacer

3. la tercera y última posibilidad es justo cuando los vectores (a_{11}, a_{12}, a_{13}) y (a_{21}, a_{22}, a_{23}) no son paralelos, y por lo tanto el conjunto de soluciones del sistema está formado por todos los puntos de la recta en la que se intersectan ambos planos.

Dado que los dos primeros casos ya están resueltos, analizaremos el tercero. Del hecho de que los vectores (a_{11}, a_{12}, a_{13}) y (a_{21}, a_{22}, a_{23}) no sean paralelos podemos concluir que el producto cruz de éstos no es el vector $\hat{0}$, es decir que

$$\begin{aligned} (a_{11}, a_{12}, a_{13}) \times (a_{21}, a_{22}, a_{23}) &= (a_{12}a_{23} - a_{13}a_{22}, a_{13}a_{21} - a_{11}a_{23}, a_{11}a_{22} - a_{12}a_{21}) \\ &\neq \hat{0} \end{aligned} \tag{5.10}$$

y por lo tanto tenemos tres posibilidades: $a_{12}a_{23} - a_{13}a_{22} \neq 0$, $a_{13}a_{21} - a_{11}a_{23} \neq 0$ o $a_{11}a_{22} - a_{12}a_{21} \neq 0$, posibilidades que a continuación nos disponemos a analizar.

Lo primero que habría que destacar es que cada una de las coordenadas del vector dado por 5.10, resulta ser el determinante de la matriz formada por los coeficientes de algunas de las incógnitas de nuestras ecuaciones. En efecto, nótese que $a_{12}a_{23} - a_{13}a_{22}$ es el determinante de la matriz que se obtiene al considerar sólo los coeficientes (de ambas ecuaciones del sistema dado por 5.9) de las incógnitas y y z; $a_{13}a_{21} - a_{11}a_{23}$ es el determinante de la matriz que se obtiene al considerar sólo los coeficientes de las incógnitas x y z, y $a_{11}a_{22} - a_{12}a_{21}$ es el determinante de la matriz que se obtiene al considerar solo los coeficientes de las incógnitas x y y.

Con base en la observación anterior, si por ejemplo se tiene que $a_{12}a_{23} - a_{13}a_{22} \neq 0$, todo parece indicar que lo más adecuado sería reescribir el sistema de ecuaciones 5.9 en la forma

$$\begin{aligned} a_{12}y + a_{13}z &= -a_{11}x - b_1 \\ a_{22}y + a_{23}z &= -a_{21}x - b_2, \end{aligned}$$

o mejor aún, en la forma matricial

$$\begin{bmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} -(a_{11}x + b_1) \\ -(a_{21}x + b_2) \end{bmatrix}.$$

De manera que, si escribimos

$$M = \begin{bmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{bmatrix},$$

como $\det(M) = a_{12}a_{23} - a_{13}a_{22} \neq 0$, sabemos que M es invertible y por lo tanto tendremos que las incógnitas y y z se pueden poner en función de la incógnita x, como

$$\begin{aligned} \begin{bmatrix} y \\ z \end{bmatrix} &= M^{-1} \begin{bmatrix} -(a_{11}x + b_1) \\ -(a_{21}x + b_2) \end{bmatrix} \\ &= \begin{bmatrix} h_1(x) \\ h_2(x) \end{bmatrix} \end{aligned}$$

y que las soluciones del sistema 5.9 estarían dadas por las ternas

$$(x, h_1(x), h_2(x)),$$

con $x \in \mathbb{R}$.

Tomando en consideración el análisis anterior, seguramente el lector estará de acuerdo en que, si ahora lo que se tiene es que $a_{13}a_{21} - a_{11}a_{23} \neq 0$, entonces las incógnitas x y z se podrán poner en función de la incógnita y ($x = h_1(y)$ y $z = h_2(y)$) y que las soluciones del sistema estarán dadas por las ternas $(h_1(y), y, h_2(y))$ con $y \in \mathbb{R}$. Y si lo que sucede es que $a_{11}a_{22} - a_{12}a_{21} \neq 0$, entonces las incógnitas x y y se podrán poner en función de la incógnita z ($x = h_1(z)$ y $y = h_2(z)$) y que las soluciones del sistema estarán dadas por las ternas $(h_1(z), h_2(z), z)$, con $z \in \mathbb{R}$.

Sin duda el caso anterior es muy ilustrativo y nos da la pauta para resolver el problema general sobre el cálculo de las soluciones del sistema de ecuaciones dado por 5.7. De esta manera, si por ejemplo se tiene que la matriz de $m \times m$ formada por los coeficientes de las incógnitas $x_1, \ldots, x_m$, dada por

$$M = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mm} \end{bmatrix}$$

tiene determinante distinto de cero (y por lo tanto tiene inversa), entonces el sistema 5.7 escrito en términos de la matriz M toma la forma

$$\begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mm} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} -\left(a_{1(m+1)}x_{m+1} + \cdots + a_{1n}x_n + b_1\right) \\ \vdots \\ -\left(a_{m(m+1)}x_{m+1} + \cdots + a_{mn}x_n + b_m\right) \end{bmatrix}$$

y por lo tanto las incógnitas $x_1, \ldots, x_m$ se podrán poner en función de las incógnitas restantes $x_{m+1}, \ldots, x_n$ como

$$\begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} = M^{-1} \begin{bmatrix} -\left(a_{1(m+1)}x_{m+1} + \cdots + a_{1n}x_n + b_1\right) \\ \vdots \\ -\left(a_{m(m+1)}x_{m+1} + \cdots + a_{mn}x_n + b_m\right) \end{bmatrix}.$$

Por lo tanto, si definimos

$$H = (h_1, \ldots, h_m) : \mathbb{R}^{n-m} \to \mathbb{R}^m$$

como

$$H(x_{m+1}, \ldots, x_n) = M^{-1} \begin{bmatrix} -\left(a_{1(m+1)}x_{m+1} + \cdots + a_{1n}x_n + b_1\right) \\ \vdots \\ -\left(a_{m(m+1)}x_{m+1} + \cdots + a_{mn}x_n + b_m\right) \end{bmatrix},$$

entonces encontramos que existen $h_1, \ldots, h_m : \mathbb{R}^{n-m} \to \mathbb{R}$ tales que

$$x_j = h_j(x_{m+1}, \ldots, x_n)$$

para $j \in \{1, \ldots, m\}$, y que las soluciones del sistema de ecuaciones 5.7 están dadas por las $n-$adas

$$(h_1(x_{m+1}, \ldots, x_n), \ldots, h_m(x_{m+1}, \ldots, x_n), x_{m+1}, \ldots, x_n) \in \mathbb{R}^n$$

con $(x_{m+1}, \ldots, x_n) \in \mathbb{R}^{n-m}$.

En el caso general se tendrá que, si los índices $1 \leq i_1 < i_2 < \cdots < i_m \leq n$, $1 \leq j_{m+1} < j_{m+2} < \cdots < j_n \leq n$, con $\{i_1, \ldots, i_m\} \cap \{j_{m+1}, \ldots, j_n\} = \varnothing$, son tales que la matriz

$$\begin{bmatrix} a_{1i_1} & \cdots & a_{1i_m} \\ \vdots & \ddots & \vdots \\ a_{mi_1} & \cdots & a_{mi_m} \end{bmatrix}$$

es invertible, entonces las incógnitas $x_{i_1}, \ldots, x_{i_m}$ se pueden poner en función de las incógnitas restantes $x_{j_{m+1}}, \ldots, x_{j_n}$. Es decir, que existen $h_{i_1}, \ldots, h_{i_m} : \mathbb{R}^{n-m} \to \mathbb{R}$ tales que

$$x_{i_k} = h_{i_k}(x_{j_{m+1}}, \ldots, x_{j_n})$$

para $k \in \{1, \ldots, m\}$, y que las soluciones del sistema de ecuaciones 5.7 están dadas por las $n-$adas formadas con los n números reales $h_{i_1}(x_{j_{m+1}}, \ldots, x_{j_n})$,..,$h_{i_m}(x_{j_{m+1}}, \ldots, x_{j_n})$,$x_{j_{m+1}}$,..,$x_{j_n}$, en donde $h_{i_k}(x_{j_{m+1}}, \ldots, x_{j_n})$ es la i_k-ésima coordenada y x_{j_l} es la j_l-ésima coordenada, para $k \in \{1, \ldots, m\}$ y $l \in \{m+1, \ldots, n\}$.

5.4.2 El caso no lineal

Una vez que hemos analizado el caso de un sistema de ecuaciones lineales, el siguiente paso será considerar un sistema de ecuaciones, no necesariamente lineales, determinado por m funciones $g_1, \ldots, g_m : \mathbb{R}^n \to \mathbb{R}$ dado por

$$\begin{aligned} g_1(x_1, \ldots, x_n) &= 0 \\ g_2(x_1, \ldots, x_n) &= 0 \\ &\vdots \\ g_m(x_1, \ldots, x_n) &= 0, \end{aligned} \tag{5.11}$$

en donde supondremos que cada función g_i es de clase C^1 en $\mathbb{R}^n$.

Denotaremos por S al conjunto de soluciones de este sistema, es decir

$$S = \{\hat{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n \mid g_1(\hat{x}) = 0, \ldots, g_m(\hat{x}) = 0\}$$

y nuestro objetivo no será encontrar y caracterizar a todos los elemento de S (como hicimos en el caso del sistema de ecuaciones lineales), lo que sin duda es un problema bastante difícil. Nuestro objetivo es algo más "modesto" y consiste en lo siguiente: si tenemos una solución $\hat{x}_0$ del sistema 5.11, es decir que $\hat{x}_0 \in S$, ¿es posible "decir algo" de las soluciones de 5.11 que están "cerca" de $\hat{x}_0$? La respuesta a esta pregunta es justo el teorema de la función implícita, el cual nos dice "algo" sobre el comportamiento de las soluciones del sistema de ecuaciones 5.11 "alrededor" del punto $\hat{x}_0$.

La idea principal detrás del teorema de la función implícita es la siguiente: dado que nuestro objetivo es conocer el comportamiento de las soluciones del sistema de ecuaciones 5.11 "cerca" o "alrededor" del punto $\hat{x}_0$, entonces sustituyamos cada ecuación $g_i(x_1, \dots, x_n) = 0$ por su "mejor aproximación lineal en $\hat{x}_0$", es decir por la ecuación

$$Dg_i(\hat{x}_0)(\hat{x} - \hat{x}_0) + g_i(\hat{x}_0) = Dg_i(\hat{x}_0)(\hat{x} - \hat{x}_0) = 0,$$

o lo que es lo mismo, por la ecuación

$$\frac{\partial g_i}{\partial x_1}(\hat{x}_0)x_1 + \frac{\partial g_i}{\partial x_2}(\hat{x}_0)x_2 + \cdots + \frac{\partial g_i}{\partial x_n}(\hat{x}_0)x_n + b_i = 0,$$

en donde $b_i = -Dg_i(\hat{x}_0)(\hat{x}_0)$, para cada $i \in \{1, \dots, m\}$.

De esta forma, si para el sistema de ecuaciones lineales dado por

$$\begin{aligned}
\frac{\partial g_1}{\partial x_1}(\hat{x}_0)x_1 + \frac{\partial g_1}{\partial x_2}(\hat{x}_0)x_2 + \cdots + \frac{\partial g_1}{\partial x_n}(\hat{x}_0)x_n + b_1 &= 0\\
\frac{\partial g_2}{\partial x_1}(\hat{x}_0)x_1 + \frac{\partial g_2}{\partial x_2}(\hat{x}_0)x_2 + \cdots + \frac{\partial g_2}{\partial x_n}(\hat{x}_0)x_n + b_2 &= 0\\
&\vdots\\
\frac{\partial g_m}{\partial x_1}(\hat{x}_0)x_1 + \frac{\partial g_m}{\partial x_2}(\hat{x}_0)x_2 + \cdots + \frac{\partial g_m}{\partial x_n}(\hat{x}_0)x_n + b_m &= 0
\end{aligned} \tag{5.12}$$

se tiene que la matriz

$$\begin{bmatrix} \frac{\partial g_1}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial g_1}{\partial x_m}(\hat{x}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial g_m}{\partial x_m}(\hat{x}_0) \end{bmatrix} \tag{5.13}$$

es invertible, por lo visto para el caso de los sistemas de ecuaciones lineales, sabemos que las incógnitas $x_1, \dots, x_m$ se pueden poner en función de las incógnitas $x_{m+1}, \dots, x_n$. Es decir que existe $h : \mathbb{R}^{n-m} \to \mathbb{R}^m$ tal que $(x_1, \dots, x_m) = h(x_{m+1}, \dots, x_n)$, y que sus soluciones están dadas por $(h(x_{m+1}, \dots, x_n), x_{m+1}, \dots, x_n)$ para todo $(x_{m+1}, \dots, x_n) \in \mathbb{R}^{n-m}$.

Pues bien, como el sistema de ecuaciones dado por 5.12 se "parece" mucho al sistema de ecuaciones dado por 5.11 "alrededor" o "cerca" de $\hat{x}_0$, el teorema de la función implícita asegura que, "alrededor" o "cerca" de $\hat{x}_0$, las incógnitas $x_1, \dots, x_m$ también se pueden poner en función de las incógnitas $x_{m+1}, \dots, x_n$; es decir, si

$$\hat{x}_0 = \left(x_1^{(0)}, \dots, x_m^{(0)}, x_{m+1}^{(0)}, \dots, x_n^{(0)}\right),$$

este teorema nos asegura que existen $\delta > 0$, $V \subset \mathbb{R}^{n-m}$ un conjunto abierto, y una función (que resultará ser única)

$$h : V \subset \mathbb{R}^{n-m} \to \mathbb{R}^m$$

de clase C^1, tales que $\left(x_{m+1}^{(0)}, \dots, x_n^{(0)}\right) \in V$,

$$h\left(x_{m+1}^{(0)}, \dots, x_n^{(0)}\right) = \left(x_1^{(0)}, \dots, x_m^{(0)}\right)$$

y las $n-$adas $(h(x_{m+1}, \dots, x_n), x_{m+1}, \dots, x_n)$ también son soluciones del sistema 5.11. Es decir, que

$$(h(x_{m+1}, \dots, x_n), x_{m+1}, \dots, x_n) \in B_\delta(\hat{x}_0) \cap S$$

para cada $(x_{m+1}, \dots, x_n) \in V$.

Antes de dar la formulación más precisa del teorema de la función implícita, haremos unas observaciones importantes. La primera de ellas tiene que ver con la elección de la matriz dada por 5.13. Como en el caso del sistema de ecuaciones lineales, si en general los índices $1 \leq i_1 < i_2 < \cdots < i_m \leq n$, $1 \leq j_{m+1} < j_{m+2} < \cdots < j_n \leq n$, con $\{i_1, \ldots, i_m\} \cap \{j_{m+1}, \ldots, j_n\} = \varnothing$, son tales que la matriz

$$\begin{bmatrix} \frac{\partial g_1}{\partial x_{i_1}}(\hat{x}_0) & \cdots & \frac{\partial g_1}{\partial x_{i_m}}(\hat{x}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_{i_1}}(\hat{x}_0) & \cdots & \frac{\partial g_m}{\partial x_{i_m}}(\hat{x}_0) \end{bmatrix}$$

es invertible, entonces lo que el teorema de la función implícita afirma es que "alrededor" de $\hat{x}_0 \in S$ las incógnitas $x_{i_1}, \ldots, x_{i_m}$ se pueden poner en función de las incógnitas restantes $x_{j_{m+1}}, \ldots, x_{j_n}$, es decir que existen $\delta > 0$, $V \subset \mathbb{R}^{n-m}$ un conjunto abierto, y funciones

$$h_{i_1}, \ldots, h_{i_m} : V \subset \mathbb{R}^{n-m} \to \mathbb{R}$$

(que serán únicas) tales que $\left(x_{j_{m+1}}^{(0)}, \ldots, x_{j_n}^{(0)}\right) \in V$,

$$x_{i_k}^{(0)} = h_{i_k}\left(x_{j_{m+1}}^{(0)}, \ldots, x_{j_n}^{(0)}\right)$$

para $k \in \{1, \ldots, m\}$, y que las $n-$adas formadas con los n números reales

$$h_{i_1}(x_{j_{m+1}}, \ldots, x_{j_n}), \ldots, h_{i_m}(x_{j_{m+1}}, \ldots, x_{j_n}), x_{j_{m+1}}, \ldots, x_{j_n},$$

en donde $h_{i_k}(x_{j_{m+1}}, \ldots, x_{j_n})$ es la i_k-ésima coordenada y x_{j_l} es la j_l-ésima coordenada, para $k \in \{1, \ldots, m\}$ y $l \in \{m+1, \ldots, n\}$, pertenecen al conjunto $B_\delta(\hat{x}_0) \cap S$ para cada $\left(x_{j_{m+1}}, \ldots, x_{j_n}\right) \in V$.

Las otras observaciones que haremos en realidad son interpretaciones geométricas del teorema de la función implícita para los casos de $\mathbb{R}^2$ y $\mathbb{R}^3$. El caso más sencillo es el de $\mathbb{R}^2$, en el que sólo se puede tener una restricción de la forma

$$g(x, y) = 0,$$

que no es más que el conjunto de nivel 0 de g ($N_0(g)$).

De acuerdo con lo visto anteriormente, si $\hat{x}_0 = (x_0, y_0) \in \mathbb{R}^2$ es tal que $g(x_0, y_0) = 0$ y $\frac{\partial g}{\partial x}(\hat{x}_0) \neq 0$ (es decir, que la matriz de 1×1 dada por $\left[\frac{\partial g}{\partial x}(\hat{x}_0)\right]$ ¡es invertible!), entonces para puntos $(x, y) \in N_0(g)$ "cercanos" al punto $\hat{x}_0 = (x_0, y_0)$, su coordenada (o variable) x se puede poner en función de su coordenada (o variable) y, es decir que existen $\delta > 0$, $I \subset \mathbb{R}$ abierto, y $h : I \subset \mathbb{R} \to \mathbb{R}$ tales que $y_0 \in I$, $h(y_0) = x_0$ y $(h(y), y) \in B_\delta(\hat{x}_0)) \cap N_0(g)$ para toda $y \in I$. Análogamente, si $\frac{\partial g}{\partial y}(\hat{x}_0) \neq 0$, entonces para puntos de $N_0(g)$ "cercanos" al punto $\hat{x}_0 = (x_0, y_0)$ su coordenada (o variable) y se puede poner en función de la coordenada (o variable) x, es decir que existen $\delta > 0$, $I \subset \mathbb{R}$ abierto, y $h : I \subset \mathbb{R} \to \mathbb{R}$ tales que $x_0 \in I$, $h(x_0) = y_0$ y $(x, h(x)) \in B_\delta(\hat{x}_0) \cap N_0(g)$ para toda $x \in I$.

En términos geométricos, lo anterior significa que si el vector $\nabla g(\hat{x}_0)$ no es horizontal ($\frac{\partial g}{\partial y}(\hat{x}_0) \neq 0$), es decir no es paralelo al eje X, entonces un "sector" de la curva de nivel 0 de g "alrededor" del punto $\hat{x}_0 = (x_0, y_0)$ se puede "ver" como la gráfica de una función de la forma $y = h(x)$ (ver figura 5.4 (a)). Y si el vector $\nabla g(\hat{x}_0)$ no es vertical ($\frac{\partial g}{\partial x}(\hat{x}_0) \neq 0$), es decir no es paralelo al eje Y, entonces un sector de la curva de nivel 0 de g alrededor del punto $\hat{x}_0 = (x_0, y_0)$ se puede ver como la gráfica de una función de la forma $x = h(y)$ (ver figura 5.4 (b)).

Para el caso de $\mathbb{R}^3$, si sólo se tiene una "restricción" de la forma

$$g(x, y, z) = 0$$

y $\hat{x}_0 = (x_0, y_0, z_0) \in \mathbb{R}^3$ es tal que $g(x_0, y_0, z_0) = 0$ y $\frac{\partial g}{\partial x}(\hat{x}_0) \neq 0$ (lo que significa que el vector $\nabla g(\hat{x}_0)$ no está en el plano YZ), entonces para puntos $(x, y, z) \in N_0(g)$ "cercanos" al punto $\hat{x}_0$, la coordenada (o variable) x se puede poner en función de las correspondientes coordenadas (o variables) y y z, es decir que existen $\delta > 0$, $V \subset \mathbb{R}^2$ abierto, y $h : V \subset \mathbb{R}^2 \to \mathbb{R}$ tales que $(y_0, z_0) \in V$, $h(y_0, z_0) = x_0$ y $(h(y, z), y, z) \in B_\delta(\hat{x}_0) \cap N_0(g)$ para toda $(y, z) \in V$.

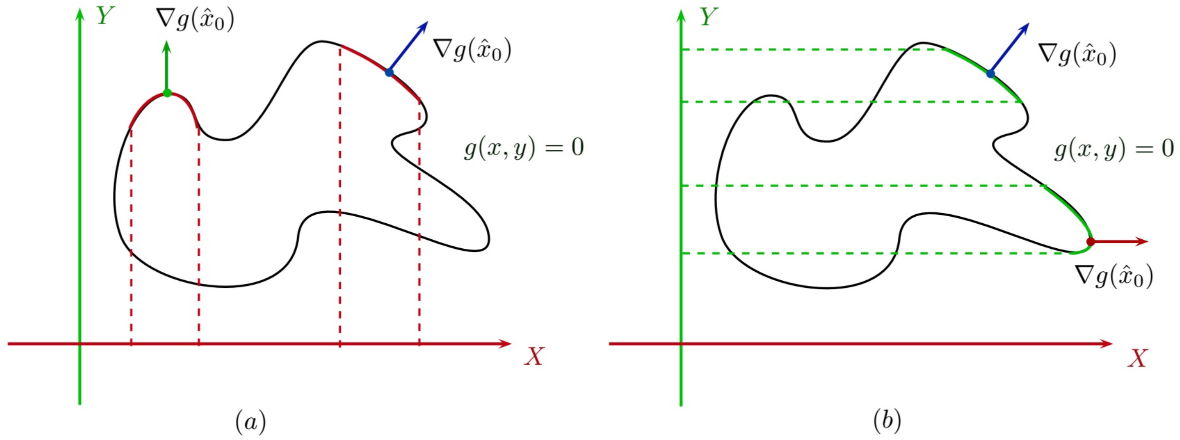

Figura 5.4: Si el vector $\nabla g(\hat{x}_0)$ no es horizontal (vector verde y azul en (a)), entonces un sector de la curva de nivel 0 de g alrededor del punto $\hat{x}_0$ se puede ver como la gráfica de una función de la forma $x = h(y)$. Y si el vector $\nabla g(\hat{x}_0)$ no es vertical (vector rojo y azul en (b)), entonces un sector de la curva de nivel alrededor del punto $\hat{x}_0$ se puede ver como la gráfica de una función de la forma $y = h(x)$.

Lo anterior, nuevamente en términos geométricos, significa que un "pedazo" del conjunto de nivel 0 de g que contiene al punto $\hat{x}_0 = (x_0, y_0, z_0)$ se puede "ver" como la gráfica de una función de la forma $x = h(y, z)$ (ver figura 5.5). Los otros casos ($\frac{\partial g}{\partial y}(\hat{x}_0) \neq 0$ y $\frac{\partial g}{\partial z}(\hat{x}_0) \neq 0$) tienen interpretaciones geométricas semejantes.

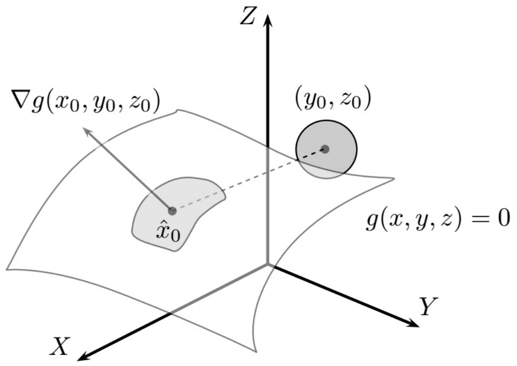

Figura 5.5: Si $\frac{\partial g}{\partial x}(\hat{x}_0) \neq 0$ (lo que significa que el vector $\nabla g(\hat{x}_0)$ no está en el plano YZ), entonces un sector de la superficie de nivel 0 de g alrededor del punto $\hat{x}_0 = (x_0, y_0, z_0)$ se puede ver como la gráfica de una función de la forma $x = h(y, z)$.

La otra posibilidad en $\mathbb{R}^3$ es cuando tenemos dos restricciones de la forma

$$\begin{aligned} g_1(x,y,z) &= 0 \\ g_2(x,y,z) &= 0, \end{aligned}$$

en cuyo caso el conjunto S de soluciones de este sistema se ve como una curva.

Si $\hat{x}_0 = (x_0, y_0, z_0) \in S$, nótese que el sistema de ecuaciones dado por 5.12 se reduce al sistema de dos ecuaciones

$$\begin{aligned} \frac{\partial g_1}{\partial x}(\hat{x}_0)x + \frac{\partial g_1}{\partial y}(\hat{x}_0)y + \frac{\partial g_1}{\partial y}(\hat{x}_0)z - \nabla g_1(\hat{x}_0) \cdot \hat{x}_0 = 0 \\ \frac{\partial g_2}{\partial x}(\hat{x}_0)x + \frac{\partial g_2}{\partial y}(\hat{x}_0)y + \frac{\partial g_2}{\partial y}(\hat{x}_0)z - \nabla g_2(\hat{x}_0) \cdot \hat{x}_0 = 0 \end{aligned}$$

y el determinante de cada una de las tres posibles submatrices de 2×2 que se pueden construir a partir de este sistema coinciden (salvo posiblemente por el signo), con las coordenadas del vector

$$\nabla g_1(\hat{x}_0) \times \nabla g_2(\hat{x}_0) = \det \begin{bmatrix} \frac{\partial g_1}{\partial y}(\hat{x}_0) & \frac{\partial g_1}{\partial z}(\hat{x}_0) \\ \frac{\partial g_2}{\partial y}(\hat{x}_0) & \frac{\partial g_2}{\partial z}(\hat{x}_0) \end{bmatrix} \hat{e}_1 - \det \begin{bmatrix} \frac{\partial g_1}{\partial x}(\hat{x}_0) & \frac{\partial g_1}{\partial z}(\hat{x}_0) \\ \frac{\partial g_2}{\partial x}(\hat{x}_0) & \frac{\partial g_2}{\partial z}(\hat{x}_0) \end{bmatrix} \hat{e}_2 + \det \begin{bmatrix} \frac{\partial g_1}{\partial x}(\hat{x}_0) & \frac{\partial g_1}{\partial y}(\hat{x}_0) \\ \frac{\partial g_2}{\partial x}(\hat{x}_0) & \frac{\partial g_2}{\partial y}(\hat{x}_0) \end{bmatrix} \hat{e}_3$$

De esta forma, si la matriz

$$\begin{bmatrix} \frac{\partial g_1}{\partial y}(\hat{x}_0) & \frac{\partial g_1}{\partial z}(\hat{x}_0) \\ \frac{\partial g_2}{\partial y}(\hat{x}_0) & \frac{\partial g_2}{\partial z}(\hat{x}_0) \end{bmatrix}$$

tiene determinante distinto de 0 (es decir, si la primera coordenada del vector $\nabla g_1(\hat{x}_0) \times \nabla g_2(\hat{x}_0)$ es distinta de 0, o equivalentemente, que este vector no pertenece al plano YZ), entonces las coordenadas (o variables) y y z de puntos de S cercanos al punto $\hat{x}_0$ se pueden poner en función de la coordenada (o variable) x. Es decir, existen $\delta > 0$, $I \subset \mathbb{R}$ abierto, y $h_1, h_2 : I \subset \mathbb{R} \to \mathbb{R}$ tales que $x_0 \in I$, $h_1(x_0) = y_0$, $h_2(x_0) = z_0$ y $(x, h_1(x), h_2(x)) \in B_\delta(\hat{x}_0) \cap S$ para toda $x \in I$. Lo anterior significa que la función de $\mathbb{R}$ en $\mathbb{R}^3$ dada por $h(x) = (x, h_1(x), h_2(x))$ (para $x \in I$) es una parametrización de un "pedazo" de la curva determinada por las restricciones g_1 y g_2 (ver figura 5.6). Las otras dos posibilidades, correspondientes a las otras dos matrices de 2×2 que se pueden obtener con las derivadas parciales de g_1 y g_2, se interpretan de manera análoga.

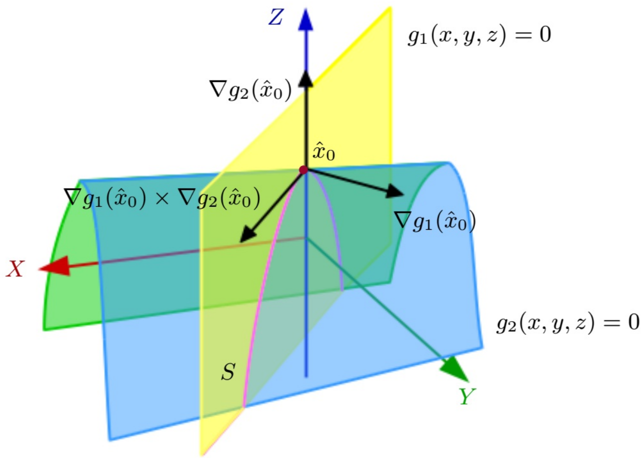

Figura 5.6: Si la primera coordenada del vector $\nabla g_1(\hat{x}_0) \times \nabla g_2(\hat{x}_0)$ es distinta de 0, entonces las variables y y z alrededor del punto $\hat{x}_0 = (x_0, y_0, z_0)$ se pueden poner en función de la variable x. Esto significa que un sector de la curva S (determinada por la intersección de las restricciones g_1 y g_2) se puede parametrizar en términos de la variable x.

Las dos primeras interpretaciones geométricas que acabamos de hacer, los casos en que el conjunto S coincide con un conjunto de nivel en $\mathbb{R}^2$ o uno en $\mathbb{R}^3$, son justo los hechos geométricos que se mencionaron en los comentarios que hicimos posteriores al ejemplo 4.34 del capítulo 4. En ese ejemplo calculamos la recta tangente de un cierto conjunto de nivel en $\mathbb{R}^2$, y el plano tangente de un cierto conjunto de nivel en $\mathbb{R}^3$, de acuerdo con las definiciones 4.33 y 4.18. Y precisamente, adelantándonos a las interpretaciones geométricas del teorema de la función implícita que acabamos de hacer, comentamos que esa recta y ese plano también se podían obtener como la recta tangente a la gráfica de una función de $\mathbb{R}$ en $\mathbb{R}$, y el plano tangente a la gráfica de una función de $\mathbb{R}^2$ en $\mathbb{R}$, respectivamente. Lo que ahora nos proponemos es tomar esos mismos conjuntos y mostrar que estas afirmaciones son ciertas.

Ejemplo 5.14

1. *Consideremos el conjunto de nivel* 0 *($N_0(g)$) de la función de clase C^1 en* $\mathbb{R}^2$

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

que, como se mencionó en el inciso 1 del ejemplo 4.34 del capítulo 4, es la curva cardioide.

Como

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

y

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

para cada $(x,y) \in \mathbb{R}^2$, en particular para el punto $(0,2) \in N_0(g)$ se tiene que $\frac{\partial g}{\partial x}(0,2) = -16 \neq 0$ y $\frac{\partial g}{\partial y}(0,2) = 16 \neq 0$.

Dado que ambas derivadas parciales son distintas de 0*, podemos aplicar el teorema de la función implícita para los dos casos.*

Si consideramos el hecho de que $\frac{\partial g}{\partial x}(0,2) \neq 0$, el teorema nos asegura que existen $\delta > 0$, $I \subset \mathbb{R}$ abierto, y $h : I \subset \mathbb{R} \to \mathbb{R}$ de clase C^1 tales que $2 \in I$, $h(2) = 0$ y $(h(t), t)) \in B_\delta((0,2)) \cap N_0(g)$ para toda $t \in I$.

En particular se tiene que

$$g(h(t), t) = 0 \tag{5.14}$$

para toda $t \in I$. Derivando esta identidad usando la regla de la cadena obtenemos que

$$\frac{\partial g}{\partial x}(h(t), t)\, h'(t) + \frac{\partial g}{\partial y}(h(t), t) = 0$$

para toda $t \in I$.

Así, evaluando para $t = 2$, se tiene que

$$\begin{aligned}\frac{\partial g}{\partial x}(h(2), 2)\, h'(2) &= \frac{\partial g}{\partial x}(0,2)\, h'(2)\\ &= -\frac{\partial g}{\partial y}(0,2)\end{aligned}$$

y por lo tanto, como $\frac{\partial g}{\partial x}(0,2) \neq 0$, concluimos que

$$\begin{aligned}h'(2) &= -\frac{\frac{\partial g}{\partial y}(0,2)}{\frac{\partial g}{\partial x}(0,2)}\\ &= -\frac{16}{-16}\\ &= 1.\end{aligned}$$

Una vez que llegamos a este punto, es importante hacer la siguiente observación: en estricto sentido, la gráfica de la función h es el conjunto de parejas de la forma $(t, h(t)) \in \mathbb{R}^2$, que no son *las parejas que pertenecen al conjunto de nivel* 0 *de la función g.*

De acuerdo con la identidad 5.14, hay que permutar las coordenadas de los elementos de la gráfica de h para que dichas parejas pertenezcan al conjunto $N_0(g)$.

La explicación de este hecho es que, en este caso, es la coordenada x a la que pusimos en función de la coordenada y. Para ser congruentes con lo anterior y no caer en errores, al momento de escribir la ecuación de la recta tangente a la gráfica de h en el punto $(0,2)$, habrá que hacer lo mismo, escribir a la coordenada x en función de la coordenada y. Es decir, escribir que

$$\begin{aligned}x &= h'(2)(y-2) + h(2)\\ &= y - 2 + 0,\end{aligned}$$

o equivalentemente

$$y = x + 2.$$

Procediendo de esta forma, sí obtenemos la misma recta que en el inciso 1 del ejemplo 4.34 del capítulo 4.

2. *Consideremos el conjunto de nivel* 0 *(*$N_0(g)$*) de la función de clase* C^1 *en* $\mathbb{R}^3$

$$g(x,y,z) = \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} - 1$$

que, como se mencionó en el inciso 2 del ejemplo 4.34 del capítulo 4, es un elipsoide.

En este caso tenemos que

$$\begin{aligned} \frac{\partial g}{\partial x}(x,y,z) &= 2\frac{x}{a^2} \\ \frac{\partial g}{\partial y}(x,y,z) &= 2\frac{y}{b^2} \\ \frac{\partial g}{\partial z}(x,y,z) &= 2\frac{z}{c^2}, \end{aligned}$$

de tal forma que si elegimos el punto $(0,b,0) \in N_0(g)$*, entonces*

$$\begin{aligned} \frac{\partial g}{\partial x}(0,b,0) &= 0 \\ \frac{\partial g}{\partial y}(0,b,0) &= \frac{2}{b} \\ \frac{\partial g}{\partial z}(0,b,0) &= 0. \end{aligned}$$

Por tanto, como en este caso la $\frac{\partial g}{\partial y}(0,b,0) = \frac{2}{b}$ *es la única derivada parcial que es distinta de* 0 *en el punto* $(0,b,0)$*, el teorema de la función implícita sólo nos permite asegurar que para puntos* $(x,y,z) \in N_0(g)$ *que están "cercanos" al punto* $(0,b,0)$*, su coordenada* y *se puede poner en función de sus coordenadas* x *y* z*. Es decir, que existen* $\delta > 0$*,* $V \subset \mathbb{R}^2$*, y* $h : V \subset \mathbb{R}^2 \to \mathbb{R}$ *tales que* $(0,0) \in V$*,* $h(0,0) = b$ *y* $(x, h(x,z), z) \in B_\delta((0,b,0)) \cap N_0(g)$ *para toda* $(x,z) \in V$*.*

Nótese nuevamente, como en el inciso anterior, que en estricto sentido las ternas que pertenecen a la gráfica de la función h *son las de la forma* $(x,z,h(x,z))$*. Las ternas* $(x,h(x,z),z)$ *se obtienen al hacer una permutación de las coordenadas de las ternas* $(x,z,h(x,z))$ *(intercambiando la segunda coordenada con la tercera), y son las que pertenecen al conjunto* $N_0(g)$*.*

Dado que para calcular el plano tangente a la gráfica de la función h *en el punto* $(0,b,0)$ *es necesario calcular sus derivadas parciales en el punto* $(0,0)$*, hacemos esto derivando nuevamente la identidad anterior por medio de la regla de la cadena.*

En efecto, como el lado izquierdo de la identidad anterior es la composición de la función $H : V \subset \mathbb{R}^2 \to \mathbb{R}^3$ *dada por*

$$H(x,z) = (x, h(x,z), z)$$

seguida de la función g*, y esta composición nos da la función constante* 0*, aplicando la regla de la cadena obtenemos que*

$$\begin{aligned} D(g \circ H)(x,z) &= Dg(H(x,z))\, DH(x,z) \\ &= \begin{bmatrix} \frac{\partial g}{\partial x}(H(x,z)) & \frac{\partial g}{\partial y}(H(x,z)) & \frac{\partial g}{\partial z}(H(x,z)) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \frac{\partial h}{\partial x}(x,z) & \frac{\partial h}{\partial z}(x,z) \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \end{bmatrix} \end{aligned}$$

para toda $(x,z) \in V$*.*
Por lo tanto, evaluando esta identidad de matrices en el punto $(0,0)$ *se obtiene que*

$$0 = \frac{\partial g}{\partial x}(0,b,0) + \frac{\partial g}{\partial y}(0,b,0)\frac{\partial h}{\partial x}(0,0) = \frac{2}{b}\frac{\partial h}{\partial x}(0,0),$$

y

$$0 = \frac{\partial g}{\partial y}(0,b,0)\frac{\partial h}{\partial z}(0,0) + \frac{\partial g}{\partial z}(0,b,0) = \frac{2}{b}\frac{\partial h}{\partial z}(0,0)$$

de donde

$$\frac{\partial h}{\partial x}(0,0) = 0 = \frac{\partial h}{\partial z}(0,0).$$

Si ahora calculamos el plano tangente a la gráfica de h *en el punto* $(0,0)$*, de acuerdo con la definición 4.18 del capítulo 4 y, como en el inciso anterior, recordando que la coordenada* y *es la que se debe escribir en términos de las coordenadas* x *y* z*, se tiene que*

$$\begin{aligned} y &= \frac{\partial h}{\partial x}(0,0)(x-0) + \frac{\partial h}{\partial z}(0,0)(z-0) + h(0,0) \\ &= b. \end{aligned}$$

Como el lector podrá comprobar, esta ecuación es la misma que se obtiene en el inciso 2. del ejemplo 4.34 del capítulo 4, tomando $(x_0, y_0, z_0) = (0,b,0)$*.*

Una vez dicho y hecho todo lo anterior, escribiremos el tan mencionado teorema de la función implícita. Con el fin de hacer sencilla su redacción, dado que en la primera observación que hicimos ya mencionamos cuál sería su formulación más general, supondremos que la matriz de $m \times m$ que se necesita que sea invertible (como parte de las hipótesis), es la correspondiente a las primeras m variables (la matriz dada en 5.13). Para simplificar aún más esta redacción, escribiremos al espacio $\mathbb{R}^n$ como $\mathbb{R}^m \times \mathbb{R}^k$.

Teorema 5.15 (de la función implícita) *Sean* $g_1, \ldots, g_m : U \subset \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}$ *de clase* C^1 *en* U*. Si*

$$S = \left\{ (\hat{x}, \hat{y}) = (x_1, \ldots, x_m, y_1, \ldots, y_k) \in U \subset \mathbb{R}^m \times \mathbb{R}^k \mid g_i(\hat{x}, \hat{y}) = 0 \text{ para } i \in \{1, \ldots, m\} \right\}$$

y $(\hat{x}_0, \hat{y}_0) = \left(x_1^{(0)}, \ldots, x_m^{(0)}, y_1^{(0)}, \ldots, y_k^{(0)} \right) \in S$ *es tal que la matriz*

$$\begin{bmatrix} \frac{\partial g_1}{\partial x_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_1}{\partial x_m}(\hat{x}_0, \hat{y}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_m}{\partial x_m}(\hat{x}_0, \hat{y}_0) \end{bmatrix}$$

es invertible, entonces existen $\delta > 0$*,* $V \subset \mathbb{R}^k$ *un conjunto abierto, y una función* $h : V \subset \mathbb{R}^k \to \mathbb{R}^m$ *de clase* C^1 *en* V*, tales que* $B_\delta((\hat{x}_0, \hat{y}_0)) \subset U$*,* $\hat{y}_0 \in V$*,* $h(\hat{y}_0) = \hat{x}_0$ *y* $\{(h(\hat{y}), \hat{y}) \in \mathbb{R}^m \times \mathbb{R}^k \mid \hat{y} \in V\} = S \cap B_\delta((\hat{x}_0, \hat{y}_0))$*. La función* h *es única con esta última propiedad.*

Por razones de comodidad, dejaremos pendiente la prueba de este teorema hasta la siguiente sección, en la que probaremos el teorema de la función inversa y con base en el cual demostraremos el teorema anterior. A cambio de esta prueba, apoyados en el teorema de la función implícita haremos una prueba que en el capítulo 4 dejamos pendiente: la prueba del teorema de los multiplicadores de Lagrange.

La formulación que daremos a continuación de este teorema, aunque totalmente equivalente a la que dimos en el capítulo 4, será escrita de tal forma que esté más acorde con la formulación que acabamos de dar del teorema de la función implícita.

Teorema 5.16 (de los multiplicadores de Lagrange) *Sean:*

1. $g_1, \ldots, g_m : U \subset \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}$ *funciones de clase* C^1 *en* U*.*

2. $S \subset \mathbb{R}^m \times \mathbb{R}^k$ *el conjunto dado por*

$$S = \left\{(\hat{x}, \hat{y}) \in U \subset \mathbb{R}^m \times \mathbb{R}^k \mid g_1(\hat{x}, \hat{y}) = 0, \ldots, g_m(\hat{x}, \hat{y}) = 0\right\}$$

3. $(\hat{x}_0, \hat{y}_0) \in S$ *tal que* $\nabla g_1(\hat{x}_0, \hat{y}_0), \ldots, \nabla g_m(\hat{x}_0, \hat{y}_0)$ *son linealmente independientes.*

Si $f : U \subset \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}$ *es una función de clase* C^1 *en* U *tal que* f *tiene un máximo o mínimo (local) en* $(\hat{x}_0, \hat{y}_0)$ *sobre* S*, entonces existen* $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ *tales que*

$$\nabla f(\hat{x}_0, \hat{y}_0) = \lambda_1 \nabla g_1(\hat{x}_0, \hat{y}_0) + \cdots + \lambda_m \nabla g_m(\hat{x}_0, \hat{y}_0).$$

Demostración. Dado que los vectores $\nabla g_1(\hat{x}_0, \hat{y}_0), \ldots, \nabla g_m(\hat{x}_0, \hat{y}_0)$ son linealmente independientes, se tiene que el rango por renglones de la matriz

$$\begin{bmatrix} \frac{\partial g_1}{\partial x_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_1}{\partial x_m}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_1}{\partial x_n}(\hat{x}_0, \hat{y}_0) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_m}{\partial x_m}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_m}{\partial x_n}(\hat{x}_0, \hat{y}_0) \end{bmatrix}$$

debe ser m, y como éste debe ser igual a su rango por columnas (inciso (c) del Corolario 2 del Teorema 3.6 de la referencia [2]), dicha matriz debe tener m columnas linealmente independientes.

Supondremos, sin pérdida de generalidad, que estas columnas son las primeras m de tal forma que la matriz de $m \times m$

$$\begin{bmatrix} \frac{\partial g_1}{\partial x_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_1}{\partial x_m}(\hat{x}_0, \hat{y}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_m}{\partial x_m}(\hat{x}_0, \hat{y}_0) \end{bmatrix}$$

será invertible. De esta forma, por el teorema de la función implícita, sabemos que existen $\delta > 0$, $V \subset \mathbb{R}^k$ y

$$h = (h_1, \ldots, h_m) : V \subset \mathbb{R}^k \to \mathbb{R}^m$$

de clase C^1 en V tales que $\hat{y}_0 \in V$, $h(\hat{y}_0) = \hat{x}_0$ y $(h(\hat{y}), \hat{y}) \in B_\delta((\hat{x}_0, \hat{y}_0)) \cap S$ para toda $\hat{y} \in V$.

Definamos ahora la función

$$H : V \subset \mathbb{R}^k \to \mathbb{R}^m \times \mathbb{R}^k$$

como $H(\hat{y}) = (h(\hat{y}), \hat{y})$, la cual es de clase C^1 en V. Ahora, como $(h(\hat{y}), \hat{y}) \in S$ para toda $\hat{y} \in V$, se tiene que $(g_i \circ H)(\hat{y}) = 0$ para toda $\hat{y} \in V$ de modo que en particular

$$\begin{aligned} D(g_i \circ H)(\hat{y}_0) &= Dg_i(H(\hat{y}_0))\, DH(\hat{y}_0) \\ &= Dg_i(\hat{x}_0, \hat{y}_0)\, DH(\hat{y}_0) \\ &= \begin{bmatrix} 0 & \cdots & 0 \end{bmatrix} \in M_{1\times k}(\mathbb{R}). \end{aligned}$$

Es decir, que el producto de matrices

$$\begin{bmatrix} \frac{\partial g_i}{\partial x_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_i}{\partial x_m}(\hat{x}_0, \hat{y}_0) & \frac{\partial g_i}{\partial y_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_i}{\partial y_k}(\hat{x}_0, \hat{y}_0) \end{bmatrix} \begin{bmatrix} \frac{\partial h_1}{\partial y_1}(\hat{y}_0) & \cdots & \frac{\partial h_1}{\partial y_k}(\hat{y}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m}{\partial y_1}(\hat{y}_0) & \cdots & \frac{\partial h_m}{\partial y_k}(\hat{y}_0) \\ 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix} \tag{5.15}$$

$$= \begin{bmatrix} 0 & \cdots & 0 \end{bmatrix} \in M_{1\times k}(\mathbb{R})$$

para cada $i \in \{1, \ldots, m\}$.

Si ahora definimos los vectores

$$\hat{w}_j = \left(\left(\frac{\partial h_1}{\partial y_j}(\hat{y}_0), \ldots, \frac{\partial h_m}{\partial y_j}(\hat{y}_0)\right), \hat{e}_j\right) \in \mathbb{R}^m \times \mathbb{R}^k$$

para $j \in \{1, \ldots, k\}$ (en donde $\hat{e}_1, \ldots, \hat{e}_k \in \mathbb{R}^k$ son los vectores canónicos), se tiene que $\hat{w}_1, \ldots, \hat{w}_k$ son linealmente independientes, de modo que si $W \subset \mathbb{R}^m \times \mathbb{R}^k$ es el subespacio generado por estos vectores (es decir, que $W = \langle \hat{w}_1, \ldots, \hat{w}_k \rangle$), entonces W es de dimensión k ($\dim(W) = k$).

Ahora nótese que del hecho de que el producto de matrices 5.15 sea la matriz (de $1 \times k$) idénticamente cero se concluye que

$$\nabla g_i\left(\hat{x}_0, \hat{y}_0\right) \cdot \hat{w}_j = 0$$

para cada $i \in \{1, \ldots, m\}$ y cada $j \in \{1, \ldots, k\}$. De esta forma, se tiene que $\nabla g_i\left(\hat{x}_0, \hat{y}_0\right) \in W^{\perp}$ (el complemento ortogonal de W en $\mathbb{R}^m \times \mathbb{R}^k$, de acuerdo con la definición de la página 349 de la referencia [2]) para cada $i \in \{1, \ldots, m\}$. Por otra parte, por el inciso (c) del Teorema 6.7 de la referencia [2], se tiene que

$$\begin{aligned}
\dim(W^{\perp}) &= \dim\left(\mathbb{R}^m \times \mathbb{R}^k\right) - \dim\left(W\right) \\
&= m + k - k \\
&= m
\end{aligned}$$

y como $\nabla g_1(\hat{x}_0, \hat{y}_0), \ldots, \nabla g_m(\hat{x}_0, \hat{y}_0)$ son linealmente independientes, concluimos que estos vectores son una base de $W^{\perp}$.

Finalmente, si ahora consideramos la función $f \circ H$ (la cual está definida en alguna vecindad del punto $\hat{y}_0 \in \mathbb{R}^k$), dado que f tiene un valor extremo (local) en el punto $(\hat{x}_0, \hat{y}_0) = H(\hat{y}_0)$, entonces $f \circ H$ tiene un valor extremo (local) en el punto $\hat{y}_0$ (problema 44 del capítulo 4), de modo que $\hat{y}_0$ es un punto crítico de $f \circ H$ y por lo tanto

$$\begin{aligned}
D(f \circ H)\left(\hat{y}_0\right) &= Df\left(H(\hat{y}_0)\right) DH(\hat{y}_0) \\
&= Df\left(\hat{x}_0, \hat{y}_0\right) DH(\hat{y}_0) \\
&= \begin{bmatrix} 0 & \cdots & 0 \end{bmatrix} \in M_{1\times k}(\mathbb{R}).
\end{aligned}$$

Es decir, que

$$\begin{bmatrix} \frac{\partial f}{\partial x_1}\left(\hat{x}_0, \hat{y}_0\right) & \cdots & \frac{\partial f}{\partial x_m}\left(\hat{x}_0, \hat{y}_0\right) & \frac{\partial f}{\partial y_1}\left(\hat{x}_0, \hat{y}_0\right) & \cdots & \frac{\partial f}{\partial y_k}\left(\hat{x}_0, \hat{y}_0\right) \end{bmatrix} \begin{bmatrix} \frac{\partial h_1}{\partial y_1}\left(\hat{y}_0\right) & \cdots & \frac{\partial h_1}{\partial y_k}\left(\hat{y}_0\right) \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m}{\partial y_1}\left(\hat{y}_0\right) & \cdots & \frac{\partial h_m}{\partial y_k}\left(\hat{y}_0\right) \\ 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & \cdots & 0 \end{bmatrix} \in M_{1\times k}(\mathbb{R}).$$

Por tanto, también se tiene que

$$\nabla f\left(\hat{x}_0, \hat{y}_0\right) \cdot \hat{w}_j = 0$$

para cada $j \in \{1, \ldots, k\}$, de modo que $\nabla f\left(\hat{x}_0, \hat{y}_0\right)$ también pertenece a $W^{\perp} = \langle \hat{w}_1, \ldots, \hat{w}_k \rangle^{\perp}$, y como $\nabla g_1(\hat{x}_0, \hat{y}_0), \ldots, \nabla g_m(\hat{x}_0, \hat{y}_0)$ son una base de $W^{\perp}$, entonces $\nabla f\left(\hat{x}_0, \hat{y}_0\right)$ debe ser una combinación lineal de estos vectores, es decir, existen $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ tales que

$$\nabla f(\hat{x}_0, \hat{y}_0) = \lambda_1 \nabla g_1(\hat{x}_0, \hat{y}_0) + \cdots + \lambda_m \nabla g_m(\hat{x}_0, \hat{y}_0),$$

con lo cual concluimos la prueba. ∎

5.5 El teorema de la función inversa

Además del papel importante que juega en la prueba del teorema de la función implícita, el teorema de la función inversa es relevante por méritos propios. De hecho, este teorema no debe ser ajeno al lector, pues el caso particular de este teorema para funciones de $\mathbb{R}$ en $\mathbb{R}$ forma parte de los resultados importantes de un primer curso de cálculo.

En esta sección haremos uso de todo el material desarrollado en las primeras secciones de este capítulo, pero adicionalmente necesitaremos algunos resultados relacionados con las funciones lineales de $\mathbb{R}^n$ en $\mathbb{R}^m$ y su representación matricial, los cuales introduciremos a continuación.

Si $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ es una función de clase C^1 en su dominio U, sabemos que para cada $\hat{x} \in U$ existe la derivada de f, que la derivada es una función lineal de $\mathbb{R}^n$ en $\mathbb{R}^m$, que ésta se representa por una matriz (la cual depende de las bases que se elijan para $\mathbb{R}^n$ en $\mathbb{R}^m$, y a la que llamamos matriz jacobiana), y que para cada $\hat{x} \in U$, esta matriz está dada por

$$Df(\hat{x}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\hat{x}) & \cdots & \frac{\partial f_1}{\partial x_n}(\hat{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\hat{x}) & \cdots & \frac{\partial f_m}{\partial x_n}(\hat{x}) \end{bmatrix}.$$

Como se podrá observar (y era de esperarse), para cada $\hat{x} \in U$ tenemos asociada una matriz cuyas entradas son funciones continuas de $\hat{x}$, y una cuestión importante es determinar de qué forma se refleja este hecho en las funciones lineales que representan estas matrices. Para ello, observemos que si evaluamos dos de estas funciones lineales (es decir, la derivada de f en dos puntos $\hat{x}, \hat{x}_0 \in U$) en un punto arbitrario $\hat{z} = (z_1, \ldots, z_n) \in \mathbb{R}^n$, se tiene que la diferencia entre estos valores está dada por

$$\begin{aligned} Df(\hat{x})(\hat{z}) - Df(\hat{x}_0)(\hat{z}) &= (Df(\hat{x}) - Df(\hat{x}_0))(\hat{z}) \\ &= \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\hat{x}) - \frac{\partial f_1}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial f_1}{\partial x_n}(\hat{x}) - \frac{\partial f_1}{\partial x_n}(\hat{x}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\hat{x}) - \frac{\partial f_m}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial f_m}{\partial x_n}(\hat{x}) - \frac{\partial f_m}{\partial x_n}(\hat{x}_0) \end{bmatrix} \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} \\ &= ((\nabla f_1(\hat{x}) - \nabla f_1(\hat{x}_0)) \cdot \hat{z}, \ldots, (\nabla f_m(\hat{x}) - \nabla f_m(\hat{x}_0)) \cdot \hat{z}), \end{aligned} \tag{5.16}$$

de tal forma que su distancia satisface que

$$\begin{aligned} \|Df(\hat{x})(\hat{z}) - Df(\hat{x}_0)(\hat{z})\| &= \|((\nabla f_1(\hat{x}) - \nabla f_1(\hat{x}_0)) \cdot \hat{z}, \ldots, (\nabla f_m(\hat{x}) - \nabla f_m(\hat{x}_0)) \cdot \hat{z})\| \\ &\leq |(\nabla f_1(\hat{x}) - \nabla f_1(\hat{x}_0)) \cdot \hat{z}| + \cdots + |(\nabla f_m(\hat{x}) - \nabla f_m(\hat{x}_0)) \cdot \hat{z}| \\ &\leq \|\nabla f_1(\hat{x}) - \nabla f_1(\hat{x}_0)\| \|\hat{z}\| + \cdots + \|\nabla f_m(\hat{x}) - \nabla f_m(\hat{x}_0)\| \|\hat{z}\| \\ &= \|\hat{z}\| \sum_{j=1}^{m} \|\nabla f_j(\hat{x}) - \nabla f_j(\hat{x}_0)\| \\ &= \|\hat{z}\| \sum_{j=1}^{m} \sqrt{\sum_{i=1}^{n} \left(\frac{\partial f_j}{\partial x_i}(\hat{x}) - \frac{\partial f_j}{\partial x_i}(\hat{x}_0) \right)^2}. \end{aligned}$$

De lo anterior se desprende que, si

$$\left| \frac{\partial f_j}{\partial x_i}(\hat{x}) - \frac{\partial f_j}{\partial x_i}(\hat{x}_0) \right| < \varepsilon',$$

para toda $i \in \{1, \ldots, n\}$ y toda $j \in \{1, \ldots, m\}$, entonces

$$\begin{aligned} \|Df(\hat{x})(\hat{z}) - Df(\hat{x}_0)(\hat{z})\| &\leq \|\hat{z}\| \sum_{j=1}^{m} \sqrt{\sum_{i=1}^{n} \left(\frac{\partial f_j}{\partial x_i}(\hat{x}) - \frac{\partial f_j}{\partial x_i}(\hat{x}_0) \right)^2} \\ &< \|\hat{z}\| \sum_{j=1}^{m} \sqrt{\sum_{i=1}^{n} (\varepsilon')^2} \\ &= \|\hat{z}\| \sum_{j=1}^{m} \varepsilon' \sqrt{\sum_{i=1}^{n} 1} \end{aligned}$$

$$
\begin{aligned}
&= \|\hat{z}\| \, \varepsilon' \sqrt{n} \sum_{j=1}^{m} 1 \\
&= \|\hat{z}\| \, m\sqrt{n}\varepsilon' \qquad (5.17)
\end{aligned}
$$

para toda $\hat{z} \in \mathbb{R}^n$.

Esta última desigualdad lo que nos dice es que, si la distancia entre las entradas correspondientes de $Df(\hat{x})$ y $Df(\hat{x}_0)$ es "pequeña", entonces la distancia entre los valores de $Df(\hat{x})$ y $Df(\hat{x}_0)$ en cualquier $\hat{z} \in \mathbb{R}^n$ es "proporcionalmente pequeña" con respecto a la norma de $\hat{z}$.

Lo interesante de la condición anterior es que lo recíproco también es cierto. Es decir, si $\hat{x}, \hat{x}_0 \in U$ son tales que $\|Df(\hat{x})\,(\hat{z}) - Df(\hat{x}_0)\,(\hat{z})\| \leq \|\hat{z}\|\,\varepsilon$ para toda $\hat{z} \in \mathbb{R}^n$, entonces se tiene que

$$
\left| \frac{\partial f_j}{\partial x_i}(\hat{x}) - \frac{\partial f_j}{\partial x_i}(\hat{x}_0) \right| \leq \varepsilon
$$

para cada $i \in \{1, \ldots, n\}$ y cada $j \in \{1, \ldots, m\}$. En efecto, dado que

$$
Df(\hat{x})\,(\hat{z}) - Df(\hat{x}_0)\,(\hat{z}) = ((\nabla f_1(\hat{x}) - \nabla f_1(\hat{x}_0)) \cdot \hat{z}, \ldots, (\nabla f_m(\hat{x}) - \nabla f_m(\hat{x}_0)) \cdot \hat{z})
$$

para toda $\hat{z} \in \mathbb{R}^n$, entonces para cada $i \in \{1, \ldots, n\}$ se tiene que

$$
Df(\hat{x})\,(\hat{e}_i) - Df(\hat{x}_0)\,(\hat{e}_i) = \left(\frac{\partial f_1}{\partial x_i}(\hat{x}) - \frac{\partial f_1}{\partial x_i}(\hat{x}_0), \ldots, \frac{\partial f_m}{\partial x_i}(\hat{x}) - \frac{\partial f_m}{\partial x_i}(\hat{x}_0) \right)
$$

y por lo tanto

$$
\begin{aligned}
\left| \frac{\partial f_j}{\partial x_i}(\hat{x}) - \frac{\partial f_j}{\partial x_i}(\hat{x}_0) \right| &\leq \|Df(\hat{x})\,(\hat{e}_i) - Df(\hat{x}_0)\,(\hat{e}_i)\| \\
&\leq \|\hat{e}_i\|\,\varepsilon \\
&= \varepsilon
\end{aligned}
$$

para cada $j \in \{1, \ldots, m\}$.

La discusión anterior da lugar a la siguiente proposición, que nos aporta una caracterización de las funciones de clase C^1 en una región U.

Proposición 5.17 *Sea $f = (f_1, \ldots, f_m) : U \subset \mathbb{R}^n \to \mathbb{R}^m$. La función f es de clase C^1 en U si y sólo si*

1. *f es derivable para cada $\hat{x} \in U$, y*

2. *para cada $\hat{x}_0 \in U$ y cada $\varepsilon > 0$ existe $\delta > 0$ tal que $B_\delta(\hat{x}_0) \subset U$, y si $\hat{x} \in B_\delta(\hat{x}_0)$, entonces*

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

para toda $\hat{z} \in \mathbb{R}^n$.

Demostración. ($\Longrightarrow$) Como f es de clase C^1 en U, por la proposición 5.10 sabemos que f es derivable para cada $\hat{x} \in U$. Sean ahora $\hat{x}_0 \in U$ y $\varepsilon > 0$. Como f es de clase C^1 en U, entonces $\frac{\partial f_j}{\partial x_i}$ es continua en $\hat{x}_0$ para cada $i \in \{1, \ldots, n\}$ y cada $j \in \{1, \ldots, m\}$, de tal forma que existe $\delta > 0$ tal que $B_\delta(\hat{x}_0) \subset U$ y si $\hat{x} \in B_\delta(\hat{x}_0) \subset U$ entonces

$$
\left| \frac{\partial f_j}{\partial x_i}(\hat{x}) - \frac{\partial f_j}{\partial x_i}(\hat{x}_0) \right| < \frac{\varepsilon}{m\sqrt{n}}
$$

para toda $i \in \{1, \ldots, n\}$ y toda $j \in \{1, \ldots, m\}$. Por tanto, tomando $\varepsilon' = \frac{\varepsilon}{m\sqrt{n}}$ en la desigualdad 5.17, para toda $\hat{z} \in \mathbb{R}^n$ se tiene que

$$
\begin{aligned}
\|Df(\hat{x})\,(\hat{z}) - Df(\hat{x}_0)\,(\hat{z})\| &\leq \|\hat{z}\|\,m\sqrt{n}\varepsilon' \\
&= \|\hat{z}\|\,\varepsilon
\end{aligned}
$$

($\Longleftarrow$) Recíprocamente, para $\hat{x}_0 \in U$ y $\varepsilon > 0$ sabemos que existe $\delta > 0$ tal que si $\hat{x} \in B_\delta(\hat{x}_0) \subset U$, entonces

$$\|Df(\hat{x})(\hat{z}) - Df(\hat{x}_0)(\hat{z})\| \leq \|\hat{z}\| \varepsilon$$

para toda $\hat{z} \in \mathbb{R}^n$.

Ahora, por la identidad 5.16, tomando $z = \hat{e}_i$ para cada $i \in \{1, \ldots, n\}$, se tiene que

$$\left(\frac{\partial f_1}{\partial x_i}(\hat{x}) - \frac{\partial f_1}{\partial x_i}(\hat{x}_0), \ldots, \frac{\partial f_m}{\partial x_i}(\hat{x}) - \frac{\partial f_m}{\partial x_i}(\hat{x}_0)\right) = Df(\hat{x})(\hat{e}_i) - Df(\hat{x}_0)(\hat{e}_i),$$

de modo que

$$\begin{aligned}
\left|\frac{\partial f_j}{\partial x_i}(\hat{x}) - \frac{\partial f_j}{\partial x_i}(\hat{x}_0)\right| &\leq \left\|\left(\frac{\partial f_1}{\partial x_i}(\hat{x}) - \frac{\partial f_1}{\partial x_i}(\hat{x}_0), \ldots, \frac{\partial f_m}{\partial x_i}(\hat{x}) - \frac{\partial f_m}{\partial x_i}(\hat{x}_0)\right)\right\| \\
&= \|Df(\hat{x})(\hat{e}_i) - Df(\hat{x}_0)(\hat{e}_i)\| \\
&\leq \|\hat{e}_i\| \varepsilon \\
&= \varepsilon
\end{aligned}$$

para cada $j \in \{1, \ldots, m\}$, lo que prueba que $\frac{\partial f_j}{\partial x_i}$ es continua en $\hat{x}_0 \in U$, y por lo tanto que f es de clase C^1 en U. ■

El tema principal de esta sección se centra en la búsqueda de condiciones para que una función f de $\mathbb{R}^n$ en $\mathbb{R}^n$ sea invertible en una vecindad de un punto $\hat{x}_0$. Dado que la derivabilidad de f en $\hat{x}_0$ nos garantiza que ésta se parece mucho (en una vecindad de $\hat{x}_0$) a una función lineal de $\mathbb{R}^n$ en $\mathbb{R}^n$, la intuición nos dice que, si esta función lineal es invertible, entonces f también lo será (al menos en una vecindad del punto $\hat{x}_0$). Por lo anterior, empezaremos por dar condiciones bajo las cuales una función lineal L de $\mathbb{R}^n$ en $\mathbb{R}^n$ es invertible. Antes de hacer esto, como seguramente el lector recordará de su curso de Álgebra Lineal, decimos que una función lineal $L : \mathbb{R}^n \to \mathbb{R}^n$ es invertible si existe otra función lineal $\tilde{L} : \mathbb{R}^n \to \mathbb{R}^n$ tal que $\left(L \circ \tilde{L}\right)(\hat{x}) = \hat{x} = \left(\tilde{L} \circ L\right)(\hat{x})$ para toda $\hat{x} \in \mathbb{R}^n$, y que cuando esta función lineal $\tilde{L}$ existe, entonces es única; por esta razón se le suele denotar por L^{-1}. También es oportuno tener presente que en este mismo curso de Álgebra Lineal se debieron haber probado dos condiciones necesarias y suficientes para que una función lineal $L : \mathbb{R}^n \to \mathbb{R}^n$ sea invertible. Una, que dice que L es invertible si y sólo si $L(\hat{x}) = \hat{0}$ sólo si $\hat{x} = \hat{0}$; y dos, que L es invertible si y sólo si la matriz M asociada a L (en cualesquiera bases de ambos $\mathbb{R}^n$, el dominio y el contradominio) es invertible, lo que a su vez es equivalente a que $\det(M) \neq 0$.

Una vez dicho lo anterior, lo siguiente que haremos será formular (apoyados en algunas de las condiciones mencionadas) otra condición necesaria y suficiente para que una función lineal $L : \mathbb{R}^n \to \mathbb{R}^n$ sea invertible.

Proposición 5.18 *Sea $L : \mathbb{R}^n \to \mathbb{R}^n$ una función lineal. L tiene inversa (o L es invertible) si y sólo si existe $m > 0$ tal que*

$$\|L(\hat{x})\| \geq m \|\hat{x}\|$$

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

Demostración. Si L tiene inversa, sabemos que $L(\hat{x}) = \hat{0}$ sólo si $\hat{x} = \hat{0}$, de tal forma que, si consideramos el conjunto

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

entonces $\|L(\hat{x})\| > 0$ para toda $\hat{x} \in S^{n-1}$.

Ahora, dado que S^{n-1} es un conjunto cerrado y acotado y L es una función continua (en $\mathbb{R}^n$), entonces $\|L\|$ alcanza un valor mínimo sobre S^{n-1}, es decir, existe $\hat{x}_0 \in S^{n-1}$ tal que

$$\|L(\hat{x})\| \geq \|L(\hat{x}_0)\| > 0$$

para toda $\hat{x} \in S^{n-1}$. Por tanto, si hacemos $m = \|L(\hat{x}_0)\| > 0$ y tomamos $\hat{x} \in \mathbb{R}^n$, con $\hat{x} \neq \hat{0}$ (pues claramente $\left\|L\left(\hat{0}\right)\right\| = 0 = m\left\|\hat{0}\right\|$), entonces

$$\frac{\hat{x}}{\|\hat{x}\|} \in S^{n-1}$$

y por lo tanto

$$L\left(\frac{\hat{x}}{\|\hat{x}\|}\right) \geq \|L(\hat{x}_0)\| = m,$$

de modo que

$$\|L(\hat{x})\| \geq m\|\hat{x}\|,$$

lo cual prueba la primera implicación.

Recíprocamente, si existe $m > 0$ tal que

$$\|L(\hat{x})\| \geq m\|\hat{x}\|$$

para toda $\hat{x} \in \mathbb{R}^n$, entonces se tiene que

$$\|L(\hat{x})\| \geq m\|\hat{x}\| > 0$$

para toda $\hat{x} \in \mathbb{R}^n$, con $\hat{x} \neq \hat{0}$, es decir que $L(\hat{x}) = \hat{0}$ sólo si $\hat{x} = \hat{0}$ lo que implica que L es invertible. ■

Con lo anterior ya tenemos las herramientas de Álgebra Lineal necesarias para probar el teorema principal de esta sección: el Teorema de la Función Inversa. Antes de hacer esto, e incluso antes de formular el teorema, daremos un par de lemas con los cuales iremos "preparando el terreno" para enunciarlo y probarlo. El primero de estos dos lemas nos permitirá probar que, si la derivada de una función f (de clase C^1 en un conjunto abierto U) es invertible en un punto $\hat{x}_0 \in U$ (en donde $Df(\hat{x}_0)$ se toma como función lineal o como matriz), entonces la derivada de f sigue siendo invertible para todo punto en una vecindad del punto $\hat{x}_0$.

Lema 5.19 *Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}^n$ de clase C^1 en el conjunto abierto U, y $\hat{x}_0 \in U$. Si $Df(\hat{x}_0)$ es invertible, entonces existen $\delta > 0$ y $m > 0$ tales que $B_\delta(\hat{x}_0) \subset U$ y*

$$\|Df(\hat{x})(\hat{z})\| \geq m\|\hat{z}\|$$

para toda $\hat{x} \in B_\delta(\hat{x}_0)$ y para toda $\hat{z} \in \mathbb{R}^n$.

Demostración. Por la proposición 5.18, dado que $Df(\hat{x}_0)$ es invertible, sabemos que existe $m' > 0$ tal que

$$\|Df(\hat{x}_0)(\hat{z})\| \geq m'\|\hat{z}\|$$

para toda $\hat{z} \in \mathbb{R}^n$.

Por otra parte, como f es de clase C^1 en U (que es un abierto), por la proposición 5.17, tomando $\varepsilon = m'/2 > 0$, sabemos que existe $\delta > 0$ tal que $B_\delta(\hat{x}_0) \subset U$ y

$$\|Df(\hat{x})(\hat{z}) - Df(\hat{x}_0)(\hat{z})\| \leq \|\hat{z}\|\frac{m'}{2}$$

para toda $\hat{x} \in B_\delta(\hat{x}_0)$ y para toda $\hat{z} \in \mathbb{R}^n$. Por lo tanto, de la desigualdad del triángulo tenemos que

$$\begin{aligned}
\|Df(\hat{x})(\hat{z})\| &\geq \|Df(\hat{x}_0)(\hat{z})\| - \|Df(\hat{x})(\hat{z}) - Df(\hat{x}_0)(\hat{z})\| \\
&\geq m'\|\hat{z}\| - \|\hat{z}\|\frac{m'}{2} \\
&= \|\hat{z}\|\frac{m'}{2}
\end{aligned}$$

para toda $\hat{x} \in B_\delta(\hat{x}_0) \subset U$ y para toda $\hat{z} \in \mathbb{R}^n$, de tal manera que tomando $m = m'/2 > 0$ logramos el resultado deseado. ■

Como una consecuencia inmediata de este lema y de la proposición 5.18, obtenemos el siguiente

Corolario 5.20 *Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}^n$ de clase C^1 en U y $\hat{x}_0 \in U$. Si $Df(\hat{x}_0)$ es invertible, entonces existe $\delta > 0$ tal que $B_\delta(\hat{x}_0) \subset U$ y $Df(\hat{x})$ es invertible para toda $\hat{x} \in B_\delta(\hat{x}_0)$.*

El segundo lema que probaremos también es fundamental en la prueba del teorema de la función inversa, pues a partir de éste podremos asegurar que una función f es localmente invertible, y que la función inversa que se puede definir es continua en su dominio. Antes, sólo recordemos que en el capítulo 1 definimos otras normas para los elementos de $\mathbb{R}^n$, una de ellas llamada la norma infinito (que utilizaremos en la prueba del siguiente lema), y que está definida (y es denotada) como

$$\|\hat{x}\|_\infty := \max\{|x_1|, \ldots, |x_n|\}$$

para cada $\hat{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$, y que con relación a la norma euclideana satisface la desigualdad

$$\|\hat{x}\| \leq \sqrt{n}\, \|\hat{x}\|_\infty,$$

la cual es válida para toda $\hat{x} \in \mathbb{R}^n$.

Lema 5.21 *Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}^n$ de clase C^1 en el conjunto abierto U, y $\hat{x}_0 \in U$. Si $Df(\hat{x}_0)$ es invertible, entonces existen $\delta > 0$ y $m > 0$ tales que $B_\delta(\hat{x}_0) \subset U$ y*

$$m\, \|\hat{y} - \hat{x}\| \leq \|f(\hat{y}) - f(\hat{x})\| \tag{5.18}$$

para toda $\hat{x}, \hat{y} \in B_\delta(\hat{x}_0)$.

Demostración. Primero recordemos que si $f = (f_1, \ldots, f_n)$, entonces

$$Df(\hat{x})(\hat{z}) = (\nabla f_1(\hat{x}) \cdot \hat{z}, \ldots, \nabla f_n(\hat{x}) \cdot \hat{z})$$

para cada $\hat{x} \in U$ y cada $\hat{z} \in \mathbb{R}^n$.

Como $Df(\hat{x}_0)$ es invertible, por la proposición 5.18 sabemos que existe $m' > 0$ tal que

$$\|Df(\hat{x}_0)(\hat{z})\| \geq m'\, \|\hat{z}\|$$

para toda $\hat{z} \in \mathbb{R}^n$.

Por otra parte, por la proposición 5.17 sabemos que para $m'/(2\sqrt{n}) > 0$ existe $\delta > 0$ tal que si $\hat{x} \in B_\delta(\hat{x}_0) \subset U$, entonces

$$\|Df(\hat{x})(\hat{z}) - Df(\hat{x}_0)(\hat{z})\| \leq \|\hat{z}\| \frac{m'}{2\sqrt{n}}$$

para toda $\hat{z} \in \mathbb{R}^n$.

Probaremos que esta $\delta > 0$ y esta $m = m'/(2\sqrt{n}) > 0$ son las cantidades para las que se satisface la desigualdad 5.18, para toda $\hat{x}, \hat{y} \in B_\delta(\hat{x}_0) \subset U$.

Dados $\hat{x}, \hat{y} \in B_\delta(\hat{x}_0)$, sea $k \in \{1, \ldots, n\}$ tal que $\|Df(\hat{x}_0)(\hat{y} - \hat{x})\|_\infty = |\nabla f_k(\hat{x}_0) \cdot (\hat{y} - \hat{x})|$. Entonces

$$\begin{aligned} |\nabla f_k(\hat{x}_0) \cdot (\hat{y} - \hat{x})| &= \|Df(\hat{x}_0)(\hat{y} - \hat{x})\|_\infty \\ &\geq \frac{1}{\sqrt{n}} \|Df(\hat{x}_0)(\hat{y} - \hat{x})\|. \end{aligned}$$

Ahora, como $B_\delta(\hat{x}_0)$ es un conjunto convexo, sabemos que el segmento

$$[\hat{x}, \hat{y}] = \{\hat{x} + t(\hat{y} - \hat{x}) \in \mathbb{R}^n \mid t \in [0, 1]\}$$

está totalmente contenido en $B_\delta(\hat{x}_0) \subset U$, de tal forma que la función $\gamma_k : [0, 1] \subset \mathbb{R} \to \mathbb{R}$ definida como

$$\gamma_k(t) = f_k(\hat{x} + t(\hat{y} - \hat{x}))$$

es derivable para toda $t \in [0, 1]$.

De esta forma, aplicando el teorema del valor medio para funciones de $\mathbb{R}$ en $\mathbb{R}$, sabemos que existe $\xi_k \in (0, 1)$ tal que

$$\begin{aligned} f_k(\hat{y}) - f_k(\hat{x}) &= \gamma_k(1) - \gamma_k(0) \\ &= (1 - 0)\, \gamma_k'(\xi_k) \end{aligned}$$

$$
\begin{aligned}
&= \gamma_k'(\xi_k) \\
&= \nabla f_k(\hat{x} + \xi_k(\hat{y} - \hat{x})) \cdot (\hat{y} - \hat{x})\,.
\end{aligned}
$$

Si recordamos ahora que para $\hat{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ se tiene que $|x_i| \leq \|\hat{x}\|$ para toda $i \in \{1, \ldots, n\}$, entonces también tenemos que

$$\|f(\hat{y}) - f(\hat{x})\| \geq |f_k(\hat{y}) - f_k(\hat{x})|$$

y

$$|\nabla f_k(\hat{x}) \cdot (\hat{y} - \hat{x}) - \nabla f_k(\hat{x}_0) \cdot (\hat{y} - \hat{x})| \leq \|Df(\hat{x})\,(\hat{y} - \hat{x}) - Df(\hat{x}_0)\,(\hat{y} - \hat{x})\|\,.$$

Con base en las desigualdades anteriores (y la desigualdad del triángulo), se tiene que

$$
\begin{aligned}
\|f(\hat{y}) - f(\hat{x})\| &\geq |f_k(\hat{y}) - f_k(\hat{x})| \\
&= |\gamma_k'(\xi_k)| \\
&= |\nabla f_k(\hat{x} + \xi_k(\hat{y} - \hat{x})) \cdot (\hat{y} - \hat{x})| \\
&\geq |\nabla f_k(\hat{x}_0) \cdot (\hat{y} - \hat{x})| - |\nabla f_k(\hat{x} + \xi_k(\hat{y} - \hat{x})) \cdot (\hat{y} - \hat{x}) - \nabla f_k(\hat{x}_0) \cdot (\hat{y} - \hat{x})| \\
&\geq \frac{1}{\sqrt{n}}\,\|Df(\hat{x}_0)\,(\hat{y} - \hat{x})\| - \|Df(\hat{x} + \xi_k(\hat{y} - \hat{x}))\,(\hat{y} - \hat{x}) - Df(\hat{x}_0)\,(\hat{y} - \hat{x})\| \\
&\geq \frac{1}{\sqrt{n}}\,(m'\,\|\hat{y} - \hat{x}\|) - \frac{m'}{2\sqrt{n}}\,\|\hat{y} - \hat{x}\| \\
&= \frac{m'}{2\sqrt{n}}\,\|\hat{y} - \hat{x}\| \\
&= m\,\|\hat{y} - \hat{x}\|\,,
\end{aligned}
$$

que es la desigualdad que se deseaba probar. ■

Como mencionamos anteriormente, de este lema se obtiene el siguiente

Corolario 5.22 *Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}^n$ de clase C^1 en el conjunto abierto U, y $\hat{x}_0 \in U$. Si $Df(\hat{x}_0)$ es invertible, entonces existe $\delta > 0$ tal que $B_\delta(\hat{x}_0) \subset U$, f es inyectiva en $B_\delta(\hat{x}_0)$, y además*

$$f^{-1} : f(B_\delta(\hat{x}_0)) \subset \mathbb{R}^n \to \mathbb{R}^n$$

es uniformemente continua en su dominio.

Demostración. Por el lema anterior sabemos que existen $\delta > 0$ y $m > 0$ tales que

$$\|f(\hat{x}') - f(\hat{x})\| \geq m\,\|\hat{x}' - \hat{x}\|$$

para todo $\hat{x}, \hat{x}' \in B_\delta(\hat{x}_0) \subset U$.

Ahora, si tomamos $\hat{x}, \hat{x}' \in B_\delta(\hat{x}_0)$, con $\hat{x} \neq \hat{x}'$, entonces $\|f(\hat{x}') - f(\hat{x})\| \geq m\,\|\hat{x}' - \hat{x}\| > 0$, de modo que $f(\hat{x}') \neq f(\hat{x})$ y por lo tanto se tiene que f es inyectiva en $B_\delta(\hat{x}_0)$.

Por otra parte, dado $\varepsilon > 0$, si tomamos $\delta' = \varepsilon m$ y $\hat{y}, \hat{y}' \in f(B_\delta(\hat{x}_0))$ tales que $\|\hat{y}' - \hat{y}\| < \delta'$, si $\hat{x}, \hat{x}' \in B_\delta(\hat{x}_0)$ son tales que $\hat{y}' = f(\hat{x}')$ y $\hat{y} = f(\hat{x})$, se tiene que

$$
\begin{aligned}
\left\|f^{-1}(\hat{y}') - f^{-1}(\hat{y})\right\| &= \|\hat{x}' - \hat{x}\| \\
&\leq \frac{1}{m}\,\|f(\hat{x}') - f(\hat{x})\| \\
&= \frac{1}{m}\,\|\hat{y}' - \hat{y}\| \\
&< \frac{1}{m}\delta' \\
&= \varepsilon,
\end{aligned}
$$

lo que prueba que f^{-1} es uniformemente continua (y por tanto continua) en $f(B_\delta(\hat{x}_0))$. ■

Con base en todo el trabajo realizado previamente ya estamos en condiciones de formular y probar el teorema de la función inversa.

Teorema 5.23 (de la función inversa) *Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}^n$ de clase C^1 en el conjunto abierto U, y $\hat{x}_0 \in U$. Si $Df(\hat{x}_0)$ es invertible, entonces existe $\delta > 0$ tal que:*

1. *$B_\delta(\hat{x}_0) \subset U$ y f es inyectiva en $B_\delta(\hat{x}_0)$,*
2. *$f^{-1} : f(B_\delta(\hat{x}_0)) \subset \mathbb{R}^n \to \mathbb{R}^n$ es continua en $f(B_\delta(\hat{x}_0))$,*
3. *$f(B_\delta(\hat{x}_0)) \subset \mathbb{R}^n$ es un conjunto abierto, y*
4. *f^{-1} es de clase C^1 en $f(B_\delta(\hat{x}_0))$ y además, si $\hat{y} = f(\hat{x}) \in f(B_\delta(\hat{x}_0))$, entonces*

$$Df^{-1}(\hat{y}) = Df^{-1}(f(\hat{x})) = (Df(\hat{x}))^{-1}.$$

Demostración. Por los lemas 5.19 y 5.21 sabemos que existen $\delta > 0$ y $m > 0$ tales que

$$m \|\hat{x}' - \hat{x}\| \leq \|f(\hat{x}') - f(\hat{x})\| \tag{5.19}$$

para toda $\hat{x}, \hat{x}' \in B_\delta(\hat{x}_0) \subset U$, y que

$$\|Df(\hat{x})(\hat{z})\| \geq m \|\hat{z}\| \tag{5.20}$$

para toda $\hat{x} \in B_\delta(\hat{x}_0)$ y toda $\hat{z} \in \mathbb{R}^n$.

Ahora, por el corolario 5.22 se tiene que f es inyectiva en $B_\delta(\hat{x}_0)$ y que f^{-1} es continua en $f(B_\delta(\hat{x}_0))$, con lo cual se tiene la prueba de los incisos 1 y 2 del enunciado.

Probaremos ahora que $f(B_\delta(\hat{x}_0)) \subset \mathbb{R}^n$ es un conjunto abierto. Sea $\hat{y}' = f(\hat{x}') \in f(B_\delta(\hat{x}_0))$, con $\hat{x}' \in B_\delta(\hat{x}_0)$. Como $B_\delta(\hat{x}_0)$ es un conjunto abierto, existe $\delta' > 0$ tal que el conjunto

$$A = \{\hat{x} \in \mathbb{R}^n \mid \|\hat{x} - \hat{x}'\| \leq \delta'\}$$

se queda contenido en $B_\delta(\hat{x}_0)$. Probaremos que existe $r > 0$ tal que $B_r(\hat{y}') \subset f(A) \subset f(B_\delta(\hat{x}_0))$.

Para "justificar" de manera intuitiva el valor de r que vamos a tomar, obsérvese que si $\hat{y}'' \in B_r(\hat{y}')$ fuera tal que $\hat{y}'' = f(\hat{x})$ para alguna $\hat{x} \in A$, por la desigualdad 5.19 se debería tener que

$$\begin{aligned}
\|\hat{x} - \hat{x}'\| &= \left\|f^{-1}(\hat{y}'') - f^{-1}(\hat{y}')\right\| \\
&\leq \frac{1}{m} \|\hat{y}'' - \hat{y}'\| \\
&< \frac{r}{m},
\end{aligned}$$

de tal forma que, para no caer en contradicción con el supuesto de que $\hat{x} \in A$, se deberá elegir r de tal forma que $r/m \leq \delta'$. Con base en el razonamiento anterior, tomamos $r = m\delta'/2 > 0$ y $\hat{y}'' \in B_r(\hat{y}')$. Para probar que existe $\hat{x}'' \in A$ tal que $f(\hat{x}'') = \hat{y}''$, definimos $h : U \subset \mathbb{R}^n \to \mathbb{R}$ como

$$\begin{aligned}
h(\hat{x}) &= \|f(\hat{x}) - \hat{y}''\|^2 \\
&= (f(\hat{x}) - \hat{y}'') \cdot (f(\hat{x}) - \hat{y}'').
\end{aligned}$$

Dado que $A \subset U$ es un conjunto cerrado y acotado, y h es continua en U, sabemos que h alcanza un valor mínimo sobre A, es decir, que existe $\hat{x}'' \in A$ tal que

$$\begin{aligned}
\|f(\hat{x}'') - \hat{y}''\|^2 &= h(\hat{x}'') \\
&\leq h(\hat{x}) \\
&= \|f(\hat{x}) - \hat{y}''\|^2
\end{aligned}$$

para toda $\hat{x} \in A$. Nótese que ahora nuestro objetivo es probar que $h(\hat{x}'') = 0$, pues de este hecho se concluye que $f(\hat{x}'') = \hat{y}''$.

Ahora, como $\hat{x}'' \in A$, se tiene que $\|\hat{x}'' - \hat{x}'\| \leq \delta'$, de modo que, aún cuando h alcanza un valor mínimo en $\hat{x}''$, no podemos asegurar que éste sea un punto crítico de h.

Con el fin de probar que $\hat{x}''$ sí es un punto crítico de h, descartaremos la posibilidad de que $\|\hat{x}'' - \hat{x}'\| = \delta'$. Si este fuera el caso, por la desigualdad del triángulo se tendría que

$$\begin{aligned}
\sqrt{h(\hat{x}'')} &= \|f(\hat{x}'') - \hat{y}''\| \\
&\geq \|f(\hat{x}'') - \hat{y}'\| - \|\hat{y}' - \hat{y}''\| \\
&= \|f(\hat{x}'') - f(\hat{x}')\| - \|\hat{y}' - \hat{y}''\| \\
&\geq m\|\hat{x}'' - \hat{x}'\| - \|\hat{y}' - \hat{y}''\| \\
&> m\delta' - \frac{m\delta'}{2} \\
&= \frac{m\delta'}{2} \\
&> \|\hat{y}' - \hat{y}''\| \\
&= \|f(\hat{x}') - \hat{y}''\| \\
&= \sqrt{h(\hat{x}')},
\end{aligned}$$

lo que contradice el hecho de que $h(\hat{x}'')$ es el valor mínimo de h sobre A. De esta forma, se debe tener que $\|\hat{x}'' - \hat{x}'\| < \delta'$ y por tanto $\hat{x}''$ es un punto crítico de h, es decir que

$$\begin{aligned}
Dh(\hat{x}'') &= 2(f(\hat{x}'') - \hat{y}'') \cdot Df(\hat{x}'') \\
&= \hat{0}
\end{aligned}$$

Si ahora recordamos que, por la desigualdad 5.20 y la proposición 5.18, se tiene que $Df(\hat{x})$ es invertible (y por tanto suprayectiva) para toda $\hat{x} \in B_\delta(\hat{x}_0)$, y como $\hat{x}'' \in A \subset B_\delta(\hat{x}_0)$ entonces debe existir $\hat{z} \in \mathbb{R}^n$ tal que

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

de tal forma que en particular se tiene que

$$\begin{aligned}
0 &= ((f(\hat{x}'') - \hat{y}'') \cdot Df(\hat{x}''))(\hat{z}) \\
&= (f(\hat{x}'') - \hat{y}'') \cdot (Df(\hat{x}'')(\hat{z})) \\
&= (f(\hat{x}'') - \hat{y}'') \cdot (f(\hat{x}'') - \hat{y}'') \\
&= \|f(\hat{x}'') - \hat{y}''\|^2
\end{aligned}$$

y por lo tanto que $f(\hat{x}'') = \hat{y}''$, lo que prueba que $B_r(\hat{y}') \subset f(B_\delta(\hat{x}_0))$, es decir que $f(B_\delta(\hat{x}_0))$ es un conjunto abierto.

Para probar el último inciso, nuestro primer paso será demostrar que f^{-1} es derivable en todo punto $\hat{y}' = f(\hat{x}') \in f(B_\delta(\hat{x}_0))$. Usando nuevamente que $Df(\hat{x}')$ es invertible para toda $\hat{x}' \in B_\delta(\hat{x}_0)$, sólo nos restará mostrar que la función lineal

$$L = (Df(\hat{x}'))^{-1}$$

satisface la definición 5.1, es decir, que

$$\lim_{\hat{y} \to \hat{y}'} \frac{f^{-1}(\hat{y}) - f^{-1}(\hat{y}') - L(\hat{y} - \hat{y}')}{\|\hat{y} - \hat{y}'\|} = \hat{0}. \tag{5.21}$$

Para ello, recurriremos a la función $g : B_\delta(\hat{x}_0) \subset \mathbb{R}^n \to \mathbb{R}^n$ dada por

$$g(\hat{x}) = \begin{cases} \frac{\hat{x} - \hat{x}' - L(f(\hat{x}) - f(\hat{x}'))}{\|f(\hat{x}) - f(\hat{x}')\|} & \text{si } \hat{x} \neq \hat{x}' \\ \\ \hat{0} & \text{si } \hat{x} = \hat{x}' \end{cases}$$

la cual está bien definida, puesto que f es inyectiva en la bola $B_\delta(\hat{x}_0)$.

Obsérvese que la composición $g \circ f^{-1}$ también está bien definida sobre el conjunto abierto $f\left(B_\delta(\hat{x}_0)\right) \setminus \{\hat{y}'\}$ y que

$$\begin{aligned}\left(g \circ f^{-1}\right)(\hat{y}) &= g\left(f^{-1}(\hat{y})\right) \\ &= \begin{cases} \frac{f^{-1}(\hat{y}) - f^{-1}(\hat{y}') - L(\hat{y}-\hat{y}')}{\|\hat{y}-\hat{y}'\|} & \text{si } \hat{y} \neq \hat{y}' \\ \\ \hat{0} & \text{si } \hat{y} = \hat{y}' \end{cases}\end{aligned}$$

para toda $\hat{y} \in f\left(B_\delta(\hat{x}_0)\right) \setminus \{\hat{y}'\}$.

En virtud de lo anterior, tenemos que

$$\lim_{\hat{y} \to \hat{y}'} \frac{f^{-1}(\hat{y}) - f^{-1}(\hat{y}') - L(\hat{y}-\hat{y}')}{\|\hat{y}-\hat{y}'\|} = \lim_{\hat{y} \to \hat{y}'} \left(g \circ f^{-1}\right)(\hat{y}),$$

de tal forma que para obtener la identidad 5.21, como f^{-1} es continua en $\hat{y}'$, por el inciso 7 de la proposición 2.44 del capítulo 2, bastará mostrar que g es continua en $\hat{x}' = f^{-1}(\hat{y}')$, lo que es equivalente a probar que

$$\lim_{\hat{x} \to \hat{x}'} g(\hat{x}) = \lim_{\hat{x} \to \hat{x}'} \frac{\hat{x} - \hat{x}' - \left(Df(\hat{x}')\right)^{-1}\left(f(\hat{x}) - f(\hat{x}')\right)}{\|f(\hat{x}) - f(\hat{x}')\|} = \hat{0}.$$

Para probar este último límite, nótese que

$$\begin{aligned}&\frac{\hat{x} - \hat{x}' - L(f(\hat{x}) - f(\hat{x}'))}{\|f(\hat{x}) - f(\hat{x}')\|} \\ &= \frac{\hat{x} - \hat{x}' - \left(Df(\hat{x}')\right)^{-1}\left(f(\hat{x}) - f(\hat{x}')\right)}{\|f(\hat{x}) - f(\hat{x}')\|} \\ &= \frac{\left(Df(\hat{x}')\right)^{-1}\left[Df(\hat{x}')(\hat{x} - \hat{x}') - \left(f(\hat{x}) - f(\hat{x}')\right)\right]}{\|f(\hat{x}) - f(\hat{x}')\|} \\ &= -\frac{\|\hat{x} - \hat{x}'\|}{\|f(\hat{x}) - f(\hat{x}')\|}\left(Df(\hat{x}')\right)^{-1}\left(\frac{f(\hat{x}) - f(\hat{x}') - Df(\hat{x}')(\hat{x} - \hat{x}')}{\|\hat{x} - \hat{x}'\|}\right)\end{aligned}$$

de tal forma que, como $\left(Df(\hat{x}')\right)^{-1}$ es una función continua (toda función lineal es continua) y f es derivable en $\hat{x}'$, se tiene que

$$\begin{aligned}&\lim_{\hat{x} \to \hat{x}'} \left(Df(\hat{x}')\right)^{-1}\left(\frac{f(\hat{x}) - f(\hat{x}') - Df(\hat{x}')(\hat{x} - \hat{x}')}{\|\hat{x} - \hat{x}'\|}\right) \\ &= \left(Df(\hat{x}')\right)^{-1}\left(\lim_{\hat{x} \to \hat{x}'} \frac{f(\hat{x}) - f(\hat{x}') - Df(\hat{x}')(\hat{x} - \hat{x}')}{\|\hat{x} - \hat{x}'\|}\right) \\ &= \left(Df(\hat{x}')\right)^{-1}\left(\hat{0}\right) \\ &= \hat{0}.\end{aligned}$$

Ahora, como por la desigualdad 5.19 sabemos que

$$\frac{\|\hat{x} - \hat{x}'\|}{\|f(\hat{x}) - f(\hat{x}')\|} \leq \frac{1}{m}$$

para toda $\hat{x}, \hat{x}' \in B_\delta(\hat{x}_0)$, con $\hat{x} \neq \hat{x}'$, entonces

$$\begin{aligned}&\lim_{\hat{x} \to \hat{x}'} \frac{\hat{x} - \hat{x}' - \left(Df(\hat{x}')\right)^{-1}\left(f(\hat{x}) - f(\hat{x}')\right)}{\|f(\hat{x}) - f(\hat{x}')\|} \\ &= -\lim_{\hat{x} \to \hat{x}'} \frac{\|\hat{x} - \hat{x}'\|}{\|f(\hat{x}) - f(\hat{x}')\|}\left(Df(\hat{x}')\right)^{-1}\left(\frac{f(\hat{x}) - f(\hat{x}') - Df(\hat{x}')(\hat{x} - \hat{x}')}{\|\hat{x} - \hat{x}'\|}\right) \\ &= \hat{0}.\end{aligned}$$

Lo anterior prueba la identidad 5.21. Por lo tanto f^{-1} es derivable en $\hat{y}' = f(\hat{x}') \in f(B_\delta(\hat{x}_0))$, y además

$$\begin{aligned} Df^{-1}(\hat{y}') &= Df^{-1}(f(\hat{x}')) \\ &= (Df(\hat{x}'))^{-1} \\ &= \left(Df(f^{-1}(\hat{y}'))\right)^{-1} \end{aligned}$$

para toda $\hat{y}' \in f(B_\delta(\hat{x}_0))$.

Finalmente, probaremos que f^{-1} es de clase C^1 en $f(B_\delta(\hat{x}_0))$ y para ello usaremos el criterio de la proposición 5.17. Antes, observemos que de la desigualdad 5.20 y el hecho de que

$$Df^{-1}(f(\hat{x})) = (Df(\hat{x}))^{-1} \tag{5.22}$$

para toda $\hat{x} \in B_\delta(\hat{x}_0)$, se tiene que

$$\begin{aligned} \|\hat{z}\| &= \left\| Df(\hat{x}) \left((Df(\hat{x}))^{-1} (\hat{z}) \right) \right\| \\ &\geq m \left\| (Df(\hat{x}))^{-1} (\hat{z}) \right\| \end{aligned}$$

y por lo tanto

$$\begin{aligned} \left\| Df^{-1}(f(\hat{x}))(\hat{z}) \right\| &= \left\| (Df(\hat{x}))^{-1} (\hat{z}) \right\| \\ &\leq \frac{1}{m} \|\hat{z}\| \end{aligned} \tag{5.23}$$

para toda $\hat{z} \in \mathbb{R}^n$.

También notemos que, si $A, B \in \mathbb{R}$ son diferentes de 0, entonces

$$\begin{aligned} A^{-1} - B^{-1} &= \frac{1}{A} - \frac{1}{B} \\ &= \frac{B-A}{AB} \\ &= \frac{1}{A}(B-A)\frac{1}{B} \\ &= A^{-1}(B-A)B^{-1} \end{aligned}$$

es decir, que

$$A^{-1} - B^{-1} = A^{-1}(B-A)B^{-1}.$$

Ahora observe que esta última identidad sigue siendo válida si A y B son matrices de $n \times n$ invertibles, de modo que si $L_1, L_2 : \mathbb{R}^n \to \mathbb{R}^n$ son funciones lineales que tienen inversa, entonces también se cumple que

$$L_1^{-1} - L_2^{-1} = L_1^{-1} \circ [L_2 - L_1] \circ L_2^{-1}$$

De esta última identidad y usando 5.22 para $\hat{x}, \hat{x}' \in B_\delta(\hat{x}_0)$, se deduce que

$$\begin{aligned} Df^{-1}(f(\hat{x})) - Df^{-1}(f(\hat{x}')) &= (Df(\hat{x}))^{-1} - (Df(\hat{x}'))^{-1} \\ &= (Df(\hat{x}))^{-1} \circ [Df(\hat{x}') - Df(\hat{x})] \circ (Df(\hat{x}'))^{-1} \\ &= Df^{-1}(f(\hat{x})) \circ [(Df(\hat{x}') - Df(\hat{x}))] \circ Df^{-1}(f(\hat{x}')) \end{aligned}$$

y por la desigualdad 5.23, se tiene que

$$\begin{aligned} \left\| Df^{-1}(f(\hat{x}))(\hat{z}) - Df^{-1}(f(\hat{x}'))(\hat{z}) \right\| &= \left\| \left(Df^{-1}(f(\hat{x})) \circ [Df(\hat{x}') - Df(\hat{x})] \circ Df^{-1}(f(\hat{x}')) \right)(\hat{z}) \right\| \\ &= \left\| Df^{-1}(f(\hat{x})) \left[[Df(\hat{x}') - Df(\hat{x})] \left(Df^{-1}(f(\hat{x}'))(\hat{z}) \right) \right] \right\| \\ &\leq \frac{1}{m} \left\| [Df(\hat{x}') - Df(\hat{x})] \left(Df^{-1}(f(\hat{x}'))(\hat{z}) \right) \right\| \end{aligned} \tag{5.24}$$

para todas $\hat{x}, \hat{x}' \in B_\delta(\hat{x}_0)$ y toda $\hat{z} \in \mathbb{R}^n$.

Una vez dicho lo anterior, sean $\hat{y}' = f(\hat{x}') \in f(B_\delta(\hat{x}_0))$ y $\varepsilon > 0$. Como f es de clase C^1 en $B_\delta(\hat{x}_0)$, por la proposición 5.17 ($\Longrightarrow$), y el hecho de que $B_\delta(\hat{x}_0)$ es un abierto, sabemos que para la cantidad $m^2\varepsilon > 0$ existe $\varepsilon' > 0$ tal que $B_{\varepsilon'}(\hat{x}') \subset B_\delta(\hat{x}_0)$, y si $\hat{x} \in B_{\varepsilon'}(\hat{x}')$, entonces

$$\|[Df(\hat{x}') - Df(\hat{x})](\hat{z})\| \leq m^2\varepsilon \|\hat{z}\| \tag{5.25}$$

para toda $\hat{z} \in \mathbb{R}^n$.

Finalmente, como f^{-1} es continua en $f(B_\delta(\hat{x}_0))$ (que ya probamos que es un conjunto abierto), sabemos que para $\varepsilon' > 0$ existe $\delta' > 0$ tal que $B_{\delta'}(\hat{y}') \subset f(B_\delta(\hat{x}_0))$ y si $\hat{y} = f(\hat{x}) \in B_{\delta'}(\hat{y}')$, entonces

$$\left\|f^{-1}(\hat{y}) - f^{-1}(\hat{y}')\right\| = \|\hat{x} - \hat{x}'\| < \varepsilon'. \tag{5.26}$$

Por lo tanto, aplicando las desigualdades 5.24, 5.25 y 5.23, concluimos que

$$\begin{aligned}
\left\|Df^{-1}(\hat{y})(\hat{z}) - Df^{-1}(\hat{y}')(\hat{z})\right\| &= \left\|Df^{-1}(f(\hat{x}))(\hat{z}) - Df^{-1}(f(\hat{x}'))(\hat{z})\right\| \\
&\leq \frac{1}{m}\left\|[Df(\hat{x}') - Df(\hat{x})]\left(Df^{-1}(f(\hat{x}'))(\hat{z})\right)\right\| \\
&\leq \frac{1}{m}\left(m^2\varepsilon\left\|Df^{-1}(f(\hat{x}'))(\hat{z})\right\|\right) \\
&= m\varepsilon\left\|Df^{-1}(f(\hat{x}'))(\hat{z})\right\| \\
&\leq m\varepsilon\left(\frac{1}{m}\|\hat{z}\|\right) \\
&= \|\hat{z}\|\,\varepsilon
\end{aligned}$$

para toda $\hat{y} = f(\hat{x}) \in B_{\delta'}(\hat{y}')$ y para toda $\hat{z} \in \mathbb{R}^n$, de tal forma que nuevamente por la proposición 5.17 ($\Longleftarrow$) tenemos que f^{-1} es de clase C^1 en $f(B_\delta(\hat{x}_0))$, con lo cual terminamos la prueba. ■

El teorema de la función inversa es un resultado de caracter teórico muy importante, pero también lo es desde un punto de vista práctico. Como prueba de ello, mostraremos cómo se emplea en el siguiente problema de cambio de coordenadas: si una función f de $\mathbb{R}^3$ en $\mathbb{R}$ está dada en términos de coordenadas esféricas, nuestro problema será encontrar la expresión de su derivada en términos de las coordenadas cartesianas, sin necesidad de escribir a f en términos de éstas últimas.

Ejemplo 5.24 *Sea $g : \mathbb{R}^3 \to \mathbb{R}^3$ definida como*

$$g(\rho, \theta, \varphi) = (\rho \operatorname{sen}(\varphi)\cos(\theta), \rho\operatorname{sen}(\varphi)\operatorname{sen}(\theta), \rho\cos(\varphi))$$

Seguramente el lector estará de acuerdo en que esta función es la función de cambio de coordenadas esféricas a coordenadas cartesianas *para puntos de $\mathbb{R}^3$.*

También podemos concluir con toda certeza que g es de clase C^k en $\mathbb{R}^3$ para toda $k \in \mathbb{N}$, y por lo tanto derivable en todo punto de $\mathbb{R}^3$, con

$$Dg(\rho, \theta, \varphi) = \begin{bmatrix} \operatorname{sen}(\varphi)\cos(\theta) & -\rho\operatorname{sen}(\varphi)\operatorname{sen}(\theta) & \rho\cos(\varphi)\cos(\theta) \\ \operatorname{sen}(\varphi)\operatorname{sen}(\theta) & \rho\operatorname{sen}(\varphi)\cos(\theta) & \rho\cos(\varphi)\operatorname{sen}(\theta) \\ \cos(\varphi) & 0 & -\rho\operatorname{sen}(\varphi) \end{bmatrix}.$$

Si ahora observamos que

$$\det(Dg(\rho, \theta, \varphi)) = -\rho^2\operatorname{sen}(\varphi),$$

tendremos que $Dg(\rho, \theta, \varphi)$ es invertible para toda $(\rho, \theta, \varphi) \in \mathbb{R}^3$, si $\rho \neq 0$ y $\varphi \neq k\pi$, con $k \in \mathbb{Z}$, y su inversa (calculada por el método que mejor conozca el lector) estará dada por

$$(Dg(\rho, \theta, \varphi))^{-1} = \begin{bmatrix} \operatorname{sen}(\varphi)\cos(\theta) & \operatorname{sen}(\varphi)\operatorname{sen}(\theta) & \cos(\varphi) \\ -\frac{\operatorname{sen}(\theta)}{\rho\operatorname{sen}(\varphi)} & \frac{\cos(\theta)}{\rho\operatorname{sen}(\varphi)} & 0 \\ \frac{1}{\rho}\cos(\theta)\cos(\varphi) & \frac{1}{\rho}\cos(\varphi)\operatorname{sen}(\theta) & -\frac{1}{\rho}\operatorname{sen}(\varphi) \end{bmatrix}.$$

De esta forma, el teorema de la función inversa nos asegura que, si $(\rho_0, \theta_0, \varphi_0) \in \mathbb{R}^3$ *es tal que* $\rho_0 \neq 0$ *y* $k\pi < \varphi_0 < (k+1)\pi$ *para alguna* $k \in \mathbb{Z}$*, entonces existe* $\delta > 0$ *tal que la función* g *es inyectiva en la bola* $B_\delta(\rho_0, \theta_0, \varphi_0)$*, que la función* g^{-1} *está definida en el abierto* $g(B_\delta(\rho_0, \theta_0, \varphi_0))$ *y que para toda* $(\rho, \theta, \varphi) \in B_\delta(\rho_0, \theta_0, \varphi_0)$ *se tiene que*

$$\begin{aligned} Dg^{-1}(g(\rho, \theta, \varphi)) &= (Dg(\rho, \theta, \varphi))^{-1} \\ &= \begin{bmatrix} \operatorname{sen}(\varphi)\cos(\theta) & \operatorname{sen}(\varphi)\operatorname{sen}(\theta) & \cos(\varphi) \\ -\frac{\operatorname{sen}(\theta)}{\rho\operatorname{sen}(\varphi)} & \frac{\cos(\theta)}{\rho\operatorname{sen}(\varphi)} & 0 \\ \frac{1}{\rho}\cos(\theta)\cos(\varphi) & \frac{1}{\rho}\cos(\varphi)\operatorname{sen}(\theta) & -\frac{1}{\rho}\operatorname{sen}(\varphi) \end{bmatrix}. \end{aligned}$$

Dado que g *es la función de cambio de coordenadas esféricas a coordenadas cartesianas, entonces* g^{-1} *será la función de cambio de coordenadas cartesianas a coordenadas esféricas, función que no es tan sencilla de calcular explícitamente. Pero aún cuando no tengamos una expresión explícita para* g^{-1}*, el hecho de que tengamos una expresión para su derivada nos resulta de mucha ayuda. En efecto, si* $f : U \subset \mathbb{R}^3 \to \mathbb{R}$ *es una función de clase* C^1 *la cual está dada en términos de coordenadas esféricas, de la misma forma que hicimos en la subsección 5.3.1 con la función de cambio de coordenadas polares a cartesianas, podemos asumir que la función* $f \circ g^{-1}$ *es la misma función* f *sólo que expresada en términos de coordenadas cartesianas.*

Por tanto, recurriendo nuevamente a la regla de la cadena, tendremos que

$$\begin{aligned} \nabla\left(f \circ g^{-1}\right)(x, y, z) &= D\left(f \circ g^{-1}\right)(x, y, z) \\ &= Df\left(g^{-1}(x, y, z)\right) Dg^{-1}(x, y, z) \\ &= \nabla f\left(g^{-1}(x, y, z)\right) Dg^{-1}(x, y, z) \end{aligned}$$

de donde concluimos que

$$\begin{aligned} &\frac{\partial\left(f \circ g^{-1}\right)}{\partial x}(x, y, z) \\ &= \operatorname{sen}(\varphi)\cos(\theta)\frac{\partial f}{\partial \rho}\left(g^{-1}(x, y, z)\right) - \frac{\operatorname{sen}(\theta)}{\rho\operatorname{sen}(\varphi)}\frac{\partial f}{\partial \theta}\left(g^{-1}(x, y, z)\right) + \frac{1}{\rho}\cos(\theta)\cos(\varphi)\frac{\partial f}{\partial \varphi}\left(g^{-1}(x, y, z)\right) \\ &= \operatorname{sen}(\varphi)\cos(\theta)\left(\frac{\partial f}{\partial \rho} \circ g^{-1}\right)(x, y, z) - \frac{\operatorname{sen}(\theta)}{\rho\operatorname{sen}(\varphi)}\left(\frac{\partial f}{\partial \theta} \circ g^{-1}\right)(x, y, z) \\ &+ \frac{1}{\rho}\cos(\theta)\cos(\varphi)\left(\frac{\partial f}{\partial \varphi} \circ g^{-1}\right)(x, y, z) \end{aligned}$$

$$\begin{aligned} &\frac{\partial\left(f \circ g^{-1}\right)}{\partial y}(x, y, z) \\ &= \operatorname{sen}(\varphi)\operatorname{sen}(\theta)\frac{\partial f}{\partial \rho}\left(g^{-1}(x, y, z)\right) + \frac{\cos(\theta)}{\rho\operatorname{sen}(\varphi)}\frac{\partial f}{\partial \theta}\left(g^{-1}(x, y, z)\right) + \frac{1}{\rho}\cos(\varphi)\operatorname{sen}(\theta)\frac{\partial f}{\partial \varphi}\left(g^{-1}(x, y, z)\right) \\ &= \operatorname{sen}(\varphi)\operatorname{sen}(\theta)\left(\frac{\partial f}{\partial \rho} \circ g^{-1}\right)(x, y, z) + \frac{\cos(\theta)}{\rho\operatorname{sen}(\varphi)}\left(\frac{\partial f}{\partial \theta} \circ g^{-1}\right)(x, y, z) \\ &+ \frac{1}{\rho}\cos(\varphi)\operatorname{sen}(\theta)\left(\frac{\partial f}{\partial \varphi} \circ g^{-1}\right)(x, y, z) \end{aligned}$$

y

$$\frac{\partial\left(f \circ g^{-1}\right)}{\partial z}(x, y, z) = \cos(\varphi)\left(\frac{\partial f}{\partial \rho} \circ g^{-1}\right)(x, y, z) - \frac{1}{\rho}\operatorname{sen}(\varphi)\left(\frac{\partial f}{\partial \varphi} \circ g^{-1}\right)(x, y, z).$$

Nótese que las identidades anteriores también se pueden escribir en términos de las coordenadas esféricas (ρ, θ, φ) *(que son de las que estamos suponiendo que depende* f*). En efecto, si recordamos que* $(x, y, z) = g(\rho, \theta, \varphi)$*, (o que* $(\rho, \theta, \varphi) = g^{-1}(x, y, z)$*) se tiene que*

$$\frac{\partial\left(f \circ g^{-1}\right)}{\partial x}(x, y, z) = \operatorname{sen}(\varphi)\cos(\theta)\frac{\partial f}{\partial \rho}(\rho, \theta, \varphi) - \frac{\operatorname{sen}(\theta)}{\rho\operatorname{sen}(\varphi)}\frac{\partial f}{\partial \theta}(\rho, \theta, \varphi) + \frac{1}{\rho}\cos(\theta)\cos(\varphi)\frac{\partial f}{\partial \varphi}(\rho, \theta, \varphi)$$

$$\frac{\partial\left(f\circ g^{-1}\right)}{\partial y}(x,y,z)=\operatorname{sen}(\varphi)\operatorname{sen}(\theta)\frac{\partial f}{\partial\rho}(\rho,\theta,\varphi)+\frac{\cos(\theta)}{\rho\operatorname{sen}(\varphi)}\frac{\partial f}{\partial\theta}(\rho,\theta,\varphi)+\frac{1}{\rho}\cos(\varphi)\operatorname{sen}(\theta)\frac{\partial f}{\partial\varphi}(\rho,\theta,\varphi)$$

y

$$\frac{\partial\left(f\circ g^{-1}\right)}{\partial z}(x,y,z)=\cos(\varphi)\frac{\partial f}{\partial\rho}(\rho,\theta,\varphi)-\frac{1}{\rho}\operatorname{sen}(\varphi)\frac{\partial f}{\partial\varphi}(\rho,\theta,\varphi).$$

Si no usamos ningún tipo de coordenadas para el punto $\hat{x}$, y cometemos un abuso de notación escribiendo que $g^{-1}(\hat{x})=\hat{x}$, entonces $f=f\circ g^{-1}$, de modo que las identidades anteriores se pueden escribir como

$$\begin{aligned}\frac{\partial f}{\partial x}(\hat{x})&=\frac{\partial\left(f\circ g^{-1}\right)}{\partial x}(\hat{x})\\&=\operatorname{sen}(\varphi)\cos(\theta)\frac{\partial f}{\partial\rho}(\hat{x})-\frac{\operatorname{sen}(\theta)}{\rho\operatorname{sen}(\varphi)}\frac{\partial f}{\partial\theta}(\hat{x})+\frac{1}{\rho}\cos(\theta)\cos(\varphi)\frac{\partial f}{\partial\varphi}(\hat{x})\end{aligned}$$

$$\begin{aligned}\frac{\partial f}{\partial y}(\hat{x})&=\frac{\partial\left(f\circ g^{-1}\right)}{\partial y}(\hat{x})\\&=\operatorname{sen}(\varphi)\operatorname{sen}(\theta)\frac{\partial f}{\partial\rho}(\hat{x})+\frac{\cos(\theta)}{\rho\operatorname{sen}(\varphi)}\frac{\partial f}{\partial\theta}(\hat{x})+\frac{1}{\rho}\cos(\varphi)\operatorname{sen}(\theta)\frac{\partial f}{\partial\varphi}(\hat{x})\end{aligned}$$

y

$$\begin{aligned}\frac{\partial f}{\partial z}(\hat{x})&=\frac{\partial\left(f\circ g^{-1}\right)}{\partial z}(\hat{x})\\&=\cos(\varphi)\frac{\partial f}{\partial\rho}(\hat{x})-\frac{1}{\rho}\operatorname{sen}(\varphi)\frac{\partial f}{\partial\varphi}(\hat{x}).\end{aligned}$$

Como mencionamos en la sección anterior, con base en el teorema de la función inversa podemos dar la prueba del teorema de la función implícita, y lo más interesante es que también podemos hacer lo recíproco, es decir, probar el teorema de la función inversa a partir del teorema de la función implícita (lo que el lector hará en el problema 25 de este capítulo). Este hecho muestra que ambos teoremas son equivalentes, razón por la cual es suficiente dar la prueba de sólo uno de ellos.

Teorema 5.25 (de la función implícita) *Sean $g_1,\dots,g_m:U\subset\mathbb{R}^m\times\mathbb{R}^k\to\mathbb{R}$ de clase C^1 en U. Si*

$$S=\left\{(\hat{x},\hat{y})=(x_1,\dots,x_m,y_1,\dots,y_k)\in U\subset\mathbb{R}^m\times\mathbb{R}^k\mid g_i(\hat{x},\hat{y})=0\ \textit{para}\ i\in\{1,\dots,m\}\right\}$$

y $(\hat{x}_0,\hat{y}_0)=\left(x_1^{(0)},\dots,x_m^{(0)},y_1^{(0)},\dots,y_k^{(0)}\right)\in S$ es tal que la matriz

$$\begin{bmatrix}\frac{\partial g_1}{\partial x_1}(\hat{x}_0,\hat{y}_0)&\cdots&\frac{\partial g_1}{\partial x_m}(\hat{x}_0,\hat{y}_0)\\\vdots&\ddots&\vdots\\\frac{\partial g_m}{\partial x_1}(\hat{x}_0,\hat{y}_0)&\cdots&\frac{\partial g_m}{\partial x_m}(\hat{x}_0,\hat{y}_0)\end{bmatrix}$$

es invertible, entonces existen $\delta>0$, $V\subset\mathbb{R}^k$ un conjunto abierto, y una función $h:V\subset\mathbb{R}^k\to\mathbb{R}^m$ de clase C^1 en V, tales que $B_\delta((\hat{x}_0,\hat{y}_0))\subset U$, $\hat{y}_0\in V$, $h(\hat{y}_0)=\hat{x}_0$ y $\{(h(\hat{y}),\hat{y})\in\mathbb{R}^m\times\mathbb{R}^k\mid\hat{y}\in V\}=S\cap B_\delta((\hat{x}_0,\hat{y}_0))$. La función h es única con esta última propiedad.

Demostración. Definamos $g:U\subset\mathbb{R}^m\times\mathbb{R}^k\to\mathbb{R}^m$ como

$$g(\hat{x},\hat{y})=(g_1(\hat{x},\hat{y}),\dots,g_m(\hat{x},\hat{y}))$$

y $f:U\subset\mathbb{R}^m\times\mathbb{R}^k\to\mathbb{R}^m\times\mathbb{R}^k$ como

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

Es inmediato que f y g son de clase C^1 en U y que además:

1. como $(\hat{x}_0, \hat{y}_0) \in S$, entonces $f(\hat{x}_0, \hat{y}_0) = (g(\hat{x}_0, \hat{y}_0), \hat{y}_0) = (\hat{0}, \hat{y}_0)$, y
2. la matriz

$$Df(\hat{x}_0, \hat{y}_0) = \begin{bmatrix} \frac{\partial g_1}{\partial x_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_1}{\partial x_m}(\hat{x}_0, \hat{y}_0) & \frac{\partial g_1}{\partial y_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_1}{\partial y_k}(\hat{x}_0, \hat{y}_0) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_m}{\partial x_m}(\hat{x}_0, \hat{y}_0) & \frac{\partial g_m}{\partial y_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_m}{\partial y_k}(\hat{x}_0, \hat{y}_0) \\ 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 1 \end{bmatrix}$$

   es invertible.

Por tanto, por el teorema de la función inversa sabemos que existe $\delta > 0$ tal que:

1. f es inyectiva en $B_\delta(\hat{x}_0, \hat{y}_0) = \left\{(\hat{x}, \hat{y}) \in \mathbb{R}^m \times \mathbb{R}^k \mid \|(\hat{x}, \hat{y}) - (\hat{x}_0, \hat{y}_0)\| < \delta\right\} \subset U$,
2. $f(B_\delta(\hat{x}_0, \hat{y}_0)) \subset \mathbb{R}^m \times \mathbb{R}^k$ es un conjunto abierto, y
3. $f^{-1} : f(B_\delta(\hat{x}_0, \hat{y}_0)) \subset \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}^m \times \mathbb{R}^k$ es de clase C^1 en $f(B_\delta(\hat{x}_0, \hat{y}_0))$.

Definimos ahora $V \subset \mathbb{R}^k$ como

$$V = \left\{\hat{y} \in \mathbb{R}^k \mid (\hat{x}, \hat{y}) \in S \cap B_\delta(\hat{x}_0, \hat{y}_0) \text{ para alguna } \hat{x} \in \mathbb{R}^m\right\}.$$

Nótese que $\hat{y}_0 \in V$, puesto que $(\hat{x}_0, \hat{y}_0) \in S \cap B_\delta(\hat{x}_0, \hat{y}_0)$, y además se tiene que

$$f(S \cap B_\delta(\hat{x}_0, \hat{y}_0)) = \left\{\hat{0}\right\} \times V \subset \mathbb{R}^m \times \mathbb{R}^k. \tag{5.27}$$

En efecto, si $f(\hat{x}, \hat{y}) \in f(S \cap B_\delta(\hat{x}_0, \hat{y}_0))$, como $(\hat{x}, \hat{y}) \in S \cap B_\delta(\hat{x}_0, \hat{y}_0)$ se tiene que $\hat{y} \in V$ y como $(\hat{x}, \hat{y}) \in S$ entonces $f(\hat{x}, \hat{y}) = (\hat{0}, \hat{y}) \in \left\{\hat{0}\right\} \times V$. Por otra parte, si $(\hat{0}, \hat{y}) \in \left\{\hat{0}\right\} \times V$, como $\hat{y} \in V$, entonces existe $\hat{x} \in \mathbb{R}^m$ tal que $(\hat{x}, \hat{y}) \in S \cap B_\delta(\hat{x}_0, \hat{y}_0)$ de modo que $(\hat{0}, \hat{y}) = f(\hat{x}, \hat{y}) \in f(S \cap B_\delta(\hat{x}_0, \hat{y}_0))$.

De la identidad 5.27 obtenemos fácilmente que el conjunto $V \subset \mathbb{R}^m$ es abierto. En efecto, si $\hat{y} \in V$, entonces $(\hat{0}, \hat{y}) \in f(S \cap B_\delta(\hat{x}_0, \hat{y}_0)) \subset f(B_\delta(\hat{x}_0, \hat{y}_0))$, y como este último conjunto es abierto, existe $r > 0$ tal que $B_r(\hat{0}, \hat{y}) \subset f(B_\delta(\hat{x}_0, \hat{y}_0))$. De esta forma, si $\hat{y}' \in B_r(\hat{y})$ se tiene que

$$\left\|(\hat{0}, \hat{y}') - (\hat{0}, \hat{y})\right\| = \|\hat{y}' - \hat{y}\| < r,$$

de modo que $(\hat{0}, \hat{y}') \in B_r(\hat{0}, \hat{y}) \subset f(B_\delta(\hat{x}_0, \hat{y}_0))$, lo cual implica que existe $\hat{x}' \in \mathbb{R}^m$ tal que $(\hat{x}', \hat{y}') \in B_\delta(\hat{x}_0, \hat{y}_0)$ y

$$(\hat{0}, \hat{y}') = f(\hat{x}', \hat{y}') = (g(\hat{x}', \hat{y}'), \hat{y}').$$

De lo anterior se deduce que $g(\hat{x}', \hat{y}') = \hat{0}$, de modo que $(\hat{x}', \hat{y}') \in S \cap B_\delta(\hat{x}_0, \hat{y}_0)$. De la definición de V se tiene que $\hat{y}' \in V$ y por lo tanto que $B_r(\hat{y}) \subset V$. Esta contención prueba que V es abierto.

Ahora, por la inyectividad de f en $B_\delta(\hat{x}_0, \hat{y}_0)$ también concluimos que para cada $\hat{y} \in V$ existe una única $\hat{x} \in \mathbb{R}^m$ tal que $(\hat{x}, \hat{y}) \in S \cap B_\delta(\hat{x}_0, \hat{y}_0)$. En efecto, si $\hat{x}, \hat{x}' \in \mathbb{R}^m$ son tales que $(\hat{x}, \hat{y}), (\hat{x}', \hat{y}) \in S \cap B_\delta(\hat{x}_0, \hat{y}_0)$, entonces $f(\hat{x}, \hat{y}) = (\hat{0}, \hat{y}) = f(\hat{x}', \hat{y})$ de donde, por la inyectividad de f en la vecindad $B_\delta(\hat{x}_0, \hat{y}_0)$, se tiene que $\hat{x} = \hat{x}'$.

Con base en lo anterior, definimos $h : V \subset \mathbb{R}^k \to \mathbb{R}^m$ de la siguiente forma: dada $\hat{y} \in V$ hacemos $h(\hat{y}) = \hat{x}$, en donde $\hat{x} \in \mathbb{R}^m$ es el único elemento de $\mathbb{R}^m$ para el cual se satisface que $(\hat{x}, \hat{y}) \in S \cap B_\delta(\hat{x}_0, \hat{y}_0)$. Como el lector podrá notar fácilmente, de la discusión anterior se tiene que h está bien definida y además es la única que se puede definir sobre el conjunto V con las propiedades de que $h(\hat{y}_0) = \hat{x}_0$, y $\{(h(\hat{y}), \hat{y}) \in \mathbb{R}^m \times \mathbb{R}^k \mid \hat{y} \in V\} = S \cap B_\delta((\hat{x}_0, \hat{y}_0))$.

Ahora sólo resta probar que h es de clase C^1 en V. Para ello, dado que la función f^{-1} sí está definida y es de clase C^1 sobre el conjunto $f(B_\delta(\hat{x}_0, \hat{y}_0))$, si escribimos que $f^{-1} = \left(\left(f^{-1}\right)_1, \ldots, \left(f^{-1}\right)_{m+k}\right)$, bastará mostrar que

$$h(\hat{y}) = \left(\left(f^{-1}\right)_1(\hat{0}, \hat{y}), \ldots, \left(f^{-1}\right)_m(\hat{0}, \hat{y})\right)$$

para cada $\hat{y} \in V$.

La identidad anterior se obtiene inmediatamente si observamos que para cada $\hat{y} \in V$ se cumple que

$$
\begin{aligned}
(h(\hat{y}), \hat{y}) &= f^{-1}(f(h(\hat{y}), \hat{y})) \\
&= f^{-1}\left((\hat{0}, \hat{y})\right) \\
&= \left(\left(f^{-1}\right)_1(\hat{0}, \hat{y}), \ldots, \left(f^{-1}\right)_m(\hat{0}, \hat{y}), \left(f^{-1}\right)_{m+1}(\hat{0}, \hat{y}), \ldots, \left(f^{-1}\right)_{m+k}(\hat{0}, \hat{y})\right),
\end{aligned}
$$

con lo cual terminamos nuestra prueba. ■

Con este teorema concluimos esta sección, este capítulo ¡y este texto!

5.6 Problemas

1. Sea $f : U \subset \mathbb{R}^2 \to \mathbb{R}$ derivable en cada punto $\hat{x} \in U$. Pruebe que:
   (a) la gráfica de f (G_f) es una superficie suave en todos sus puntos
   (b) si $\hat{x}_0 = (x_0, y_0, f(x_0, y_0)) \in G_f$, el plano tangente a G_f en $\hat{x}_0$ calculado de acuerdo con la definición 4.18 es el mismo que se obtiene si se calcula de acuerdo con la definición 5.4.

2. Calcule una ecuación cartesiana del plano tangente en un punto, de una superficie que está parametrizada por una función $\sigma : U \subset \mathbb{R}^2 \to \mathbb{R}^3$ (definición 5.4).

3. Pruebe que la proposición 5.6 es independiente de las funciones coordenadas de f que se tomen.

4. Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ derivable en el punto $\hat{x}_0 \in U$. Pruebe que existen $r > 0$ y $M > 0$ tales que
$$\frac{\|f(\hat{x}) - f(\hat{x}_0)\|}{\|\hat{x} - \hat{x}_0\|} \leq M$$
para toda $\hat{x} \in (B_r(\hat{x}_0) \setminus \{\hat{x}_0\}) \subset U$.

5. Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$, $g : V \subset \mathbb{R}^m \to \mathbb{R}^k$ y $\hat{x}_0 \in U$ tales que $f(U) \subset V$. Definimos $\varphi : U \subset \mathbb{R}^n \to \mathbb{R}^k$ como
$$\varphi(\hat{x}) = \begin{cases} \frac{g(f(\hat{x})) - g(f(\hat{x}_0)) - Dg(f(\hat{x}_0))(f(\hat{x}) - f(\hat{x}_0))}{\|f(\hat{x}) - f(\hat{x}_0)\|} & \text{si } f(\hat{x}) - f(\hat{x}_0) \neq \hat{0} \\ \hat{0} & \text{si } f(\hat{x}) - f(\hat{x}_0) = \hat{0} \end{cases}$$
Pruebe que, si f es continua en $\hat{x}_0$ y g es derivable en $\hat{y}_0 = f(\hat{x}_0)$, entonces φ es continua en $\hat{x}_0$.

6. Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$ y $g : U \subset \mathbb{R}^n \to \mathbb{R}^m$ derivables en el punto $\hat{x}_0 \in U$. Pruebe que la función $(fg)(\hat{x}) = f(\hat{x})g(\hat{x})$ es derivable en $\hat{x}_0 \in U$ y dé una fórmula para la $D(fg)(\hat{x}_0)$.

7. Si $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \end{bmatrix} \in M_{1\times 3}(\mathbb{R})$ y
$$B = \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ b_{21} & \cdots & b_{2n} \\ b_{31} & \cdots & b_{3n} \end{bmatrix} \in M_{3\times n}(\mathbb{R}),$$
definimos el producto cruz de la matriz A por la matriz B (que denotaremos por $A \times B$), como la matriz de $3 \times n$ (con entradas reales) cuyas entradas de su $j-$ésima columna coinciden con las coordenadas del vector
$$(a_{11}, a_{12}, a_{13}) \times (b_{1j}, b_{2j}, b_{3j}) = (a_{12}b_{3j} - a_{13}b_{2j}, a_{13}b_{1j} - a_{11}b_{3j}, a_{11}b_{2j} - a_{12}b_{1j})$$
es decir,
$$A \times B := \begin{bmatrix} a_{12}b_{31} - a_{13}b_{21} & \cdots & a_{12}b_{3n} - a_{13}b_{2n} \\ a_{13}b_{11} - a_{11}b_{31} & \cdots & a_{13}b_{1n} - a_{11}b_{3n} \\ a_{11}b_{21} - a_{12}b_{11} & \cdots & a_{11}b_{2n} - a_{12}b_{1n} \end{bmatrix} \in M_{3\times n}(\mathbb{R})$$

Sean $f = (f_1, f_2, f_3)$, $g = (g_1, g_2, g_3) : U \subset \mathbb{R}^n \to \mathbb{R}^3$ derivables en el punto $\hat{x}_0 \in U$. Pruebe que la función $(f \times g)(\hat{x}) := f(\hat{x}) \times g(\hat{x})$ es derivable en $\hat{x}_0 \in U$ y dé una fórmula para $D(f \times g)(\hat{x}_0)$ en términos del producto cruz de matrices definido en el párrafo anterior.

8. Encuentre todas las funciones $f : \mathbb{R}^n \to \mathbb{R}^n$ tales que $Df(\hat{x})$ es una matriz diagonal para toda $\hat{x} \in \mathbb{R}^n$. Pruebe su respuesta.

9. Abusar de la notación (como de cualquier otra cosa) suele causar problemas. En particular, usar letras (que casi siempre denotan variables) para referirse también a funciones nos puede llevar a errores. Sea $w = f(x, y, z)$ y $z = g(x, y)$. Por la regla de la cadena, se tiene que:

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

Como x y y son variables independientes, entonces $\frac{\partial y}{\partial x} = 0$ y como $\frac{\partial x}{\partial x} = 1$, se tiene que:

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

Así, $\frac{\partial w}{\partial z}\frac{\partial z}{\partial x} = 0$. Si $f(x, y, z) = 2x + y + 3z$ y $g(x, y) = 5x + 18$, entonces $\frac{\partial w}{\partial z} = 3$ y $\frac{\partial z}{\partial x} = 5$ y por lo tanto $0 = 15$. ¿Cuál es el error?

10. Sean f y g definidas como

$$f(u, v) = (u\cos(v), u\,\mathrm{sen}(v))$$

con $0 < u$ y $-\pi/2 < v < \pi/2$, y

$$g(x, y) = (\sqrt{x^2 + y^2}, \arctan(y/x))$$

con $0 < x$. Calcule $D(f \circ g)(x, y)$ y $D(g \circ f)(u, v)$.

11. (a) Suponga que la variable w está en función de las variables x, y, z y t (es decir: $w = f(x, y, z, t)$), que $x = g(u, z, t)$ y que $z = h(u, t)$. Tomando en cuenta todas estas relaciones, calcule $\frac{\partial w}{\partial t}$

    (b) si

$$\begin{aligned} f(x, y, z, t) &= 2xy + 3z + t^2 \\ g(u, z, t) &= ut\,\mathrm{sen}(z) \\ h(u, t) &= 2u + t \end{aligned}$$

    calcule $\frac{\partial w}{\partial t}$ para $u = 1, t = 2$ y $y = 3$.

12. Suponga que $f : \mathbb{R}^2 \to \mathbb{R}$ es tal que $\frac{\partial f}{\partial x}(2, 1) = 3$, $\frac{\partial f}{\partial y}(2, 1) = -2$, $\frac{\partial^2 f}{\partial x^2}(2, 1) = 0$, $\frac{\partial^2 f}{\partial y \partial x}(2, 1) = \frac{\partial^2 f}{\partial x \partial y}(2, 1) = 1$ y $\frac{\partial^2 f}{\partial y^2}(2, 1) = 2$. Si $g(u, v) = (u^2 + v, uv)$, calcule $\frac{\partial^2 (f \circ g)}{\partial v \partial u}(1, 1)$.

13. Sean $f, g : U \subset \mathbb{R}^2 \to \mathbb{R}$ de clase C^1 en U. Definimos $F : U \subset \mathbb{R}^2 \to \mathbb{R}^2$ y $H : U \subset \mathbb{R}^2 \to \mathbb{R}$ como

$$F(x, y) = (f(x, y), g(x, y)) \qquad \text{y} \qquad H(x, y) = \|F(x, y)\|^2$$

Demuestre que no existe $\hat{x}_0 \in U$ que satisfaga las siguientes dos propiedades:

    (a) H tiene un máximo local en $\hat{x}_0$

    (b) $DF(\hat{x}_0)$ es invertible.

14. Sea $f : \mathbb{R}^n \to \mathbb{R}^n$ derivable en un punto $\hat{x}_0 \in \mathbb{R}^n$ y además supóngase que $\hat{x}_0$ es un punto fijo de f (es decir: $f(\hat{x}_0) = \hat{x}_0$). Si A denota a la matriz $Df(\hat{x}_0)$ y $k \in \mathbb{N}$, encuentre una función $g : \mathbb{R}^n \to \mathbb{R}^n$ derivable en $\hat{x}_0$ tal que $g(\hat{x}_0) = \hat{x}_0$ y $Dg(\hat{x}_0) = A^k$.

15. Sea $f : U \subset \mathbb{R}^3 \to \mathbb{R}$ una función que está expresada en términos de las coordenadas cartesianas (x, y, z) de cada punto $\hat{x} \in U$. Use la función de cambio de coordenadas cilíndricas (ρ, θ, z) a coordenadas cartesianas (x, y, z), y la regla de la cadena, para encontrar (en cada punto $\hat{x} \in U$) una base ortonormal de $\mathbb{R}^3$ en la cual se pueda expresar a la derivada de f en $\hat{x}$ $(Df(\hat{x}))$ en términos de las derivadas parciales $\frac{\partial f}{\partial \rho}$, $\frac{\partial f}{\partial \theta}$ y $\frac{\partial f}{\partial z}$. Compare con lo obtenido en el problema 24 del capítulo 4.

16. Repita el problema anterior ahora para las coordenadas esféricas (ρ, θ, φ) y compare con lo obtenido en el problema 25 del capítulo 4.

17. Sean $f : U \subset \mathbb{R}^2 \to \mathbb{R}$ de clase C^2 en U tal que $\frac{\partial f}{\partial x}(x, y) > 0$ para toda $(x, y) \in U$, y $(x_0, y_0) \in U$ tal que $\frac{\partial^2 f}{\partial y^2}(x_0, y_0) \neq 0$. Haga $I = \{y \in \mathbb{R} \mid (x_0, y) \in U\}$.

    (a) Pruebe que existen $\delta > 0$ y $h : (y_0 - \delta, y_0 + \delta) \subset I \to \mathbb{R}$ de clase C^1 tal que $f(h(y), y) = f(x_0, y_0)$ para toda $y \in (y_0 - \delta, y_0 + \delta)$.

    (b) Si definimos $g : I \subset \mathbb{R} \to \mathbb{R}$ como $g(y) = f(x_0, y)$, pruebe que: g tiene un máximo (mínimo) local en y_0 si y sólo si h tiene un mínimo (máximo) local en y_0.

    Interprete geométricamente.

18. Sean $g_1, \ldots, g_m : U \subset \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}$, S y $(\hat{x}_0, \hat{y}_0) \in S$ como en el teorema de la función implícita. Si

$$A = \begin{bmatrix} \frac{\partial g_1}{\partial x_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_1}{\partial x_m}(\hat{x}_0, \hat{y}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_m}{\partial x_m}(\hat{x}_0, \hat{y}_0) \end{bmatrix}$$

    y

$$B = \begin{bmatrix} \frac{\partial g_1}{\partial y_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_1}{\partial y_k}(\hat{x}_0, \hat{y}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial y_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_m}{\partial y_k}(\hat{x}_0, \hat{y}_0) \end{bmatrix}$$

    pruebe que:

$$Dh(\hat{y}_0) = -A^{-1}B,$$

    donde h es la función cuya existencia es garantizada en el mencionado teorema.

19. Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}$ de clase C^1 en U y $S = \{\hat{x} \in U \mid f(\hat{x}) = cte\}$. Si $\hat{x} \in \mathbb{R}^n$, denotamos por $\hat{x}^{(-i)}$ el elemento de $\mathbb{R}^{n-1}$ que se obtiene de $\hat{x}$ al "eliminarle" su $i-$ésima coordenada, con $i \in \{1, \ldots, n\}$. Sea $\hat{x}_0 = (x_0^{(1)}, \ldots, x_0^{(n)}) \in S$.

    (a) pruebe que, si $\frac{\partial f}{\partial x_i}(\hat{x}_0) \neq 0$, entonces existe $h_i : U_i \subset \mathbb{R}^{n-1} \to \mathbb{R}$ de clase C^1 en U_i tal que $\hat{x}_0^{(-i)} \in U_i$, $h_i(\hat{x}_0^{(-i)}) = x_0^{(i)}$ y

$$(x_1, \ldots, x_{i-1}, h_i(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n), x_{i+1}, \ldots, x_n) \in S$$

    para todo $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \in U_i$

    (b) si h_i es la función del inciso anterior, calcule $\frac{\partial h_i}{\partial x_{i+1}}(\hat{x}_0^{(-i)})$ (para $i = n$ calcule $\frac{\partial h_n}{\partial x_1}(\hat{x}_0^{(-n)})$) en términos de derivadas parciales de f

    (c) si $\frac{\partial f}{\partial x_i}(\hat{x}_0) \neq 0$ para cada $i \in \{1, \ldots, n\}$, pruebe que

$$\frac{\partial h_1}{\partial x_2}(\hat{x}_0^{(-1)})\frac{\partial h_2}{\partial x_3}(\hat{x}_0^{(-2)}) \cdots \frac{\partial h_{n-1}}{\partial x_n}(\hat{x}_0^{(-(n-1))})\frac{\partial h_n}{\partial x_1}(\hat{x}_0^{(-n)}) = (-1)^n$$

20. Sea $g : U \subset \mathbb{R}^n \to \mathbb{R}$ de clase C^1 (en U) y $\hat{x}_0 \in U$ tal que $g(\hat{x}_0) = 0$ y $\frac{\partial g}{\partial x_n}(\hat{x}_0) \neq 0$. Si $\hat{v} \neq \hat{0}$ es tal que $\hat{v} \cdot \nabla g(\hat{x}_0) = 0$, pruebe que existen $\delta > 0$ y

$$\gamma : (-\delta, \delta) \subset \mathbb{R} \to \mathbb{R}^n$$

derivable tales que $g(\gamma(t)) = 0$ para toda $t \in (-\delta, \delta)$, $\gamma(0) = \hat{x}_0$ y $\gamma'(0) = \hat{v}$. Interprete geométricamente.

21. Sea $g : U \subset \mathbb{R}^3 \to \mathbb{R}$ de clase C^1 (en U) y $\hat{x}_0 = (x_0, y_0, z_0) \in U$ tal que $g(\hat{x}_0) = 0$ y $\frac{\partial g}{\partial z}(\hat{x}_0) \neq 0$. Pruebe que:

    (a) existe $V \subset \mathbb{R}^2$ abierto, y $\sigma : V \subset \mathbb{R}^2 \to \mathbb{R}^3$ de clase C^1 en V, tales que $(x_0, y_0) \in V$ y $g(\sigma(x, y)) = 0$ para toda $(x, y) \in V$. Interprete geométricamente

    (b) el plano tangente al conjunto de nivel 0 de g ($N_0(g)$) calculado usando la parametrización σ del inciso anterior es el mismo si se calcula usando la definición 4.33

22. Sean $g_1, g_2 : U \subset \mathbb{R}^3 \to \mathbb{R}$ de clase C^1 (en U) y $\hat{x}_0 = (x_0, y_0, z_0) \in U$ tales que $g_1(\hat{x}_0) = g_2(\hat{x}_0) = 0$ y $\nabla g_1(\hat{x}_0) \times \nabla g_2(\hat{x}_0) \neq \hat{0}$. Pruebe que:

    (a) existe $\delta > 0$ y

    $$\gamma : (-\delta, \delta) \subset \mathbb{R} \to \mathbb{R}^3$$

    derivable tal que $g_1(\gamma(t)) = g_2(\gamma(t)) = 0$ para toda $t \in (-\delta, \delta)$, $\gamma(0) = \hat{x}_0$ y $\gamma'(0) \neq \hat{0}$

    (b) la recta tangente en el punto $\gamma(0)$ a la curva descrita por γ, es la misma recta que se obtiene al intersectar al plano tangente en el punto $\hat{x}_0$ del conjunto de nivel 0 de la función g_1, con el plano tangente en el punto $\hat{x}_0$ del conjunto de nivel 0 de la función g_2. Interprete geométricamente

23. Considere el conjunto de soluciones de las ecuaciones:

$$\begin{aligned} 2x + y + 2z + u - v - 1 &= 0 \\ xy + z - u + 2v - 1 &= 0 \\ yz + xz + u^2 + v &= 0 \end{aligned}$$

    (a) Muestre que, en una vecindad del punto $(1, 1, -1, 1, 1)$, las variables x, y y z (del conjunto de soluciones) se pueden poner en función de las variables u y v.

    (b) Calcule la derivada de la función del inciso anterior en el punto $(1, 1)$.

    (c) Encuentre, usando el teorema de la función implícita, todas la ternas de variables que se puedan poner en función de las restantes dos en una vecindad del mismo punto.

24. Sea

$$f(x) = \begin{cases} x^2 \operatorname{sen}(1/x) + \frac{x}{2} & \text{si } x \neq 0 \\ 0 & \text{si } x = 0 \end{cases}$$

    (a) calcule $f'(x)$ para toda $x \in \mathbb{R}$

    (b) pruebe que para toda $\delta > 0$ existen $x, y \in (-\delta, \delta)$ tales que $f'(x) < 0$ y $f'(y) > 0$

    (c) pruebe que f no es invertible en ninguna vecindad del cero. ¿Este ejemplo contradice el Teorema de la Función Inversa?

25. Pruebe el teorema de la función inversa a partir del teorema de la función implícita.

26. Sean $f : \mathbb{R}^n \to \mathbb{R}^n$ de clase C^1 y $c > 0$ tales que para toda $\hat{x}, \hat{y} \in \mathbb{R}^n$ se tiene que

$$\|f(\hat{x}) - f(\hat{y})\| \geq c \, \|\hat{x} - \hat{y}\|$$

    Pruebe que:

(a) para toda $\hat{x} \in \mathbb{R}^n$, $Df(\hat{x})$ es invertible
(b) $f(\mathbb{R}^n)$ es abierto
(c) $f(\mathbb{R}^n)$ es cerrado
(d) f es biyectiva y $f^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ es de clase C^1 (en $\mathbb{R}^n$).

27. Sea $g : U \subset \mathbb{R}^n \to \mathbb{R}^m$ de clase C^1 (en U) con $n > m$. Pruebe que g no es inyectiva.

28. Sea $g = (g_1, \dots, g_m) : A \subset \mathbb{R}^n \to \mathbb{R}^m$ de clase C^1 (en A) con $n > m$. Sea $\hat{x}_0 \in A$ tal que $g(\hat{x}_0) = \hat{0}$.

    (a) Si la matriz
$$M = \begin{bmatrix} \frac{\partial g_1}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial g_1}{\partial x_m}(\hat{x}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial g_m}{\partial x_m}(\hat{x}_0) \end{bmatrix}$$
es de rango m, pruebe que existen $U, V \subset \mathbb{R}^n$ abiertos y $f : U \to V$ una biyección de clase C^1 en U (y f^{-1} de clase C^1 en V) tales que $\hat{x}_0 \in V \subset A$ y $g(f(x_1, \dots, x_n)) = (x_1, \dots, x_m)$ para todo $(x_1, \dots, x_n) \in U$

    (b) Describa el conjunto $(g \circ f)^{-1}(\{\hat{0}\})$

    (c) Use el primer inciso para demostrar que si $Dg(\hat{x}_0)$ tiene rango máximo (m), entonces $Dg(\hat{x})$ tiene rango máximo para $\hat{x}$ en una vecindad de $\hat{x}_0$

    (d) Si $Dg(\hat{x}_0)$ no tiene rango máximo (es decir, si el rango de $Dg(\hat{x}_0)$ es $k < m$) ¿se sigue cumpliendo un resultado análogo al del primer inciso? ¿al del tercer inciso? (es decir, si $Dg(\hat{x}_0)$ tiene rango $k < m$ ¿existe una vecindad de $\hat{x}_0$ en la que Dg sigue teniendo rango k?). Pruebe sus respuestas

    (e) Interprete geométricamente.

29. Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ de clase C^1 (en U) con $n < m$.

    (a) Si $\hat{x}_0 = \left(x_1^{(0)}, \dots, x_n^{(0)}\right) \in U$ es tal que las primeras n columnas de la matriz $Df(\hat{x}_0)$ son linealmente independientes, pruebe que existen $\tilde{U}, \tilde{V} \subset \mathbb{R}^m$ abiertos y $h : \tilde{U} \to \tilde{V}$ una biyección de clase C^1 en $\tilde{U}$ (y h^{-1} de clase C^1 en $\tilde{V}$) tales que $\left(x_1^{(0)}, \dots, x_n^{(0)}, 0, \dots, 0\right) \in \tilde{V}, f(\hat{x}_0) \in \tilde{U}$ y $h(f(x_1, \dots, x_n)) = (x_1, \dots, x_n, 0, \dots, 0) \in \mathbb{R}^m$ para todo $(x_1, \dots, x_n) \in f^{-1}(\tilde{U})$ tal que $(x_1, \dots, x_n, 0, \dots, 0) \in \tilde{V}$

    (b) Use el primer inciso para demostrar que si $Df(\hat{x}_0)$ tiene rango máximo (n), entonces $Df(\hat{x})$ tiene rango máximo para $\hat{x}$ en una vecindad de $\hat{x}_0$

    (c) Si $Df(\hat{x}_0)$ no tiene rango máximo (es decir, si el rango de $Df(\hat{x}_0)$ es $k < n$) ¿se sigue cumpliendo un resultado equivalente al del primer inciso? ¿al del segundo inciso? (es decir, si $Df(\hat{x}_0)$ tiene rango $k < n$ ¿existe una vecindad de $\hat{x}_0$ en la que Df sigue teniendo rango k?). Pruebe sus respuestas

    (d) Interprete geométricamente.

30. Sea $f(x, y) = (e^x \cos(y), e^x \operatorname{sen}(y))$.

    (a) Pruebe que f no es inyectiva en $\mathbb{R}^2$
    (b) Pruebe que para cualquier $\hat{x} = (x, y) \in \mathbb{R}^2$ existe $\delta > 0$ tal que f es invertible en $B_\delta(\hat{x})$
    (c) Si $(u_0, y_0) = f(x_0, y_0)$, calcule $Df^{-1}(u_0, y_0)$ en términos de u_0 y v_0
    (d) ¿Existe una función de $\mathbb{R}$ en $\mathbb{R}$ que tenga las dos propiedades anteriores? Pruebe su respuesta.

31. Sean $f, g : U \subset \mathbb{R}^3 \to \mathbb{R}$ de clase C^1 en U. Pruebe que la función
$$F(x, y, z) = (f(x, y, z), g(x, y, z), f(x, y, z) + g(x, y, z))$$
es tal que, si tuviera inversa en alguna vecindad de algún punto de U, ésta no sería derivable.

32. Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}^n$ de clase C^1 (en U) y $A \subset U$. Pruebe que, si $\det(Df(\hat{x})) \neq 0$ para toda $\hat{x} \in int(A)$, entonces $f(int(A)) \subset int(f(A))$.

33. Sea $f : U \subset \mathbb{R}^3 \to \mathbb{R}$ de clase C^2 en U. Si f está expresada en términos de coordenadas esféricas, calcule $\frac{\partial^2 f}{\partial z^2}(\hat{x})$ en términos de estas mismas coordenadas (*sugerencia*: proceda como en el ejemplo 5.24).