Sección de Matemáticas
Universidad de La Laguna

Amadou Diadie Coulibaly

Los Teoremas de la Función Inversa y de la Función Implícita

Inverse Function and Implicit Function Theorems

Trabajo Fin de Grado
Grado en Matemáticas
La Laguna, septiembre de 2022

Dirigido por
Jorge Betancor Pérez
Lourdes Rodríguez Mesa

Jorge Betancor Pérez
Departamento de Análisis Matemático
Universidad de La Laguna
38200 La Laguna, Tenerife

Lourdes Rodríguez Mesa
Departamento de Análisis Matemático
Universidad de La Laguna
38200 La Laguna, Tenerife

Resumen · Abstract

Resumen

En esta memoria analizamos dos teoremas fundamentales del Cálculo Diferencial: El Teorema de la Función Inversa y el Teorema de la Función Implícita. Consideramos primero algunos aspectos generales sobre espacios normados y métricos, y en particular, sobre el espacio de las matrices $M_n(\mathbb{R})$, $n \in \mathbb{N}$. *Asimismo, abordamos el estudio del Teorema del punto fijo de Banach, que resulta ser un elemento fundamental en la demostración del Teorema de la función inversa, así como también estudiamos el Teorema del valor medio en el contexto vector-valuado. Por último enunciamos y probamos los dos teoremas principales de esta memoria, presentando algunas de sus consecuencias, entre las que figura la versión del teorema de la función inversa para funciones holomorfas.*

Palabras clave: *Función inversa local – Función implícita – Teorema del punto fijo de Banach – Teorema del valor medio.*

Abstract

In this work we analyze two essential theorems in Differential Calculus: The Inverse Function Theorem and the Implicit Function Theorem. We first consider some general aspects on normed vector and metric spaces, and particularly, on matrix spaces $M_n(\mathbb{R})$, $n \in \mathbb{N}$. *Likewise, we address the study of Banach Fixed Point Theorem, which turns out to be a fundamental element in the proof of the Inverse Function Theorem, and we also study the Mean Value Theorem in the vector-valued context. Finally we state and prove the two main aforementioned theorems, showing some of their consequences, among which is the version of the Inverse Function Theorem for holomorphic functions.*

Keywords: *Local inverse function – Implicit function – Banach fixed-point theorem – Mean Value Theorem.*

Contenido

Introducción

La presente memoria está dedicada a dos resultados relevantes dentro del Cálculo Diferencial: el Teorema de la función inversa y el Teorema de la función implícita. Ambos teoremas constituyen unos de los más importantes, y de los más antiguos, paradigmas en las matemáticas modernas. Si nos remontamos al siglo XVII podemos encontrar textos de Isaac Newton (1642-1727) donde aparece el germen de la idea que se encuentra detrás del teorema de la función implícita, incluso en el trabajo de Gottfried Leibniz (1646-1716) se muestra explícitamente un ejemplo de derivación implícita. Más tarde, ya entre los siglos XVIII y XIX, mientras que Joseph Louis Lagrange (1736-1813) presenta un resultado que puede verse esencialmente como una versión del teorema de la función inversa, el matemático Augustin-Louis Cauchy (1789-1857) aborda el estudio del teorema de la función implícita con rigor matemático, hecho por el cual se le atribuye su descubrimiento. Estos teoremas siguen siendo fundamentales en la actualidad, constituyen, por ejemplo, una herramienta teórica básica para el estudio de la variedades diferenciables.

Ambos resultados pueden pensarse como formulaciones complementarias equivalentes de la misma idea. De hecho, pueden deducirse uno a partir del otro. El Teorema de la función inversa es técnicamente un teorema de existencia local de la inversa de una función. Determinar explícitamente la expresión de la función inversa, cuando existe, no suele ser tarea fácil. Es deseable poder encontrar un criterio que permita garantizar, por una parte, que existe la inversa de una función, al menos localmente, y por otra parte, que esta función inversa hereda las buenas propiedades de la función de partida (continuidad, diferenciabilidad,...). El Teorema de la función inversa proporciona condiciones suficientes, en términos de las derivadas en un punto, que aseguran que una función sea invertible en un entorno de dicho punto.

Por otro lado el Teorema de la función implícita se relaciona con la cuestión de tratar de resolver un sistema de ecuaciones. El asunto se puede formular como un problema de existencia de una función definida implícitamente por una ecuación. El Teorema de la función implícita proporciona condiciones suficientes para abordar con éxito este problema.

Hemos dividido el trabajo en dos capítulos. En el primero presentamos los elementos fundamentales de la teoría que son necesarios para el desarrollo del siguiente capítulo. En concreto, se introducen los conceptos de espacio normado y espacio métrico, se abordan sus principales propiedades y se proporcionan algunos ejemplos, prestando una atención particular al espacio de las matrices al que dedicamos un apartado. Es también en este primer capítulo donde presentamos y demostramos el Teorema del punto fijo de Banach, una herramienta de gran utilidad para la prueba del teorema de la función inversa.

En el segundo capítulo se desarrolla el estudio de los dos teoremas centrales de la memoria. En un primer apartado tratamos el Teorema de la función inversa en el contexto de las funciones reales de variable real, presentando previamente diversos resultados en relación a la monotonía y la continuidad de las funciones. Para el análisis del Teorema de la función inversa en el contexto vector-valuado abordamos primero algunos resultados en relación al Teorema del valor medio para funciones de varias variables. Asimismo, analizamos la existencia de la función inversa, tanto global como local, para algunas funciones particulares. Con estos ejemplos podemos examinar las condiciones que se exigen después en el Teorema de la función inversa. Y terminamos la memoria con el enunciado y demostración de los Teoremas de la función inversa y de la función implícita, aportando algunas consecuencias, como la versión del teorema de la función inversa para funciones holomorfas.

Para el desarrollo de este trabajo nos hemos apoyado en las referencias indicadas en la bibliografía, de las que destacamos las obras de Edwards [1], Fleming [2] y Flores y Sadarangani [3].

1

Espacios normados y métricos

1.1. Espacios normados

Sea X un espacio vectorial sobre $\mathbb{R}$. Una *norma* en X es una aplicación $\|\cdot\| : X \longrightarrow \mathbb{R}$, que verifica las condiciones siguientes:

(a) *no degeneración*: si $x \in X$, $\|x\| = 0$, entonces $x = 0$;

(b) *homogeneidad*: $\|\lambda x\| = |\lambda|\|x\|$, $\lambda \in \mathbb{R}$, $x \in X$;

(c) *desigualdad triangular*: $\|x + y\| \leq \|x\| + \|y\|$, $x, y \in X$.

Observamos, en virtud de (b), que $\|0\| = 0$ y que $\|-x\| = \|x\|$. Luego, usando la desigualdad triangular, podemos deducir que $\|x\| \geq 0$, $x \in X$. Es decir, una norma es una aplicación de X en $[0,\infty)$.

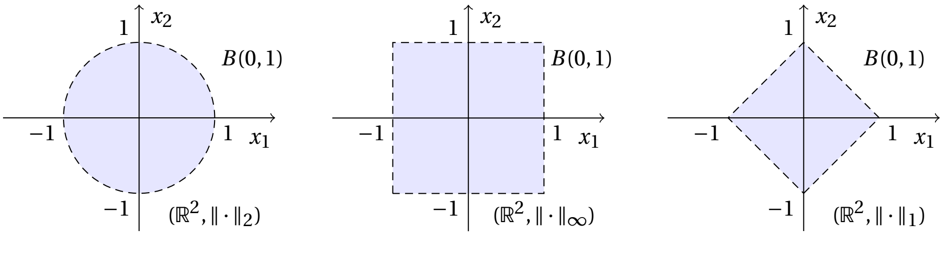
Podemos definir sobre X una topología relativa a la norma $\|\cdot\|$ de la forma siguiente. Para cada $x \in X$ y $r > 0$, denotamos por $B(x,r)$ al conjunto, asociado a $\|\cdot\|$, dado por $B(x,r) = \{y \in X : \|y - x\| < r\}$. Decimos que un subconjunto $A \subset X$ es abierto si $A = \emptyset$, o bien, cuando para cada $a \in A$ existe $r > 0$ tal que $B(a,r) \subset A$.

Como primer ejemplo consideramos el espacio $X = \mathbb{R}^n$. Para cada $1 \leq p \leq \infty$ podemos definir $\|\cdot\|_p$ mediante

$$\|x\|_p = \Big(\sum_{i=1}^{n} |x_i|^p\Big)^{1/p}, \quad 1 \leq p < \infty, \quad \text{y} \quad \|x\|_\infty = \max_{i=1,\dots,n} |x_i|,$$

siendo $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. Para cada $1 \leq p \leq \infty$, $\|\cdot\|_p$ es una norma en $\mathbb{R}^n$.

La norma de cada vector x de un espacio normado X se interpreta siempre como la longitud de x. En particular, en $\mathbb{R}^n$, si consideramos otra norma diferente a la euclídea, $\|\cdot\|_2$, estamos cambiando la forma de medir la longitud de $x \in \mathbb{R}^n$. En la siguiente figura representamos el conjunto $B(0,1)$ en $\mathbb{R}^2$, cuando consideramos $\|\cdot\|_p$, para $p = 1$ $p = 2$ y $p = \infty$.



Sin embargo, podemos probar que las normas que hemos definido para $\mathbb{R}^n$ son equivalentes. Recordamos que, dado un espacio vectorial X y dos normas $\|\cdot\|_a$ y $\|\cdot\|_b$ en X, decimos que son equivalentes cuando existe $c > 1$ de manera que

$$\frac{1}{c}\|x\|_a \le \|x\|_b \le c\|x\|_a, \quad x \in X, \tag{1.1}$$

lo cual equivale a decir que las dos normas generan en X la misma topología.

Proposición 1.1. *Sean $1 \le p, q \le \infty$, $\|\cdot\|_p$ y $\|\cdot\|_q$ son normas equivalentes en $\mathbb{R}^n$.*

Demostración. Mostramos, en primer lugar, que si $1 \le p < \infty$, entonces

$$\|x\|_\infty \le \|x\|_p \le n^{1/p}\|x\|_\infty, \quad x \in \mathbb{R}^n. \tag{1.2}$$

Sea $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. Que $\|x\|_\infty \le \|x\|_p$, se deduce fácilmente ya que

$$|x_k| = (|x_k|^p)^{1/p} \le \Big(\sum_{j=1}^n |x_j|^p\Big)^{1/p} = \|x\|_p, \quad k = 1, ..., n.$$

Por otro lado, podemos escribir,

$$\|x\|_p = \big(\sum_{j=1}^n |x_j|^p\big)^{1/p} \le \big(\sum_{j=1}^n (\max_{k=1,...,n} |x_k|)^p\big)^{1/p} = \max_{k=1,...,n} |x_k| (\sum_{j=1}^n 1)^{1/p} = n^{1/p}\|x\|_\infty.$$

Supongamos ahora, $1 \le p, q < \infty$. De (1.2) podemos deducir que

$$\|x\|_p \le n^{1/p}\|x\|_\infty \le n^{1/p}\|x\|_q \le n^{1/p+1/q}\|x\|_\infty \le n^{1/p+1/q}\|x\|_p.$$

Luego, $n^{-1/p}\|x\|_p \le \|x\|_q \le n^{1/q}\|x\|_p$, esto es, concluimos que las normas son equivalentes (podemos tomar $c = n^{1/p+1/q}$ en (1.1)). □

En relación al resultado recogido en esta proposición podemos afirma más aún: dos normas cualesquiera en $\mathbb{R}^n$ son equivalentes.

Proposición 1.2. *Toda norma en $\mathbb{R}^n$ es equivalente a $\|\cdot\|_\infty$.*

Demostración. Sea $\|\cdot\|$ una norma en $\mathbb{R}^n$. Consideramos la base canónica $\{e_j\}_{j=1}^n$, siendo $e_j = (e_{j1}, e_{j2}, \dots, e_{jn})$, donde $e_{jk} = 1$, si $k = j$ y $e_{jk} = 0$, cuando $k \neq j$.

Sea $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. Podemos escribir $x = \sum_{j=1}^n x_j e_j$, y, aplicando la desigualdad triangular obtenemos

$$\|x\| \le \sum_{j=1}^n |x_j|\|e_j\| \le \max_{k=1,...,n} |x_k| \sum_{j=1}^n \|e_j\| = c\|x\|_\infty, \quad \text{con } c = \sum_{j=1}^n \|e_j\|.$$

Por otro lado, sea $S = \{x \in \mathbb{R}^n : \|x\|_\infty = 1\}$. El conjunto S es compacto en $\mathbb{R}^n$ con la topología usual (que, en virtud de la Proposición 1.1, coincide con la topología

generada por la norma $\|\cdot\|_\infty$). Recordamos que $A \subset \mathbb{R}$ es compacto en $\mathbb{R}^n$ con la topología usual si, y solo si, A es cerrado y acotado.
Denotamos por F la función definida por la norma $\|\cdot\|$, esto es, $F(x) = \|x\|$, $x \in \mathbb{R}^n$. Se tiene que F es continua como función de $(\mathbb{R}^n, \|\cdot\|_\infty)$ en $([0,\infty), \tau_u)$ (Aquí, τ_u representa la topología usual). En efecto, sea $x_0 \in \mathbb{R}^n$ y $(x_n)_{n=1}^\infty \subset \mathbb{R}^n$ una sucesión tal que $x_n \to x_0$, cuando $n \to \infty$, en la norma $\|\cdot\|_\infty$. Entonces,

$$|F(x_n) - F(x_0)| = |\|x_n\| - \|x_0\|| \leq \|x_n - x_0\| \leq c\|x_n - x\|_\infty,$$

y, por tanto, $F(x_n) \longrightarrow F(x_0)$, cuando $n \to \infty$.

Puesto que S es compacto en $(\mathbb{R}^n, \|\cdot\|_\infty)$, podemos encontrar $a, b \in S$ de manera que

$$\min_{y\in S} F(y) = F(a) \leq F(x) \leq F(b) = \max_{y\in S} F(y), \quad x \in S.$$

Sea $x \in \mathbb{R}^n \setminus \{0\}$. Es claro que $y = \frac{x}{\|x\|_\infty} \in S$ ($\|y\|_\infty = 1$). Luego, $F(a) \leq F(y)$, esto es, $\|a\| \leq \|y\| = \frac{\|x\|}{\|x\|_\infty}$, lo que implica que $\|a\|\|x\|_\infty \leq \|x\|$.

Además, teniendo en cuenta que $a \in S$, se sigue que $\|a\|_\infty = 1$ y necesariamente $a \neq 0$. Por tanto, $\|a\| \neq 0$ y $\|x\|_\infty \leq \frac{1}{\|a\|}\|x\|$.
Cuando $x = 0$ la desigualdad anterior es clara (de hecho, se trata de una igualdad).

□

Observación 1.3. La propiedad que se recoge en la Proposición 1.2 es cierta para cualquier espacio vectorial de dimensión finita.

De manera natural surge entonces la pregunta de qué ocurre cuando un espacio vectorial es de dimensión infinita. Vamos a ver que la equivalencia de las normas es una propiedad que distingue los espacios vectoriales de dimensión finita de los de dimensión infinita.

Sea X un espacio vectorial de dimensión infinita y tomamos una base $\mathfrak{B}$ de X, de la que sabemos que tiene cardinal no finito. Cada $x \in X$ tiene una representación, que es única, en términos de los elementos de $\mathfrak{B}$, esto es, existen únicos $b_1, \ldots, b_k \in \mathfrak{B}$ y $\alpha_1, \ldots, \alpha_k \in \mathbb{R}$ tales que $x = \alpha_1 b_1 + \ldots + \alpha_k b_k$.

Podemos entonces definir dos normas $\|\cdot\|_1$ y $\|\cdot\|_\infty$ en X mediante

$$\|x\|_1 = \sum_{j=1}^k |\alpha_j| \quad \text{y} \quad \|x\|_\infty = \max_{j=1,\ldots,k} |\alpha_j|, \quad x = \sum_{j=1}^k \alpha_j b_j.$$

Veamos primero que, efectivamente, son normas en X.

(a) **No degeneradas**. Sea $x \in X$ tal que $\|x\|_1 = 0$ (o bien, $\|x\|_\infty = 0$). Si $x = \sum_{j=1}^k \alpha_j b_j$, $\alpha_j \in \mathbb{R}$ y $b_j \in \mathfrak{B}$, es la representación (única) de x respecto a la base $\mathfrak{B}$, entonces

$$\sum_{j=1}^k |\alpha_j| = 0 \quad \Big(\text{o bien,} \quad \max_{j=1,\ldots,k} |\alpha_j| = 0\Big),$$

lo que implica claramente que $\alpha_j = 0$, para todo $j = 1, \ldots, k$, y por tanto, $x = 0$.

(b) **Homogéneas**. Sean $\lambda \in \mathbb{R}$ y $x = \sum_{j=1}^{k} \alpha_j b_j \in X$, donde $\alpha_j \in \mathbb{R}$ y $b_j \in \mathfrak{B}$, $j = 1,...,k$. Entonces, $\lambda x = \sum_{j=1}^{k} \lambda \alpha_j b_j$ y, esta es la única expresión de λx como combinación lineal finita de elementos de $\mathfrak{B}$. Tenemos así que

$$\|\lambda x\|_1 = \sum_{j=1}^{k} |\lambda \alpha_j| = |\lambda| \sum_{j=1}^{n} |\alpha_j| = |\lambda| \|x\|_1.$$

y

$$\|\lambda x\|_\infty = \max_{j=1,...,k} |\lambda \alpha_j| = |\lambda| \max_{j=1,...,k} |\alpha_j| = |\lambda| \|x\|_\infty.$$

(c) **Desigualdad triangular**. Sean $x, y \in X$ tales que $x = \sum_{i=1}^{k} \alpha_i b_i$ y $y = \sum_{j=1}^{m} \alpha'_j b'_j$, donde $\alpha_i, \alpha'_j \in \mathbb{R}$ y $b_i, b'_j \in \mathfrak{B}$, $i = 1,\dots,k$, $j = 1,\dots,m$.
Supongamos que $b_{i_0} = b'_{j_1}$, para ciertos $i_0 \in 1,...,k$, y $j_1 \in 1,...,m$. Entonces, en la combinación lineal que define a $x + y$ aparece el sumando $(\alpha_{i_0} + \alpha'_{j_1}) b_{i_0}$ y entonces, tanto en $\|x+y\|_1$ como en $\|x+y\|_\infty$ tenemos el término $|\alpha_{i_0} + \alpha'_{j_1}|$. De la desigualdad triangular en $\mathbb{R}$ se sigue que $|\alpha_{i_0} + \alpha'_{j_1}| \le |\alpha_{i_0}| + |\alpha'_{j_0}|$. Por tanto, se sigue que

$$\|x + y\|_1 \le \sum_{i=1}^{k} |\alpha_j| + \sum_{j=1}^{m} |\alpha'_j| = \|x\|_1 + \|y\|_1,$$

y

$$\|x + y\|_\infty \le \max_{i=1,...,k} |\alpha_i| + \max_{j=1,...,m} |\alpha'_j| = \|x\|_\infty + \|y\|_\infty.$$

Veamos ahora que las dos normas definidas no son equivalentes en X. Concretamente, vamos a mostrar que no existe $c > 0$ de modo que $\|x\|_1 \le c\|x\|_\infty$, $x \in X$. Ya que el cardinal de $\mathfrak{B}$ no es finito, podemos encontrar una sucesión $(b_n)_{n=1}^{\infty} \subset \mathfrak{B}$ siendo $b_n \neq b_m$, $n \neq m$. Si, para cada $n \in \mathbb{N}$, consideramos $a_n = \sum_{j=1}^{n} b_j$, es claro que $\|a_n\|_1 = \sum_{j=1}^{n} 1 = n$, mientras que $\|a_n\|_\infty = \max_{j=1,...,k} 1 = 1$, $n \in \mathbb{N}$. Luego, si existiera $c > 0$ tal que $\|x\|_1 \le c\|x\|_\infty$, $x \in X$, en particular, tomando $x = a_n$, $n \in \mathbb{N}$, se llegaría a que $n \le c$, para todo $n \in \mathbb{N}$. Y esto, evidentemente, es falso.

Hemos probado así que $\|\cdot\|_1$ y $\|\cdot\|_\infty$ no son equivalentes en X.

1.1.1. Operadores lineales entre espacios normados

Dedicamos esta sección a analizar algunos aspectos básicos en relación a las aplicaciones lineales entre dos espacios normados que tienen un buen comportamiento algebraico y también topológico: los operadores lineales y continuos.

Un *operador lineal* es sencillamente una aplicación lineal entre dos espacios vectoriales, ambos sobre el mismo cuerpo $\mathbb{K}$. Para hablar de continuidad consideramos $(X, \|\cdot\|_X)$ e $(Y, \|\cdot\|_Y)$ dos espacios normados. Sea $T : X \longrightarrow Y$ un operador lineal. Decimos que T es *un operador continuo en* $x_0 \in X$ cuando para todo $\varepsilon > 0$, existe $\delta > 0$ de manera que si $\|x - x_0\|_X < \delta$ entonces $\|T(x) - T(x_0)\|_Y < \varepsilon$.

La continuidad de un operador lineal entre espacios normados puede caracterizarse de varias formas, como las que recogemos en el siguiente teorema.

Teorema 1.4. *Sean* $(X, \|.\|_X)$ *e* $(Y, \|.\|_Y)$ *espacios normados y* T *un operador lineal de* X *en* Y. *Son equivalentes:*

(i) T *es continuo en* 0.

(ii) T *es continuo en* x_0, *para algún* $x_0 \in X$.

(iii) T *es continuo en cada* $x_0 \in X$.

(iv) Existe $c > 0$ *tal que* $\|T(x)\|_Y \leq c\|x\|_X$, $x \in X$.

Demostración. Las implicaciones $(i) \Rightarrow (ii)$ y $(iii) \Rightarrow (i)$ son evidentes.

Veamos que $(ii) \Rightarrow (i)$. Supongamos que T es continua en $x_0 \in X$ y fijamos $\varepsilon > 0$. Por hipótesis, existe $\delta_0 > 0$ tal que $\|T(y) - T(x_0)\| < \varepsilon$, cuando $\|y - x_0\|_X < \delta_0$. De aquí podemos deducir que $\|Tx\|_Y < \varepsilon$, cuando $\|x\|_X < \delta_0$. En efecto, sea $x \in X$ tal que $\|x\|_X < \delta_0$. Luego, si $y = x + x_0$, se tiene que $\|y - x_0\|_X < \delta_0$ y, por tanto, $\|T(y) - T(x_0)\|_Y < \varepsilon$. Dado que T es lineal, podemos concluir que

$$\|T(x)\|_Y = \|T(x - y) + T(y)\|_Y = \|T(y) - T(x_0)\|_Y < \varepsilon,$$

esto es, T es continuo en 0 (nótese que $T(0) = 0$ por ser T lineal).

$(i) \Rightarrow (iii)$ Supongamos que T es continuo en 0 y sea $x_0 \in X$. De nuevo fijamos $\varepsilon > 0$. Sabemos que existe $\delta_0 > 0$ verificando que $\|T(z)\|_Y < \varepsilon$, cuando $\|z\|_X < \delta_0$. Entonces, usando de nuevo la linealidad de T, se tiene que, para todo $x \in X$ tal que $\|x - x_0\|_X < \delta_0$ se verifica $\|T(x) - T(x_0)\|_Y = \|T(x - x_0)\|_Y < \varepsilon_0$, y (iii) queda así establecido.

$(i) \Rightarrow (iv)$ Supongamos que T es continua en 0. Tomando $\varepsilon = 1$, encontramos $\delta > 0$ de modo que $\|T(x)\|_Y < 1$ cuando $\|x\|_X < \delta$.

Ya que $T(0) = 0$ la desigualdad en (iv) se sigue trivialmente para cualquier valor de $c > 0$. Consideramos ahora $x \in X \setminus \{0\}$ y tomamos $z = \frac{\delta}{2}\frac{x}{\|x\|_X}$. Es claro que $\|z\|_X = \frac{\delta}{2} < \delta$, por lo que $\|T(z)\|_Y < 1$. La linealidad de T y la homogeneidad de la norma $\|\cdot\|_Y$ permiten concluir que

$$\|T(x)\|_Y = \left\|T\left(\frac{2\|x\|_X}{\delta} z\right)\right\|_Y = \frac{2\|x\|_X}{\delta}\|T(z)\|_Y < \frac{2}{\delta}\|x\|_X.$$

Ya que $\delta > 0$ es una constante independiente de $x \in X$, hemos obtenido (iv) con $c = 2/\delta$.

Finalmente probamos $(iv) \Rightarrow (i)$. Supongamos que existe $c > 0$ de forma que $\|T(x)\|_Y \leq c\|x\|_X$, $x \in X$. Sea $\varepsilon > 0$. Tomando $\delta = \varepsilon/c$, se tiene que, para todo $x \in X$ tal que $\|x\|_X < \delta$ se cumple que $\|T(x)\|_Y \leq c\|x\|_X < c\delta = \varepsilon$ y (i) queda probado. □

Sean $(X, \|\cdot\|_X)$ e $(Y, \|\cdot\|_Y)$ dos espacios normados. Denotamos por $L(X, Y)$ al espacio vectorial de los operadores lineales y continuos de X en Y. En $L(X, Y)$ podemos definir una norma mediante

$$\|T\|_{X\to Y} = \sup_{x \in X\setminus\{0\}} \frac{\|T(x)\|_Y}{\|x\|_X}, \quad T \in L(X, Y).$$

Veamos que, efectivamente, es una norma en $L(X, Y)$.

Proposición 1.5. *Sean $(X, \|\cdot\|_X)$ e $(Y, \|\cdot\|_Y)$ dos espacios normados. La aplicación $\|\cdot\|_{X\to Y}$ es una norma en $L(X,Y)$.*

Demostración. Probamos las tres condiciones que definen una norma.

(a)**No degenerada**. Sea $T \in L(X,Y)$ tal que $\|T\|_{X\to Y} = 0$. Entonces, $\frac{\|T(x)\|_Y}{\|x\|_X} = 0$, $x \in X\backslash\{0\}$, esto es, $\|T(x)\|_Y = 0$, $x \in X\backslash\{0\}$. Puesto que $\|\cdot\|_Y$ es una norma, se sigue que $T(x) = 0$, $x \in X \setminus \{0\}$. Como $T(0) = 0$, se concluye que $T = 0$.

(b)**Homogénea**. Sean $T \in L(X,Y)$ y $\lambda \in \mathbb{R}$. Usando la homogeneidad de la norma $\|\cdot\|_Y$ podemos escribir,

$$\|\lambda T\|_{X\to Y} = \sup_{x\in X\setminus\{0\}} \frac{\|\lambda T(x)\|_Y}{\|x\|_X} = \sup_{x\in X\setminus\{0\}} \frac{|\lambda|\|T(x)\|_Y}{\|x\|_X} = |\lambda|\|T\|_{X\to Y}.$$

(c) **Desigualdad triangular**. Sean $T, S \in L(X,Y)$. Tenemos que

$$\begin{aligned}\|T+S\|_{X\to Y} &= \sup_{x\in X\setminus\{0\}} \frac{\|T(x)+S(x)\|_Y}{\|x\|_X} \le \sup_{x\in X\setminus\{0\}} \left(\frac{\|T(x)\|_Y}{\|x\|_X} + \frac{\|S(x)\|_Y}{\|x\|_X}\right)\\ &\le \sup_{x\in X\setminus\{0\}} \frac{\|T(x)\|_Y}{\|x\|_X} + \sup_{x\in X\setminus\{0\}} \frac{\|S(x)\|_Y}{\|x\|_X} = \|T\|_{X\to Y} + \|S\|_{X\to Y}.\end{aligned}$$

□

Podemos expresar la norma $\|T\|_{X\to Y}$, para $T \in L(X,Y)$ como indicamos a continuación.

Observación 1.6. Sean $(X, \|\cdot\|_X)$ e $(Y, \|\cdot\|_Y)$ dos espacios normados. Para cada operador $T \in L(X,Y)$ se tiene que

$$\|T\|_{X\to Y} = \sup_{\|x\|_X=1} \|T(x)\|_Y = \text{mín}\big\{c > 0 : \|T(x)\|_Y \le c\|x\|_X,\ x \in X\big\}.$$

Demostración. Sea $T \in L(X,Y)$. La linealidad de T y la homogeneidad de $\|\cdot\|_Y$ nos permite escribir

$$\|T\|_{X\to Y} = \sup_{x\in X\setminus\{0\}} \frac{\|T(x)\|_Y}{\|x\|_X} = \sup_{x\in X\setminus\{0\}} \left\|T\Big(\frac{x}{\|x\|_X}\Big)\right\|.$$

Puesto que, si $x \in X\backslash\{0\}$, entonces $z = \frac{x}{\|x\|_X}$ verifica que $\|z\|_X = 1$, concluimos que $\|T\|_{X\to Y} \le \sup_{\|z\|_X=1} \|T(z)\|_Y$. La desigualdad inversa es trivial.
Por otro lado, si $c > 0$ es tal que $\|T(x)\|_Y \le c\|x\|_X$, $x \in X$, entonces es trivial que

$$\sup_{x\in X\setminus\{0\}} \frac{\|T(x)\|_Y}{\|x\|_X} \le c,$$

y por tanto, $\|T\|_{X\to Y} \le \text{ínf}\big\{c > 0 : \|T(x)\|_Y \le c\|x\|_X,\ x \in X\big\}$.

Ahora bien, por ser T continuo, se tiene que existe $c_0 > 0$ de manera que $\|T(x)\|_Y \le c_0\|x\|_X$, $x \in X$. Luego,

$$\|T\|_{X\to Y} = \sup_{x\in X\setminus\{0\}} \frac{\|T(x)\|_Y}{\|x\|_X} < \infty,$$

de donde se sigue que $\|T(x)\|_Y \le \|T\|_{X\to Y}\|x\|_X$, $x \in X$. Concluimos de esta forma que $\|T\|_{X\to Y} = \text{mín}\big\{c > 0 : \|T(x)\|_Y \le c\|x\|_X,\ x \in X\big\}$. □

Terminamos la sección mostrando otra propiedad que marca la diferencia entre los espacios de dimensión finita y los de dimensión infinita.

Proposición 1.7. *Sean* $(X, \|\cdot\|_X)$ *e* $(Y, \|\cdot\|_Y)$ *dos espacios normados. Supongamos además que* X *tiene dimensión finita. Entonces, todo operador lineal* $T : X \longrightarrow Y$ *es continuo.*

Demostración. Es claro que el operador $T = 0$ es continuo. Asumimos entonces que $T : X \longrightarrow Y$ es un operador lineal no nulo.

Pongamos que $\dim X = n$ y elegimos una base $\{e_1, ..., e_n\}$ de X. Para establecer que T es continuo, en virtud del Teorema 1.4, es suficiente ver que existe $c > 0$ de forma que $\|T(x)\|_Y \leq c\|x\|_X$, $x \in X$.

Sea $x \in X$. Sabemos que existen únicos $\alpha_1, ..., \alpha_n \in \mathbb{R}$ tales que $x = \sum_{j=1}^n \alpha_j e_j$. Luego, teniendo en cuenta que T es lineal y usando la desigualdad triangular obtenemos que

$$\|T(x)\|_Y = \left\| \sum_{j=1}^n \alpha_j T(e_j) \right\|_Y \leq \sum_{j=1}^n |\alpha_j| \|T(e_j)\|_Y \leq \max_{k=1,...,n} |\alpha_k| \sum_{j=1}^n \|T(e_j)\|_Y.$$

Recordamos que la aplicación $\|\cdot\|_\infty$ definida en X por

$$\|z\|_\infty = \max_{j=1,...,n} |\lambda_j|, \quad z = \sum_{j=1}^n \lambda_j e_j \in X,$$

es una norma en X. Además, por ser X de dimensión finita, esta norma es equivalente a $\|\cdot\|_X$. Por tanto, existe $M > 1$ de manera que

$$\frac{1}{M}\|x\|_X \leq \|x\|_\infty \leq M\|x\|_X, \quad x \in X.$$

Con todo ello, si definimos $C = \sum_{j=1}^n \|T(e_j)\|_Y$ (nótese que $C > 0$, pues $T \neq 0$) podemos concluir que

$$\|T(x)\|_Y \leq C\|x\|_\infty \leq CM\|x\|_X,$$

que es lo que queríamos demostrar. □

Proposición 1.8. *Sean* $(X, \|\cdot\|_X)$ *e* $(Y, \|\cdot\|_Y)$ *espacios normados. Supongamos que* $\dim X = \infty$ *y que* $Y \neq \{0\}$. *Entonces, existe un operador* $T : X \to Y$ *lineal que no es continuo.*

Demostración. Sea $\mathfrak{B}$ una base de X. Podemos suponer que $\|b\|_X = 1$ para todo $b \in \mathfrak{B}$. De no ser así, sustituimos b por $\frac{b}{\|b\|_X}$ en la base $\mathfrak{B}$ (nótese que $b \neq 0$) y el conjunto obtenido sigue siendo una base de X.

Ya que $\mathrm{card}(\mathfrak{B}) = \infty$, existe una sucesión $\{b_k\}_{k=1}^\infty \subset \mathfrak{B}$ siendo $b_k \neq b_m$, para $k \neq m$. Para encontrar un operador $T : X \longrightarrow Y$ lineal que no sea continuo, elegimos $v \in Y \setminus \{0\}$ (por hipótesis, $Y \neq \{0\}$).

Definimos $T(b_n) = nv$, $n \in \mathbb{N}$, y $T(b) = 0$, cuando $b \in \mathfrak{B}\setminus\{b_k\}_{k=1}^{\infty}$. Para cada $x \in X$ podemos escribir $x = \sum_{j=1}^{m} \alpha_j u_j$, donde $\alpha_j \in \mathbb{R}$ y $u_j \in \mathfrak{B}$, $j = 1, \ldots, m$. Además, esta representación es única. Definimos $T(x) = \sum_{j=1}^{m} \alpha_j T(u_j)$.

De esta forma, T es lineal. En efecto, sean $x, y \in X$ y $a, b \in \mathbb{R}$. Pongamos

$$x = \sum_{i=1}^{m} \alpha_i u_i \quad \text{e} \quad y = \sum_{j=1}^{r} \beta_j w_j,$$

con $\alpha_i, \beta_j \in \mathbb{R}$ y $u_i, w_j \in \mathfrak{B}$, $i = 1, \ldots, m$, $j = 1, \ldots, r$. Tenemos que $T(x) = \sum_{i=1}^{m} \alpha_i T(u_i)$ y $T(y) = \sum_{j=1}^{r} \beta_j T(w_j)$.

Por otra parte,

$$ax + by = \sum_{i=1}^{m} a\alpha_i u_i + \sum_{j=1}^{r} b\beta_j w_j.$$

Observamos que si $u_i = w_j$, para ciertos i, j, $i = 1, \ldots, m$, $j = 1, \ldots, r$, entonces en la suma anterior aparece el sumando $(a\alpha_i + b\beta_j)u_i$. Teniendo esto presente, obtenemos que $T(ax + by) = aT(x) + bT(y)$.

Veamos que T no es un operador continuo. Puesto que $Tb_n = nv$, $n \in \mathbb{N}$ se sigue que $\|T(b_n)\|_Y = n\|v\|_Y$, $n \in \mathbb{N}$. Pero $\|b_n\| = 1$, $n \in \mathbb{N}$. Luego, si existiera c verificando que $\|T(x)\|_Y \leq c\|x\|_X$, $x \in X$, entonces, en particular, se cumpliría que $n\|v\|_Y = \|Tb_n\|_Y \leq c\|b_n\|_X \leq c$ para todo $n \in \mathbb{N}$. Y esto no es posible pues $\mathbb{N}$ no es un conjunto acotado. Concluimos que T no es continuo. $\square$

1.2. El espacio de las matrices

Denotamos por $M_n(\mathbb{R})$, $n \in \mathbb{N}$, el espacio de las matrices cuadradas de orden n. Sabemos que $(M_n(\mathbb{R}), +, \cdot_{\mathbb{R}})$ es un espacio vectorial de dimensión n^2 y una base para este espacio es $\{e^{ij}\}_{i,j=1,\ldots,n}$ donde, para cada $i, j = 1, \ldots, n$, e^{ij} es la matriz $e^{ij} = (a_{k,m}^{ij})_{k,m=1,\ldots,n}$, siendo $a_{k,m}^{ij} = 1$, cuando $k = i$ y $m = j$ y $a_{k,m}^{ij} = 0$, en el resto.

Ya que $\dim M_n(\mathbb{R}) < \infty$, todas las normas definidas en $M_n(\mathbb{R})$ son equivalentes. Veamos algunas de las normas que podemos asociar a $M_n(\mathbb{R})$.

En primer lugar consideramos la norma $\|\cdot\|_p$, $1 \leq p \leq \infty$, definida por

$$\|A\|_p = \Big(\sum_{i,j=1}^{n} |a_{ij}|^p\Big)^{\frac{1}{p}}, \text{ si } 1 \leq p < \infty, \quad \text{y} \quad \|A\|_\infty = \max_{i,j=1,\ldots,n} |a_{ij}|,$$

para cada $A = (a_{ij})_{i,j=1,\ldots,n} \in M_n(\mathbb{R})$.

En efecto, se tiene que, para cada $1 \leq p \leq \infty$, $\|\cdot\|_p$ es una norma en $M_n(\mathbb{R})$. Y en particular, la norma $\|\cdot\|_2$ está asociada a un producto escalar en $M_n(\mathbb{R})$, que hace de $(M_n(\mathbb{R}), \|\cdot\|_2)$ un espacio de Hilbert. Nótese que podemos considerar esta norma como la correspondiente $\|\cdot\|_p$ sobre el espacio $\mathbb{R}^{n^2}$.

Por otra parte si $A \in M_n(\mathbb{R})$ podemos definir la aplicación lineal $T_A : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ dada por $T_A(x) = Ax^T$, $x \in \mathbb{R}^n$. (Aquí, como es habitual, B^T representa la matriz traspuesta de la matriz B). De esta forma, si $A = (a_{ij})_{i,j=1,\dots,n} \in M_n(\mathbb{R})$ y $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, se tiene que

$$T_A(x) = \Big(\sum_{j=1}^n a_{ij} x_j\Big)_{i=1}^n = \big(T_A^i(x)\big)_{i=1}^n, \quad x \in \mathbb{R}^n.$$

Observamos que para cada $i = 1, \dots, n$, $T_A^i(x) = \sum_{j=1}^n a_{ij} x_j$ es una aplicación lineal de $\mathbb{R}^n$ en $\mathbb{R}$. Luego, por la Proposición 1.7 se deduce que T_A es un operador continuo.

La relación entre las matrices y los operadores asociados a ellas nos permite definir otra norma en $M_n(\mathbb{R})$ como sigue. Sean $\|\cdot\|_a$ y $\|\cdot\|_b$ dos normas en $\mathbb{R}^n$. Sabemos que estas normas son equivalentes. Definimos

$$\|A\|_{\|\cdot\|_a \to \|\cdot\|_b} = \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{\|T_A(x)\|_b}{\|x\|_a} = \sup_{\|x\|_a = 1} \|T_A(x)\|_b, \quad A \in M_n(\mathbb{R}), \tag{1.3}$$

esto es, $\|A\|_{\|\cdot\|_a \to \|\cdot\|_b} = \|T_A\|_{(\mathbb{R}^n, \|\cdot\|_a) \to (\mathbb{R}^n, \|\cdot\|_b)}$. En virtud de la Proposición 1.5 se tiene que $\|\cdot\|_{\|\cdot\|_a \to \|\cdot\|_b}$ es una norma en $M_n(\mathbb{R})$ y además, ya que $\dim(M_n(\mathbb{R})) < \infty$, es equivalente a $\|\cdot\|_p$, $1 \le p \le \infty$.

Analizamos ahora algunos casos particulares.

1. $\|\cdot\|_a = \|\cdot\|_b = \|\cdot\|_\infty$. En este caso, $\|A\|_{\infty,\infty} := \|A\|_{\|\cdot\|_\infty \to \|\cdot\|_\infty} = \sup_{\|x\|_\infty = 1} \|T_A(x)\|_\infty$, $A \in M_n(\mathbb{R})$.

   Vamos a demostrar que podemos expresar esta norma mediante

$$\|A\|_{\infty,\infty} = \max_{i=1,\dots,n} \sum_{j=1}^n |a_{ij}|, \quad A = (a_{ij})_{i,j=1}^n \in M_n(\mathbb{R}). \tag{1.4}$$

   Sea $A = (a_{ij})_{i,j=1}^n \in M_n(\mathbb{R})$. Para cada $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ tenemos que

$$\|T_A(x)\|_\infty = \max_{i=1,\dots,n} |T_A^i(x)| = \max_{i=1,\dots,n} \Big|\sum_{j=1}^n a_{ij} x_j\Big|.$$

   Se sigue entonces que

$$\begin{aligned} \|T_A(x)\|_\infty &\le \max_{i=1,\dots,n} \sum_{j=1}^n |a_{ij}||x_j| \le \max_{k=1,\dots,n} |x_k| \max_{i=1,\dots,n} \sum_{j=1}^n |a_{ij}| \\ &= \|x\|_\infty \max_{i=1,\dots,n} \sum_{j=1}^n |a_{ij}|, \quad x = (x_1, \dots, x_n) \in \mathbb{R}^n. \end{aligned}$$

   Luego, $\|A\|_{\infty,\infty} \le \max_{i=1,\dots,n} \sum_{j=1}^n |a_{i,j}|$.

Para establecer la desigualdad inversa, definimos, para cada $i = 1, \dots, n$, el vector $x^i = (x^i_j)_{j=1}^n \in \mathbb{R}^n$ donde

$$x^i_j = \begin{cases} \frac{a_{ij}}{|a_{ij}|}, & \text{cuando } a_{ij} \neq 0, \\ 1, & \text{cuando } a_{ij} = 0, \end{cases} \qquad j = 1, \dots, n.$$

Es claro que, para cada $i = 1, \dots, n$ se tiene que $\|x^i\|_\infty = \text{máx}_{j=1,\dots,n} |x^i_j| = 1$ y, además,

$$\|T_A(x^i)\|_\infty = \max_{k=1,\dots,n} \Big| \sum_{j=1}^n a_{kj} x^i_j \Big| \geq \Big| \sum_{j=1}^n a_{ij} x^i_j \Big| = \Big| \sum_{\substack{j=1 \\ a_{ij} \neq 0}}^n a_{ij} x^i_j \Big|$$
$$= \sum_{\substack{j=1 \\ a_{ij} \neq 0}}^n \frac{a_{ij}^2}{|a_{ij}|} = \sum_{j=1}^n |a_{ij}|.$$

Por tanto, $\|A\|_{\infty,\infty} = \sup_{\|x\|_\infty = 1} \|T_A(x)\|_\infty \geq \|T_A(x^i)\|_\infty \geq \sum_{j=1}^n |a_{ij}|$, $i = 1, \dots, n$, esto es, $\|A\|_{\infty,\infty} \geq \text{máx}_{i=1,\dots,n} \sum_{j=1}^n |a_{ij}|$.
La igualdad (1.4) queda entonces establecida.

2. $\|\cdot\|_a = \|\cdot\|_b = \|\cdot\|_1$. Se tiene que $\|A\|_{1,1} = \sup_{\|x\|_1 = 1} \|T_A(x)\|_1$, $A \in M_n(\mathbb{R})$.
Sea $A = (a_{ij})_{i,j=1}^n \in M_n(\mathbb{R})$. En este caso, para cada $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, podemos escribir

$$\|T_A(x)\|_1 = \sum_{i=1}^n \Big| \sum_{j=1}^n a_{ij} x_j \Big| \leq \sum_{i=1}^n \sum_{j=1}^n |a_{ij}||x_j| = \sum_{j=1}^n |x_j| \sum_{i=1}^n |a_{ij}|$$
$$\leq \max_{j=1,\dots,n} \sum_{i=1}^n |a_{ij}| \sum_{j=1}^n |x_j| = \max_{j=1,\dots,n} \sum_{j=1}^n |a_{ij}| \|x\|_1.$$

De aquí se sigue que $\|A\|_{1,1} \leq \text{máx}_{j=1,\dots,n} \sum_{i=1}^n |a_{ij}|$.
Ahora consideramos, para cada $j = 1, \dots, n$, el vector $e^j = (e^j_i)_{i=1}^n$, donde, para cada $i = 1, \dots, n$, $e^j_i = 1$, si $i = j$ y $e^j_i = 0$, cuando $i \neq j$. De esta forma obtenemos que

$$T_A(e^j) = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{pmatrix}, \quad j = 1, \dots, n.$$

y, entonces, $\|T_A(e^j)\|_1 = \sum_{i=1}^n |a_{ij}|$, $j = 1, \dots, n$. Ya que $\|e^j\|_1 = 1$, $j = 1, \dots, n$, se tiene que $\|A\|_{1,1} \geq \|T_A(e^j)\|_1 = \sum_{i=1}^n |a_{ij}|$, $j = 1, \dots, n$, lo que equivale a que $\|A\|_{1,1} \geq \text{máx}_{j=1,\dots,n} \sum_{i=1}^n |a_{ij}|$. Se concluye que

$$\|A\|_{1,1} = \max_{k=1,\dots,n} \sum_{j=1}^n |a_{jk}|.$$

3. $\|\cdot\|_a = \|\cdot\|_b = \|\cdot\|_2$. Ahora tenemos $\|A\|_{2,2} = \sup_{\|x\|_2=1} \|T_A(x)\|_2$, $A \in M_n(\mathbb{R})$. Vamos a demostrar que se verifica

$$\|A\|_{2,2} = \sqrt{\text{máx}\{\text{ autovalores de } A^T A\}}. \tag{1.5}$$

Sea $A \in M_n(\mathbb{R})$. La matriz $A^T A$ es simétrica real, y por tanto, es diagonalizable y todos sus autovalores son reales y no negativos. Escribimos $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ para los autovalores $A^T A$. Además, podemos considerar $\mathfrak{B} = \{u^1, \ldots, u^n\}$ una base ortonormal de $\mathbb{R}^n$ constituida por los autovectores de $A^T A$ asociados a los autovalores $\{\lambda_j\}_{j=1}^n$. Así, para cada $k = 1, \ldots, n$, $A^T A(u^k) = \lambda_k u^k$, $\|u^k\|_2 = 1$ y $\langle u^k, u^\ell \rangle = 0$, $k \neq \ell$, $\ell = 1, \ldots, n$.
Sea $x \in \mathbb{R}^n$. Podemos expresar x, de manera única, en términos de los elementos de $\mathfrak{B}$ como $x = \sum_{k=1}^n \alpha_k u^k$, para ciertos $\alpha_k \in \mathbb{R}$, $k = 1, \ldots, n$. Por ser $\mathfrak{B}$ ortonormal, se tiene entonces que $\|x\|_2 = \sum_{k=1}^n |\alpha_k|^2$. Además,

$$A^T A x = \sum_{k=1}^n \alpha_k A^T A u^k = \sum_{k=1}^n \alpha_k \lambda_k u^k,$$

y de nuevo la ortonormalidad de $\mathfrak{B}$ lleva a que $\|A^T A x\|_2^2 = \sum_{k=1}^n |\alpha_k \lambda_k|^2$.
Podemos escribir entonces que

$$\lambda_1^2 \|x\|_2^2 = \lambda_1^2 \sum_{k=1}^n |\alpha_k|^2 \leq \|A^T A x\|_2^2 \leq \lambda_n^2 \sum_{j=1}^n |\alpha_j|^2 = \lambda_n^2 \|x\|_2^2.$$

Teniendo ahora en cuenta que $\|Ax\|_2^2 = \langle Ax, Ax \rangle = \langle A^T A x, x \rangle \leq \|A^T A x\|_2 \|x\|_2$, podemos afirmar que $\|T_A(x)\|_2 = \|Ax\|_2 \leq \sqrt{\lambda_n} \|x\|_2$, por lo que $\|A\|_{2,2} \leq \sqrt{\lambda_n}$.
Por otra parte, se tiene que, para cada $k = 1, \ldots, n$,

$$\|T_A(u^k)\|_2 = \|Au^k\|_2^2 = < A^T A u^k, u^k >= \lambda_k < u^k, u^k >= \lambda_k \|u^k\|_2 = \lambda_k.$$

En particular, $\|T_A(u^n)\|_2 = \sqrt{\lambda_n}$, y ya que $\|u^n\|_2 = 1$, se sigue que $\|A\|_{2,2} \geq \|T_A(u^n)\|_2 = \sqrt{\lambda_n}$. Concluimos que

$$\|A\|_{2,2} = \sqrt{\lambda_n} = \sqrt{\text{máx}\{\text{ autovalores de } A^T A\}}.$$

Terminamos esta sección considerando una clase particular de normas en $M_n(\mathbb{R})$, las normas submultiplicativas. Se dice que una norma $\|\cdot\|$ en $M_n(\mathbb{R})$ es *submultiplicativa* cuando se verifica que $\|AB\| \leq \|A\| \|B\|$, $A, B \in M_n(\mathbb{R})$.

Veamos, en primer lugar, qué podemos decir de las normas $\|\cdot\|_p$, cuando $1 \leq p \leq \infty$.

Proposición 1.9. *Las normas* $\|\cdot\|_1$ *y* $\|\cdot\|_2$ *son submultiplicativas en* $M_n(\mathbb{R})$.

Demostración. Sean $A=(a_{ij})_{i,j=1}^n$ y $B=(b_{ij})_{i,j=1}^n\in M_n(\mathbb{R})$. Se tiene que

$$AB=\Big(\sum_{\ell=1}^n a_{i\ell}b_{\ell j}\Big)_{i,j=1}^n.$$

Luego,

$$\|AB\|_1=\sum_{i,j=1}^n\Big|\sum_{\ell=1}^n a_{i\ell}b_{\ell j}\Big| \quad \text{y} \quad \|AB\|_2=\Big(\sum_{i,j=1}^n\Big|\sum_{\ell=1}^n a_{i\ell}b_{\ell j}\Big|^2\Big)^{1/2}.$$

Entonces, por un lado

$$\begin{aligned}\|AB\|_1&\le\sum_{i,j=1}^n\sum_{\ell=1}^n|a_{i\ell}||b_{\ell j}|\le\sum_{i,j=1}^n\sum_{\ell=1}^n|a_{i\ell}|\sum_{k=1}^n|b_{kj}|\\&=\sum_{i=1}^n\sum_{\ell=1}^n|a_{i\ell}|\sum_{j=1}^n\sum_{k=1}^n|b_{kj}|=\|A\|_1\|B\|_1,\end{aligned}$$

y por otro, aplicando la desigualdad de Cauchy-Schwarz[1] obtenemos

$$\begin{aligned}\|AB\|_2&\le\Big(\sum_{i,j=1}^n\sum_{\ell=1}^n|a_{i\ell}|^2\sum_{k=1}^n|b_{\ell j}|^2\Big)^{1/2}\\&=\Big(\sum_{i=1}^n\sum_{\ell=1}^n|a_{i\ell}|^2\sum_{j=1}^n\sum_{k=1}^n|b_{kj}|^2\Big)^{1/2}=\|A\|_2\|B\|_2,\end{aligned}$$

y el resultado queda establecido. □

Cuando $2<p\le\infty$ la situación es diferente.

Proposición 1.10. *Si $n>1$, entonces se tiene que $\|\cdot\|_p$, $2<p\le\infty$, no es una norma submultiplicativa en $M_n(\mathbb{R})$.*

Demostración. Supongamos que $n>1$. Consideramos las matrices $A=(a_{ij})_{i,j=1}^n$ y $B=A^T$, donde $a_{1j}=1$, $j=1,\dots,n$, y $a_{ij}=0$, en el resto. Entonces,

$$AB=\begin{pmatrix} n & 0 & \cdots & 0\\ 0 & 0 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 0\end{pmatrix}.$$

Es claro que $\|A\|_\infty=1=\|B\|_\infty$ y que $\|AB\|_\infty=n$. Luego, $\|AB\|_\infty>\|A\|_\infty\|B\|_\infty$.

Y si $2<p<\infty$, entonces $\|A\|_p=n^{1/p}=\|B\|_p$ y $\|AB\|_p=n$, y se concluye que $\|AB\|_p=n>n^{2/p}=\|A\|_p\|B\|_p$. □

[1] Desigualdad de Cauchy-Schwarz: $\big|\sum_{i=1}^n x_iy_i\big|\le\big(\sum_{i=1}^n x_i^2\big)^{1/2}\big(\sum_{i=1}^n y_i^2\big)^{1/2}$, $(x_i)_{i=1}^n,(y_i)_{i=1}^n\in\mathbb{R}^n$.

Por útimo, establecemos un resultado para normas del tipo (1.3).

Proposición 1.11. *Sea $\|\cdot\|$ una norma en $\mathbb{R}^n$. Definimos*

$$|||A||| = \sup_{x\in\mathbb{R}^d\setminus\{0\}} \frac{\|Ax\|}{\|x\|}, \quad A \in M_n(\mathbb{R}).$$

Entonces, $|||\cdot|||$ es una norma submultiplicativa en $M_n(\mathbb{R})$.

Demostración. Como ya comentamos, $|||\cdot|||$ es una norma en $M_n(\mathbb{R})$. Vamos a ver que es submultiplicativa.

Sean $A, B \in M_n(\mathbb{R})$. Para cada $x \in \mathbb{R}^n$ tal que $Bx \neq 0$ (nótese que entonces también $x \neq 0$) podemos escribir

$$\frac{\|ABx\|}{\|x\|} = \frac{\|A(Bx)\|}{\|Bx\|}\frac{\|Bx\|}{\|x\|} \leq \sup_{z\in\mathbb{R}^n\setminus\{0\}} \frac{\|Az\|}{\|z\|} \sup_{y\in\mathbb{R}^n\setminus\{0\}} \frac{\|By\|}{\|y\|} = |||A|||\,|||B|||.$$

Por otra parte si $Bx = 0$, siendo $x \neq 0$, entonces $\frac{\|ABx\|}{\|x\|} = 0 \leq |||A|||\,|||B|||$. Concluimos así que $|||AB||| \leq |||A|||\,|||B|||$. □

Observación 1.12. En virtud de las Proposiciones 1.10 y 1.11, podemos afirmar que si $2 < p \leq \infty$, no existe una norma $\|\cdot\|$ en $\mathbb{R}^n$ tal que $|||\cdot||| = \|\cdot\|_p$.

1.3. Espacios métricos. El Teorema del punto fijo de Banach

Dado un conjunto X, una *distancia* o *métrica* es una aplicación $d : X \times X \longrightarrow [0,\infty)$ que a cada par de puntos $x, y \in X$ le asocia un número no negativo $d(x, y)$ que verifica las siguientes condiciones:

(a) *separación*: $d(x, y) = 0$, si y solo si, $x = y$;

(b) *simetría*: $d(x, y) = d(y, x)$, $x, y \in X$;

(c) *desigualdad triangular*: $d(x, y) \leq d(x, z) + d(z, y)$, $x, y, z \in X$.

En tal caso se dice que (X, d) es un espacio métrico.

Sobre un espacio métrico (X, d) podemos considerar una topología asociada a la distancia d. De la misma forma que para los espacios normados, para cada $x \in X$ y $r > 0$, se define $B_d(x, r) = \{y \in X : d(x, y) < r\}$. La topología asociada a d en X es aquella generada por la familia $\{B_d(x, r)\}_{x\in X, r>0}$. Si denotamos por τ_d a esta topología, entonces $A \in \tau_d$ si, y solo si, $A = \emptyset$, o bien, para cada $a \in A$ existe $r > 0$ tal que $B_d(a, r) \subset A$. En adelante, cuando no sea necesario especificar d, escribiremos $B(x, r)$ en lugar de $B_d(x, r)$.

Consideramos a continuación algunos espacios métricos particulares.

1. **Métrica discreta**. Sea $X \neq \emptyset$. Definimos $d : X \times X \longrightarrow [0,\infty)$ mediante

$$d(x, y) = \begin{cases} 1, & x \neq y, \\ 0, & x = y. \end{cases}, \quad x, y \in X.$$

Es sencillo ver que d verifica las condiciones de una métrica sobre X. Además, es claro que $B(x,r) = \{x\}$, cuando $0 < r \leq 1$ y $B(x,r) = X$, si $r > 1$. También se verifica que $\overline{B(x,r)} = B(x,r)$, $x \in X$, $r > 0$.
La topología definida por esta métrica en X se conoce como *topología discreta*. Se observa que $\tau_d = \mathscr{P}(X)$, la clase de todos los subconjuntos de X.

2. **Métrica generada por una norma**. Sea $(X, \|\cdot\|)$ un espacio normado. Definimos $d(x,y) = \|x-y\|$, $x,y \in X$. De las propiedades para la norma se deducen fácilmente las correspondientes para d.
Una característica de esta distancia es que es *invariante por traslaciones*, es decir, se verifica que $d(x+z, y+z) = d(x,y)$, $x,y,z \in X$. Además d es *homogénea* en el sentido de que se tiene que $d(\lambda x, \lambda y) = |\lambda| d(x,y)$, $x,y \in X$ y $\lambda \in \mathbb{R}$.

3. **Métricas no equivalentes que generan la misma topología**. Sea (X,d) un espacio métrico. Definimos la aplicación d_1 mediante
$$d_1(x,y) = \frac{d(x,y)}{1+d(x,y)}, \quad x,y \in X.$$
Veamos que d_1 es una métrica sobre X. Es claro que $d_1 : X \times X \longrightarrow [0,\infty)$ y que $d_1(x,y) = 0$ si, y solo si, $d(x,y) = 0$, lo que equivale, por ser d una métrica, a que $x = y$. Asimismo, también resulta evidente que $d_1(x,y) = d_1(y,x)$, $x,y \in X$.
Probamos ahora que se verifica la desigualdad triangular. Para ello, consideramos la función $h(t) = \frac{t}{1+t}$, $t \in (-1,\infty)$. Un cálculo directo conduce a que
$$h'(t) = \frac{1}{(1+t)^2} > 0, \quad t \in (-1,\infty),$$
por lo que h es una función creciente en $(-1,+\infty)$.
Sean $x,y,z \in X$. Como d es una distancia en X se tiene que $d(x,y) \leq d(x,z) + d(z,y)$. Entonces, $h(d(x,y)) \leq h(d(x,z)+d(z,y))$, y se obtiene que
$$\begin{aligned} d_1(x,y) &= \frac{d(x,y)}{1+d(x,y)} \leq \frac{d(x,z)+d(z,y)}{1+d(x,z)+d(z,y)} \\ &\leq \frac{d(x,z)}{1+d(x,z)} + \frac{d(z,y)}{1+d(z,y)} = d_1(x,z) + d_1(z,y). \end{aligned}$$
Las métricas d y d_1 generan la misma topología. En efecto, denotamos por τ_d y τ_{d_1} las topologías definidas en X por d y d_1, respectivamente. Vamos a establecer que $\tau_d = \tau_{d_1}$.
Consideramos primero $A \in \tau_d \backslash \{\emptyset\}$. Para que $A \in \tau_{d_1}$ debemos asegurar que para cada $x \in A$ existe $r > 0$ de manera que $B_{d_1}(x,r) \subset A$. Así que vamos a fijar $x \in A$. Como A es abierto en τ_d, sabemos que existe $r_0 > 0$ de modo que $B_d(x,r_0) \subset A$. Luego, basta encontrar $r > 0$ tal que $B_{d_1}(x,r) \subset B_d(x,r_0)$.

Observamos que si $y \in B_{d_1}(x,r)$, con $0<r<1$, entonces

$$d_1(x,y) = \frac{d(x,y)}{1+d(x,y)} < r \quad \text{si , y solo si,} \quad d(x,y) < \frac{r}{1-r}.$$

Luego, basta elegir $r \in (0,1)$ de manera que $\frac{r}{1-r} \leq r_0$, o lo que es equivalente, $0<r\leq \frac{r_0}{r_0+1}$. De esta forma, $B_{d_1}(x,r) \subseteq B_d(x,r_0)$.
Supongamos ahora que $A \in \tau_{d_1} \setminus \{\emptyset\}$. Ya que $d_1(x,y) \leq d(x,y)$, $x,y \in X$, se verifica que $B_d(x,r) \subset B_{d_1}(x,r)$, $x \in X$, $r>0$. Luego, dados $x \in A$ y $r_0>0$ tales que $B_{d_1}(x,r_0) \subset A$, tomando $r=r_0$ se obtiene que $B_d(x,r) \subset A$.
Si consideramos, en particular, $X=\mathbb{R}$ y d la distancia euclídea, teniendo en cuenta lo que hemos probado se tiene que las métricas d y $d_1(x,y) = \frac{|x-y|}{1+|x-y|}$, $x,y \in \mathbb{R}$, definen la misma topología en $\mathbb{R}$. Sin embargo, no son métricas equivalentes en $\mathbb{R}$.
Observamos que siempre se cumple que $d_1(x,y) \leq d(x,y)$, $x,y \in \mathbb{R}$, pero no podemos encontrar $c>0$ de forma que $d(x,y) \leq c d_1(x,y)$, $x,y \in \mathbb{R}$. Basta notar que para cada $n \in \mathbb{N}$, $d(0,n)=n$ y $d_1(0,n) = \frac{n}{n+1} \leq 1$ y no es posible encontrar $c>0$ de manera que $n \leq c$, $n \in \mathbb{N}$.
Una consecuencia de este hecho es que d_1 no es una métrica asociada a una norma. Es decir, no existe una norma $\|\cdot\|$ sobre $\mathbb{R}$ de manera que se cumpla que $d_1(x,y) = \|x-y\|$, $x,y \in \mathbb{R}$. Supongamos que, por el contrario, sí existe tal norma. Recordando que en $\mathbb{R}$ todas las normas son equivalentes, podemos entonces encontrar $c \geq 1$ tal que $\frac{1}{c}\|x\| \leq |x| \leq c\|x\|$, $x \in \mathbb{R}$. Y esto conduce a que d y d_1 son métricas equivalentes, lo cual hemos visto que no es cierto.

Una de las propiedades relevantes para un espacio métrico es la completitud.

Sea (X,d) un espacio métrico. Se dice que una sucesión $(x_n)_{n=1}^{\infty} \subset X$ es *de Cauchy* cuando se verifica que para todo $\varepsilon>0$ existe $n_0 \in \mathbb{N}$ tal que $d(x_k,x_\ell) < \varepsilon$, para todo $k,\ell \geq n_0$. Por otro lado, una sucesión $(x_n)_{n=1}^{\infty} \subset X$ es *convergente a* $x \in X$ cuando para todo $\varepsilon>0$ podemos encontrar $n_0 \in \mathbb{N}$ de modo que $d(x_n,x)<\varepsilon$, cuando $n \geq n_0$.

Proposición 1.13. *Sean (X,d) un espacio métrico y $(x_n)_{n=1}^{\infty}$ una sucesión convergente en X. Entonces $(x_n)_{n=1}^{\infty}$ es de Cauchy en X.*

Demostración. Supongamos que $(x_n)_{n=1}^{\infty}$ converge a $x \in X$. Fijamos $\varepsilon>0$. Sabemos que existe $n_0 \in \mathbb{N}$ tal que $d(x_n,x) < \frac{\varepsilon}{2}$, cuando $n \geq n_0$. Usando la desigualdad triangular se deduce entonces que

$$d(x_k,x_\ell) \leq d(x_k,x) + d(x,x_\ell) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon, \quad k,\ell \geq n_0.$$

□

El resultado inverso no siempre es cierto, lo que da pie a definir el concepto de completitud. Decimos que un espacio métrico (X,d) es *completo* cuando se cumple que toda sucesión de Cauchy es convergente en X.

Mencionamos ahora algunos espacios métricos concretos. El primer ejemplo de espacio métrico completo que suele aparecer en cualquier texto es $(\mathbb{R}, |\cdot|)$. En general, el espacio $(\mathbb{R}^n, d_{\|\cdot\|})$, $n \geq 1$, con $d_{\|\cdot\|}$ la métrica asociada a la norma euclídea $\|\cdot\|$ en $\mathbb{R}^n$ es un espacio métrico completo. Y, dado que todas las normas en $\mathbb{R}^n$ son equivalentes (por ser $\mathbb{R}^n$ de dimensión finita) también es completo con respecto a la distancia generada por cualquier norma en $\mathbb{R}^n$.

El espacio $(\mathbb{R}^n, d_{\|\cdot\|})$ es un caso particular de los conocidos como espacios de Banach. Sea $(X, \|\cdot\|)$ un espacio vectorial normado. Se dice que $(X, \|\cdot\|)$ es *un espacio de Banach* cuando $(X, d_{\|\cdot\|})$ es completo, siendo $d_{\|\cdot\|}$ la métrica asociada a la norma $\|\cdot\|$. Se tiene que cualquier espacio normado de dimensión finita es de Banach.

Veamos, por último, un ejemplo de espacio métrico que no es completo. Denotamos por $P[x]$ el espacio vectorial de los polinomios en una variable con coeficientes reales definidos en [0.1]. Observamos que $\mathfrak{B} = \{x^n\}_{n=0}^{\infty}$ es una base para el espacio vectorial $P[x]$. Por tanto, $\dim P[x] = \infty$.

Sobre $P[x]$ consideramos la norma $\|\cdot\|_\infty$ dada por $\|p\|_\infty = \text{máx}_{x\in[0,1]} |p(x)|$, para cada $p \in P[x]$. De este modo $\|\cdot\|_\infty$ es una norma en $P[x]$.

Probamos que $(P[x], \|\cdot\|_\infty)$ no es completo. Para ello, consideramos la función continua $f(x) = e^x$, $x \in [0,1]$. El teorema de Weierstrass nos dice que existe una sucesión de polinomios $(p_n)_{n=1}^{\infty} \subset P[x]$ tal que

$$\lim_{n\to\infty} \text{máx}_{x\in[0,1]} |p_n(x) - f(x)| = 0. \tag{1.6}$$

Veamos que $(p_n)_{n=1}^{\infty}$ es de Cauchy en $(P[x], \|\cdot\|_\infty)$. Sea $\varepsilon > 0$. De (1.6) se sigue que existe $n_0 \in \mathbb{N}$ tal que $\text{máx}_{x\in[0,1]} |p_n(x) - f(x)| < \frac{\varepsilon}{2}$, $n \geq n_0$. Por tanto, si $k, \ell \in \mathbb{N}$, $k, \ell \geq n_0$, podemos escribir

$$\|p_k - p_\ell\|_\infty \leq \underset{x\in[0,1]}{\text{máx}} |p_k(x) - f(x)| + \underset{x\in[0,1]}{\text{máx}} |p_\ell(x) - f(x)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Supongamos ahora que, para cierto polinomio $p \in P[x]$, se tiene que $p_n \longrightarrow p$, cuando $n \to \infty$, en $(P[x], \|\cdot\|_\infty)$, esto es,

$$\|p_n - p\|_\infty = \underset{x\in[0,1]}{\text{máx}} |p_n(x) - p(x)| \longrightarrow 0, \text{ cuando } n \to \infty.$$

Pero, por (1.6), y teniendo en cuenta que

$$0 \leq \underset{x\in[0,1]}{\text{máx}} |f(x) - p(x)| \leq \underset{x\in[0,1]}{\text{máx}} |f(x) - p_n(x)| + \underset{x\in[0,1]}{\text{máx}} |p_n(x) - p(x)|, \quad n \in \mathbb{N},$$

se concluye entonces que $\text{máx}_{x\in[0,1]} |f(x) - p(x)| = 0$, esto es, $f(x) = p(x)$, $x \in [0,1]$. Y esto es un absurdo pues f no es un polinomio.

Terminamos este apartado presentando el resultado central de esta sección. Este teorema será de utilidad en la demostración del Teorema de la función inversa (Teorema 2.18).

Teorema 1.14 (Teorema del punto fijo de Banach). *Sea* (X,d) *un espacio métrico completo. Supongamos que* $T : X \longrightarrow X$ *es una aplicación contractiva, esto es, que existe* $0 < \alpha < 1$ *tal que:* $d(T(x),T(y)) \leq \alpha d(x,y)$, $x, y \in X$. *Entonces, existe un único* $x_0 \in X$ *verificando* $T(x_0) = x_0$.

Demostración. Probamos en primer lugar la unicidad. Para ello supongamos que $x_1, x_2 \in X$ son dos puntos fijos para T, esto es, $Tx_1 = x_1$ y $Tx_2 = x_2$. Entonces, $d(x_1,x_2) = d(T(x_1),T(x_2)) \leq \alpha d(x_1,x_2)$. Ya que $0 < \alpha < 1$ se tiene que $d(x_1,x_2) = 0$ y por tanto, $x_1 = x_2$.

Analizamos ahora la existencia. Tomamos $x_0 \in X$ y definimos, para cada $n \in \mathbb{N}$, el elemento $x_n = T^n(x_0) = T \circ T \circ \overset{n}{\cdots} \circ T(x_0)$. Veamos que $(x_n)_{n\in\mathbb{N}}$ es una sucesión de Cauchy en (X,d). Sea $\varepsilon > 0$. Haciendo uso de la desigualdad triangular, para cada $n, m \in \mathbb{N}$, con $m \geq n$, podemos escribir

$$d(x_n,x_m) = d(T^n(x_0),T^m(x_0)) \leq \sum_{k=n}^{m-1} d(T^k(x_0),T^{k+1}(x_0)).$$

Ya que, para cada $\ell \in \mathbb{N}$,

$$d(T^\ell(x_0),T^{\ell+1}(x_0)) \leq \alpha d(T^{\ell-1}(x_0),T^\ell(x_0)) \leq \ldots \leq \alpha^\ell d(x_0,T(x_0)),$$

obtenemos que $d(x_n,x_m) \leq \sum_{k=n}^{m-1} \alpha^k d(x_0,T(x_0))$, $m \geq n$. Puesto que $\alpha \in (0,1)$, la serie $\sum_{k=0}^{\infty} \alpha^k$ es convergente y, por tanto, de Cauchy. Podemos encontrar así $n_0 \in \mathbb{N}$ de manera que $|\sum_{k=n}^{m-1} \alpha^k| < \varepsilon_0$, $m > n \geq n_0$, siendo $\varepsilon_0 = \varepsilon(1 + d(x_0,T(x_0)))^{-1}$.

Se sigue que

$$d(x_n,x_m) < \frac{\varepsilon}{1 + d(x_0,T(x_0))} d(x_0,T(x_0)) < \varepsilon, \quad m > n \geq n_0.$$

Teniendo en cuenta que (X,d) es completo, se concluye que la sucesión $(x_n)_{n\in\mathbb{N}}$ es convergente, esto es, existe $\widetilde{x} \in X$ tal que $x_n \longrightarrow \widetilde{x}$, cuando $n \to \infty$.

Además $d(T(x_n),T(\widetilde{x})) \leq \alpha d(x_n,\widetilde{x})$, $n \in \mathbb{N}$, lo que nos permite concluir que $T(x_n) \longrightarrow T(\widetilde{x})$, cuando $n \to \infty$. Por otro lado, para cada $n \in \mathbb{N}$, $T(x_n) = T(T^n(x_0)) = T^{n+1}(x_0) = x_{n+1}$. Por tanto,

$$\lim_{n\to\infty} T(x_n) = \lim_{n\to\infty} x_{n+1} = \widetilde{x}.$$

Ya que el límite de una sucesión en (X,d), si existe, es único, concluimos que $T(\widetilde{x}) = \widetilde{x}$. □

2

Teoremas de la función inversa e implícita

2.1. Teorema de la función inversa en el contexto real

Como paso previo al estudio del Teorema de la función inversa en el contexto habitual vector-valuado, vamos a analizar algunos resultados sobre la existencia y propiedades de la función inversa para funciones reales de variable real.

Nuestros primeros resultados establecen la relación entre los conceptos de continuidad y monotonía.

Dado $\Omega \subset \mathbb{R}$ denotamos por $\mathrm{Ais}(\Omega)$ el conjunto de todos los *puntos aislados* en Ω, esto es, los valores $a \in \Omega$ para los que existe $r > 0$ de manera que se tiene que $(a-r, a+r) \cap \Omega = \{a\}$ (a es el único punto de Ω en el intervalo $(a-r, a+r)$). Por otro lado, el *conjunto derivado* de Ω, $\mathrm{Der}(\Omega)$, está constituido por todos sus *puntos de acumulación*, lo que significa que $a \in \mathrm{Der}(\Omega)$ si, y solo si, para cada $r > 0$, $(a-r, a+r) \setminus \{a\} \cap \Omega \neq \emptyset$.

Sea $f : \Omega \subset \mathbb{R} \to \mathbb{R}$. Diremos que f es continua en $a \in \Omega$ cuando $a \in \mathrm{Ais}(\Omega)$, o bien, cuando $a \in \mathrm{Der}(\Omega)$ y se verifica que, para todo $\varepsilon > 0$, existe $\delta > 0$ de tal manera que $|f(x) - f(a)| < \varepsilon$ cuando $x \in \Omega$ y $|x - a| < \delta$.

Recordamos también que se dice que f es *monótona* cuando es *creciente* ($f(x_1) \leq f(x_2)$, si $x_1 < x_2$) o bien *decreciente* ($f(x_1) \geq f(x_2)$, si $x_1 < x_2$). Y hablamos de *estrictamente monótona* cuando las desigualdades anteriores son estrictas.

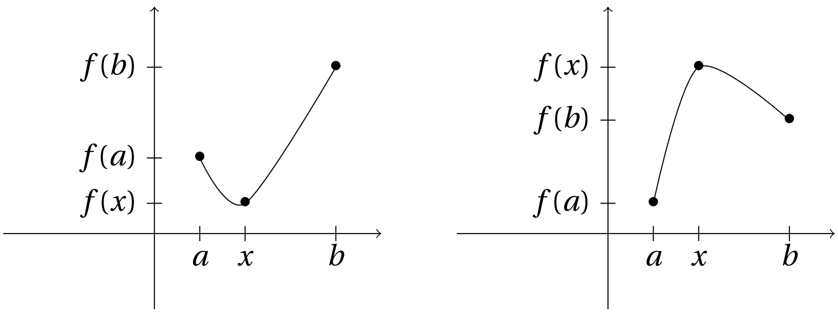
Como establecemos a continuación, cuando Ω es un intervalo, la continuidad y la inyectividad implican monotonía.

Proposición 2.1. *Sean I un intervalo y $f : I \longrightarrow \mathbb{R}$ una función continua e inyectiva. Entonces, f es estrictamente monótona.*

Demostración. Del hecho de que f sea inyectiva, si probamos que f es monótona podermos concluir que es estrictamente monótona.

Asumimos primero que I es un intervalo de la forma $I = [a, b]$, siendo $-\infty < a < b < +\infty$. Como $a \neq b$ y f es inyectiva, necesariamente $f(a) \neq f(b)$. Podemos suponer, sin pérdida de generalidad, que $f(a) < f(b)$. Nuestro objetivo es probar que $f(a) \leq f(x) \leq f(b)$, $x \in [a, b]$. Es evidente para $x = a$ y $x = b$. Fijamos entonces

$x \in (a,b)$. Ya que $x \neq a$ y $x \neq b$, se tiene que $f(x) \neq f(a)$ y $f(x) \neq f(b)$ y entonces nos encontramos en alguna de las dos situaciones reflejadas en la figura siguiente.



Supongamos que $f(x) < f(a)$. En este caso tenemos que f es continua en $[x,b]$ y $f(x) < f(a) < f(b)$. En virtud del teorema del valor intermedio[1] podemos encontrar $c \in (a,b)$ tal que $f(c) = f(a)$. Pero esto no es posible puesto que f es una función inyectiva. De la misma forma, si $f(x) > f(b)$ tenemos que f es continua en $[a,x]$ y $f(a) < f(b) < f(x)$ y, de nuevo, por el teorema del valor intermedio se deduce que existe $c \in (a,b)$ tal que $f(c) = f(b)$, lo que no es posible al ser f inyectiva. Por tanto, $f(a) < f(x) < f(b)$.

Sean ahora $x, y \in [a,b]$ tales que $x < y$. Como f es continua e inyectiva en $[x,b]$ y además, $y \in [x,b]$, lo que acabamos de establecer conduce a que $f(x) \leq f(y) \leq f(b)$.

Consideramos ahora un intervalo I cualquiera y vamos a suponer que f no es monótona. Esto significa que existen $x_1, y_1, x_2, y_2 \in I$ de manera que

$$x_1 < y_1,\ f(x_1) < f(y_1), \quad \text{y} \quad x_2 < y_2,\ f(x_2) > f(y_2). \tag{2.1}$$

Tomamos $x_0 = \text{mín}\{x_1, x_2\}$ e $y_0 = \text{máx}\{y_1, y_2\}$. Es claro que $x_0 < y_0$ y que $[x_0, y_0] \subset I$. Luego, ya que f es una función continua e inyectiva en $[x_0, y_0]$, lo que probamos antes nos dice que f es estrictamente monótona en $[x_0, y_0]$.

Pero $x_1, y_1, x_2, y_2 \in [x_0, y_0]$. Luego por (2.1) tendríamos que f no es monótona en $[x_0, y_0]$. Esta contradicción permite concluir que f es monótona en I. □

Podemos establecer además que el conjunto imagen de un intervalo por una función continua y monótona es también un intervalo.

Proposición 2.2. *Sean I un intervalo y $f : I \longrightarrow \mathbb{R}$ una función continua y monótona en I. Entonces $f(I)$ es un intervalo.*

Demostración. Asumimos, sin pérdida de generalidad, que f es creciente en I.

Si el intervalo es cerrado y acotado, esto es, $I = [a,b]$, con $-\infty < a < b < +\infty$, entonces podemos asegurar que $f([a,b]) = [f(a), f(b)]$.

[1] **Teorema del valor intermedio**. Sean $-\infty < a < b < +\infty$ y $f : [a,b] \longrightarrow \mathbb{R}$ una función continua. Si $\lambda \in \mathbb{R}$ verifica $f(a) < \lambda < f(b)$ ó $f(b) < \lambda < f(a)$, entonces existe $c \in (a,b)$ tal que $f(c) = \lambda$.

En efecto, es claro que por ser f creciente, $f(a) \leq f(x) \leq f(b)$, para todo $x \in [a,b]$. Luego $f([a,b]) \subseteq [f(a), f(b)]$. Por otro lado, dado $y \in [f(a), f(b)]$, puesto que f es continua en $[a,b]$, el teorema del valor intermedio nos dice que podemos encontrar $c \in [a,b]$ tal que $f(c) = y$. Por tanto, $y \in f([a,b])$.

Supongamos ahora que $I = (a,b)$, donde $-\infty \leq a < b \leq +\infty$. Nuestro objetivo es comprobar que se satisface la siguiente relación

$$\Big(\inf_{x\in(a,b)} f(x), \sup_{x\in(a,b)} f(x)\Big) \subset f((a,b)) \subset \Big[\inf_{x\in(a,b)} f(x), \sup_{x\in(a,b)} f(x)\Big]. \tag{2.2}$$

De esta manera, podemos asegurar que $f((a,b))$ es un intervalo. Este puede ser finito o infinito y abierto, cerrado, semiabierto a la derecha o semiabierto a la izquierda.

Es claro que dado $z \in (a,b)$ se tiene que

$$\inf_{x\in(a,b)} f(x) \leq f(z) \leq \sup_{x\in(a,b)} f(x),$$

por lo que la segunda inclusión en (2.2) se satisface trivialmente.

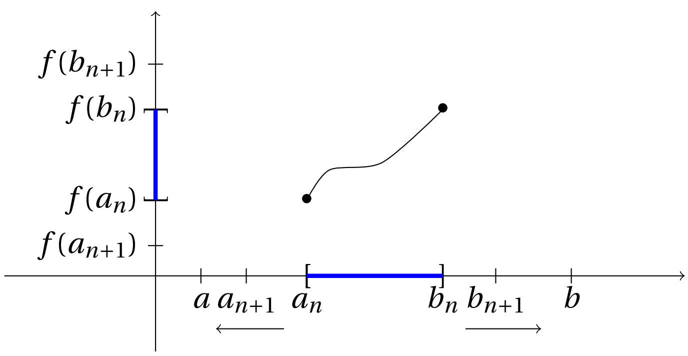
Consideramos ahora dos sucesiones $(a_n)_{n=1}^{\infty}$ y $(b_n)_{n=1}^{\infty}$ tales que $a_n \longrightarrow a$ y $b_n \longrightarrow b$, cuando $n \to \infty$, y de forma que

$$a < a_{n+1} < a_n < a_1 < b_1 < b_n < b_{n+1} < b, \quad n \in \mathbb{N}.$$

Tenemos que, para cada $n \in \mathbb{N}$, $[a_n, b_n] \subset (a,b)$, y puesto que f es continua y creciente en $[a_n, b_n]$, el argumento anterior para los intervalos cerrados y acotados conduce a que $f([a_n, b_n]) = [f(a_n), f(b_n)]$, $n \in \mathbb{N}$.

Por otro lado, de las condiciones dadas para las sucesiones $(a_n)_{n=1}^{\infty}$ y $(b_n)_{n=1}^{\infty}$ se sigue que $(a,b) = \cup_{n\in\mathbb{N}}[a_n, b_n]$. Por tanto,

$$f((a,b)) = f\Big(\bigcup_{n=1}^{\infty}[a_n, b_n]\Big) = \bigcup_{n=1}^{\infty} f([a_n, b_n]) = \bigcup_{n=1}^{\infty}[f(a_n), f(b_n)]. \tag{2.3}$$



Observamos que dado que f es creciente se tiene que $f(a_{n+1}) \leq f(a_n) \leq f(b_n) \leq f(b_{n+1})$, $n \in \mathbb{N}$ y, por tanto, la colección de conjuntos $\big\{[f(a_n), f(b_n)]\big\}_{n=1}^{\infty}$ es una sucesión creciente de intervalos encajados.

Con el fin de establecer la primera desigualdad en (2.2) veamos primero que

$$(\inf_{n\in\mathbb{N}} f(a_n), \sup_{n\in\mathbb{N}} f(b_n)) \subset f((a,b)). \tag{2.4}$$

Sea $y \in (\text{ínf}_{n\in\mathbb{N}} f(a_n), \sup_{n\in\mathbb{N}} f(b_n))$. Afirmamos que existe $n_0 \in \mathbb{N}$ tal que $f(a_{n_0}) < y < f(b_{n_0})$. En efecto, si $y \leq f(a_n)$ ó $y \geq f(b_n)$, para todo $n \in \mathbb{N}$, entonces, $y \leq \text{ínf}_{n\in\mathbb{N}} f(a_n)$, o bien, $y \geq \sup_{n\in\mathbb{N}} f(b_n)$, lo que contradice nuestra hipótesis sobre y. Por tanto, podemos encontrar $n_1, n_2 \in \mathbb{N}$ tales que $f(a_{n_1}) < y < f(b_{n_2})$. Tomando $n_0 = \text{máx}\{n_1, n_2\}$ y teniendo en cuenta que f es creciente, que $a_n \downarrow a$ y que $b_n \uparrow b$, obtenemos entonces que $f(a_{n_0}) \leq f(a_{n_1}) < y < f(b_{n_2}) \leq f(b_{n_0})$ y nuestra afirmación queda demostrada.

Hemos establecido que $y \in [f(a_{n_0}), f(b_{n_0})]$. Luego, en virtud de (2.3) se deduce que $y \in f((a,b))$ y así (2.4) queda probado.

Para terminar la demostración de (2.2) bastaría entonces con justificar que

$$\text{ínf}_{n\in\mathbb{N}} f(a_n) = \text{ínf}_{x\in(a,b)} f(x) \quad \text{y} \quad \sup_{n\in\mathbb{N}} f(b_n) = \sup_{x\in(a,b)} f(x). \tag{2.5}$$

Observamos primero que, ya que $a_n, b_n \in (a,b)$, $n \in \mathbb{N}$, se tiene que

$$\text{ínf}_{x\in(a,b)} f(x) \leq f(a_n) \quad \text{y} \quad f(b_n) \leq \sup_{x\in(a,b)} f(x), \quad n \in \mathbb{N}.$$

Por tanto,

$$\text{ínf}_{x\in(a,b)} f(x) \leq \text{ínf}_{n\in\mathbb{N}} f(a_n) \quad \text{y} \quad \sup_{n\in\mathbb{N}} f(b_n) \leq \sup_{x\in(a,b)} f(x).$$

Por otro lado, dado $x \in (a,b)$, como $a_n \downarrow a$ y $b_n \uparrow b$, cuando $n \to \infty$, podemos encontrar $n_x \in \mathbb{N}$ tal que $a < a_{n_x} < x < b_{n_x} < b$. La monotonía de f lleva a que

$$\text{ínf}_{n\in\mathbb{N}} f(a_n) \leq f(a_{n_x}) \leq f(x) \leq f(b_{n_x}) \leq \sup_{n\in\mathbb{N}} f(b_n).$$

Obtenemos así que

$$\text{ínf}_{n\in\mathbb{N}} f(a_n) \leq \text{ínf}_{x\in(a,b)} f(x) \quad \text{y} \quad \sup_{x\in(a,b)} f(x) \leq \sup_{n\in\mathbb{N}} f(b_n),$$

y podemos deducir entonces que se verifica (2.5) y, como consecuencia, (2.2) queda establecido. □

Observación 2.3. Si $f : (a,b) \longrightarrow \mathbb{R}$ es continua y creciente como en la prueba del resultado anterior, podemos asegurar además que

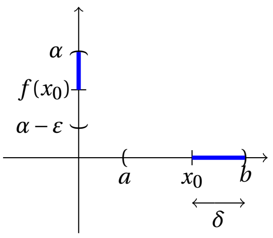
(a) $\sup_{x\in(a,b)} f(x) = \lim_{x\to b^-} f(x)$;

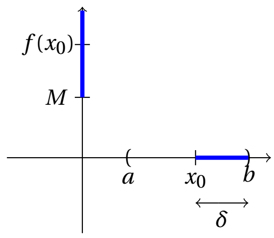
(b) $\text{ínf}_{x\in(a,b)} f(x) = \lim_{x\to a^+} f(x)$.

Para justificar (a) supongamos en primer lugar que $\sup_{x\in(a,b)} f(x) = \alpha < \infty$. Sea $\varepsilon > 0$. Queremos determinar $\delta > 0$ tal que $|f(x) - \alpha| = \alpha - f(x) < \varepsilon$, cuando $x \in (a,b)$ y $0 < b - x < \delta$, es decir, $\alpha - \varepsilon < f(x)$, $x \in (a,b)$, $x \in (b-\delta, b)$.

Por definición de supremo sabemos que existe $x_0 \in (a, b)$ tal que $\alpha - \varepsilon < f(x_0) \leq \alpha$. Ya que f es creciente se tiene que $f(x_0) \leq f(x)$, $x \in (x_0, b)$. Por tanto, basta tomar $\delta = b - x_0$ para obtener que $\alpha - \varepsilon < f(x) \leq \alpha$ cuando $x \in (b - \delta, b)$.

Supongamos ahora que $\sup_{x \in (a,b)} f(x) = +\infty$. En este caso, $\lim_{x \to b^-} f(x) = +\infty$. En efecto, fijamos $M > 0$. Debemos encontrar $\delta > 0$ de modo que para todo $x \in (a, b)$ tal que $0 < b - x < \delta$ se tiene que $f(x) > M$. Ya que $\sup_{x \in (a,b)} f(x) = +\infty$, existe $x_0 \in (a, b)$ tal que $f(x_0) > M$. Como f es monótona creciente se sigue que $f(x) \geq f(x_0) > M$ cuando $x \in (x_0, b)$. Si tomamos $\delta = b - x_0$ se obtiene que para todo $x \in (b - \delta, b) = (x_0, b)$ se verifica $f(x) > M$. De esta forma, $\lim_{x \to b^-} f(x) = +\infty$.

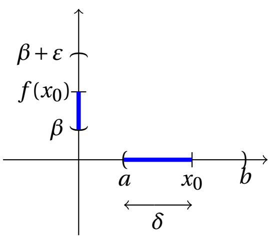


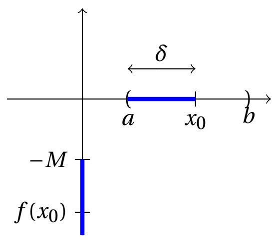


(b) Suponemos primero que $\inf_{x \in (a,b)} f(x) = \beta < \infty$ y fijamos $\varepsilon > 0$. Nuestro objetivo ahora es determinar $\delta > 0$ tal que $|\beta - f(x)| = f(x) - \beta < \varepsilon$, cuando $x \in (a, b)$ y $0 < x - a < \delta$, o de forma equivalente, que $f(x) < \beta + \varepsilon$, $x \in (a, b)$ y $x \in (a, a + \delta)$.

Ahora, por definicion de ínfimo, existe $x_0 \in (a, b)$ tal que $\beta \leq f(x_0) < \beta + \varepsilon$. Como f es creciente se tiene que $f(x) \leq f(x_0)$ cuando $a < x < x_0$. Y por tanto, $\beta \leq f(x) < \beta + \varepsilon$, $x \in (a, x_0)$. Tomando $\delta = x_0 - a > 0$ tenemos que $\beta \leq f(x) < \beta + \varepsilon$ cuando $x \in (a, a + \delta)$. Entonces $\lim_{x \to a^+} f(x) = \beta$.

Si $\inf_{x \in (a,b)} f(x) = -\infty$, entonces se tiene que $\lim_{x \to a^+} f(x) = -\infty$. En efecto, sea $M > 0$. Ya que $\inf_{x \in (a,b)} f(x) = -\infty$ podemos encontrar $x_0 \in (a, b)$ tal que $f(x_0) < -M$. Y como f es creciente se sigue que $f(x) \leq f(x_0) < -M$ cuando $a < x < x_0$. Tomando $\delta = x_0 - a > 0$ llegamos a que, para todo $x \in (a, a + \delta) = (a, x_0)$ se satisface $f(x) < -M$ y, por tanto, $\lim_{x \to a^+} f(x) = -\infty$.





$\square$

Un resultado, en cierto sentido, inverso al anterior, nos dice que el hecho de que el conjunto imagen de una función monótona sea un intervalo resulta ser una condición suficiente para que la función sea continua.

Proposición 2.4. *Sea $f:\Omega\subset\mathbb{R}\to\mathbb{R}$ una función monótona. Si $f(\Omega)$ es un intervalo, entonces f es continua en Ω.*

Demostración. Suponemos que f es creciente y que $f(\Omega)$ es un intervalo. Cuando f es decreciente se procede de forma análoga.

Sea $a\in\Omega$. Si $a\in\mathrm{Ais}(\Omega)$, entonces, por definición, f es continua en a. Asumamos entonces que $a\in\mathrm{Der}(\Omega)$. Para probar que f es continua en a, es suficiente ver que para cada sucesión monótona creciente $(a_n)_{n=1}^{\infty}\subset\Omega\setminus\{a\}$ tal que $\lim_{n\to\infty}a_n=a$, se tiene que $\lim_{n\to\infty}f(a_n)=f(a)$ (Ver Observación 2.5).
Fijamos una sucesión monótona creciente $(a_n)_{n=1}^{\infty}\subset\Omega\setminus\{a\}$ tal que $\lim_{n\to\infty}a_n=a$. Vamos a probar que $\lim_{n\to\infty}f(a_n)=f(a)$.

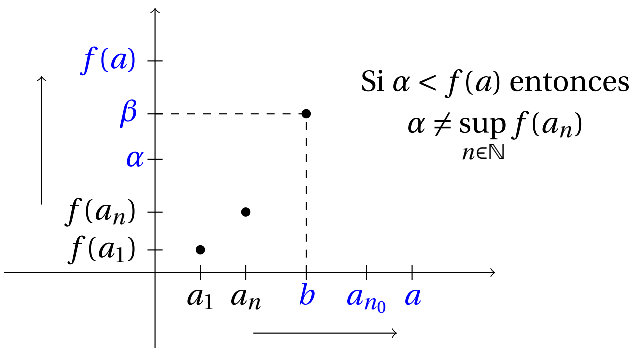
Ya que $(a_n)_{n=1}^{\infty}$ es creciente y $a_n<a$, $n\in\mathbb{N}$, la monotonía de f conduce a que $(f(a_n))_{n=1}^{\infty}$ es una sucesión creciente y

$$\lim_{n\to\infty}f(a_n)=\sup_{n\in\mathbb{N}}f(a_n)\le f(a).$$

Veamos que $\alpha=\sup_{n\in\mathbb{N}}f(a_n)=f(a)$. Para ello, supongamos que $\alpha<f(a)$.

Ya que $f(\Omega)$ es un intervalo y $f(a_1),f(a)\in f(\Omega)$, siendo $f(a_1)<f(a)$, se tiene que $[f(a_1),f(a)]\subset f(\Omega)$. Además, como $f(a_1)\le\alpha$, el intervalo $(\alpha,f(a))\subset f(\Omega)$. Por tanto, tomando $\beta\in(\alpha,f(a))$, podemos encontrar $b\in\Omega$ tal que $f(b)=\beta$.

Observamos que $b<a$ pues f es creciente y $f(b)=\beta\in(\alpha,f(a))$. Entonces, como $a_n\uparrow a$, cuando $n\to\infty$, existe $n_0\in\mathbb{N}$ de forma que $b<a_{n_0}<a$. Y, por tanto, $\beta=f(b)\le f(a_{n_0})$. Pero entonces, $\beta\le\alpha$ lo que contradice nuestra hipótesis inicial.



Se concluye así que no existe $\beta\in(\alpha,f(a))$, y por tanto, $\alpha=f(a)$. □

Observación 2.5. En la prueba anterior hemos usado la siguiente propiedad: *Sea $f:\Omega\subset\mathbb{R}\longrightarrow\mathbb{R}$ una función y $a\in\mathrm{Der}(\Omega)$. Si para toda sucesión $(a_n)_{n=1}^{\infty}\subset\Omega\setminus\{a\}$ monótona tal que $\lim_{n\to\infty}a_n=a$ se tiene que $\lim_{n\to\infty}f(a_n)=f(a)$, entonces f es continua en a.*

Demostración. Vamos a suponer que f no es continua en a. Nuestro objetivo es encontrar una sucesión $(a_n)_{n=1}^{\infty}\subset\Omega\setminus\{a\}$ monótona tal que $a_n\longrightarrow a$, $n\to\infty$, mientras que $(f(a_n))_{n=1}^{\infty}$ no converge a $f(a)$.

Ya que f no es continua en a, existe $\varepsilon > 0$ tal que, para cada $n \in \mathbb{N}$ podemos encontrar $a_n \in \Omega$ siendo $|a_n - a| < 1/n$ y $|f(a_n) - f(a)| \geq \varepsilon$. Nótese que $a_n \neq a$, $n \in \mathbb{N}$.

Definimos los conjuntos W_+ y W_- como sigue

$$W_+ = \{n \in \mathbb{N} : a_n > a\} \quad \text{y} \quad W_- = \{n \in \mathbb{N} : a_n < a\}.$$

Entonces, $\mathbb{N} = W_+ \cup W_-$, y por tanto, al menos uno de los dos conjuntos W_+ ó W_- es infinito.

Supongamos que W_- es infinito y tomamos $n_1 \in W_-$. Aseguramos que existe $n_2 \in W_-$ tal que $a_{n_1} < a_{n_2} < a$. Basta tener en cuenta que $W_- \setminus \{n \in \mathbb{N} : n < n_1\}$ es un conjunto infinito, que $a_{n_1} < a$ y que $\lim_{n\to\infty} a_n = a$. Repitiendo este proceso se consigue una sucesión $(a_{n_k})_{k=1}^{\infty}$ de manera que $n_k < n_{k+1}$, $a_{n_k} < a_{n_{k+1}} < a$ y verificando $|a_{n_k} - a| < \frac{1}{n_k}$ y $|f(a_{n_k}) - f(a)| \geq \varepsilon$, $k \in \mathbb{N}$.

De esta forma, obtenemos una sucesión $(a_{n_k})_{k=1}^{\infty}$ estrictamente creciente tal que $a_{n_k} \to a$, cuando $k \to \infty$ y sin embargo, $(f(a_{n_k}))_{k=1}^{\infty}$ no converge a $f(a)$.

En el caso de que W_+ sea un conjunto infinito basta proceder de la misma manera para obtener una sucesión $(a_{n_k})_{k=1}^{\infty}$ estrictamente decreciente que converge a a y tal que $(f(a_{n_k}))_{k=1}^{\infty}$ no converge a $f(a)$.

$\square$

Nos preguntamos ahora qué se puede decir de las funciones monótonas que no son continuas. Un resultado en este sentido es que una función monótona tiene, a lo sumo, un conjunto numerable de puntos de discontinuidad.

Proposición 2.6. *Sea $f : \Omega \subset \mathbb{R} \to \mathbb{R}$ una función monótona. El conjunto de los puntos de discontinuidad de f es vacío, finito o infinito numerable.*

Demostración. Supongamos que f es monótona creciente. Si $a \in \operatorname{Ais}(\Omega)$, entonces f es continua en a. Luego, como estamos interesados en los puntos donde f es discontinua examinamos los puntos de acumulación, esto es, los elementos de $\operatorname{Der}(\Omega)$. Recordamos que

$$\operatorname{Der}(\Omega) = \{a \in \mathbb{R} : \ (a - r, a + r) \cap \Omega \neq \emptyset, \text{ para todo } r > 0\}.$$

Consideramos los conjuntos $\operatorname{Der}_+(\Omega)$ y $\operatorname{Der}_-(\Omega)$ de los puntos de acumulación por la derecha y por la izquierda, respectivamente, que se definen como

$$\operatorname{Der}_+(\Omega) = \{a \in \mathbb{R} : \ (a, a + r) \cap \Omega \neq \emptyset, \text{ para todo } r > 0\},$$

y

$$\operatorname{Der}_-(\Omega) = \{a \in \mathbb{R} : \ (a - r, a) \cap \Omega \neq \emptyset, \text{ para todo } r > 0\}.$$

Nótese que un punto es de acumulación en Ω si, y solo si, lo es por la derecha o por la izquierda, es decir, $\operatorname{Der}(\Omega) = \operatorname{Der}_+(\Omega) \cup \operatorname{Der}_-(\Omega)$. Esto implica que un punto de acumulación puede ser de tres clases: un punto de $\operatorname{Der}_+(\Omega) \setminus \operatorname{Der}_-(\Omega)$, de

$\mathrm{Der}_-(\Omega) \setminus \mathrm{Der}_+(\Omega)$ o bien de $\mathrm{Der}_+(\Omega) \cap \mathrm{Der}_-(\Omega)$. Observamos además, que por ser f monótona creciente, si $a \in \mathrm{Der}_+(\Omega)$ y $f_+(a) = \lim_{x\to a^+} f(x)$, entonces

$$f_+(a) = \inf_{x\in\Omega\cap(a,+\infty)} f(x), \tag{2.6}$$

y en el caso de que $a \in \mathrm{Der}_-(\Omega)$ y $f_-(a) = \lim_{x\to a^-} f(x)$, se tiene que

$$f_-(a) = \sup_{x\in\Omega\cap(-\infty,a)} f(x). \tag{2.7}$$

Denotamos por $\mathrm{Dis}(f)$ el conjunto de puntos de discontinuidad de f. Podemos escribir

$$\begin{aligned}
&\mathrm{Dis}(f) = \Omega \cap \mathrm{Der}(\Omega) \cap \mathrm{Dis}(f)\\
&= \Omega \cap \Big[\big(\mathrm{Der}_+(\Omega) \setminus \mathrm{Der}_-(\Omega)\big) \cup \big(\mathrm{Der}_-(\Omega) \setminus \mathrm{Der}_+(\Omega)\big) \cup \big(\mathrm{Der}_+(\Omega) \cap \mathrm{Der}_-(\Omega)\big)\Big] \cap \mathrm{Dis}(f)\\
&\subseteq \big[\Omega \setminus \mathrm{Der}_-(\Omega)\big] \cup \big[\Omega \setminus \mathrm{Der}_+(\Omega)\big] \cup \big[\Omega \cap \mathrm{Der}_+(\Omega) \cap \mathrm{Der}_-(\Omega) \cap \mathrm{Dis}(f)\big]\\
&=: A_1 \cup A_2 \cup A_3.
\end{aligned}$$

De esta forma, para mostrar que $\mathrm{Dis}(f)$ es a lo sumo numerable basta probar que A_1, A_2 y A_3 lo son. La idea es la siguiente: si queremos establecer que un conjunto A es a lo sumo numerable, basta encontrar para cada $i \in A$ un intervalo I_i de manera que $I_k \cap I_\ell = \emptyset$, $k \neq \ell$. Vamos a ver esto. Sea $\{(a_i, b_i)\}_{i\in A}$ una colección de intervalos disjuntos dos a dos. Por ser $\mathbb{Q}$ denso en $\mathbb{R}$, se tiene que para cada $i \in A$ existe $q_i \in \mathbb{Q}$ en (a_i, b_i). Además, como los intervalos son disjuntos dos a dos, podemos asegurar que $q_i \neq q_j$, $i, j \in A$, $i \neq j$. Ya que $\mathbb{Q}$ es numerable podemos concluir entonces que la colección es a lo sumo numerable.

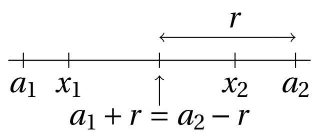
Analizamos entonces la propiedad para cada A_k, $k = 1, 2, 3$.

(a) $A_1 = \Omega \setminus \mathrm{Der}_-(\Omega)$. Sea $a \in A_1$. Como $a \notin \mathrm{Der}_-(\Omega)$, existe $r_a > 0$ de manera que $(a - r_a, a) \cap \Omega = \emptyset$. Además, si $b \in A_1$ y $a \neq b$ se tiene que $(a - r_a, a) \cap (b - r_b, b) = \emptyset$. En efecto, asumamos que, por ejemplo, $b < a$. Si $(a - r_a, a) \cap (b - r_b, b) \neq \emptyset$, entonces, siendo $b < a$, se tendría que $b \in (a - r_a, a)$ y de aquí, puesto que $b \in \Omega$, se llegaría a que $(a - r_a, a) \cap \Omega \neq \emptyset$, lo que contradice la hipótesis. Luego, se concluye que $\{(a - r_a, a)\}_{a\in A_1}$ es una colección de intervalos disjuntos dos a dos.

(b) $A_2 = \Omega \setminus \mathrm{Der}_+(\Omega)$. Se procede de manera análoga al caso anterior. Sea $a \in A_2$. Ya que $a \notin \mathrm{Der}_+(\Omega)$, existe $r_a > 0$ de manera que $(a, a + r_a) \cap \Omega = \emptyset$. Además, si $b \in A_2$ y $a \neq b$ se tiene, razonando como antes, que $(a, a + r_a) \cap (b, b + r_b) = \emptyset$. Se obtiene así que $\{(a, a + r_a)\}_{a\in A_2}$ es una familia de intervalos disjuntos dos a dos.

(c) $A_3 = \Omega \cap \mathrm{Der}_+(\Omega) \cap \mathrm{Der}_-(\Omega) \cap \mathrm{Dis}(f)$.
Sea $a \in A_3$. Ya que $a \in \Omega \cap \mathrm{Der}_+(\Omega) \cap \mathrm{Der}_-(\Omega)$ se tiene, por (2.6) y (2.7), que $f_-(a) \leq f(a) \leq f_+(a)$. Observamos que $f_-(a) \neq f_+(a)$, pues en caso contrario, $f_-(a) = f(a) = f_+(a)$ y entonces f sería continua en a, y por hipótesis, $a \in \mathrm{Dis}(f)$. Se tiene así que la discontinuidad en a es de salto finito y $f_-(a) < f_+(a)$.

Por otra parte, si $a_1, a_2 \in A_3$ y $a_1 < a_2$ se cumple que $f_+(a_1) \leq f_-(a_2)$. En efecto, sean $a_1, a_2 \in \mathrm{Der}_+(\Omega) \cap \mathrm{Der}_-(\Omega)$ de manera que $a_1 < a_2$. Tomamos $r = \frac{a_2 - a_1}{2}$. Puesto que $a_1 \in \mathrm{Der}_+(\Omega)$, podemos elegir $x_1 \in \Omega$ tal que $x_1 \in (a_1, a_1 + r)$. De la misma forma, teniendo en cuenta que $a_2 \in \mathrm{Der}_-(\Omega)$, tomamos $x_2 \in \Omega$ tal que $x_2 \in (a_2 - r, a_2)$. Nótese que $x_1 < x_2$.



Entonces, usando (2.6) y (2.7) y teniendo en cuenta que f es creciente, se obtiene que $f_+(a_1) \leq f(x_1) \leq f(x_2) \leq f_-(a_2)$. Luego, para todo $a_1, a_2 \in A_3$ con $a_1 < a_2$, se verifica que

$$f_-(a_1) < f_+(a_1) \leq f_-(a_2) < f_+(a_2).$$

Por tanto, $(f_-(a_1), f_+(a_1)) \cap (f_-(a_2), f_+(a_2)) = \emptyset$. Esto es, $\{(f_-(a), f_+(a))\}_{a \in A_3}$ es una colección de intervalos disjuntos dos a dos.

Concluimos que A_1, A_2 y A_3 son conjuntos a lo sumo numerables, y por tanto también el conjunto de puntos de discontinuidad de f, $\mathrm{Dis}(f)$, es a lo sumo numerable. $\square$

Dedicamos la última parte de esta sección a establecer algunas propiedades para las funciones inversas, donde incluimos el Teorema de la función inversa para funciones reales de variable real.

Recordamos que si $f : A \to B$ es una función biyectiva, siendo $A, B \subset \mathbb{R}$, la *función inversa de f*, que se denota por f^{-1}, es la aplicación que a cada $y \in B$ le asigna el valor $f^{-1}(y) = x$, siendo x el único elemento de A tal que $f(x) = y$. De esta forma, $(f^{-1} \circ f)(x) = x$, $x \in A$, y $(f \circ f^{-1})(y) = y$, $y \in B$.

Veamos primero que la monotonía se conserva por la operación inversa.

Proposición 2.7. *Sea $f : A \to B$ una función biyectiva, donde $A, B \subseteq \mathbb{R}$. Si f es monótona creciente (o decreciente) entonces f^{-1} es también monótona creciente (o decreciente).*

(Nótese que como f es biyectiva, si f es monótona entonces f es estrictamente monótona.)

Demostración. Supongamos que f es estrictamente creciente. Fijamos $y_1, y_2 \in B$ tales que $y_1 < y_2$. Sean $x_1, x_2 \in A$ tales que $f(x_1) = y_1$ y $f(x_2) = y_2$. Queremos ver que $x_1 < x_2$.

Supongamos que $x_2 \leq x_1$. En el caso de que $x_1 = x_2$ se tiene que $f(x_1) = f(x_2)$, esto es, $y_1 = y_2$, lo que no es cierto. Por otro lado, si $x_2 < x_1$, como f es estrictamente creciente, se cumple que $y_2 = f(x_2) < f(x_1) = y_1$, lo que también contradice la hipótesis.

Análogamente puede procederse cuando f es estrictamente decreciente. $\square$

Recogemos ahora un resultado sobre la continuidad de la función inversa.

Proposición 2.8. *Sea $f : I \longrightarrow f(I)$ una función biyectiva, donde I es un intervalo en $\mathbb{R}$. Si f es monótona entonces $f^{-1} : f(I) \longrightarrow I$ es una función continua.*

Demostración. Supongamos que f es estrictamente creciente. En virtud de la Proposición 2.7, $f^{-1} : f(I) \longrightarrow I$ es estrictamente creciente. Ya que $f^{-1}(f(I)) = I$ es un intervalo, de acuerdo a la Proposición 2.4 deducimos que f^{-1} es continua en I. □

Asimismo, se pueden dar condiciones que aseguran la derivabilidad de f^{-1}. Este resultado resulta además fundamental para establecer el Teorema de la función inversa que abordaremos después.

Teorema 2.9. *Sea $f : I \to f(I)$ una función biyectiva y monótona, siendo I un intervalo en $\mathbb{R}$. Supongamos que f es derivable en $a \in I$ y $f'(a) \neq 0$. Entonces, f^{-1} es derivable en $f(a)$ y $(f^{-1})'(f(a)) = 1/f'(a)$.*

Demostración. Tenemos que ver que

$$\lim_{h\to 0} \frac{f^{-1}(f(a)+h) - f^{-1}(f(a))}{h}$$

existe y vale $(f'(a))^{-1}$.
Tomamos un sucesión $(b_n)_{n\in\mathbb{N}} \subseteq f(I)\setminus\{f(a)\}$ tal que $b_n \longrightarrow f(a)$, cuando $n \to \infty$.
Para cada $n \in \mathbb{N}$ existe un único $a_n \in I \setminus \{a\}$ tal que $f(a_n) = b_n$.
De la Proposición 2.8 sabemos que f^{-1} es continua en $f(a)$ y, ya que $b_n \longrightarrow f(a)$, se sigue que $a_n = f^{-1}(b_n) \longrightarrow f^{-1}(f(a)) = a$.

Como f es derivable en a se tiene que

$$\lim_{n\to\infty} \frac{f(a_n) - f(a)}{a_n - a} = \lim_{x\to a} \frac{f(x) - f(a)}{x - a} = f'(a).$$

Entonces,

$$\lim_{n\to\infty} \frac{f^{-1}(b_n) - f^{-1}(f(a))}{b_n - f(a)} = \lim_{n\to\infty} \frac{a_n - a}{f(a_n) - f(a)} = \lim_{n\to\infty} \frac{1}{\frac{f(a_n)-f(a)}{a_n - a}} = \frac{1}{f'(a)}.$$

Recordamos que $f'(a) \neq 0$. Se concluye así que $(f^{-1})'(f(a)) = \frac{1}{f'(a)}$. □

A continuación enunciamos y probamos el Teorema de la función inversa para funciones reales de variable real.

Como es usual, dado un abierto $\Omega \subset \mathbb{R}$, diremos que una función $f : \Omega \longrightarrow \mathbb{R}$ es de clase C^k en Ω, con $k \in \mathbb{N}$, y escribiremos $f \in C^k(\Omega)$, si existen todas sus derivadas hasta orden k y son continuas en Ω.

Teorema 2.10. *Sean $\Omega \subset \mathbb{R}$ un abierto y $f : \Omega \longrightarrow \mathbb{R}$ una función en $C^1(\Omega)$. Supongamos que $a \in \Omega$ y $f'(a) \neq 0$. Entonces, existe $r > 0$ tal que $[a-r, a+r] \subset \Omega$, $f((a-r, a+r))$ es un intervalo abierto, la función*

$$f_{|(a-r,a+r)} : (a-r, a+r) \longrightarrow f((a-r, a+r))$$

es biyectiva y $(f_{|(a-r,a+r)})^{-1}$ es de clase C^1 en $f((a-r, a+r))$. Además,

$$(f^{-1})'(y) = \frac{1}{f'(f^{-1}(y))}, \quad y \in f(a-r, a+r).$$

Demostración. Vamos a asumir que $f'(a) > 0$. Si $f'(a) < 0$ podemos proceder de manera similar.

Ya que f' es continua en Ω, la conservación de signo de las funciones continuas nos permite encontrar $r > 0$ tal que $[a-r, a+r] \subset \Omega$ y $f'(x) > 0$, $x \in [a-r, a+r]$.

Vamos a ver que f es creciente en el intervalo $[a-r, a+r]$. Para ello consideramos $a - r \leq x_1 < x_2 \leq a + r$. Tenemos que f es continua en $[x_1, x_2]$ y derivable en (x_1, x_2). Del teorema del valor medio[2] se sigue que existe $c \in (x_1, x_2)$ tal que $f(x_2) - f(x_1) = f'(c)(x_2 - x_1)$. Luego, $f(x_2) - f(x_1) > 0$ y se obtiene que f es estrictamente creciente en $(a-r, a+r)$.

Por tanto, $f((a-r, a+r)) = (f(a-r), f(a+r))$. Tenemos entonces que

$$f_{|(a-r,a+r)} : (a-r, a+r) \longrightarrow (f(a-r), f(a+r)),$$

es una función biyectiva. Además es derivable en $(a-r, a+r)$ y se cumple que $f'(x) > 0$, $x \in (a-r, a+r)$. En virtud del Teorema 2.9 $(f_{|(a-r,a+r)})^{-1}$ es derivable en $(f(a-r), f(a+r))$ y

$$((f^{-1})'(y) = \frac{1}{f'(x)}, \text{ siendo } f(x) = y, \quad y \in (f(a-r), f(a+r)).$$

Probamos para terminar que $(f^{-1})'$ es continua en $(f(a-r), f(a+r))$.

Sean $y_0 \in (f(a-r), f(a+r)) = f((a-r, a+r))$ y $x_0 \in (a-r, a+r)$ tal que $y_0 = f(x_0)$ (nótese que x_0 es único).

Consideramos una sucesión $(y_n)_{n=1}^{\infty} \subseteq (f(a-r), f(a+r))$ tal que $y_n \longrightarrow y_0$, cuando $n \to \infty$. Para cada $n \in \mathbb{N}$ consideramos el único $x_n \in (a-r, a+r)$ tal que $f(x_n) = y_n$.

Ya que $(f_{|(a-r,a+r)})^{-1}$ es continua en $(f(a-r), f(a+r))$, se verifica entonces que $x_n \longrightarrow x_0$, cuando $n \to \infty$. Además, $(f^{-1})'(y_n) = (f'(x_n))^{-1}$, $n \in \mathbb{N} \cup \{0\}$. Ya que f' es continua en x_0 se sigue que

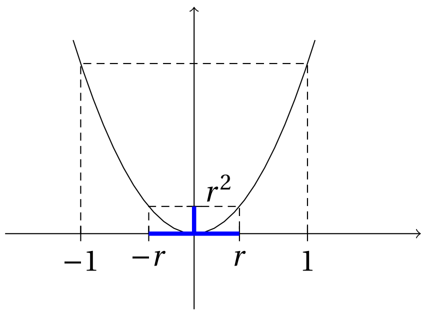
$$\lim_{n\to\infty} (f^{-1})'(y_n) = \lim_{n\to\infty} \frac{1}{f'(x_n)} = \frac{1}{f'(x_0)} = (f^{-1})'(y_0).$$

Concluimos así que $(f^{-1})'$ es continua en y_0. □

[2] **Teorema del valor medio**. Si $f : [a,b] \longrightarrow \mathbb{R}$, con $-\infty < a < b < +\infty$ es continua en $[a,b]$ y derivable en (a,b), entonces existe $c \in (a,b)$ tal que $f(b) - f(a) = f'(c)(b-a)$.

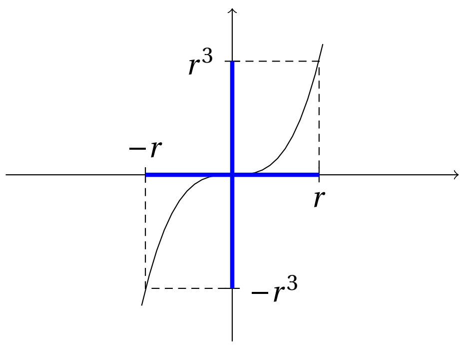
Terminamos mostrando con dos ejemplos dos situaciones diferentes en relación a la existencia de inversa local cuando $f'(a) = 0$.

Ejemplo 2.11. Sea $f(x) = x^2$, $x \in [-1,1]$.



Esta función no tiene inversa local en el punto $a = 0$. Basta observar que si $r > 0$ entonces $f((-r,r)) = [0,r^2)$ que no es un intervalo abierto y $f : (-r,r) \longrightarrow [0,r^2)$ no es biyectiva, ya que $f(\frac{r}{2}) = f(-\frac{r}{2}) = \frac{r^2}{4}$. Podemos deducir del Teorema de la función inversa, ya que se cumple que $f \in C^1(-1,1)$ (de hecho es de clase C^∞), que no se satisface la condición $f'(0) \neq 0$, circunstancia que también podemos comprobar directamente pues $f'(x) = 2x$, $x \in (-1,1)$, por lo que, en particular, $f'(0) = 0$.

Ejemplo 2.12. Consideramos la función $f(x) = x^3$, $x \in \mathbb{R}$.



Tenemos que f es derivable en $\mathbb{R}$ siendo $f'(x) = 3x^2$, $x \in \mathbb{R}$. De nuevo, en particular, $f'(0) = 0$ y no se cumple la hipótesis del teorema anterior para $a = 0$.

En este caso, sí podemos hablar de función inversa en un entorno del origen. Sea $r > 0$. Se verifica que $f((-r,r)) = (-r^3,r^3)$ y la función f es biyectiva, siendo $f^{-1} : (-r^3,r^3) \longrightarrow (-r,r)$, con $f^{-1}(y) = \sqrt[3]{y}$, $y \in (-r^3,r^3)$.

Observamos que, sin embargo, esta inversa local no es derivable en el origen. De hecho, f^{-1} es derivable en $\mathbb{R} \setminus \{0\}$ y $(f^{-1})'(y) = (3\sqrt[3]{y^2})^{-1}$, $y \neq 0$.

Observación 2.13. Si $f : (a,b) \longrightarrow (c,d)$ es biyectiva, derivable y además la función inversa $f^{-1} : (c,d) \longrightarrow (a,b)$ también es derivable, entonces se verifica que $(f^{-1})'(y) = (f'(x))^{-1}$, siendo $f(x) = y$.

En efecto, basta tener en cuenta que $(f^{-1} \circ f)(x) = x$, $x \in (a,b)$. Luego,

$$(f^{-1} \circ f)'(x) = (f^{-1})'(f(x))f'(x) = 1, \quad x \in (a,b).$$

De aquí se deduce que $f'(x) \neq 0$, $x \in (a,b)$, y $(f^{-1})'(f(x)) = (f'(x))^{-1}$, $x \in (a,b)$. □

2.2. Teorema de la función inversa en el contexto vector-valuado

Recordamos, en primer lugar, los conceptos básicos que manejaremos. Consideramos $f : \Omega \subset \mathbb{R}^n \longrightarrow \mathbb{R}$, donde Ω es un conjunto abierto en $\mathbb{R}^n$. Se dice que f es *diferenciable en* $a \in \Omega$ cuando existe una aplicacion lineal $L : \mathbb{R}^n \longrightarrow \mathbb{R}$ tal que

$$\lim_{h \to 0} \frac{f(a+h) - f(a) - L(h)}{\|h\|} = 0. \tag{2.8}$$

Aquí $\|h\| = \|h\|_2 = \sqrt{h_1^2 + \cdots + h_n^2}$, siendo $h = (h_1, \ldots, h_n) \in \mathbb{R}^n$.

Un hecho importante es que la aplicación lineal L en (2.8) es única, esto es, si $L_j : \mathbb{R}^n \longrightarrow \mathbb{R}$ es lineal, $j = 1,2$, y satisface (2.8), entonces, $L_1 = L_2$ ([6, Theorem 9.12]).

Si f es diferenciable en $a \in \Omega$ se llama *diferencial de f en a*, y la denotamos por $df(a)$, a la aplicación $L : \mathbb{R}^n \longrightarrow \mathbb{R}$ que verifica (2.8).

Por otro lado, dado $u \in \mathbb{R}^n$ tal que $\|u\| = 1$, se conoce como *derivada de f en a en la dirección de u* (denotada por $d_u f(a)$) al límite, si existe,

$$\lim_{t \longrightarrow 0} \frac{f(a+tu) - f(a)}{t}.$$

Se puede ver que si f es diferenciable en $a \in \Omega$ entonces existe la derivada de f en a en la dirección de u para cada $u \in \mathbb{R}^n$, $\|u\| = 1$, y se verifica que $d_u f(a) = (df(a))(u)$ ([2, p. 83]). En particular, cuando tomamos $u = e_k$, $k = 1, ..., n$, donde $e_k = (e_i^k)_{i=1}^n$, con $e_i^k = 0$, $i = 1, ..., n$, $i \neq k$, y $e_k^k = 1$, la derivada de f en a en la dirección de e_k se denota por $\frac{\partial f}{\partial x_k}(a)$ y se le llama *derivada parcial de f en a respecto de x_k*. Al vector

$$\nabla f(a) = \Big(\frac{\partial f}{\partial x_1}(a), \ldots, \frac{\partial f}{\partial x_n}(a)\Big)$$

se le conoce como *gradiente de f en a* .

Si f es diferenciable en $a \in \Omega$ podemos expresar la diferencial de f en a en términos del vector gradiente, si más que tener en cuenta que $df(a)$ es una aplicación lineal y, por tanto, si $h = (h_1, ..., h_n) \in \mathbb{R}^n$ se tiene que

$$df(a)(h) = df(a)\Big(\sum_{k=1}^n h_k e_k\Big) = \sum_{k=1}^n h_k df(a)(e_k) = \sum_{k=1}^n h_k \frac{\partial f}{\partial x_k}(a) = \nabla f(a) \cdot h.$$

Además, para cada $u \in \mathbb{R}^n$, $\|u\| = 1$, se tiene que

$$d_u f(a) = \lim_{t\to 0} \frac{f(a+tu)-f(a)}{t} = \lim_{t\to 0}\Big(\frac{f(a+tu)-f(a)-t\nabla f(a)\cdot u}{t\|u\|} + \nabla f(a)\cdot u\Big),$$

y, si f es diferenciable en a,

$$\lim_{t\to 0} \frac{f(a+tu)-f(a)-\nabla f(a)\cdot(tu)}{t\|u\|} = \lim_{h\to 0}\frac{f(a+h)-f(a)-df(a)(h)}{\|h\|} = 0,$$

por lo que $d_u f(a) = \nabla f(a)\cdot u = df(a)(u)$.

Observación 2.14. Sean $f : \Omega \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ una función definida en el abierto Ω y $a \in \Omega$. La existencia de las derivadas parciales de f en a no garantiza que la función f sea diferenciable en a. No obstante, si las derivadas parciales de f son continuas en a, entonces sí podemos asegurar la diferenciabilidad de f en a y que $df(a)(h) = \nabla f(a)\cdot h$, $h \in \mathbb{R}^n$ ([2, Theorem 3.2]).

Generalizamos ahora estos conceptos al contexto de las funciones definidas en un abierto $\Omega \subset \mathbb{R}^n$ y con valores en $\mathbb{R}^m$, para $m \in \mathbb{N}$, $m \geq 1$, esto es,

$$\begin{aligned} f : \Omega \subset \mathbb{R}^n &\longrightarrow \mathbb{R}^m \\ x &\longmapsto f(x) = (f_1(x),\dots,f_m(x)), \end{aligned}$$

donde, para cada $j = 1,\dots,m$, f_j es una función definida en Ω que toma valores reales.

Sean $f : \Omega \subset \mathbb{R}^n \longrightarrow \mathbb{R}^m$ y $a \in \Omega$, siendo Ω un abierto. Decimos que *f es diferenciable en a* cuando f_j es diferenciable en a, para cada $j = 1,\dots,m$. En este caso, *la diferencial de f en a*, $d(f)(a)$ es la aplicación lineal determinada por la matriz de orden $m\times n$ dada por

$$Df(a) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \frac{\partial f_1}{\partial x_2}(a) & \cdots & \frac{\partial f_1}{\partial x_n}(a) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(a) & \frac{\partial f_m}{\partial x_2}(a) & \cdots & \frac{\partial f_m}{\partial x_n}(a) \end{pmatrix}.$$

Observamos que si f es diferenciable en a entonces,

$$\lim_{h\to 0}\frac{f(a+h)-f(a)-df(a)(h)}{\|h\|} = 0.$$

Nótese además que

$$\begin{aligned} df(a)(h) &= \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \frac{\partial f_1}{\partial x_2}(a) & \cdots & \frac{\partial f_1}{\partial x_n}(a) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(a) & \frac{\partial f_m}{\partial x_2}(a) & \cdots & \frac{\partial f_m}{\partial x_n}(a) \end{pmatrix}\begin{pmatrix} h_1 \\ \vdots \\ h_n\end{pmatrix} = \Big(\sum_{k=1}^n \frac{\partial f_1}{\partial x_k}(a)h_k,\dots,\sum_{k=1}^n \frac{\partial f_n}{\partial x_k}(a)h_k\Big) \\ &= (\nabla f_1(a)\cdot h,\dots,\nabla f_n(a)\cdot h), \quad h = (h_1,\dots,h_n) \in \mathbb{R}^n. \end{aligned}$$

Cuando $n = m$ se tiene que $Df(a)$ es una matriz cuadrada, y en este caso, llamamos *Jacobiano de f en a* al determinante $Jf(a) = \det(Df(a))$. Por tanto, cuando $n = m$, la matriz $Df(a)$ es invertible si y solo si $Jf(a) \neq 0$.

2.2.1. Teorema del valor medio para funciones de varias variables

En la demostración del Teorema de la función inversa para funciones reales de una variable usamos el Teorema del valor medio (ver Observación 2.1). Una pregunta natural que podemos plantearnos es si existe un teorema análogo para funciones de varias variables.

Presentamos en primer lugar un resultado para funciones reales de varias variables.

Teorema 2.15 (Teorema del valor medio). *Sea Ω un abierto y convexo de $\mathbb{R}^n$. Consideramos $f : \Omega \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ una función diferenciable en $C^1(\Omega)$, esto es, $\frac{\partial f}{\partial x_k}$ es continua en Ω para cada $k = 1,\ldots,n$. Se tiene que, dados $a, b \in \Omega$ existe $c \in [a,b]$ (segmento que une a y b) de manera que*

$$f(b) - f(a) = \nabla f(c) \cdot (b-a) = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(c)(b_k - a_k).$$

Aquí $a = (a_1,\ldots,a_n)$ y $b = (b_1,\ldots,b_n)$.

Demostración. Sean $a = (a_1,\ldots,a_n)$ y $b = (b_1,\ldots,b_n) \in \Omega$. Ya que Ω es convexo, el segmento $[a,b]$ está contenido en Ω. Consideramos la parametrización del segmento $[a,b]$ dada por $g(t) = tb + (1-t)a$, $t \in [0,1]$, y escribimos $g = (g_1,..., g_n)$, siendo $g_i(t) = tb_i + (1-t)a_i$, $t \in [0,1]$, para $i = 1,..., n$. Nótese que $g(0) = a$, $g(1) = b$ y $g'(t) = b - a$.

Definimos entonces la función $h : [0,1] \longrightarrow \mathbb{R}$ como $h(t) = f(g(t))$, $t \in [0,1]$. Ya que f es continua en Ω y g es continua en $[0,1]$, se sigue que h es continua en $[0,1]$. Además, puesto que f es diferenciable en Ω y g es derivable en $(0,1)$, también podemos asegurar que h es derivable en $(0,1)$ y, aplicando la regla de la cadena obtenemos

$$h'(t) = df(g(t))(g'(t)) = (\nabla f)(g(t)) \cdot g'(t), \quad t \in (0,1).$$

El Teorema de valor medio para funciones reales de una variable (Observación 2.1) garantiza la existencia de $t_0 \in (0,1)$ tal que $h(1) - h(0) = h'(t_0)$, esto es,

$$f(g(1)) - f(g(0)) = (\nabla f)(g(t_0)) \cdot g'(t_0),$$

y, por tanto, $f(b) - f(a) = \nabla f(c) \cdot (b-a)$, siendo $c = g(t_0) \in [a,b] \setminus \{a,b\}$. □

Nos planteamos ahora si existe una versión del Teorema del valor medio para funciones vector-valuadas, es decir, funciones de la forma $f : \Omega \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^m$, con $m > 1$. La respuesta es que no, como mostramos con el siguiente ejemplo.

Sea $f : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ la función definida por $f(x,y) = (e^x \cos y, e^x \operatorname{sen} y)$ y escribimos $f_1(x,y) = e^x \cos y$, $f_2(x,y) = e^x \operatorname{sen} y$, $(x,y) \in \mathbb{R}^2$.

Tomamos $a = (0,0)$ y $b = (0,2\pi)$. Es claro que $f(a) = f(b) = (1,0)$. Por otra parte, la función f es diferenciable en $\mathbb{R}^2$ (más aún, $f \in C^\infty(\mathbb{R}^2)$) y se tiene

$$Df(x,y) = \begin{pmatrix} \dfrac{\partial f_1}{\partial x}(x,y) & \dfrac{\partial f_1}{\partial y}(x,y) \\ \dfrac{\partial f_2}{\partial x}(x,y) & \dfrac{\partial f_2}{\partial y}(x,y) \end{pmatrix} = \begin{pmatrix} e^x \cos y & -e^x \operatorname{sen} y \\ e^x \operatorname{sen} y & e^x \cos y \end{pmatrix}, \quad (x,y) \in \mathbb{R}^2.$$

Vamos a suponer que existe $c = (c_1, c_2) \in [a,b]$ de manera que $f(b) - f(a)) = df(c)(b-a)$, esto es, se verifica

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} e^{c_1} \cos c_2 & -e^{c_1} \operatorname{sen} c_2 \\ e^{c_1} \operatorname{sen} c_2 & e^{c_1} \cos c_2 \end{pmatrix} \begin{pmatrix} 0 \\ 2\pi \end{pmatrix} = \begin{pmatrix} -2\pi e^{c_1} \operatorname{sen} c_2 \\ 2\pi e^{c_1} \cos c_2 \end{pmatrix}.$$

Pero esto no es cierto puesto que $2\pi e^{c_1} \operatorname{sen} c_2 = 0$ y $2\pi e^{c_1} \cos c_2 = 0$ implica que $\operatorname{sen} c_2 = \cos c_2 = 0$, y sabemos que no existe $c_2 \in \mathbb{R}$ que cumpla tales condiciones.

Este ejemplo nos muestra que la propiedad recogida en el Teorema del valor medio no es una propiedad que podamos pasar de dimensión uno a dimensión mayor que uno de una manera simple. Además nos lleva a pensar que para establecer el teorema de la función inversa en el caso de funciones $f : \Omega \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^n$, con $n > 1$, vamos a necesitar de estrategias diferentes a la usadas cuando $n = 1$.

Precisamente, para la demostración del Teorema de la función inversa (Teorema 2.18) haremos uso del siguiente resultado.

Proposición 2.16. *Sean Ω un abierto convexo de $\mathbb{R}^n$ y $f : \Omega \subset \mathbb{R}^n \longrightarrow \mathbb{R}^n$ una función diferenciable en Ω. Supongamos que para cierta C_0 se cumple que $\|Df(x)\| \leq C_0$, $x \in \Omega$, siendo $\|\cdot\|$ una norma en $M_n(\mathbb{R})$. Entonces existe $c > 0$ tal que, para cada $a, b \in \Omega$, $\|f(b) - f(a)\|_2 \leq c\|b - a\|_2$.*

Demostración. Escribimos $f = (f_1, \ldots, f_n)$. Sabemos que todas las normas en $M_n(\mathbb{R})$ son equivalentes, por lo que podemos encontrar $C > 0$ de modo que

$$\max_{j,k=1,\ldots,n} \left| \frac{\partial f_j}{\partial x_k}(x) \right| \leq C\|Df(x)\|, \quad x \in \Omega. \tag{2.9}$$

Sean $a, b \in \Omega$. Como Ω es convexo, el segmento $[a,b] \subset \Omega$. Consideramos, como antes, la parametrización del segmento $[a,b]$ dada por $g(t) = tb + (1-t)a$, $t \in [0,1]$, y definimos la función $h : [0,1] \longrightarrow \mathbb{R}$ mediante

$$h(t) = f(g(t)) \cdot (f(b) - f(a)) = \sum_{j=1}^{n} f_j(g(t))(f_j(b) - f_j(a)), \quad t \in [0,1].$$

Es claro que h es continua en $[0,1]$ y derivable en $(0,1)$, siendo

$$\begin{aligned}
h'(t) &= \sum_{j=1}^{n} (\nabla f_j)(g(t)) \cdot g'(t)(f_j(b) - f_j(a)) \\
&= \sum_{j=1}^{n} \sum_{k=1}^{n} \frac{\partial f_j}{\partial x_k}(g(t))(b_k - a_k)(f_j(b) - f_j(a)), \quad t \in (0,1).
\end{aligned}$$

De acuerdo con el Teorema de valor medio (Observación 2.1), existe $t_0 \in (0,1)$ tal que $h(1) - h(0) = h'(t_0)$. Ya que $h(1) = f(b) \cdot (f(b) - f(a))$ y $h(0) = f(a) \cdot (f(b) - f(a))$, se sigue entonces que

$$(f(b) - f(a)) \cdot (f(b) - f(a)) = \sum_{j=1}^{n} \sum_{k=1}^{n} \frac{\partial f_j}{\partial x_k}(g(t_0))(b_k - a_k)(f_j(b) - f_j(a)).$$

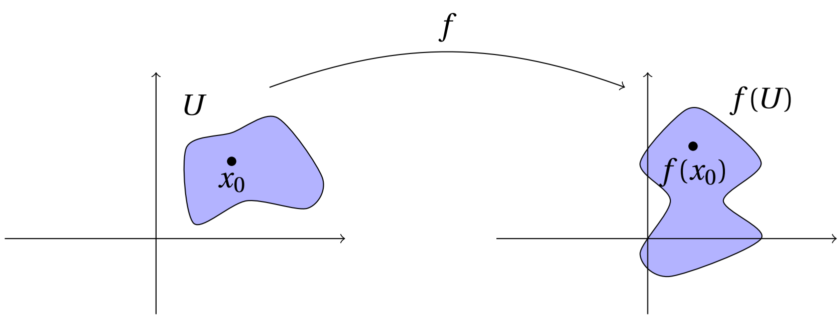
Luego, teniendo en cuenta (2.9)

$$\begin{aligned}
\|f(b) - f(a)\|^2 &\le \sum_{j=1}^{n} \sum_{k=1}^{n} \left| \frac{\partial f_j}{\partial x_k}(g(t_0)) \right| |b_k - a_k| |f_j(b) - f_j(a)| \\
&\le C \|Df(g(t_0))\| \sum_{k=1}^{m} |b_k - a_k| \sum_{j=1}^{n} |f_j(b) - f_j(a)| \\
&= CC_0 \|b - a\|_1 \|f(b) - f(a)\|_1 \le CC_0 n^2 \|b - a\|_2 \|f(b) - f(a)\|_2,
\end{aligned}$$

donde hemos usado que $\|x\|_1 \le n\|x\|_2$, $x \in \mathbb{R}^n$. Concluimos que si $f(b) \neq f(a)$ entonces $\|f(b) - f(a)\|_2 \le c\|b - a\|_2$, con $c = CC_0 n^2$. Cuando $f(b) = f(a)$ la desigualdad es clara. □

2.2.2. Funciones global y localmente invertibles. Ejemplos

En esta sección vamos a analizar la existencia de la función inversa global o local para algunas funciones particulares $f : \Omega \subset \mathbb{R}^n \longrightarrow \mathbb{R}^n$, como paso previo al Teorema de la función inversa, que proporciona condiciones bajo las cuales se asegura la existencia de funciones localmente invertibles, esto es, funciones F para las cuales, si $x_0 \in \mathbb{R}^n$, es posible encontrar un entorno $U \subset \Omega$ de x_0 de manera que la aplicación $f_{|U} : U \longrightarrow f(U)$ es biyectiva.



1. Vamos a comenzar estudiando el comportamiento de las aplicaciones lineales. Sea $f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ una aplicación lineal. Denotamos por A la matriz en $M_n(\mathbb{R})$ asociada a f respecto de la base canónica en $\mathbb{R}^n$. Entonces, $f(x) = Ax$, $x \in \mathbb{R}^n$. Si $A = (a_{ij})_{i,j=1,...,n}$, podemos escribir, para cada $x = (x_1,...,x_n) \in \mathbb{R}^n$,

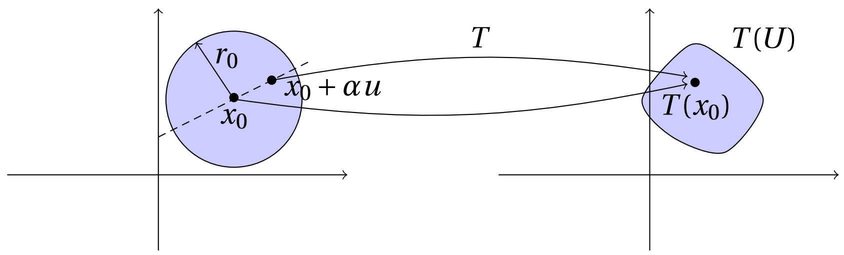
$$f(x) = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n \end{pmatrix}.$$

Definimos $f = (f_1,...,f_n)$, donde $f_i(x) = \sum_{j=1}^n a_{ij}x_j$, $x = (x_1,...,x_n) \in \mathbb{R}^n$. Es claro que, para todo $i, j = 1,...,n$, $\frac{\partial f_i}{\partial x_j}(x) = a_{ij}$, $x \in \mathbb{R}^n$. Por tanto, $Df(x) = A$, $x \in \mathbb{R}^n$.
Observamos también que, ya que $n = \dim(\ker f) + \dim(\operatorname{Im} f)$, se verifica que $\ker f = \{0\}$ si, y solo si, $\operatorname{Im} f = \mathbb{R}^n$. Dicho de otro modo, f es inyectiva, si, y solo si, f es sobre, y por tanto, equivalente a que f sea biyectiva.
Veamos que f es biyectiva si, y solo si, A es una matriz invertible. Notamos primero que por ser f lineal, si es biyectiva, entonces su inversa es también lineal. En efecto, supongamos que f es biyectiva. Sean $u, v \in \mathbb{R}^d$ y $\alpha, \beta \in \mathbb{R}$. Definimos $x = f^{-1}(u)$ e $y = f^{-1}(v)$, esto es, $f(x) = u$ y $f(y) = v$. Por tanto, ya que f es lineal, $f(\alpha x + \beta y) = \alpha f(x) + \beta f(y) = \alpha u + \beta v$. Luego, $\alpha x + \beta y = f^{-1}(\alpha u + \beta v)$, es decir, $f^{-1}(\alpha u + \beta v) = \alpha f^{-1}(u) + \beta f^{-1}(v)$ y concluimos que f^{-1} es lineal.
Sea B la matriz asociada a f^{-1} en la base canónica. Luego, $f^{-1}(y) = By$, $y \in \mathbb{R}^n$. Ya que $x = f^{-1}(f(x)) = BAx$, $x \in \mathbb{R}^n$ e $y = f(f-1(y)) = ABy$, $y \in \mathbb{R}^n$, se sigue que $AB = BA = I$, esto es, la matriz A es invertible y $B = A^{-1}$.
Por otra parte, si A es una matriz invertible, entonces claramente la función f es biyectiva y $f^{-1}(y) = A^{-1}y$.
Concluimos que f es globalmente invertible si, y solo si, $\det A \neq 0$, lo que equivale también a que el sistema de ecuaciones $Ax = b$ tiene una única solución para cada $b \in \mathbb{R}^n$.
Nos planteamos ahora qué ocurre cuando $\det(A) = 0$. Sabemos que en este caso f no es biyectiva, es decir, no admite inversa global. Nos preguntamos entonces si f es localmente invertible, esto es, si dado $x_0 \in \mathbb{R}^n$, podemos encontrar U, entorno de x_0, de modo que $f_{|U} : U \longrightarrow f(U)$ es invertible.
La respuesta, negativa, la recogemos a continuación como propiedad.

Propiedad 2.17. Si $A \in M_n(\mathbb{R})$ y $\det(A) = 0$, entonces la aplicación lineal T definida por A no es localmente invertible en ningún punto de $\mathbb{R}^n$.

Demostración. Sea $x_0 \in \mathbb{R}^d$. Veamos que la aplicación $T : B(x_0, r_0) \longrightarrow T(B(x_0, r_0))$, para $r_0 > 0$, no es inyectiva.
Ya que $\det(A) = 0$, se tiene que 0 es un autovalor de A. Denotamos por E_0 el espacio propio asociado al autovalor 0 y tomamos $u \in E_0 \setminus \{0\}$. Por ser T lineal y $Tu = 0$, deducimos que $T(\alpha u) = \alpha Tu = 0$, $\alpha \in \mathbb{R}$.

Fijamos $r_0 > 0$. Elegimos $\alpha \in \mathbb{R}$ de manera que $|\alpha| < \frac{r_0}{\|u\|}$ (nótese que $\|u\| \neq 0$). Entonces, $x_0 + \alpha u \in B(x_0, r_0)$, pues $\|(x_0 + \alpha u) - x_0\| = |\alpha|\|u\| < r_0$. Además se tiene que $T(x_0 + \alpha u) = T(x_0) + \alpha T(u) = T(x_0)$, por lo que T no es inyectiva en $B(x_0, r_0)$.



□

Hemos visto que en el caso de las aplicaciones lineales, si $f : \Omega \subset \mathbb{R}^n \longrightarrow \mathbb{R}^n$ no es globalmente invertible, entonces para ningún punto de Ω la función es localmente invertible. En general, como se muestra en algunos de los ejemplos que siguen, esta situación no se tiene cuando consideramos otro tipo de funciones.

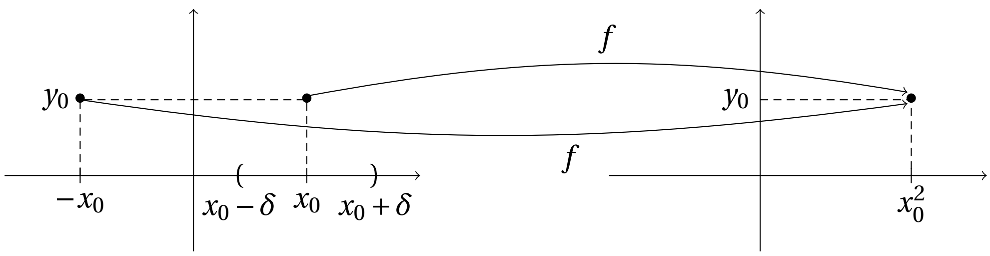
2. En el siguiente ejemplo presentamos una función no lineal globalmente invertible. Sea $f(x, y) = (x, y + x^2)$, $(x, y) \in \mathbb{R}^2$. Ya que

$$(x, y + x^2) = (u, v) \iff u = x,\ y = v - u^2, \quad x, y, u, v \in \mathbb{R},$$

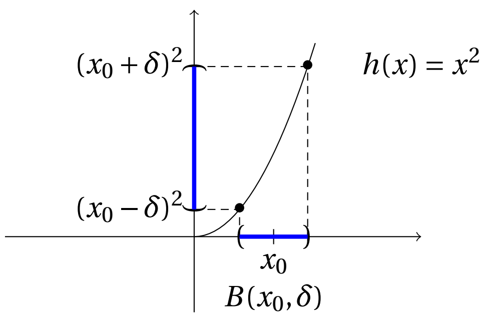
la función g definida en $\mathbb{R}^2$ como $g(u, v) = (u, v - u^2)$, $(u, v) \in \mathbb{R}^2$ resulta ser la función inversa de f. En efecto, $(g \circ f)(x, y) = g(x, y + x^2) = (x, y)$, $(x, y) \in \mathbb{R}^2$, y $(f \circ g)(u, v) = f(u, v - u^2) = (u, v)$, $(u, v) \in \mathbb{R}^2$. Por tanto, $g = f^{-1}$.
Observamos que si escribimos $f = (f_1, f_2)$, donde $f_1(x, y) = x$, y $f_2(x, y) = y + x^2$, $(x, y) \in \mathbb{R}^2$, entonces

$$\det(Df(x)) = \begin{vmatrix} 1 & 0 \\ 2x & 1 \end{vmatrix} = 1 \neq 0, \quad (x, y) \in \mathbb{R}^2.$$

3. Consideramos ahora una función que no es globalmente invertible pero sí lo es localmente en casi todos sus puntos. Tomamos $f(x, y) = (x^2, y)$, $(x, y) \in \mathbb{R}^2$. Claramente, la función f no es biyectiva ya que $f(x, y) = f(-x, y) = (x^2, y)$, $(x, y) \in \mathbb{R}^2$.



Sea $(x_0, y_0) \in \mathbb{R}^2$, tal que $x_0 \neq 0$. Elegimos $\delta > 0$ tal que $0 \notin (x_0-\delta, x_0+\delta)$ (por ejemplo, $\delta = \frac{|x_0|}{2}$) y tomamos $U_0 = (x_0-\delta, x_0+\delta) \times \mathbb{R}$.
Es fácil comprobar que $f_{|U_0} : U_0 \longrightarrow \mathbb{R}^2$ es una aplicación inyectiva. En efecto, sean (x_1, y_1) y $(x_2, y_2) \in U_0$ tales que $f(x_1, y_1) = f(x_2, y_2)$. Entonces $(x_1^2, y_1) = (x_2^2, y_2)$. Luego, es evidente que $y_1 = y_2$ y, puesto que $0 \notin (x_0-\delta, x_0+\delta)$, también se verifica que $x_1 = x_2$. Se obtiene así que f es localmente invertible en (x_0, y_0).
Por otro lado, podemos probar que $f(U_0)$ es un conjunto abierto en $\mathbb{R}^2$. Para ello, supongamos que $x_0 > 0$ (cuando $x_0 < 0$ podemos proceder del mismo modo). Basta tener en cuenta que la función $h(x) = x^2$, $x \in \mathbb{R}$, es continua y creciente en $(0,\infty)$. Entonces, $h(B(x_0,\delta)) = ((x_0-\delta)^2, (x_0+\delta)^2)$ y, por tanto, $f(U_0) = ((x_0-\delta)^2, (x_0+\delta)^2) \times \mathbb{R}$.



Por otra parte, tenemos que la función dada no es localmente invertible en $(0, y_0)$, con $y_0 \in \mathbb{R}$, pues dado $\eta > 0$ y $V_0 = (-\eta, \eta) \times (y_0-\eta, y_0+\eta) \longrightarrow f(V_0) \subseteq \mathbb{R}^2$, es claro que $f(\frac{\eta}{2}, y_0) = f(\frac{-\eta}{2}, y_0)$. Luego f no es inyectiva.
Si denotamos por $f_1(x, y) = x^2$ y $f_2(x, y) = y$, $(x, y) \in \mathbb{R}^2$, entonces $f = (f_1, f_2)$ y se tiene que

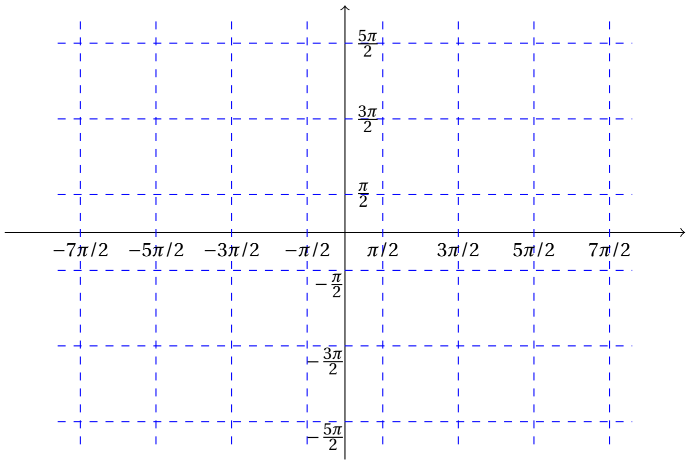
$$\det(Df(x,y)) = \begin{vmatrix} 2x & 0 \\ 0 & 1 \end{vmatrix} = 2x, \quad (x,y) \in \mathbb{R}^2.$$

Observamos que $\det(Df(x, y)) = 0$ si y solo si $x = 0$ y los puntos donde existe inversa local para la función f verifican que $\det Df(x, y) \neq 0$.

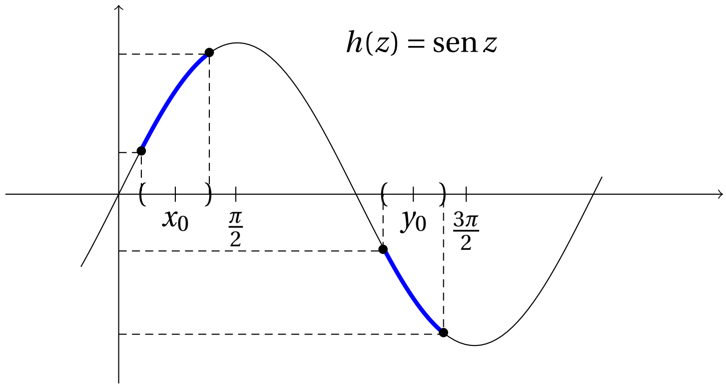
4. Otro ejemplo de función no biyectiva que es localmente integrable para ciertos puntos es el siguiente. Tomamos $f(x, y) = (f_1(x, y), f_2(x, y)) = (\operatorname{sen} x, \operatorname{sen} y)$, $(x, y) \in \mathbb{R}^2$. En este caso,

$$\det Df(x,y) = \begin{vmatrix} \cos x & 0 \\ 0 & \cos y \end{vmatrix} = \cos x \cos y, \quad (x,y) \in \mathbb{R}^2.$$

Por tanto, $\det(Df(x, y)) = 0$ si, y solo si, $\cos x = 0$ ó $\cos y = 0$, esto es, si, y solo si, $x = \frac{\pi}{2} + k\pi$, ó bien, $y = \frac{\pi}{2} + m\pi$, con $k, m \in \mathbb{Z}$.
Representamos por E al conjunto $E = \{(x, y) \in \mathbb{R}^2 : \det(Df(x, y)) = 0\}$.



La función f no es globalmente invertible ya que, por ejemplo, $f(x+2k\pi, y) = (\operatorname{sen} x, \operatorname{sen} y) = f(x, y)$, $(x, y) \in \mathbb{R}^2$, $k \in \mathbb{Z}$.
Sin embargo, f es localmente invertible en (x_0, y_0) cuando (x_0, y_0) no pertenece a E. Veámoslo. Sea $(x_0, y_0) \notin E$. Esto significa que $x_0 \neq \frac{\pi}{2} + k\pi$, para $k \in \mathbb{Z}$, y que $y \neq m\pi + \frac{\pi}{2}$, $m \in \mathbb{Z}$.
Podemos encontrar así $\delta_1, \delta_2 > 0$ tales que la función $h(z) = \operatorname{sen} z$ es estrictamente monótona en los intervalos $I_1 = (x_0 - \delta_1, x_0 + \delta_1)$ e $I_2 = (y_0 - \delta_2, y_0 + \delta_2)$.



La función h es entonces biyectiva en I_1 e I_2, y $h^{-1}(w) = \operatorname{arcosen}(w)$, cuando $w \in h(I_1) \cup h(I_2)$.
También se tiene que $h(I_j)$, $j = 1, 2$, es un conjunto abierto. Concretamente, $h(I_1) = (\operatorname{sen}(x_0 - \delta_1), \operatorname{sen}(x_0 + \delta_1))$, cuando h es creciente, y $h(I_1) = (\operatorname{sen}(x_0 + \delta_1), \operatorname{sen}(x_0 - \delta_1))$, cuando h es decreciente. Análogamente ocurre para I_2.
Definiendo $U_0 = I_1 \times I_2$ se tiene que $f_{|U_0}$ es biyectiva, siendo $(f_{|U_0})^{-1}(u, v) = (\operatorname{arc sen} u, \operatorname{arc sen} v)$, $(u, v) \in f(U_0)$.

5. Una función que no es biyectiva pero que tiene inversa local para todo punto es $f(x, y) = (e^x \operatorname{sen} y, e^x \cos y)$, $(x, y) \in \mathbb{R}^2$. En este caso, ya que $f = (f_1, f_2)$, donde $f_1(x, y) = e^x \operatorname{sen} y$ y $f_2(x, y) = e^x \cos y$, $(x, y) \in \mathbb{R}^2$, vemos que

$$\det(Df(x, y)) = \begin{vmatrix} e^x \operatorname{sen} y & e^x \cos y \\ e^x \cos y & -e^x \operatorname{sen} y \end{vmatrix} = -e^{-2x} \neq 0, \quad (x, y) \in \mathbb{R}^2.$$

Puesto que $f(x,y) = f(x,y+2k\pi)$, $(x,y) \in \mathbb{R}^2$, $k \in \mathbb{Z}$, la función f no es biyectiva. Veamos que f es localmente invertible en todo punto de $\mathbb{R}^2$.
En primer lugar, notamos que

$$f(x_1,y_1) = f(x_2,y_2) \quad \text{si, y solo si,} \quad x_1 = x_2,\ y_1 = y_2 + 2k\pi,\ k \in \mathbb{Z}. \tag{2.10}$$

Ya comentamos que $f(x,y) = f(x,y+2k\pi)$, $(x,y) \in \mathbb{R}^2$, $k \in \mathbb{Z}$. Si suponemos ahora que $f(x_1,y_1) = f(x_2,y_2)$, entonces se verifican las ecuaciones

$$\begin{aligned} e^{x_1}\operatorname{sen}(y_1) &= e^{x_2}\operatorname{sen}(y_2) \\ e^{x_1}\cos(y_1) &= e^{x_2}\cos(y_2), \end{aligned}$$

de donde se deduce que $e^{2x_1} = e^{2x_2}$, y entonces, $x_1 = x_2$. De esta forma, $\operatorname{sen}(y_1) = \operatorname{sen}(y_2)$ y $\cos(y_1) = \cos(y_2)$, lo que implica que $y_1 = y_2 + 2k\pi$.
Fijamos $(x_0,y_0) \in \mathbb{R}^2$ y consideramos $U_0 = \mathbb{R} \times (y_0 - \frac{\pi}{2}, y_0 + \frac{\pi}{2})$. Si tomamos $y_1, y_2 \in (y_0 - \frac{\pi}{2}, y_0 + \frac{\pi}{2})$ es claro que $|y_1 - y_2| < \pi$, y, por tanto, de (2.10) se sigue que $f(x_1,y_1) \neq f(x_2,y_2)$, cuando $(x_1,y_1),(x_2,y_2) \in U_0$ y $(x_1,y_1) \neq (x_2,y_2)$. De esta forma, U_0 es un abierto que contiene a (x_0,y_0) y $f_{|U_0} : U_0 \longrightarrow f(U_0)$ es biyectiva.
Para dar una expresión de $(f_{|U_0})^{-1}$ basta determinar x e y en términos de u y v a partir de las ecuaciones

$$\begin{aligned} e^x \operatorname{sen} y &= u \\ e^x \cos y &= v. \end{aligned}$$

De aquí se obtiene fácilmente que $e^{2x} = u^2 + v^2$ y, entonces, $x = \ln\sqrt{u^2+v^2}$ (nótese que $(u,v) \neq (0,0)$). Para la variable y nos queda

$$\operatorname{sen} y = \frac{u}{e^x} = \frac{u}{\sqrt{u^2+v^2}} \quad \text{y} \quad \cos y = \frac{v}{e^x} = \frac{v}{\sqrt{u^2+v^2}},$$

de donde se sigue que $y = \operatorname{arcsen}\left(\frac{u}{\sqrt{u^2+v^2}}\right) = \arccos\left(\frac{v}{\sqrt{u^2+v^2}}\right)$.
Luego podemos expresar $(f_{|U_0})^{-1}$ mediante

$$(f_{|U_0})^{-1}(u,v) = \left(\ln\sqrt{u^2+v^2}, \operatorname{arcsen}\left(\frac{u}{\sqrt{u^2+v^2}}\right)\right), \quad (u,v) \in f(U_0).$$

2.2.3. Teorema de la función inversa

Hemos visto en los ejemplos anteriores que las funciones consideradas son localmente invertibles en los puntos $x \in \mathbb{R}^2$ donde $\det Df(x) \neq 0$. El Teorema de la función inversa, que presentamos a continuación, nos dice que esta es una condición suficiente para asegurar que una función es localmente invertible.

Teorema 2.18 (Teorema de la función inversa). *Sean $f : \Omega \subset \mathbb{R}^n \longrightarrow \mathbb{R}^n$ una función de clase C^1 en el abierto Ω y $x_0 \in \Omega$. Supongamos que $Df(x_0)$ es invertible, esto es, que* $\det(Df(x_0)) \neq 0$. *Entonces, existe un abierto $V_0 \subset \Omega$ tal que $x_0 \in V_0$ y de manera que:*

(i) $f(V_0)$ es abierto en $\mathbb{R}^n$.
(ii) $f : V_0 \longrightarrow f(V_0)$ es biyectiva.
(iii) $(f_{|V_0})^{-1}$ es una función de clase C^1 en $f(V_0)$.
(iv) $D((f_{|V_0})^{-1})(y) = (Df(x))^{-1}$, siendo $y \in f(V_0)$ y $x \in V_0$ tal que $f(x) = y$.

Demostración. Suponemos, en primer lugar, que $x_0 = 0$, $f(x_0) = 0$ y $Df(x_0) = I$.

Escribimos $f = (f_1, \dots, f_n)$. Ya que $f \in C^1(\Omega)$, podemos asegurar que $\det(Df) : \Omega \longrightarrow \mathbb{R}$ es una función continua en Ω, siendo

$$Df(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \cdots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1}(x) & \cdots & \frac{\partial f_n}{\partial x_n}(x) \end{pmatrix}, \quad x \in \Omega.$$

Puesto que $\det(Df(0)) = \det I = 1$, la continuidad de $\det(Df)$ nos permite encontrar $R_0 > 0$ de manera que $\det(Df(x)) > 0$, para todo $x \in B(0, R_0)$. Además, podemos elegir R_0 de manera que $\overline{B(0, R_0)} \subset \Omega$.

Sea $\|\cdot\|$ una norma en el espacio de las matrices $M_n(\mathbb{R})$ (recordamos que en $M_n(\mathbb{R})$ todas las normas son equivalentes).

Definimos la función $\psi(x) = x - f(x)$, $x \in \Omega$. Es claro que $D\psi(x) = I - Df(x)$, $x \in \Omega$, y, en particular, $D\psi(0) = I - Df(0) = I - I = 0$. Ya que $\psi \in C^1(\Omega)$, de la prueba de la Proposición 2.16 se sigue que, para cierta $c_0 > 0$,

$$\left\|\psi(x_1) - \psi(x_2)\right\|_2 \leq c_0 \sup_{z \in B(0,R)} \|D\psi(z)\| \left\|x_1 - x_2\right\|_2, \quad x_1, x_2 \in B(0, R),$$

para cada $R \in (0, R_0)$. Además, puesto que $D\psi(0) = 0$ y $\psi \in C^1(\Omega)$, podemos encontrar $R_1 \in (0, R_0)$ de manera que $\|D\psi(z)\| < \frac{1}{2c_0}$, $z \in \overline{B(0, R_1)}$. Por tanto, se sigue que

$$\left\|\psi(x_1) - \psi(x_2)\right\|_2 < \frac{1}{2}\left\|x_1 - x_2\right\|_2, \quad x_1, x_2 \in \overline{B(0, R_1)}. \tag{2.11}$$

De esta forma, ψ aplica $\overline{B(0, R_1)}$ en $B(0, \frac{R_1}{2})$ (obsérvese que $\psi(0) = 0$).

Vamos a probar a continuación que para cada $y \in B(0, \frac{R_1}{2})$ existe $x_y \in B(0, R_1)$ tal que $f(x_y) = y$. Sea $y \in \mathbb{R}^n$ tal que $\|y\|_2 < \frac{R_1}{2}$.

Consideramos la función Φ_y definida por $\Phi_y(x) = \psi(x) + y$, $x \in \Omega$. Teniendo en cuenta (2.11) podemos deducir que, para todo $x \in \overline{B(0, R_1)}$,

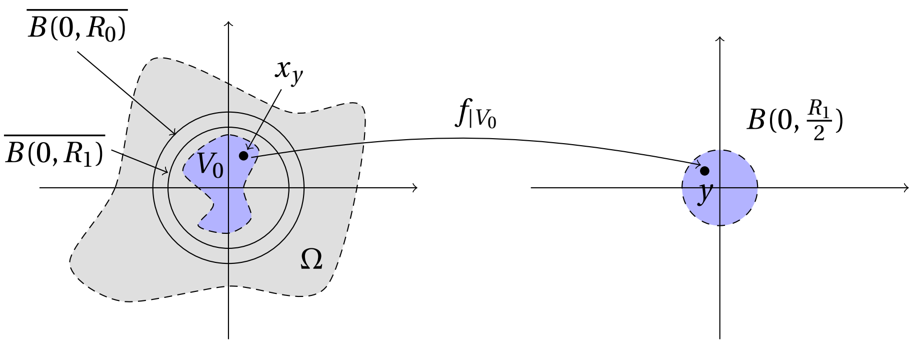
$$\begin{aligned} \|\Phi_y(x)\|_2 &\leq \|\psi(x)\|_2 + \|y\|_2 = \|\psi(x) - \psi(0)\|_2 + \|y\|_2 \\ &< \frac{\|x\|_2}{2} + \|y\|_2 < \frac{R_1}{2} + \frac{R_1}{2} = R_1. \end{aligned} \tag{2.12}$$

Por tanto, podemos decir que $\Phi_y : \overline{B(0,R_1)} \longrightarrow \overline{B(0,R_1)}$. Además, de nuevo por (2.11),

$$\left\|\Phi_y(x_1)-\Phi_y(x_2)\right\|_2 = \left\|\psi(x_1)-\psi(x_2)\right\|_2 < \frac{1}{2}\left\|x_1-x_2\right\|_2, \quad x_1,x_2\in\overline{B(0,R_1)},$$

esto es, Φ_y es una función contractiva. Ya que $(\overline{B(0,R_1)},\|\cdot\|_2)$ es completo, aplicando el Teorema del punto fijo de Banach (Teorema 1.14) se deduce que existe un único $x_y\in\overline{B(0,R_1)}$ verificando que $\Phi_y(x_y)=y$, o lo que es equivalente, tal que $f(x_y)=y$. Nótese que, de esta forma, si $z\in\Omega$, $z\neq x_y$, y $f(z)=y$, entonces $\|z\|_2>R_1$. Podemos asegurar además que $x_y\in B(0,R_1)$. Basta observar que $\|x_y\|_2=\|\Phi_y(x_y)\|_2$ y tener en cuenta (2.12).

Definimos entonces $V_0=f^{-1}(B(0,\frac{R_1}{2}))\cap B(0,R_1)$. Es claro que $0\in V_0$ y ya que f es continua en Ω, $f^{-1}(B(0,\frac{R_1}{2}))$ es un abierto, y por tanto V_0 es abierto.



Tenemos entonces que la aplicación

$$\begin{aligned} f_{|V_0} : V_0 &\longrightarrow B\Big(x_0,\frac{R_1}{2}\Big)\\ x_y &\longmapsto y \end{aligned}$$

es biyectiva y $f_{|V_0}^{-1} : B(x_0,\frac{R_1}{2}) \longrightarrow V_0$ es también biyectiva. LLamamos $g=(f_{|V_0})^{-1}$. Veamos que g verifica las propiedades (iii) y (iv) del teorema.

Observamos, en primer lugar, que g es continua en $B(x_0,\frac{R_1}{2})$. En efecto, sean $y_1,y_2\in B(x_0,\frac{R_1}{2})$ y x_1,x_2 los únicos puntos en V_0 tales que $f(x_1)=y_1$ y $f(x_2)=y_2$. Tenemos así que $g(y_1)=x_1$ y $g(y_2)=x_2$. Ya que $x_1,x_2\in B(0,R_1)$ podemos hacer uso de (2.11) para obtener

$$\begin{aligned}\left\|g(y_1)-g(y_2)\right\|_2 &= \left\|x_1-x_2\right\|_2 = \left\|x_1-f(x_1)-(x_2-f(x_2))+f(x_1)-f(x_2)\right\|_2\\ &= \left\|\psi(x_1)-\psi(x_2)+f(x_1)-f(x_2)\right\|_2 \le \left\|\psi(x_1)-\psi(x_2)\right\|_2+\left\|f(x_1)-f(x_2)\right\|_2\\ &< \frac{1}{2}\left\|x_1-x_2\right\|_2+\left\|f(x_1)-f(x_2)\right\|_2.\end{aligned}$$

Luego, $\|x_1-x_2\|_2\le 2\|f(x_1)-f(x_2)\|_2$, esto es, $\|g(y_1)-g(y_2)\|_2\le 2\|y_1-y_2\|_2$ y podemos concluir que g es continua en V_0.

Veamos a continuación que g es diferenciable en $B(0,\frac{R_1}{2})$ y que, para todo $y_0 \in B(0,\frac{R_1}{2})$ se verifica que $Dg(y_0) = (Df(x_0))^{-1}$, siendo $x_0 \in V_0$ tal que $f(x_0) = y_0$. Nótese que $\det(Df(x_0)) \neq 0$, ya que $x_0 \in V_0 \subset B(0,R_0)$.

Sea $y_0 \in B(0,\frac{R_1}{2})$. Escribimos $x_0 = g(y_0)$, esto es, $f(x_0) = y_0$ con $x_0 \in V_0$. Tenemos que establecer que

$$\lim_{h\to 0} \frac{g(y_0+h) - g(y_0) - (Df(x_0))^{-1}h}{\|h\|_2} = 0. \tag{2.13}$$

Es claro que $y_0 + h \in B(0,\frac{R_1}{2})$ siempre que $\|h\|_2 < \frac{R_1}{2} - \|y_0\|_2$. Consideramos $h \in \mathbb{R}^n$ tal que $0 < \|h\|_2 < \frac{R_1}{2} - \|y_0\|_2$ y definimos $k = g(y_0+h) - g(y_0)$ y $x_1 = g(y_0+h)$. Nótese que al ser $h \neq 0$, se tiene que $y_0 + h \neq y_0$ y, por tanto, $g(y_0+h) \neq g(y_0)$, esto es, $k \neq 0$. Además, $y_0 + h = f(x_1) = f(x_0+k)$ y

$$\|k\|_2 = \|x_1 - x_0\|_2 \leq 2\|f(x_1) - f(x_0)\|_2 = 2\|y_0 + h - y_0\| = 2\|h\|_2.$$

Entonces, podemos escribir

$$\begin{aligned}\frac{g(y_0+h) - g(y_0) - (Df(x_0))^{-1}h}{\|h\|_2} &= \frac{(Df(x_0))^{-1}(Df(x_0)k - (y_0+h-y_0))}{\|h\|_2} \\ &= \frac{\|k\|_2}{\|h\|_2}(Df(x_0))^{-1}\Big(\frac{Df(x_0)k - f(x_0+k) + f(x_0)}{\|k\|_2}\Big),\end{aligned}$$

de donde se sigue que

$$\left\|\frac{g(y_0+h) - g(y_0) - (Df(x_0))^{-1}h}{\|h\|_2}\right\|_2 \leq 2\|(Df(x_0))^{-1}\|_{2\to 2}\frac{\|f(x_0+k) - f(x_0) - Df(x_0)k\|_2}{\|k\|_2}.$$

Ya que $k \to 0$ cuando $h \to 0$ y que $\lim_{k\to 0}\frac{\|f(x_0+k)-f(x_0)-Df(x_0)k\|_2}{\|k\|_2} = 0$, (2.13) queda probado.

Veamos finalmente que g es una función de clase C^1 en $B(0,\frac{R_1}{2})$. Sea $y_0 \in B(0,R_1/2)$. Hemos probado que

$$Dg(y_0) = (Df(g(y_0))^{-1} = \frac{1}{\det(Df(g(y_0))}(\operatorname{Adj} Df(g(y_0))^T,$$

donde

$$Df(g(y_0)) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(g(y_0)) & \cdots & \frac{\partial f_1}{\partial x_n}(g(y_0)) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(g(y_0)) & \cdots & \frac{\partial f_m}{\partial x_m}(g(y_0)) \end{pmatrix}.$$

Observamos que, al ser g es continua de $B(0,\frac{R_1}{2})$ en V_0, $f \in C^1(V_0)$ y $\det Df(g(y_0)) \neq 0$, cada entrada de la matriz $Dg(y_0)$ es el resultado de la suma, producto y composición de funciones continuas en $B(0,\frac{R_1}{2})$.

Hemos establecido así el teorema bajo las hipótesis $x_0 = 0$, $f(x_0) = 0$ y $Df(0) = I$.

Asumimos ahora que $x_0 \in \Omega$ satisface $\det Df(x_0) \neq 0$ y consideramos la función $F(x) = (Df(x_0))^{-1}(f(x + x_0) - f(x_0))$, $x \in \Omega' = \Omega - x_0$.

Observamos que $0 \in \Omega'$, ya que $0 = x_0 - x_0 \in \Omega - x_0$. Además, $F(0) = 0$ y $DF(0) = (Df(x_0))^{-1}Df(x_0) = I$. Teniendo en cuenta lo que hemos probado anteriormente, sabemos que existe $R > 0$ y un abierto $\widetilde{V}_0 \subset \Omega'$ tal que $0 \in \widetilde{V}_0$, $F : \widetilde{V}_0 \to B(0, R)$ es biyectiva, $G = (F_{|\widetilde{V}_0})^{-1} : B(0, R) \to \widetilde{V}_0$ es $C^1(B(0, R))$ y, para cada $y \in B(0, R)$ se verifica que $DG(y) = (DF(G(y)))^{-1}$.

Obtenemos de aquí, V_0 abierto de Ω, siendo $x_0 \in V_0$ y $W_0 = f(V_0)$ abierto en $\mathbb{R}^n$ tales que $f : V_0 \to W_0$ es biyectiva y $g = (f_{|V_0})^{-1} : W_0 \to V_0$ es una función en $C^1(W_0)$, siendo $Dg(y) = (Df(g(y)))^{-1}$, $y \in W_0$. □

Observación 2.19. Si en el teorema anterior consideramos $f \in C^k(\Omega)$ con $k \in \mathbb{N}, k \geq 1$, entonces podemos probar que $g \in C^k(W_0)$.

Corolario 2.20. *Sea $f : \Omega \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ una función de clase C^1 en el abierto Ω. Supongamos que* $\det(Df(x)) \neq 0$ *para cada $x \in \Omega$. Entonces f es una aplicación abierta.*

Demostración. Sea $U \subset \Omega$ un conjunto abierto. Fijamos $y_0 \in f(U)$ y tomamos $x_0 \in U$ tal que $f(x_0) = y_0$. Sabemos que $\det(Df(x_0)) \neq 0$. Ya que $f \in C^1(U)$, del Teorema de la función inversa se sigue que existe $U_0 \subseteq U$ abierto, con $x_0 \in U_0$, tal que $f(U_0)$ es abierto y la aplicación $f : U_0 \longrightarrow f(U_0)$ es biyectiva, siendo $(f_{|U_0})^{-1} : f(U_0) \longrightarrow U_0$ una función en $C^1(f(U_0))$ y $D((f_{|U_0})^{-1})(y) = (Df(f^{-1}(y)))^{-1}$, $y \in f(U_0)$.

Hemos probado así que para cada $y \in f(U)$ existe U_y abierto $U_y \subset f(U)$ tal que $y \in U_y$. Por tanto, $f(U) = \cup_{y \in f(U)} U_y$. Ya que la unión arbitraria de abiertos es abierta, concluimos que $f(U)$ es un conjunto abierto. □

Como consecuencia de este resultado se obtiene el siguiente corolario.

Corolario 2.21. *Sea $f : \Omega \subset \mathbb{R}^n \longrightarrow \mathbb{R}^n$ una función inyectiva de clase C^1 en el abierto Ω. Supongamos que* $\det(Df(x)) \neq 0$, *$x \in \Omega$. Entonces, $f(\Omega)$ es abierto, $f^{-1} : f(\Omega) \longrightarrow \Omega$ es $C^1(f(\Omega))$ y $Df^{-1}(y) = (Df(f^{-1}(y)))^{-1}$, $y \in f(\Omega)$.*

Observación 2.22. Supongamos que $f : \Omega \subset \mathbb{R}^n \longrightarrow \mathbb{R}^n$ es inyectiva y diferenciable en Ω siendo Ω abierto y que $f^{-1} : f(\Omega) \longrightarrow \Omega$ es diferenciable en $f(\Omega)$, siendo $f(\Omega)$ abierto. Tenemos que $f^{-1}(f(x)) = x$, $x \in \Omega$. Aplicando la regla de la cadena se obtiene que $D(f^{-1}(f(x))) = (Df^{-1})(f(x))Df(x) = I$ Esto nos dice que $\det(Df(x)) \neq 0$, $x \in \Omega$, y $(Df(x))^{-1} = (Df^{-1})(f(x))$, $x \in \Omega$.

Haciendo uso del Teorema de la función inversa podemos también establecer el siguiente resultado para funciones holomorfas.

Corolario 2.23. *Sean $f : \Omega \subset \mathbb{C} \longrightarrow \mathbb{C}$ holomorfa, Ω abierto y $z_0 \in \Omega$. Supongamos que $f'(z_0) \neq 0$. Entonces, existe U_0 abierto en Ω tal que*

(i) $z_0 \in U_0$;

(ii) $f(U_0)$ *es abierto en* $\mathbb{C}$;

(iii) $f_{|U_0} : U_0 \longrightarrow f(U_0)$ *es biyectiva, y* $(f_{|U_0})^{-1} : f(U_0) \longrightarrow U_0$ *es holomorfa en* $f(U_0)$ *siendo* $(f^{-1})'(w) = \frac{1}{f'(f^{-1}(w))}$, $w \in f(U_0)$.

Demostración. Definimos $\widetilde{\Omega} = \{(x, y) \in \mathbb{R}^2 : x + yi \in \Omega\}$ y escribimos $f = f_1 + i f_2$, siendo $f_1 = \mathrm{Re} f$ y $f_2 = \mathrm{Im} f$. Asimismo, consideramos la función $F = (F_1, F_2)$, donde $F_j(x, y) = f_j(x + iy)$, $(x, y) \in \widetilde{\Omega}$, para $j = 1, 2$.

Ya que f es holomorfa en Ω se sigue que $F_1, F_2 \in C^\infty(\Omega)$ y

$$f'(z) = \lim_{h\to 0} \frac{f(z+h) - f(z)}{h} = \frac{\partial F_1}{\partial x}(x, y) + i \frac{\partial F_2}{\partial x}(x, y), \quad z = x + iy \in \Omega. \tag{2.14}$$

Además, F_1 y F_2 satisfacen las ecuaciones de Cauchy - Riemann, esto es,

$$\begin{aligned} \frac{\partial F_1}{\partial x}(x, y) &= \frac{\partial F_2}{\partial y}(x, y) \\ \frac{\partial F_2}{\partial x}(x, y) &= -\frac{\partial F_1}{\partial y}(x, y), \quad (x, y) \in \widetilde{\Omega}, \end{aligned}$$

por lo que, para todo $(x, y) \in \widetilde{\Omega}$,

$$\det DF(x, y) = \begin{vmatrix} \frac{\partial F_1}{\partial x}(x, y) & -\frac{\partial F_2}{\partial x}(x, y) \\ \frac{\partial F_2}{\partial x}(x, y) & \frac{\partial F_1}{\partial x}(x, y) \end{vmatrix} = \Big(\frac{\partial F_1}{\partial x}(x, y)\Big)^2 + \Big(\frac{\partial F_2}{\partial x}(x, y)\Big)^2.$$

Pongamos $z_0 = x_0 + iy_0$, con $(x_0, y_0) \in \widetilde{\Omega}$. Por hipótesis $f'(z_0) \neq 0$. De (2.14) se deduce entonces que $\frac{\partial F_1}{\partial x}(x_0, y_0) \neq 0$, o bien, $\frac{\partial F_2}{\partial x}(x_0, y_0) \neq 0$. Luego, $\det(DF(x_0, y_0)) \neq 0$.

Atendiendo a todo lo anterior tenemos una función $F : \widetilde{\Omega} \subset \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ de clase C^∞ en el abierto $\widetilde{\Omega}$ y un punto $(x_0, y_0) \in \widetilde{\Omega}$ tal que $\det(DF(x_0, y_0)) \neq 0$. El Teorema de la función inversa asegura entonces la existencia de un abierto $\widetilde{U}_0 \subset \widetilde{\Omega}$ tal que $(x_0, y_0) \in \widetilde{U}_0$, $\widetilde{W}_0 = f(\widetilde{U}_0)$ es abierto en $\mathbb{R}^2$, la función $F : \widetilde{U}_0 \longrightarrow \widetilde{W}_0$ es biyectiva siendo $G = (F_{|\widetilde{U}_0})^{-1} : \widetilde{W}_0 \longrightarrow \widetilde{U}_0$ una función en $C^\infty(\widetilde{W}_0)$ y

$$DG(u, v) = (DF(G(u, v)))^{-1}, \quad (u, v) \in \widetilde{W}_0.$$

Consideramos $G = (G_1, G_2)$. Definimos los conjuntos $W_0 = \{u + iv : (u, v) \in \widetilde{W}_0\}$ y $U_0 = \{x + iy : (x, y) \in \widetilde{U}_0\}$ y la aplicación

$$\begin{aligned} g : \ W_0 &\longrightarrow U_0 \\ u + iv &\mapsto g(u + iv) = G_1(u, v) + iG_2(u, v). \end{aligned}$$

Es claro que

$$g(f(z)) = g(f_1(z) + i f_2(z)) = g(F_1(x, y) + iF_2(x, y))$$

$$= G_1(F_1(x,y), F_2(x,y)) + iG_2(F_1(x,y), F_2(x,y)), \quad z = x+iy \in U_0.$$

Además, $G(F(x,y)) = (G_1(F(x,y)), G_2(F(x,y))) = (x,y)$, cuando $(x,y) \in \widetilde{U}_0$, esto es, $G_1(F(x,y)) = x$, y $G_2(F(x,y)) = y$, para todo $(x,y) \in \widetilde{U}_0$. Se concluye así que $g(f(z)) = x+iy = z$, siendo $z = x+iy \in U_0$.

Por otro lado, si $w = u+iv \in W_0$ se tiene que $f(g(w)) = f(G_1(u,v)+iG_2(u,v)) = F_1(G_1(u,v), G_2(u,v)) + iF_2(G_1(u,v), G_2(u,v)) = u+iv = w$. Se sigue entonces que $g = (f_{|U_0})^{-1} : W_0 \longrightarrow U_0$.

Para terminar la prueba veamos que $g'(w) = \frac{1}{f'(g(w))}$, $w \in W_0$. Sabemos que $g = g_1 + ig_2$, donde $g_j(w) = G_j(u,v)$, $w = u+iv \in W_0$, para $j = 1,2$. Luego, para establecer que g es holomorfa en W_0 basta ver que se satisfacen las ecuaciones de Cauchy-Riemann

$$\left.\begin{aligned} \frac{\partial G_1}{\partial u}(u,v) &= \frac{\partial G_2}{\partial v}(u,v) \\ \frac{\partial G_1}{\partial v}(u,v) &= -\frac{\partial G_2}{\partial u}(u,v) \end{aligned}\right\}, \quad (u,v) \in \widetilde{W}_0. \tag{2.15}$$

Puesto que $DG(u,v) = (DF(G(u,v)))^{-1}$, $(u,v) \in \widetilde{W}_0$, y

$$(DF(x,y))^{-1} = \frac{1}{\left(\frac{\partial F_1}{\partial x}(x,y)\right)^2 + \left(\frac{\partial F_2}{\partial x}(x,y)\right)^2} \begin{pmatrix} \frac{\partial F_2}{\partial y}(x,y) & -\frac{\partial F_1}{\partial y}(x,y) \\ -\frac{\partial F_2}{\partial x}(x,y) & \frac{\partial F_1}{\partial x}(x,y) \end{pmatrix}, \quad (x,y) \in \widetilde{U}_0,$$

se sigue que

$$\begin{aligned} DG(u,v) &= \begin{pmatrix} \frac{\partial G_1}{\partial u}(u,v) & \frac{\partial G_1}{\partial v}(u,v) \\ \frac{\partial G_2}{\partial u}(u,v) & \frac{\partial G_2}{\partial v}(u,v) \end{pmatrix} \\ &= \frac{1}{\left(\frac{\partial F_1}{\partial x}(G(u,v))\right)^2 + \left(\frac{\partial F_2}{\partial x}(G(u,v))\right)^2} \begin{pmatrix} \frac{\partial F_2}{\partial y}(G(u,v)) & -\frac{\partial F_1}{\partial y}(G(u,v)) \\ -\frac{\partial F_2}{\partial x}(G(u,v)) & \frac{\partial F_1}{\partial x}(G(u,v)) \end{pmatrix}. \end{aligned}$$

Por tanto,

$$\begin{aligned} \frac{\partial G_1}{\partial u}(u,v) &= \frac{1}{(\frac{\partial F_1}{\partial x}(G(u,v)))^2 + (\frac{\partial F_2}{\partial x}(G(u,v)))^2} \frac{\partial F_2}{\partial y}(G(u,v)) \\ \frac{\partial G_1}{\partial v}(u,v) &= \frac{-1}{(\frac{\partial F_1}{\partial x}(G(u,v)))^2 + (\frac{\partial F_1}{\partial x}(G(u,v)))^2} \frac{\partial F_1}{\partial y}(G(u,v)) \\ \frac{\partial G_2}{\partial u}(u,v) &= \frac{-1}{(\frac{\partial F_1}{\partial x}(G(u,v)))^2 + (\frac{\partial F_2}{\partial x}(G(u,v)))^2} \frac{\partial F_2}{\partial x}(G(u,v)) \\ \frac{\partial G_2}{\partial v}(u,v) &= \frac{1}{(\frac{\partial F_1}{\partial x}(G(u,v)))^2 + (\frac{\partial F_2}{\partial x}(G(u,v)))^2} \frac{\partial F_1}{\partial x}(G(u,v)). \end{aligned}$$

Entonces, puesto que $\frac{\partial F_1}{\partial x}(x,y) = \frac{\partial F_2}{\partial y}(x,y)$ y $\frac{\partial F_1}{\partial y}(x,y) = -\frac{\partial F_2}{\partial x}(x,y)$, $(x,y) \in \widetilde{U}_0$, obtenemos (2.15). Concluimos que g es holomorfa en W_0 y que, además,

$$\begin{aligned}
g'(w) &= \frac{\partial G_1}{\partial u}(u,v) + i\frac{\partial G_2}{\partial u}(u,v) \\
&= \frac{1}{\left(\frac{\partial F_1}{\partial x}(G(u,v))\right)^2 + \left(\frac{\partial F_2}{\partial x}(G(u,v))\right)^2}\left(\frac{\partial F_1}{\partial x}(G(u,v)) - i\frac{\partial F_2}{\partial x}(G(u,v))\right) \\
&= \frac{1}{\frac{\partial F_1}{\partial x}(G(u,v)) + i\frac{\partial F_2}{\partial x}(G(u,v))} = \frac{1}{f'(g(w))}, \quad w = u + iv \in W_0.
\end{aligned}$$

□

2.3. El Teorema de la función implícita

Es bastante probable que en algún curso de cálculo elemental hayamos escuchado la frase “la ecuación $f(x,y) = 0$ define implícitamente a y como función de x, ó a x como función de y". Este es un enunciado poco preciso pero nos da una idea de lo que formula el Teorema de la función implícita. Si queremos ser más rigurosos deberíamos añadir que se tiene en un entorno de cualquier punto (x_0, y_0) tal que $f(x_0, y_0) = 0$ y que verifique que al menos una de las derivadas parciales $\frac{\partial f}{\partial x}(x_0,y_0)$ ó $\frac{\partial f}{\partial y}(x_0,y_0)$ sea no nula.

Cuando resolvemos un sistema de ecuaciones como el siguiente

$$\left.\begin{aligned} 2x + y + z &= 5 \\ x - 3y + 2z &= 4 \end{aligned}\right\}$$

obtenemos infinitas soluciones que vienen dadas por

$$x = \frac{19}{7} - \frac{5}{7}z, \quad y = -\frac{3}{7} + \frac{3}{7}z, \quad z \in \mathbb{R}. \tag{2.16}$$

Podemos pensar este problema bajo el punto de vista del enunciado anterior, considerando la función $f : \mathbb{R} \times \mathbb{R}^2 \longrightarrow \mathbb{R}^2$, dada por

$$f(z,(x,y)) = (f_1(z,x,y), f_2(z,x,y)) = (2x + y + z - 5, x - 3y + 2z - 4).$$

Lo que nos dice (2.16) es que de la ecuación $f(z,x,y) = 0$ es posible “despejar"las variables x e y en términos de z. Observamos que en este caso,

$$\begin{vmatrix} \frac{\partial f_1}{\partial x}(z,x,y) & \frac{\partial f_1}{\partial y}(z,x,y) \\ \frac{\partial f_2}{\partial x}(z,x,y) & \frac{\partial f_1}{\partial y}(z,x,y) \end{vmatrix} = \begin{vmatrix} 2 & 1 \\ 1 & -3 \end{vmatrix} \neq 0, \quad x,y,z \in \mathbb{R}.$$

Como veremos, esta condición es un caso particular de la que se considera en el Teorema de la función implícita que enunciamos y probamos a continuación.

Teorema 2.24 (Teorema de la función implícita). *Sean Ω un abierto en $\mathbb{R}^n \times \mathbb{R}^m$,*

$$\begin{aligned} f : \mathbb{R}^n \times \mathbb{R}^m &\longrightarrow \mathbb{R}^m \\ (x,y) &\longmapsto f(x,y) = (f_1(x,y),\dots,f_m(x,y)), \end{aligned}$$

una función de clase C^k en Ω y $(a,b) \in \Omega$ tal que $f(a,b) = 0$. Supongamos además que

$$\begin{vmatrix} \frac{\partial f_1}{\partial y_1}(a,b) & \cdots & \frac{\partial f_1}{\partial y_m}(a,b) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial y_1}(a,b) & \cdots & \frac{\partial f_m}{\partial y_m}(a,b) \end{vmatrix} \neq 0.$$

Entonces, existen abiertos $A \subset \mathbb{R}^n$ y $B \subset \mathbb{R}^m$ y $g : A \longrightarrow B$ verificando las propiedades siguientes:

(i) $a \in A, b \in B$ y $A \times B \subset \Omega$.
(ii) $f(x,g(x)) = 0$, $x \in A$.
(iii) Para cada $x \in A$ existe un único $y \in B$ tal que $f(x,y) = 0$, siendo $y = g(x)$.
(iv) $g \in C^k(A)$.

Demostración. Para demostrar el resultado haremos uso del Teorema de la función inversa. Comenzamos definiendo la función $F : \Omega \subset \mathbb{R}^n \times \mathbb{R}^m \longrightarrow \mathbb{R}^n \times \mathbb{R}^m$ mediante

$$F(x,y) = (x, f(x,y)) = (F_1(x,y),\dots,F_n(x,y),F_{n+1}(x,y),\dots,F_{n+m}(x,y)), \quad (x,y) \in \Omega,$$

donde $F_j(x,y) = x_j$, $j = 1,\dots,n$, $F_{n+j}(x,y) = f_j(x,y)$, $j = 1,\dots,m$, $(x,y) \in \Omega$. Ya que $f \in C^k(\Omega)$, es claro que $F \in C^k(\Omega)$. Además,

$$DF(x,y) = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \\ \frac{\partial f_1}{\partial_{x_1}}(x,y) & & \cdots & \frac{\partial f_1}{\partial_{x_n}}(x,y) & \frac{\partial f_1}{\partial_{y_1}}(x,y) & \cdots & \frac{\partial f_1}{\partial_{y_m}}(x,y) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial_{x_1}}(x,y) & & \cdots & \frac{\partial f_m}{\partial_{x_n}}(x,y) & \frac{\partial f_m}{\partial_{y_1}}(x,y) & \cdots & \frac{\partial f_m}{\partial_{y_m}}(x,y) \end{pmatrix}$$

Luego,

$$\det(DF(x,y)) = \begin{vmatrix} \frac{\partial f_1}{\partial y_1}(x,y) & \cdots & \frac{\partial f_1}{\partial y_m}(x,y) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial y_1}(x,y) & \cdots & \frac{\partial f_m}{\partial y_m}(x,y) \end{vmatrix}, \quad (x,y) \in \Omega,$$

y en particular, $\det(DF(a,b)) \neq 0$. Estamos así en las condiciones del Teorema de función inversa. Y, por tanto, existen U abierto en $\mathbb{R}^n$ y V abierto en $\mathbb{R}^m$ tales que

(i) $a \in U$, $b \in V$ y $U \times V \subset \Omega$;

(ii) $\det(DF(x,y)) \neq 0$, $(x,y) \in U \times V$.

(iii) $F_{|U\times V} : U \times V \longrightarrow F(U \times V) = W$ es biyectiva y W es abierto. Además, la función $G = (F_{|U\times V})^{-1}$ pertenece a $C^k(W)$ y $DG(z,t) = (DF(G(z,t)))^{-1}$, $(z,t) \in W$.

Escribimos $G = (G_1, \ldots, G_n, G_{n+1}, \ldots, G_{n+m})$. Definimos el conjunto $A = \{x \in U : (x,0) \in W\}$, tomamos $B = V$, y consideramos la función

$$\begin{aligned} g : A &\longrightarrow \mathbb{R}^m \\ x &\longmapsto g(x) = (G_{n+1}(x,0), \ldots, G_{n+m}(x,0)). \end{aligned}$$

Veamos que A, B y g satisfacen las propiedades enunciadas en el teorema.

En primer lugar observamos que $a \in A$, pues $a \in U$ y $(a,0) = (a, f(a,b)) = F(a,b) \in F(U \times V) = W$. Recordamos que $b \in V$, luego $b \in B$.

Para probar que A es abierto en $\mathbb{R}^n$ basta considerar la función

$$\begin{aligned} p : \mathbb{R}^n &\longrightarrow \mathbb{R}^n \times \mathbb{R}^m \\ (x,y) &\longmapsto (x,0). \end{aligned}$$

Esta función es continua y A puede escribirse como $A = U \cap p^{-1}(W)$. El conjunto U es abierto en $\mathbb{R}^n$ y, ya que W es abierto en $\mathbb{R}^n \times \mathbb{R}^m$, se tiene que $p^{-1}(W)$ es abierto en $\mathbb{R}^n$. Luego, A es abierto en $\mathbb{R}^n$.
Por último veamos que se verifican las propiedades enunciadas para la función g.

(a) $g \in C^k(A)$, ya que $G \in C^k(W)$.

(b) $g(A) \subset B = V$. En efecto, sea $x \in A$. Tenemos que $(x,0) \in W$. Por tanto,

$$G(x,0) = (G_1(x,0), \ldots, G_n(x,0), G_{n+1}(x,0), \ldots, G_{n+m}(x,0)) \in U \times V.$$

Se sigue que

$$g(x) = (G_{n+1}(x,0), \ldots, G_{n+m}(x,0)) \in V.$$

(c) $f(x, g(x)) = 0, x \in A$. Sea $x \in A$. Tenemos que

$$f(x, g(x)) = f(x, G_{n+1}(x,0), \ldots, G_{n+m}(x,0)).$$

Por otra parte

$$\begin{aligned} G(F(x,y)) = (x,y) &= G(F_1(x,y), \ldots, F_n(x,y), F_{n+1}(x,y), \ldots, F_{n+m}(x,y)) \\ &= G(x_1, \ldots, x_n, f_1(x,y), \ldots, f_m(x,y)), \quad (x,y) \in U \times V, \end{aligned}$$

y

$$\begin{aligned} F(G(z,t)) = (z,t) &= F(G_1(z,t), \ldots, G_n(z,t), G_{n+1}(z,t), \ldots, G_{n+m}(z,t)) \\ &= (G_1(z,t)), \ldots, G_n(z,t), f_1(G(z,t)), \ldots, f_m(G(z,t))), \quad (z,t) \in W. \end{aligned}$$

Luego, $G_j(z,t)) = z_j$, $j = 1, \ldots, n$, $(z,t) \in W$ y, por tanto, $x_j = G_j(x,0)$, $j = 1, \ldots, n$ (recordamos que $(x,0) \in W$ por ser $x \in A$). Se sigue entonces que

$$f(x,g(x)) = f(G_1(x,0),\dots,G_n(x,0),G_{n+1}(x,0),\dots,G_{n+m}(x,0)) = f(G(x,0)),$$

y puesto que $F(G(x,0)) = (x, f(G(x,0))) = (x,0)$ concluimos que $f(x,g(x)) = 0$.

(d) Probamos finalmente que si $x \in A$, $y \in B$, y $f(x,y) = 0$, entonces $y = g(x)$. Sean $x \in A$, $y \in B$ tales que $f(x,y) = 0$. Tenemos que $G(x,f(x,y)) = G(F(x,y)) = (x,y)$. Luego,

$$G(x,0) = (G_1(x,0),\dots,G_n(x,0),G_{n+1}(x,0),\dots,G_{n+m}(x,0)) = (x,y),$$

y se deduce que $y = (G_{n+1}(x,0),\dots,G_{n+k}(x,0)) = g(x)$. $\square$

Bibliografía

[1] EDWARDS, C.H., Jr. *Advanced calculus of several variables.* Academic Press, Inc., New York, 1973.

[2] FLEMING, W. *Functions of several variables.* 2nd Edition. Springer-Verlag, New York, 1977.

[3] FLORES, M. y SADARANGANI, K. *Cálculo diferencial e integral de varias variables.* Servicio de Publicaciones ULL, 2011.

[4] MARSDEN, J.E. *Elementary classical analysis.* Freeman, New York, 1974.

[5] MARSDEN, J.E. y TROMBA, A. *Cálculo vectorial.* 3^a Edition. Addison-Wesley Iberoamericana, Wilmington, 1991.

[6] RUDIN, W. *Principles of Mathematical Analysis.* 3rd edition. McGraw-Hill Inc., Singapore, 1976.

Inverse Function and Implicit Function Theorems

Amadou Diadie Coulibaly
Facultad de Ciencias • Sección de Matemáticas
Universidad de La Laguna
alu0101284828@ull.edu.es

Abstract

In this work we analyze two essential theorems in Differential Calculus: The Inverse Function Theorem and the Implicit Function Theorem. We first consider some general aspects on normed vector and metric spaces, and particularly, on matrix spaces $M_n(\mathbb{R})$*,* $n \in \mathbb{N}$*. Likewise, we address the study of Banach Fixed Point Theorem, which turns out to be a fundamental element in the proof of the Inverse Function Theorem, and we also study the Mean Value Theorem in the vector-valued context. Finally we state and prove the two main aforementioned theorems, showing some of their consequences, among which is the version of the Inverse Function Theorem for holomorphic functions.*

1. Introduction

The Inverse Function Theorem and the Implicit Function Theorem constitute ones of the most important, and oldest, paradigms in modern mathematics. If we go back to the 17th century we can see the germ of the idea for the implicit function theorem in the writings of Isaac Newton (1642-1727) and, in the work of Gottfried Leibniz (1646-1716) an example of implicit derivation is shown explicitly. Later, while Joseph Louis Lagrange (1736-1813) found a result that is essentially a version of the inverse function theorem, Augustin-Louis Cauchy (1789-1857) approached the study of the implicit function theorem with mathematical rigor, a fact for which its discovery is attributed to him. These theorems are still fundamental today, constituting, for example, a basic theoretical tool for the study of differentiable manifolds.

2. Outline of the first chapter

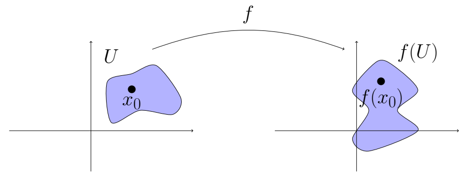
We present the fundamental elements of the theory that are necessary for the development of the next chapter. Specifically, the concepts of normed space and metric space are introduced, their main properties are addressed and some examples are provided, paying special attention to the space of matrices $M_n(\mathbb{R})$, $n \in \mathbb{N}$. We also present and prove the Banach Fixed Point Theorem which is stated as follows.

Banach Fixed Point Theorem Let (X, d) be a complete metric space. Suppose that $T : X \longrightarrow X$ is a contractive map, that is, there exists $0 < \alpha < 1$ such that: $d(T(x), T(y)) \leq \alpha d(x, y)$, $x, y \in X$. Then, there exists a unique $x_0 \in X$ verifying that $T(x_0) = x_0$.

3. Outline of the second chapter

This chapter is devoted to the study of the of the Inverse Function and Implicit Function Theorems. First we treat the Inverse Function Theorem in the context of real functions of real variable. We previously show various results in relation to the monotony and continuity of the functions. For the analysis of the Inverse Function Theorem in the vector-valued context, we first approach some results in relation to the Mean Value Theorem for functions of several variables.
One of the central results we discuss is the Inverse Function Theorem in the vector-valued setting that we formulate in the following terms.

Inverse Function Theorem
Let $\Omega \subset \mathbb{R}^n$ an open subset, $f : \Omega \longrightarrow \mathbb{R}^n$ a function in $C^1(\Omega)$ and $x_0 \in \Omega$. Suppose that $Df(x_0)$ is invertible, that is, $\det(Df(x_0)) \neq 0$. Then, there exists an open subset $U \subset \Omega$ such that
(i) $x_0 \in U$;
(ii) $f(U)$ is an open subset of $\mathbb{R}^n$;
(iii) $f_{|U} : U \longrightarrow f(U)$ is bijective and $f_{|U}^{-1} \in C^1(f(U))$. Furthermore, $D((f_{|U})^{-1})(y) = (Df(x))^{-1}$, where $y \in f(U)$ and $f(x) = y$.



One of the consequences of this theorem involves holomorphic functions as follows.

Corollary Let $f : \Omega \subset \mathbb{C} \longrightarrow \mathbb{C}$ a holomorphic function, Ω an open set and $z_0 \in \Omega$. Suppose that $f'(z_0) \neq 0$. Then, there exists an open ser $U_0 \subset \Omega$ for which $z_0 \in U_0$, $f(U_0)$ is an open set in $\mathbb{C}$ and the function $f_{|U_0} : U_0 \longrightarrow f(U_0)$ is bijective, $(f_{|U_0})^{-1} : f(U_0) \longrightarrow U_0$ is holomorphic on $f(U_0)$ and $(f^{-1})'(w) = \frac{1}{f'(f^{-1}(w))}$, $w \in f(U_0)$.

The other fundamental topic we deal with in the memory is the Implicit Function Theorem.

Implicit Function Theorem
Given an open subset $\Omega \subset \mathbb{R}^n \times \mathbb{R}^k$ and

$$\begin{aligned} f : \mathbb{R}^n \times \mathbb{R}^m &\longrightarrow \mathbb{R}^m \\ (x, y) &\longmapsto f(x, y) = (f_1(x, y), ..., f_m(x, y)), \end{aligned}$$

such that $f \in C^k(\Omega)$ and $(a, b) \in \Omega$ such that $f(a, b) = 0$. If

$$\begin{vmatrix} \frac{\partial f_1}{\partial y_1}(a,b) & \cdots & \frac{\partial f_1}{\partial y_m}(a,b) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial y_1}(a,b) & \cdots & \frac{\partial f_m}{\partial y_m}(a,b) \end{vmatrix} \neq 0.$$

then, there exist open subsets $A \subset \mathbb{R}^n$, $B \subset \mathbb{R}^m$ and $g : A \longrightarrow B$ verifying the following properties:
(i) $a \in A, b \in B$ and $A \times B \subset \Omega$.
(ii) $f(x, g(x)) = 0$, $x \in A$.
(iii) For each $x \in A$ there exists a unique $y \in B$ such that $f(x, y) = 0$, being $y = g(x)$.
(iv) $g \in C^k(A)$.

References

[1] EDWARDS, C.H., Jr. *Advanced calculus of several variables*. Academic Press, Inc., New York, 1973.

[2] FLEMING, W. *Functions of several variables*. Springer-Verlag, New York, 1977.

[3] FLORES, M. y SADARANGANI, K. *Cálculo diferencial e integral de varias variables*. Servicio de Publicaciones ULL, 2011.

[4] RUDIN, W. *Principles of Mathematical Analysis*. 3rd edition. McGraw-Hill Inc., Singapore, 1976.

Sección de Matemáticas
Universidad de La Laguna