

Cursos de Grado

Fascículo 3

Comité Editorial:

Carlos Cabrelli (Director). *Departamento de Matemática, FCEyN, Universidad de Buenos Aires.*
E-mail: cabrelli@dm.uba.ar

Guillermo Cortiñas. *Departamento de Matemática, FCEyN, Universidad de Buenos Aires.*
E-mail: gcorti@dm.uba.ar

Claudia Lederman. *Departamento de Matemática, FCEyN, Universidad de Buenos Aires.*
E-mail: cllederma@dm.uba.ar

Auxiliar editorial:

Leandro Vendramin. *Departamento de Matemática, FCEyN, Universidad de Buenos Aires.*
E-mail: lvendramin@dm.uba.ar

ISSN 1851-1317 (Versión Electrónica)

ISSN 1851-1295 (Versión Impresa)

Derechos reservados

© 2010 Departamento de Matemática, Facultad de Ciencias Exactas y Naturales,

Universidad de Buenos Aires.
Departamento de Matemática
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires
Ciudad Universitaria - Pabellón I
(1428) Ciudad de Buenos Aires
Argentina.

<http://www.dm.uba.ar>

e-mail: secre@dm.uba.ar

tel/fax: (+54-11)-4576-3335

Este libro nació como unas notas para un curso de Análisis, y de a poco se fue convirtiendo en algo más que unas notas. Traté de conservar un estilo coloquial, sumado al rigor de definiciones y teoremas, sin evitar ningún tema, aún aquellos que suelen tener (inmerecida) fama de intratables. Pienso que todos los temas que aquí se estudian son accesibles para un estudiante que haya hecho un curso de cálculo en una variable real, y otro de álgebra lineal, si se abordan de la manera adecuada. En lo posible intenté evitar el uso de coordenadas, y aunque en muchos casos es necesario volver a ellas para dar un sentido concreto a las ideas geométricas, puede decirse que en general, con modificaciones menores, las pruebas se adaptan a un contexto más amplio. Esa fue mi intención al preparar las clases, y es la intención de este libro: hacer accesibles temas sutiles del análisis y el cálculo en una y varias variables reales, y vincular estos temas de manera intrínseca con la geometría del espacio, sin permitir que el uso de coordenadas oscurezca esta relación.

Si bien el libro contiene problemas que surgen naturalmente a lo largo de la presentación de los temas, estos no suplen una verdadera práctica donde los conceptos se trabajen a partir de ejemplos progresivamente más sofisticados. En este caso, las prácticas de la materia Análisis I (dictada por el Departamento de Matemática de la Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires) son el complemento ideal para estas notas, que fueron pensadas con estas prácticas en la mano. Se puede acceder a las mismas a través de la página del [Departamento de Matemática, cms.dm.uba.ar](http://cms.dm.uba.ar).

Agradecimientos:

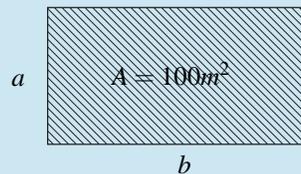
- A Sam por su entusiasmo y apoyo constante, a mis hijos por darle color a los grises.
- A Cristian Conde por hacer de editor y corrector en las sombras, en forma totalmente desinteresada.
- A Pablo Groisman porque cuando sólo había unas notas en borrador, me hizo notar que las notas podían tener algún valor para los alumnos.

Es supersticiosa y vana la costumbre de buscar sentido en los libros, equiparable a buscarlo en los sueños o en las líneas caóticas de las manos.

J. L. BORGES

En este libro queremos estudiar problemas que involucren varias variables o incógnitas, todas ellas números reales. Como aprendimos en cursos de cálculo elementales, lo más práctico es darle nombre a las variables y después pensar cuáles son las funciones que modelan el problema; estas funciones van a depender de esas variables. Es la estrategia de “nombrar y conquistar”.

Por ejemplo, si queremos saber qué forma debe tener un terreno rectangular de área 100m^2 , para gastar la menor cantidad de alambre en su borde, al perímetro lo llamamos P , el área A , y llamamos b a la base del rectángulo y a a su altura.



Se tiene $A = b \cdot a = 100$ y por otro lado $P = 2a + 2b$. Despejando de la ecuación del área nos queda una ecuación en una sola variable, $P(a) = 2a + 2 \cdot \frac{100}{a}$. Entonces aparece una función P , de variable real a , a la cual le buscamos el mínimo. Es importante entender el dominio de la función, que es $a > 0$. Por otro lado, para buscarle el mínimo, ¿qué hacemos? Simplemente le buscamos los extremos locales a P . ¡Y para eso hace falta calcular la derivada de P ! Hagámoslo: se tiene $P'(a) = 2 - 2 \frac{100}{a^2}$. Igualando a cero se tiene $a^2 = 100$, y como a debe ser un número positivo, $a = 10$, y en consecuencia $b = 10$. Es decir el terreno debe ser cuadrado.

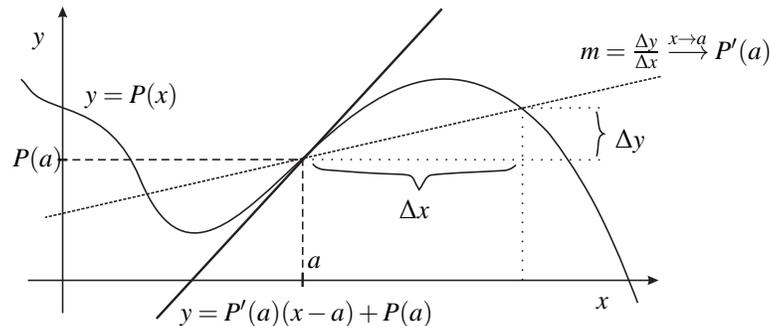
Vemos como para resolver un problema muy simple, aparecen herramientas que no son “obvias”. En primer lugar la noción de dominio, que requiere entender un poco el conjunto de los números reales. En segundo lugar, la noción de derivada, que si la pensamos un poquito era pensar en las rectas tangentes al gráfico de P , que a su vez se obtenían como límite de rectas secantes. Sí, apareció el límite. Y todo para probar que el terrenito era cuadrado.

En la próxima figura se representa el gráfico de una función P , y una recta secante al gráfico en el punto

$(a, P(a))$. La pendiente de esta recta se calcular como

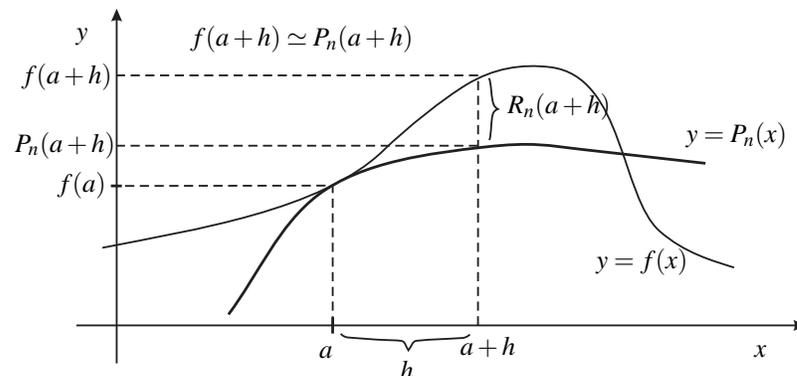
$$m = \frac{\Delta y}{\Delta x} = \frac{P(x) - P(a)}{x - a},$$

y en los casos en que esta cantidad tiene límite cuando $x \rightarrow a$, decimos que P es derivable en a y al número que nos da el límite lo llamamos $P'(a)$.



También sabemos que los límites nos dicen que pasa con el gráfico de la función en puntos “conflictivos”; que tipo de discontinuidad tiene, si tiene o no asíntotas. Por otro lado, la noción de derivada es muy útil para entender el comportamiento de la función, no necesariamente cerca de un máximo o un mínimo.

Otras cosas que aprendimos fueron que una función muy complicada se puede aproximar con la recta tangente para calcularla “cerca”, y también que usando el polinomio de Taylor, en muchos casos se la puede aproximar tanto como uno quiera, e incluso estimar el error que cometemos cuando usamos el polinomio en vez de la función original. En la siguiente figura se presenta el gráfico de f y se observa que el polinomio P_n es una buena aproximación de f para los x que están cerca de $x = a$.



Pues bien: todas estas herramientas se pueden generalizar a funciones de varias variables. Así, por ejemplo, de aquí a unos meses vamos a estar en condiciones de resolver el siguiente problema:

Notar que, en el problema que aquí se presenta, hay demasiadas variables (y pocas ecuaciones) como para intentar el truco de despejar y llevar todo el problema a una sola variable. Volveremos a este problema en el Capítulo 5, cuando dispongamos de otras herramientas. Mientras tanto, invitamos al lector a resolverlo usando argumentos de simetría en las variables a, l, p que lo modelan.

Queremos armar una caja rectangular, tal que la suma del ancho, el largo y la profundidad de la caja no excedan los tres metros ($a + l + p \leq 300cm$). ¿Cuál es el máximo volumen que puede tener la caja, y cuáles serán sus dimensiones?

Por el camino, van a surgir ideas, herramientas y ejemplos que tienen valor por sí mismos, y que no solamente van a servir para resolver problemas de máximos y mínimos como el de arriba. Van a surgir problemas que en principio sólo tienen que ver con la formulación de estas definiciones, es decir, problemas intrínsecos de la matemática. ¡Comencemos!

Índice general	IX
1. Cálculo en \mathbb{R}^n	1
1.1. Números reales	1
1.1.1. Axiomas	2
1.1.2. Sucesiones	8
1.2. El espacio como n-uplas	12
1.2.1. Distancia	13
1.2.2. Producto escalar	14
1.2.3. Entornos y conjuntos abiertos	16
1.2.4. Límites en \mathbb{R}^n	22
1.2.5. Puntos de acumulación y conjuntos cerrados	24
1.2.6. Conjuntos acotados	27
1.3. Problemas	28
2. Funciones	31
2.1. Dominio, gráfico, imagen	31
2.1.1. Funciones $f : \mathbb{R}^n \rightarrow \mathbb{R}$	31
2.1.2. Curvas	32
2.1.3. Gráfico, curvas y superficies de nivel	33
2.1.4. Límite	36
2.2. Funciones $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$	40
2.2.1. Composición	40
2.2.2. Curvas y el límite	42
2.2.3. Límite y sucesiones	43
2.3. Continuidad	45
2.3.1. Propiedades de las funciones continuas	45
2.3.2. Caracterización de las funciones continuas	48

2.3.3. Continuidad uniforme	48
2.4. Problemas	49
3. Derivadas y Diferencial	51
3.1. Derivadas	51
3.1.1. Curvas	53
3.1.2. Derivadas direccionales	55
3.1.3. Derivadas parciales	58
3.2. Plano tangente y Diferencial	59
3.2.1. Unicidad de la diferencial	64
3.2.2. Álgebra de funciones diferenciables	67
3.2.3. Repaso de los teoremas en una variable	70
3.2.4. Criterio de diferenciabilidad	73
3.2.5. Funciones $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$	74
3.3. Teoremas de Lagrange y Fermat en \mathbb{R}^n	80
3.4. NOTAS	81
4. Función inversa e implícita	85
4.1. Función Inversa	85
4.1.1. Funciones de clase C^k	86
4.1.2. Transformaciones lineales	87
4.2. Superficies de nivel y funciones implícitas	93
4.2.1. Función implícita en \mathbb{R}^n	98
4.3. NOTAS	105
5. Taylor y extremos	107
5.1. Polinomio de Taylor	107
5.1.1. Varias variables	108
5.1.2. Demostración alternativa de la fórmula de Taylor de orden 2 en \mathbb{R}^n	115
5.2. Extremos	117
5.2.1. Formas cuadráticas	118
5.2.2. El Hessiano y los extremos	122
5.3. Extremos con restricciones, multiplicadores de Lagrange	127
5.3.1. Extremos en una región	127
5.3.2. Extremos en regiones con borde que se puede parametrizar	127
5.3.3. Multiplicadores de Lagrange	129
5.3.4. Multiplicadores en \mathbb{R}	131

5.3.5. Un ejemplo elemental	132
5.3.6. Varias ligaduras	133
6. Integrales en \mathbb{R}	135
6.1. Integrales	135
6.1.1. Propiedades	141
6.2. Integrales impropias	146
6.2.1. Intervalos infinitos	148
6.2.2. Convergencia condicional y absoluta	149
6.2.3. Criterios de comparación	150
6.3. NOTAS	157
7. Integrales múltiples	159
7.1. Integrales en el plano	159
7.1.1. Rectángulos y particiones	159
7.1.2. Integral de Riemann	162
7.1.3. Integrales iteradas y el Teorema de Fubini	166
7.2. Integrales en \mathbb{R}^3 y \mathbb{R}^n	172
7.3. Propiedades	177
7.3.1. Medida de una región y Teorema del valor medio	178
7.4. NOTAS	180
8. Teorema de cambio de variables	183
8.1. El método de sustitución	183
8.2. Particiones generales	183
8.3. Transformaciones lineales	185
8.4. Cambio de variable	187
8.4.1. Cambio de variable en \mathbb{R}^3	193
Bibliografía	201

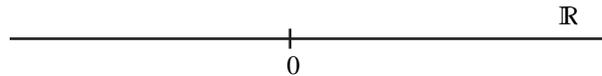
¡No soy un número! ¡Soy un hombre libre!

El Prisionero

1.1. Números reales

Lo primero que notamos en la charla del prefacio, es que es relevante entender cómo es el dominio de una función, que es un subconjunto de \mathbb{R} ; y si vamos a hablar de límites y derivadas, es bueno entender un poco mejor los números reales.

Observación 1.1.1. *¿Qué es un número real? Esta pregunta inocente es más difícil de contestar de lo que parece. Una primer aproximación intuitiva (o que aprendimos en la escuela) nos dice que es el conjunto de los números que ocupan la recta numérica, es decir, hay una correspondencia entre los números reales y los elementos de la recta.*

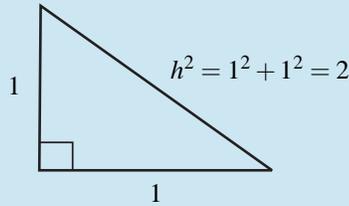


Pero esto trae otras preguntas, como por ejemplo qué es una recta y si todas las rectas son iguales o al menos equivalentes en algún sentido, y cómo es la tal correspondencia. Esto no quiere decir que la idea la tengamos que descartar, ¡por el contrario! esta idea es fundamental para poder imaginar mentalmente los números y los problemas que sobre ellos consideremos.

Si repasamos un poco lo aprendido, veremos que lo más importante no es qué son, sino que propiedades tienen. En realidad, desde un punto de vista más formal, uno podría pensar que la pregunta ¿qué son? se contesta haciendo una lista de las propiedades más importantes, y viendo que si un conjunto cumple estas propiedades entonces es \mathbb{R} . Pero ¿quién garantiza que hay algún conjunto que cumpla esas propiedades en abstracto? Si alguien dice "el conjunto que las verifica es \mathbb{R} ", pues yo le contestaría ¡que no lo puede poner como ejemplo si todavía no nos pusimos de acuerdo en qué es un número real!

Repasemos: los números naturales $\mathbb{N} = \{1, 2, 3, \dots\}$ son todos números reales. Los números enteros (que se obtienen tomando los inversos de los naturales respecto de la suma y agregando el neutro 0 de la suma), $\mathbb{Z} = \{\dots, -2, 1, 0, 1, 2, 3, \dots\}$ son todos reales. Si uno considera cocientes de números enteros (algo bastante natural al pensar en la división), obtiene los racionales $\mathbb{Q} = \{p/q : p, q \in \mathbb{Z}, q \neq 0\}$.

Sabemos que con esto NO alcanza. Por ejemplo en un triángulo de catetos unitarios, la hipotenusa es $h = \sqrt{2}$, que no es ningún número racional.



Esta última afirmación requiere una prueba: supongamos que $\sqrt{2}$ es racional. Entonces existen p, q números naturales tales que $\sqrt{2} = p/q$. Vamos a suponer que cancelamos todos los factores posibles de p y q de manera que no tengan ningún factor común. Elevando al cuadrado,

$$2 = \frac{p^2}{q^2},$$

o equivalentemente

$$2q^2 = p^2.$$

Ahora pensemos cuántas veces aparece el factor 2 en la factorización de p^2 . Como $p^2 = p \cdot p$, el factor 2 aparece un número par de veces en la factorización de p^2 , pues aparece el doble de veces que en la factorización de p . En el caso en el que no aparece (o sea si p es impar) esta afirmación sigue siendo cierta pues vamos a adoptar la convención de que 0 es un número par (pues se puede escribir como $0 = 2 \cdot 0$). Entonces a la derecha de la igualdad, tenemos un número par de factores 2. El mismo razonamiento aplicado a q^2 nos dice que a la izquierda el factor 2 aparece un número impar de veces. Esto es imposible, lo que prueba que $\sqrt{2}$ no es racional.

Todo número real que no es racional se denomina *irracional*, y al conjunto de números irracionales se lo suele denotar con la letra \mathbb{I} . Se tienen las obvias inclusiones estrictas

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}.$$

Observación 1.1.2. Como curiosidad, los irracionales también se pueden clasificar como algebraicos o trascendentes. Los primeros son los que son ceros de polinomios con coeficientes enteros, por ejemplo $\sqrt{2}$ es algebraico puesto que es raíz del polinomio $p(x) = x^2 - 2$. Ejemplos de números trascendentes son e y π . Por supuesto, ¡probar que estos números NO son algebraicos tampoco es elemental! Para una prueba a sólo un clic puede verse la Wikipedia ¹ (en inglés).

1.1.1. Axiomas

Volvamos al asunto de las cosas que hacen que \mathbb{R} sea \mathbb{R} . Veamos primero los axiomas que hacen de \mathbb{R} un cuerpo ordenado. Esto es, propiedades que no se demuestran sino que se suponen válidas.

¹http://en.wikipedia.org/wiki/Transcendental_number

Entre triviales y excesivamente abstractos para nuestra intuición, son la base sobre la que se construyen todos los resultados del análisis.

Axiomas de la suma:

- S1) Conmutatividad: si $a, b \in \mathbb{R}$, $a + b = b + a$.
- S2) Asociatividad: si $a, b, c \in \mathbb{R}$, $a + (b + c) = (a + b) + c$.
- S3) Elemento Neutro: existe y lo llamamos "0". Si $a \in \mathbb{R}$, $a + 0 = a$.
- S4) Inverso Aditivo: si $a \in \mathbb{R}$, existe un inverso aditivo, lo llamamos " $-a$ ", verifica $a + (-a) = 0$.

En general escribimos $a - b$ en vez de $a + (-b)$.

Axiomas del producto:

- P1) Conmutatividad: si $a, b \in \mathbb{R}$, $a \cdot b = b \cdot a$.
- P2) Asociatividad: si $a, b, c \in \mathbb{R}$, $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.
- P3) Elemento Neutro: existe y lo llamamos "1". Si $a \in \mathbb{R}$, $1 \cdot a = a$.
- P4) Inverso Multiplicativo: si $a \in \mathbb{R}$, $a \neq 0$, existe un inverso multiplicativo, lo llamamos " a^{-1} ", verifica $a \cdot a^{-1} = 1$.

En general omitimos el punto, es decir escribimos $a \cdot b = ab$. También es usual escribir $\frac{1}{a} = a^{-1}$.

Propiedad distributiva:

- D) Distributividad: si $a, b, c \in \mathbb{R}$, $a(b + c) = ab + ac$.

Axiomas del orden:

- O1) Tricotomía: si $a, b \in \mathbb{R}$, hay sólo tres posibilidades (excluyentes), que son $a < b$, $a = b$, $a > b$.
- O2) Transitividad: si $a, b, c \in \mathbb{R}$ son tales que $a < b$ y $b < c$, entonces $a < c$.

O3) Monotonía de la suma: Si $a, b, c \in \mathbb{R}$ son tales que $a < b$, entonces $a + c < b + c$.

O4) Monotonía del producto: si $a, b, c \in \mathbb{R}$, son tales que $a < b$ y $c > 0$, entonces $ac < bc$.

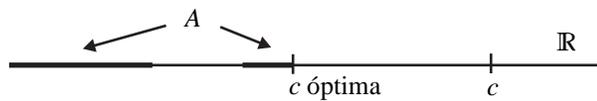
En particular dados dos números reales distintos siempre hay un tercero en el medio, y siguiendo este razonamiento se ve que hay infinitos. Si $a, b \in \mathbb{Q}$, el promedio está en \mathbb{Q} y entonces se ve que hay infinitos racionales. Estos hechos los usaremos sin otra aclaración.

Se tiene $0 < 1$ ¡justificarlo con los axiomas! Se deduce fácilmente que si $a < b$, entonces el promedio $p = \frac{a+b}{2}$ es un número real tal que $a < p < b$.

Algo más sutiles, son las pruebas de que entre dos reales siempre hay un racional, y de que entre dos racionales siempre hay un irracional, pruebas que dejaremos para el Corolario 1.1.12.

Hasta aquí las propiedades que hacen de \mathbb{R} un cuerpo ordenado. El problema es que no es el único. Es decir, \mathbb{Q} cumple todos estos axiomas. Lo que falta es el axioma de completitud, es decir, ver que no hay agujeros en \mathbb{R} , cuando por ejemplo si los hay en \mathbb{Q} .

Intentamos formalizar un poco. ¿Está claro que si $A \subset \mathbb{R}$ es no vacío, y tiene un tope por encima, entonces podemos buscar el tope más chiquito? Pensarlo en la recta: indicamos al conjunto A con líneas mas gruesas



Un tope por encima es una *cota superior*. Es decir,

Definición 1.1.3. Una *cota superior* es un número real M tal que para todo $a \in A$, se tiene $a \leq M$.

Si hay una cota superior, hay infinitas (cualquiera más grande sirve). El asunto es encontrar la óptima (la más pequeña). Esta se denomina *supremo* del conjunto A , y se anota $\sup A$. Es decir,

Definición 1.1.4. $s \in \mathbb{R}$ es el *supremo* de A si para todo $a \in A$, $a \leq s$ y además si s' es otra *cota superior*, $s \leq s'$.

Lo relevante, que hay que tener en cuenta en esta discusión es el axioma de completitud de los números reales, que dice que

Axioma. Si $A \subset \mathbb{R}$ es no vacío, y está acotado superiormente, entonces A tiene supremo s .

Se dice que " \mathbb{R} es completo en el orden". Algo obvio:

Proposición 1.1.5. El supremo de un conjunto es único.

Demostración. Si hay otro, s' , se tiene por un lado $s' \leq s$ (pues s es cota superior y s' es la menor cota superior), y por otro lado $s \leq s'$ (pues s también es la menor cota superior). \square

Análogamente se define *cota inferior* e *ínfimo*,

Definición 1.1.6. Una cota inferior es un número real m tal que para todo $a \in A$, se tiene $a \geq m$. El número $i \in \mathbb{R}$ es el ínfimo de A si para todo $a \in A$, $a \geq i$ y además si i' es otra cota inferior, $i \geq i'$. Es decir el ínfimo es la más grande de las cotas inferiores, y también si existe es único.

Definición 1.1.7. Un conjunto $A \subset \mathbb{R}$ se dice acotado si está acotado tanto superiormente como inferiormente.

El supremo (o el ínfimo) puede pertenecer o no al conjunto. Hagamos algunos ejemplos con cuidado, por más que parezcan obvios. Queda como ejercicio para el lector el siguiente problema, cuyo resultado usaremos de aquí en más:

Problema 1.1.8. Probar que si $r \in \mathbb{R}$, $r > 0$ y $r^2 = 3$, entonces $r \notin \mathbb{Q}$. Es decir, probar que $\sqrt{3}$ es irracional. Se sugiere ver la prueba de que $\sqrt{2}$ no es racional.

Con este resultado a mano, podemos hallar el supremo del siguiente conjunto.

Ejemplo 1.1.9. Hallar el supremo s del conjunto $A = \{x \in \mathbb{R} : x^2 < 3\}$.

El candidato es $s = \sqrt{3}$. Notemos que s es cota superior pues si $x \in A$, entonces $x^2 < 3$ implica $|x| < \sqrt{3}$, luego $x \leq |x| < \sqrt{3}$. ¿Cómo probamos que es la menor? Por el absurdo: supongamos que hay otra cota superior más pequeña, digamos $s' < \sqrt{3}$. Ahora entre s' y s hay al menos un número real a (hay infinitos). Se tiene $a < s = \sqrt{3}$, luego $a^2 < 3$, es decir $a \in A$. Por otro lado $s' < a$, con lo cual s' no sería cota superior. El absurdo proviene de suponer que hay otra cota superior más pequeña que s .

Observación 1.1.10. Notar que si uno busca el supremo del conjunto $A = \{x \in \mathbb{Q} : x^2 < 3\}$, el supremo va a ser nuevamente $s = \sqrt{3}$, ¡que no es un número racional! Esto nos dice que la propiedad de completitud es inherente a los números reales. Es decir, hay subconjuntos de \mathbb{Q} que están acotados superiormente en \mathbb{Q} , pero que NO tienen supremo en \mathbb{Q} .

Demostremos mejor esta última observación. Para ello necesitamos algunas herramientas técnicas.

Proposición 1.1.11. Arquimedianeidad o principio de Arquímedes: Si x es un número real, entonces existe $n \in \mathbb{N}$ tal que $n > x$. Permite comparar cualquier número con un natural. Se puede deducir del axioma de completitud.

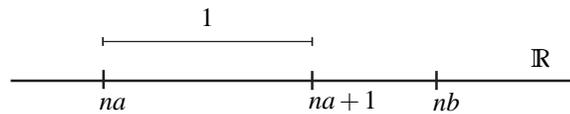
Demostración. Sea $x \in \mathbb{R}$, y consideremos $X = \{k \in \mathbb{N} : k \leq x\}$. Si $X = \emptyset$, en particular $1 > x$, luego se puede tomar $n = 1$. Si X no es vacío, como está acotado superiormente (por x), tiene un supremo $s = \sup X \in \mathbb{R}$. Como $s - 1$ no puede ser cota superior de X (pues s es la más chica), debe existir $k_0 \in X$ tal que $s - 1 < k_0$. Tomando $n = k_0 + 1$ se tiene $n \in \mathbb{N}$ y además $n > s$. Por esto último $n \notin X$, lo que nos dice que $n > x$. \square

Con esta herramienta podemos probar algo que todos sabemos que es cierto. En nuestra formación inicial y media, aprendimos algunos hechos supuestamente útiles sin ninguna pista del por qué de su validez (¡o el por qué de su relevancia!).

Corolario 1.1.12. *Entre dos números reales distintos siempre hay un número racional. Entre dos números reales distintos siempre hay un número irracional.*

Demostración. Sean $a < b$ números reales (elegimos los nombres para que queden en ese orden, es decir, llamamos a al más chico y b al más grande). Entonces $b - a > 0$, y en consecuencia $\frac{1}{b-a} > 0$. Existe por el principio de Arquímedes un número natural n tal que $n > \frac{1}{b-a}$. Observemos que

$$nb = na + n(b - a) > na + 1.$$



En consecuencia, tiene que existir un número entero k entre nb y na . Es decir, existe $k \in \mathbb{Z}$ tal que $na < k < nb$. Dividendo por n se obtiene $a < \frac{k}{n} < b$, que nos dice que $\frac{k}{n} \in \mathbb{Q}$ está entre a y b .

Ahora veamos que entre dos números reales siempre hay un irracional. Vamos a volver a suponer que $a < b$. Multiplicando por $\frac{1}{\sqrt{2}}$ obtenemos $\frac{1}{\sqrt{2}}a < \frac{1}{\sqrt{2}}b$. Por lo anterior existe un número racional $\frac{k}{n}$ entre ellos, es decir $\frac{1}{\sqrt{2}}a < \frac{k}{n} < \frac{1}{\sqrt{2}}b$. Multiplicando por $\sqrt{2}$ se obtiene $a < \frac{k}{n}\sqrt{2} < b$. Afirmamos que este número del medio es irracional: si fuera racional, existirían j, m enteros tales que $\frac{k}{n}\sqrt{2} = \frac{j}{m}$. Pero entonces $\sqrt{2} = \frac{jn}{mk}$ sería racional, lo cual es falso, como ya probamos. \square

Ejemplo 1.1.13. *Sea $A = \{x \in \mathbb{Q} : x^2 < 2 \text{ y } x > 0\}$. Entonces $A \subset \mathbb{Q}$ es no vacío y está acotado superiormente en \mathbb{Q} , pero no tiene supremo en \mathbb{Q} .*

Demostración. Que es no vacío es evidente pues $1 \in A$. Por otro lado, $C = 2 \in \mathbb{Q}$ es una cota superior de A pues si $x \in A$ entonces $x^2 < 2 < 4$, con lo cual $|x| < 2$, y como x es positivo, se tiene $x < 2$. Por último, supongamos que existe $s \in \mathbb{Q}$ tal que $s = \sup A$. Es decir, s es una cota superior, y es la más pequeña de ellas. Hay tres posibilidades: la primera $s = \sqrt{2}$ es imposible pues $\sqrt{2}$ no es racional. La segunda es que $s < \sqrt{2}$ (y podemos suponer $s \geq 1$ pues $1 \in A$). Pero si tomamos r un número racional tal que $s < r < \sqrt{2}$, llegamos a un absurdo pues elevando al cuadrado obtenemos $r^2 < 2$, y como r es positivo (por ser mayor que s), se tiene $r \in A$. Esto contradice que s es cota superior de A . La última posibilidad es $s > \sqrt{2}$. Pero entonces volvemos a elegir $t \in \mathbb{Q}$ tal que $\sqrt{2} < t < s$, y este número es una cota superior de A en \mathbb{Q} más pequeña que s , lo cual es imposible. \square

Corolario 1.1.14. *El cuerpo \mathbb{Q} no es completo en el orden (no verifica el axioma de completitud).*

Ejemplo 1.1.15. *Hallar el supremo s del conjunto $A = \{1 - \frac{1}{n} : n \in \mathbb{N}\}$. ¿ $s \in A$?*

Aquí el candidato natural es $s = 1$ (que NO es un elemento de A). Es una cota superior pues $1 - \frac{1}{n} \leq 1$ para todo n natural. Falta ver que es la menor. De nuevo: si hubiera otra cota superior, digamos $s' < 1$, tomemos $n_0 \in \mathbb{N}$ tal que $n_0 > \frac{1}{1-s'}$ que siempre se puede por el principio de Arquímedes. Entonces despejando se obtiene $s' < 1 - \frac{1}{n_0}$ lo que contradice que s' sea cota superior.

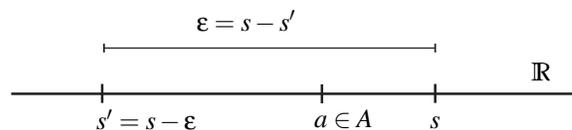
Observación 1.1.16. *Puede probarse que si A es conjunto que es un cuerpo ordenado (cumple los axiomas de arriba S, P, D, O), y A es completo en el orden, entonces A ES \mathbb{R} (o más bien se lo puede identificar con \mathbb{R} respetando las operaciones y el orden)².*

Algo que quedó en el tintero y que no vamos a tratar aquí, es como se pueden construir (si, construir a partir de los números racionales) los números reales, para de esa forma estar seguros de que hay un conjunto que verifica todo lo que queremos. ¡Recordemos que nunca dijimos quién es \mathbb{R} ! Hay varias maneras de hacer esta construcción, por ejemplo usando cortaduras de Dedekind o sucesiones de Cauchy; el lector insaciable con acceso a Internet puede leer sobre estas construcciones en la Wikipedia³, aunque hay otras fuentes en papel mejor escritas de donde leerlo, como por ejemplo, el libro de Birkhoff-McLane de Álgebra [1].

Volvamos al supremo de un conjunto $A \subset \mathbb{R}$. Una caracterización útil es la siguiente, en términos de ε (algo habitual en las definiciones de límite)

Proposición 1.1.17. *s es el supremo de A si y sólo si*

1. $s \geq a$ para todo $a \in A$
2. dado $\varepsilon > 0$, existe $a \in A$ tal que $s - \varepsilon < a$.



Demostración. Empecemos con la definición vieja, entonces obviamente se cumple 1). Por otro lado, dado $\varepsilon > 0$, se tienen $s - \varepsilon < s$ y esto nos dice que $s - \varepsilon$ no es cota superior de A (recordemos que s era la más chica). O sea, existe algún $a \in A$ tal que $s - \varepsilon < a$, lo cual prueba 2).

Si partimos de esta definición, tenemos que probar que un s que verifica 1) y 2) es el supremo. Claramente s es cota superior, falta probar que es la más chica. Si hubiera una cota superior más chica, digamos s' , entonces tomamos $\varepsilon = s - s' > 0$ y llegamos a un absurdo pues 2) nos dice que existe $a \in A$ tal que $s - (s - s') < a$, es decir, $s' < a$ lo que contradice que s' es cota superior. \square

²<http://planetmath.org/?op=getobj&from=objects&id=5607>

³http://en.wikipedia.org/wiki/Construction_of_real_numbers

Análogamente, se tiene el siguiente resultado que dejamos como ejercicio al lector.

¿Qué se puede decir de un conjunto $A \subset \mathbb{R}$ tal que $\inf(A) = \sup(A)$?

¿Y si $B \subset \mathbb{R}$ es tal que $\inf(A) \leq \inf(B)$ y $\sup(B) \leq \sup(A)$?

Problema 1.1.18. i es el ínfimo de A si y sólo si

1. $i \leq a$ para todo $a \in A$
2. dado $\varepsilon > 0$, existe $a \in A$ tal que $i + \varepsilon > a$.

1.1.2. Sucesiones

Una sucesión se puede pensar como un cierto conjunto de números reales, ordenado de alguna manera. Más precisamente, es una función $s : \mathbb{N} \rightarrow \mathbb{R}$, que en general se anota s_n en vez de $s(n)$. Es decir, si ponemos $s_n = 1/n$, lo que queremos nombrar es la sucesión cuyos primeros términos son $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots$ y así sucesivamente. Es importante observar que no es lo mismo que la sucesión $1, \frac{1}{3}, \frac{1}{2}, \frac{1}{5}, \frac{1}{4}, \dots$ (en otro orden). Estas sucesiones tienen todos sus términos distintos. Pero esto no tiene por qué ser así. Por ejemplo

$$1, 2, 1, 2, 1, 2, \dots$$

toma sólo dos valores distintos, y como caso extremo

$$3, 3, 3, 3, 3, 3, \dots$$

es una sucesión que se denomina *sucesión constante*. Las sucesiones son útiles para describir problemas de aproximación. Así, si quiero descubrir que pasa con $1/x$ cuando x tiende a infinito, basta mirar algunos términos como $1; \frac{1}{10} = 0,1; \frac{1}{100} = 0,01$ para descubrir que tiende a cero.

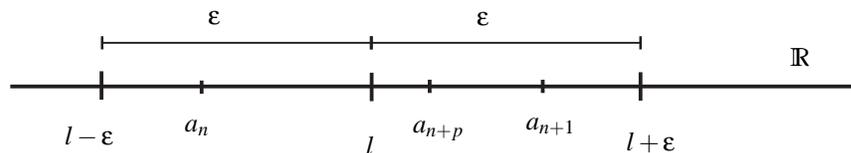
Definición 1.1.19. Se dice que una sucesión a_n tiende al número l cuando $n \rightarrow \infty$, si dado $\varepsilon > 0$ existe $n_0 \in \mathbb{N}$ tal que si $n \geq n_0$, entonces $|a_n - l| < \varepsilon$. En ese caso escribimos

$$\lim_{n \rightarrow \infty} a_n = l \text{ o bien } a_n \xrightarrow{n \rightarrow \infty} l.$$

Es decir que a_n tiende a l si todos los términos de a_n están tan cerca de l como uno quiera a partir de un término dado. Recordemos que $|a_n - l| < \varepsilon$ quiere decir

$$l - \varepsilon < a_n < l + \varepsilon,$$

como se indica en la figura:



La condición de que todos los términos de la sucesión estén suficientemente cerca del límite a partir de un n_0 se simplifica cuando sabemos *a priori* que todos los términos están a un lado de una constante dada. Dejamos el siguiente problema que pone de manifiesto esta simplificación.

Problema 1.1.20. Si $a_n \leq l$ para todo $n \in \mathbb{N}$, para probar que $\lim_{n \rightarrow \infty} a_n = l$ basta probar que $l - a_n < \varepsilon$. Análogamente, si $a_n \geq l$, basta probar que $a_n - l < \varepsilon$.

Para una exposición más detallada (pero todavía elemental) sobre sucesiones, puede verse el libro de Noriega [5], o la guía de Análisis del CBC de la UBA. Nosotros vamos a ir directamente a lo que nos interesa, que para empezar es una caracterización del supremo.

Definición 1.1.21. Recordemos que una sucesión es creciente si

$$a_1 \leq a_2 \leq a_3 \leq \dots,$$

y estrictamente creciente si

$$a_1 < a_2 < a_3 < \dots.$$

Análogamente se define sucesión decreciente y estrictamente decreciente. Una sucesión se dice monótona si es creciente o decreciente, y estrictamente monótona si es estrictamente creciente o decreciente.

Como vimos, todo conjunto no vacío $A \subset \mathbb{R}$, si está acotado superiormente, tiene supremo. Se tiene el siguiente resultado útil que relaciona el supremo con las sucesiones:

Proposición 1.1.22. Si $s = \sup A$, entonces existe una sucesión creciente $\{a_n\}$ de elementos de A tal que $a_n \xrightarrow{n \rightarrow \infty} s$. Si $s \notin A$, se puede elegir a_n estrictamente creciente.

Demostración. Si $s \in A$, basta tomar $a_n = s$ la sucesión constante (que es creciente). Si no, tomamos algún elemento $\bar{a}_1 \in A$, y ponemos $a_1 := \bar{a}_1$. Ahora consideramos $\bar{a}_2 = \frac{a_1 + s}{2}$ el promedio, que es un número menor estrictamente menor que s . Existe entonces algún elemento $a_2 \in A$ tal que $\bar{a}_2 < a_2 < s$ (sino \bar{a}_2 sería cota superior de A , lo cual es imposible pues $\bar{a}_2 < s$).

Así vamos definiendo \bar{a}_n como el promedio $\frac{a_{n-1} + s}{2}$, y tomamos como a_n cualquier elemento de A que está entre \bar{a}_n y s . Ciertamente a_n es creciente pues $a_n > \bar{a}_n > a_{n-1}$, y por otro lado $0 < s - a_n < \frac{s - a_{n-1}}{2}$. Luego, repitiendo esta acotación se consigue bajar hasta a_1 , y por cada vez que bajamos, aparece un número 2 dividiendo, con lo cual obtenemos.

$$0 < s - a_n < \frac{s - a_{n-1}}{2} < \dots < \frac{s - a_1}{2^{n-1}}.$$

Esto último prueba que $\lim_{n \rightarrow \infty} a_n = s$, usando la definición de límite, pues $s - a_1$ está fijo. \square

Dejamos como ejercicio para el lector, un resultado análogo para el ínfimo de un conjunto. Recordar que es clave verificar que *todos los términos* a_n de la sucesión deben ser elementos del conjunto A .

Problema 1.1.23. Si $i = \inf A$, entonces existe una sucesión decreciente $\{a_n\}$ de elementos de A tal que $a_n \rightarrow i$. Si $i \notin A$, se puede elegir a_n estrictamente decreciente.

Subsucesiones

Una herramienta extremadamente útil son las subsucesiones. Recordemos que son simplemente una elección de un subconjunto de la sucesión original, provisto del mismo orden que tenía la sucesión original. Así por ejemplo, si $a_n = \frac{1}{n}$, una subsucesión es $b_k = \frac{1}{2k}$, es decir, tomamos

$$\frac{1}{4}, \frac{1}{6}, \frac{1}{8}, \dots$$

que son algunos de los términos de la sucesión original. Para dar más precisión a esta idea, recordemos que una sucesión es simplemente una función $s: \mathbb{N} \rightarrow \mathbb{R}$.

Definición 1.1.24. Sea $a: \mathbb{N} \rightarrow \mathbb{R}$ una sucesión. Una subsucesión de a es una nueva sucesión $b: \mathbb{N} \rightarrow \mathbb{R}$ que se puede construir a partir de la original de la siguiente manera: existe una función estrictamente creciente $j: \mathbb{N} \rightarrow \mathbb{N}$ tal que para todo $k \in \mathbb{N}$, $b(k) = a \circ j(k)$. Más coloquialmente: si

$$a_1, a_2, \dots, a_n, \dots$$

es una sucesión, entonces una subsucesión $\{a_{n_k}\}$ de $\{a_n\}$ es una sucesión de la forma

$$a_{n_1}, a_{n_2}, \dots, a_{n_k}, \dots$$

donde $n_1 < n_2 < \dots < n_k < \dots$.

Observación 1.1.25. Se puede seguir esta idea un paso más: una subsucesión de una subsucesión se define en forma análoga. Pero si lo pensamos un poco, nos hemos quedado con un conjunto aún más reducido de la sucesión original, y los índices deben seguir siendo estrictamente crecientes. Luego una subsucesión de una subsucesión de $\{a_n\}$ no es más que otra subsucesión de la sucesión original $\{a_n\}$ (y si todavía no conseguí marearlos es porque lo dije muy despacito).

Una consecuencia de la Proposición 1.1.22, es que toda sucesión creciente y acotada tiene límite, el cual coincide con el supremo:

Proposición 1.1.26. Si $\{a_n\}$ es una sucesión creciente ($a_{n+1} \geq a_n \geq \dots \geq a_1$) y acotada superiormente, entonces si $s = \sup\{a_n\}$, se tiene $\lim_{n \rightarrow \infty} a_n = s$.

Demostración. Consideremos el conjunto de puntos que forma la sucesión, es decir $A = \{a_n\}$. Por la Proposición 1.1.22, existe una sucesión dentro de A que es creciente y tiende a $s = \sup A$. Pero esto es lo que llamamos una subsucesión, es decir la nueva sucesión consta de algunos de los elementos de $\{a_n\}$, y la anotamos $\{a_{n_k}\}$. En consecuencia, dado $\varepsilon > 0$, existe k_0 tal que $k \geq k_0$ implica $s - a_{n_k} < \varepsilon$. Tomando $n_0 = n_{k_0}$ se observa que si $n \geq n_0$, entonces por ser a_n creciente se tiene

$$s - a_n \leq s - a_{n_0} = s - a_{n_{k_0}} < \varepsilon,$$

que es lo que queríamos probar. □

Se tiene un resultado análogo con el ínfimo de una sucesión decreciente y acotada inferiormente, cuya prueba una vez más queda como ejercicio.

Proposición 1.1.27. Si $\{a_n\}$ es una sucesión decreciente ($a_{n+1} \leq a_n \leq \dots \leq a_1$) y acotada inferiormente, entonces si $i = \inf\{a_n\}$, se tiene $\lim_{n \rightarrow \infty} a_n = i$.

Volvamos a pensar en sucesiones que no son monótonas.

Hay algo bastante evidente: si una sucesión no está acotada (por ejemplo, no está acotada superiormente), entonces podemos extraer de ella una subsucesión creciente. De la misma manera, si no está acotada inferiormente, podemos extraer una sucesión decreciente.

Lo que sigue es un resultado clave que nos va a ser muy útil más adelante, que generaliza esta idea a cualquier sucesión de números reales. Su demostración es un tanto técnica (paciencia).

Proposición 1.1.28. Sea $\{a_n\}$ una sucesión de números reales. Entonces existe una subsucesión $\{a_{n_k}\}$ de $\{a_n\}$ que es monótona.

Demostración. Para armar la subsucesión, nos interesan los términos de la sucesión que son más grandes que todos los términos que les siguen, es decir, los a_n tales que $a_n \geq a_k$ para todo $k \geq n$. Los vamos a llamar términos **dominantes**. Hay dos casos: hay infinitos términos dominantes o hay finitos. Pensemos primero el caso en el que la sucesión tiene infinitos términos dominantes

$$a_{n_1}, a_{n_2}, a_{n_3}, \dots \text{ con } n_1 < n_2 < n_3 < \dots$$

Por ser todos dominantes, $a_{n_1} \geq a_{n_2} \geq a_{n_3} \geq \dots$. Hemos construido una subsucesión decreciente.

Veamos ahora el caso en el que hay finitos términos dominantes (incluyendo la posibilidad de que no haya ninguno). Por ser finitos, hay uno que es el último (es decir, de allí en adelante ningún término de la

sucesión es dominante). Tomemos a_{n_1} el primer término de la sucesión que viene después de este último dominante (si no hay ningún término dominante en la sucesión tomamos $n_1 = 1$). Existe entonces más adelante algún $n_2 > n_1$ tal que $a_{n_2} > a_{n_1}$ (¡existe sino a_{n_1} sería dominante!). Así siguiendo, vamos hallando $n_k > \dots > n_3 > n_2 > n_1$ tales que $a_{n_k} > \dots > a_{n_3} > a_{n_2} > a_{n_1}$. En este caso hemos construido una subsucesión creciente. \square

Se obtiene así un corolario fundamental, que luego generalizaremos a \mathbb{R}^n .

Corolario 1.1.29 (Bolzano-Weierstrass). *Sea $\{a_n\}$ una sucesión acotada de números reales. Entonces existe una subsucesión $\{a_{n_k}\}$ de $\{a_n\}$ que es convergente.*

Demostración. Por la proposición previa, podemos extraer de $\{a_n\}$ una subsucesión monótona (que está acotada por estar acotada la original). Por la Proposición 1.1.26, esta subsucesión tiene a su vez una subsucesión convergente, que según lo observado antes, no es más que una subsucesión de la sucesión original. \square

Por último recordemos qué quiere decir que una sucesión “tiende a infinito”.

Definición 1.1.30. *Una sucesión $\{a_n\}$ de números reales tiende a más infinito si para todo $M > 0$ existe $n_0 \in \mathbb{N}$ tal que $n \geq n_0$ implica $a_n > M$. Lo anotamos*

$$\lim_{n \rightarrow \infty} a_n = +\infty \text{ o bien } a_n \xrightarrow{n \rightarrow \infty} +\infty.$$

Análogamente, decimos que $\lim_{n \rightarrow \infty} a_n = -\infty$ si para todo $M < 0$ existe $n_0 \in \mathbb{N}$ tal que $n \geq n_0$ implica $a_n < M$.

Es decir que para cualquier cota que a mi se me ocurra, todos los términos de la sucesión son más grandes que esa cota a partir de un n_0 (en el caso $+\infty$).

1.2. El espacio como n-uplas

En esta sección vamos a establecer las herramientas básicas para poder hablar de límites en \mathbb{R}^n . Necesitamos recordar nociones de vectores, norma, producto interno y también extender la idea de entorno de un punto de \mathbb{R}^n . Muchas de las primeras definiciones (las que involucran norma, distancia y producto interno) les van a resultar familiares, aunque algunas demostraciones de propiedades tal vez no las hayan visto antes.

Dado $n \in \mathbb{N}$, una n -upla es una tira ordenada de números,

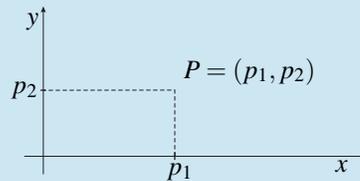
$$(x_1, x_2, x_3, \dots, x_n),$$

no necesariamente distintos.

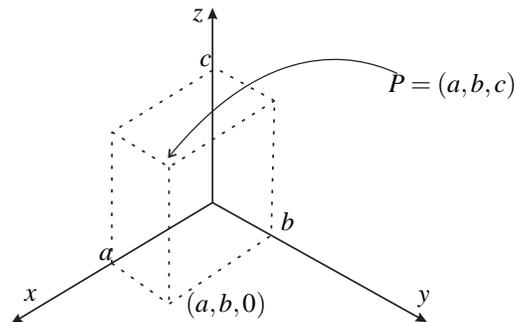
El plano en el cual estamos acostumbrados a pensar lo podemos describir con 2-uplas (que en general llamamos *pares ordenados*):

$$\mathbb{R}^2 = \{(x_1, x_2) : x_i \in \mathbb{R}\}.$$

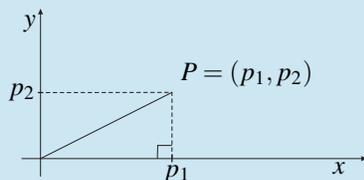
La primera coordenada la llamamos *abscisa* y la segunda *ordenada*. En un gráfico las representamos en el eje horizontal y vertical respectivamente.



Lo mismo ocurre con el espacio \mathbb{R}^3 , al cual representamos usando tres ejes:



1.2.1. Distancia



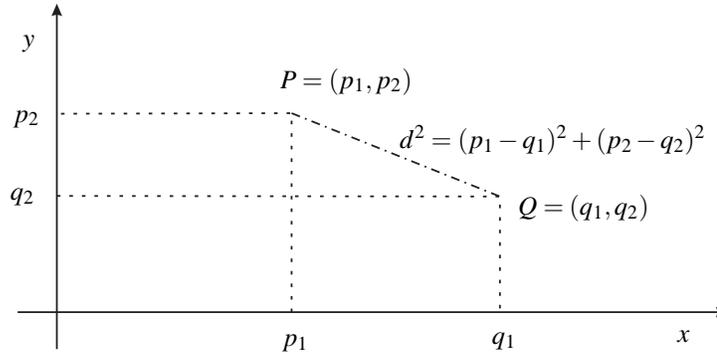
¿Cuál es la distancia de un punto $P = (p_1, p_2)$ al origen? Si observamos nuevamente la figura del plano xy , se ha formado un triángulo rectángulo, con vértices en $(0, 0)$, $(p_1, 0)$ y (p_1, p_2) .

Por lo tanto $d = \sqrt{p_1^2 + p_2^2}$. En el caso general, la distancia de $P = (p_1, p_2, \dots, p_n) \in \mathbb{R}^n$ al origen (que escribimos con un cero gordo, $\mathbf{0} \in \mathbb{R}^n$) está dada por

$$d = \sqrt{p_1^2 + p_2^2 + \dots + p_n^2},$$

como puede observarse por ejemplo en el dibujo en \mathbb{R}^3 de arriba.

¿Cómo se calcula la distancia entre dos puntos? Es sencillo, en la figura se ve que si $P = (p_1, p_2)$ y $Q = (q_1, q_2)$, entonces queda determinado un nuevo triángulo rectángulo, de lados $|p_1 - q_1|$ y $|p_2 - q_2|$



Luego

$$d(P, Q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2},$$

y en general, si $P = (p_1, p_2, \dots, p_n)$ y $Q = (q_1, q_2, \dots, q_n)$ son puntos de \mathbb{R}^n , su distancia está dada por

$$d(P, Q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2}.$$

Si a los puntos del plano los pensamos como vectores, entonces la distancia de $P = (p_1, p_2, \dots, p_n) \in \mathbb{R}^n$ al origen es la longitud del vector P , y la llamamos *norma* de P . La anotamos

$$\|P\| = \sqrt{p_1^2 + p_2^2 + \dots + p_n^2}.$$

En el caso de dos vectores P, Q , su distancia es la longitud del vector que los une, que es $Q - P$ (o $P - Q$ dependiendo de dónde lo hacemos empezar). Luego

$$d(P, Q) = \|P - Q\|.$$

1.2.2. Producto escalar

De las propiedades geométricas básicas, tenemos también la noción de ángulo entre dos vectores, para lo cuál es útil pensar primero en el producto escalar

$$\langle P, Q \rangle = p_1 q_1 + p_2 q_2 + \dots + p_n q_n.$$

Tiene las siguientes propiedades (que dejo para verificar al lector, salvo la última). Sean P, Q, R, S vectores de \mathbb{R}^n , entonces

1. $\langle P, Q \rangle = \langle Q, P \rangle$
2. $\langle P, P \rangle = \|P\|^2 \geq 0$ para todo $P \in \mathbb{R}^n$, y es cero si y sólo si $P = \mathbf{0}$.

3. $\langle \alpha P, \beta Q \rangle = \alpha\beta \langle P, Q \rangle$ para todo par de números $\alpha, \beta \in \mathbb{R}$.
4. $\langle P + R, Q \rangle = \langle P, Q \rangle + \langle R, Q \rangle$.
5. $|\langle P, Q \rangle| \leq \|P\| \|Q\|$ (desigualdad de Cauchy-Schwarz).

Demostración. (de la desigualdad de C-S). Dado $t \in \mathbb{R}$, consideremos la función real $g(t) = \|P - tQ\|^2$. Por las propiedades anteriores se tiene

$$\|P - tQ\|^2 = \langle P - tQ, P - tQ \rangle = \langle P, P \rangle - 2t \langle P, Q \rangle + t^2 \langle Q, Q \rangle.$$

Como P, Q están fijos, g es un polinomio de grado 2, $g(t) = at^2 + bt + c$ donde $a = \|Q\|^2$, $b = -2\langle P, Q \rangle$ y $c = \|P\|^2$. Por otro lado, como $g(t) = \|P - tQ\|^2 \geq 0$, este polinomio tiene a lo sumo una raíz real. Como $a \geq 0$, debe tener discriminante no positivo, es decir, se debe cumplir $b^2 - 4ac \leq 0$. Pero esto es

$$(2\langle P, Q \rangle)^2 - 4\|Q\|^2 \|P\|^2 \leq 0,$$

una expresión de la que se deduce fácilmente la desigualdad deseada pasando el término negativo a la derecha y tomando raíz cuadrada a ambos miembros. \square

Observación 1.2.1. La igualdad en la desigualdad de C-S vale si y sólo si P, Q están alineados, es decir si existe $\alpha \in \mathbb{R}$ tal que $P = \alpha Q$. En efecto, de la demostración anterior se deduce que para que valga la igualdad el discriminante del polinomio g debe ser igual a cero. Pero en ese caso, el polinomio g tiene una raíz (doble), es decir existe $\alpha \in \mathbb{R}$ tal que $g(\alpha) = \|P - \alpha Q\|^2 = 0$. En consecuencia debe ser $P - \alpha Q = 0$, puesto que la norma se anula únicamente en el origen.

Volviendo al ángulo entre dos vectores, como

$$-1 \leq \frac{\langle P, Q \rangle}{\|P\| \|Q\|} \leq 1,$$

podemos pensar que la cantidad del medio define el coseno del ángulo $\alpha(P, Q)$ entre los vectores, es decir

$$\langle P, Q \rangle = \|P\| \|Q\| \cos(\alpha).$$

Ciertamente en los casos que podemos dibujar (\mathbb{R}^2 y \mathbb{R}^3) ustedes saben que el ángulo así definido es el ángulo que los vectores P, Q sustentan cuando uno los dibuja. En el caso $n \geq 4$, siempre se puede pensar que dos vectores generan un plano $\{P, Q\}$ de \mathbb{R}^n , y la figura que corresponde es el ángulo formado en este plano (independientemente de que no nos podamos imaginar el espacio en el cual este plano está metido).

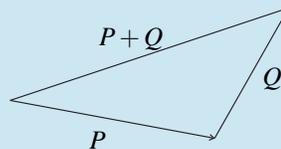
Propiedades de la norma

La norma tiene también una serie de propiedades bastante útiles, que enunciamos a continuación. Sean $P, Q \in \mathbb{R}^n$.

1. $\|P\| \geq 0$, y $\|P\| = 0$ si y sólo si $P = \mathbf{0}$.
2. $\|\lambda P\| = |\lambda| \|P\|$ para todo $\lambda \in \mathbb{R}$.
3. $\|P + Q\| \leq \|P\| + \|Q\|$ (desigualdad triangular).

Las dos últimas requieren alguna demostración. Pensemos primero qué quieren decir. La propiedad de sacar escalares simplemente refleja el hecho de que si uno multiplica un vector por un número, lo que está haciendo (si el número es positivo) es estirar o achicar el vector sin modificar su dirección. Si el número es negativo, el vector simplemente se da vuelta y la magnitud relevante es entonces el módulo del número.

Por otro lado, la última propiedad se conoce como *desigualdad triangular*, y es un simple reflejo del hecho de que en la figura de abajo, es más corto el camino yendo por la diagonal que primero recorriendo un lado y después el otro.



Demostremos la desigualdad triangular. Tenemos

$$\|P + Q\|^2 = \langle P + Q, P + Q \rangle = \|P\|^2 + \|Q\|^2 + 2\langle P, Q \rangle.$$

Por la desigualdad de C-S,

$$\langle P, Q \rangle \leq |\langle P, Q \rangle| \leq \|P\| \|Q\|.$$

Es decir, nos quedó

$$\|P + Q\|^2 \leq \|P\|^2 + \|Q\|^2 + 2\|P\| \|Q\| = (\|P\| + \|Q\|)^2.$$

Tomando raíz cuadrada a ambos términos se obtiene la desigualdad triangular.

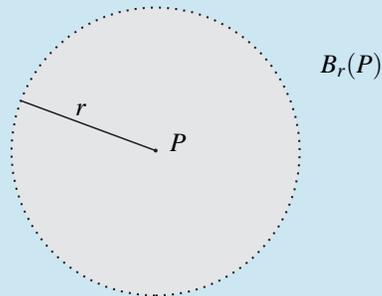
La desigualdad triangular en este punto parece un tanto técnica, pero si recordamos que queremos hablar de límites, nos dice algo fundamental: si dos puntos están cerca del origen, entonces ¡su suma también está cerca del origen!

Asimismo, si X está cerca del origen e Y está cerca de X , entonces la desigualdad triangular nos dice que Y está cerca del origen pues

$$\|Y\| \leq \|Y - X\| + \|X\|.$$

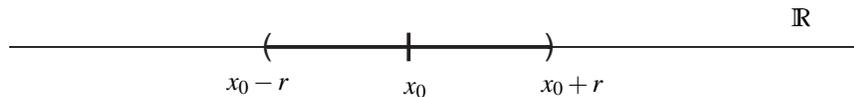
1.2.3. Entornos y conjuntos abiertos

Pensando en la norma como distancia, podemos definir un cierto entorno de un punto $P \in \mathbb{R}^n$ de la siguiente manera: nos quedamos con el conjunto de puntos que están a distancia menor que un número positivo r de un punto P dado. Es decir, si pensamos en el plano, con un disco alrededor del punto, como en la figura de la derecha.



Lo denotamos así: $B_r(P) = \{Q \in \mathbb{R}^n : d(Q, P) = \|Q - P\| < r\}$

Observemos que los puntos del borde del disco NO pertenecen a este conjunto. El conjunto se denomina *bola abierta de radio r con centro en P* . En el caso de la recta real ($n = 1$) se obtiene un intervalo: si tomamos un número $x_0 \in \mathbb{R}$ y consideramos $B_r(x_0) = \{x \in \mathbb{R} : |x - x_0| < r\}$, lo que obtenemos es el intervalo abierto de largo $2r$ (el “diámetro”) centrado en x_0 .

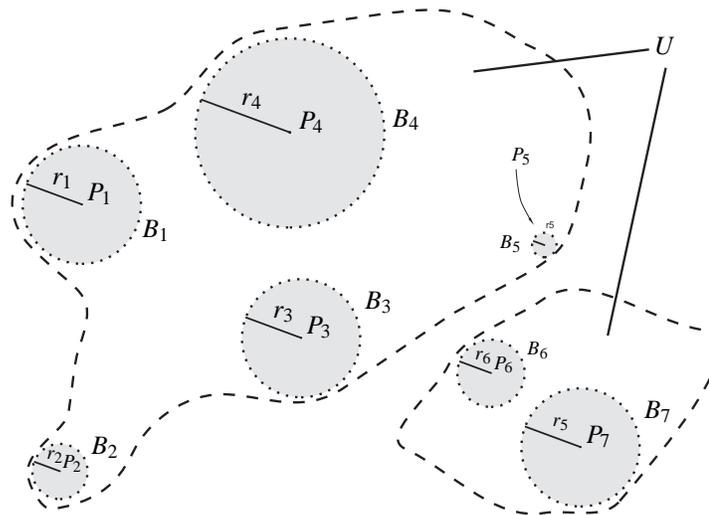


La siguiente definición generaliza esta idea.

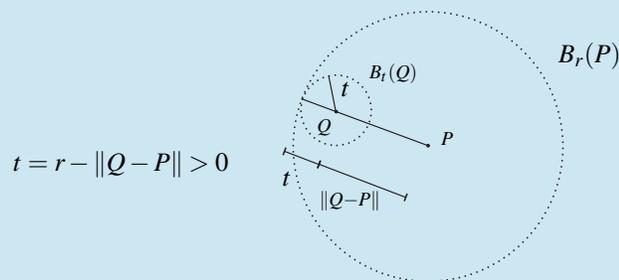
Definición 1.2.2. Un conjunto $U \subset \mathbb{R}^n$ es **abierto** si para cada punto $P \in U$ existe una bola abierta $B_r(P)$ centrada en P tal que $B_r(P) \subset U$ (el radio puede depender del punto P).

La idea intuitiva es que en un conjunto abierto, no hay puntos que estén en el borde, sino que están todos propiamente adentro. Porque los puntos del borde no tienen esta propiedad, ya cualquier bola centrada en ellos, tiene puntos tanto de fuera del conjunto como de adentro. Insisto con una parte importante de la definición: el radio r , para cada punto P del conjunto abierto U , va variando para adecuarse a la forma del conjunto y permitir que $B_r(P) \subset U$.

Un ejemplo ilustrativo se obtiene tomando algunos puntos en una figura en el plano como la siguiente:



Se imaginan que una bola abierta debería cumplir la definición de abierto (¡si no, estamos mal!). Veamos su demostración, que se deduce del dibujo



Lema 1.2.3. Sea $B_r(P)$ una bola abierta en \mathbb{R}^n . Para cada $Q \in B_r(P)$ existe $t > 0$ tal que $B_t(Q) \subset B_r(P)$. En particular una bola abierta es un conjunto abierto.

Demostración. Como $Q \in B_r(P)$, se tiene $d = \|Q - P\| < r$. Si $d = 0$, es porque $Q = P$ y tomamos $t = r$. El caso interesante es $0 < d < r$. En el dibujo se ve que $B_t(Q) \subset B_r(P)$, donde $t = r - d > 0$. Demostrémoslo. Tomemos $X \in B_{r-d}(Q)$, entonces $\|X - Q\| < r - d$. En consecuencia

$$\|X - P\| = \|X - Q + Q - P\| \leq \|X - Q\| + \|Q - P\| < (r - d) + d = r$$

donde hemos usado la desigualdad triangular. Esto nos dice que $X \in B_r(P)$, como afirmábamos. \square

Tenemos la siguiente equivalencia:

Proposición 1.2.4. Un conjunto $U \subset \mathbb{R}^n$ es abierto si y sólo si U se puede escribir como unión de bolas abiertas.

Demostración. Veamos \Rightarrow), supongamos que U es abierto. Entonces para cada punto $P \in U$ existe $r > 0$ tal que $B_r(P) \subset U$. En consecuencia $\cup_{P \in U} B_r(P) \subset U$ pues todas las bolas están en U . Y por otro lado es evidente que $U \subset \cup_{P \in U} B_r(P)$. Es decir hemos probado que

$$U \subset \cup_{P \in U} B_r(P) \subset U,$$

lo que nos asegura la igualdad $U = \cup_{P \in U} B_r(P)$, o sea U es una unión de bolas abiertas.

Veamos ahora \Leftarrow), supongamos que $U = \cup_{P_i} B_r(P_i)$ es una unión de bolas abiertas, donde los P_i son algunos de los puntos de U (pueden ser todos). Si $Q \in U$, tenemos que ver que existe $t > 0$ tal que $B_t(Q) \subset U$. Como U es unión de bolas abiertas, el punto Q debe estar en alguna de ellas, digamos $Q \in B_{r_0}(P_0)$. Pero por el Lema 1.2.3, existe $t > 0$ tal que $B_t(Q) \subset B_{r_0}(P_0)$. Por propiedad transitiva de la inclusión, se tienen $B_t(Q) \subset U$. \square

Y una propiedad fundamental:

Proposición 1.2.5. Si $\{U_i\}_{i \in I}$ es una familia de conjuntos abiertos de \mathbb{R}^n , entonces la unión $U = \cup_{i \in I} U_i$ es un conjunto abierto.

Demostración. Cada abierto U_i se puede escribir como una unión de bolas abiertas por la proposición anterior. Entonces U es la unión de todas las bolas de todos los abiertos, y resulta ser abierto de nuevo por la proposición anterior. \square

Algo similar ocurre con la intersección, pero teniendo cuidado porque sólo es cierto el resultado si son finitos conjuntos. Basta verlo para dos, ese es el contenido del siguiente lema.

Lema 1.2.6. Si U, V son conjuntos abiertos de \mathbb{R}^n , entonces su intersección $U \cap V$ es un conjunto abierto.

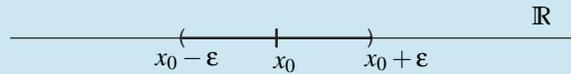
Demostración. Dado $P \in U \cap V$, existen $r_1 > 0$ y $r_2 > 0$ tales que $B_{r_1}(P) \subset U$ (por ser U abierto) y $B_{r_2}(P) \subset V$ (por ser V abierto). Si $r = \min\{r_1, r_2\}$, entonces $B_r(P) \subset U \cap V$ lo que prueba que éste último es abierto. \square

Se extiende el resultado anterior a finitos conjuntos abiertos. Sin embargo, el resultado es *falso* si consideramos infinitos abiertos.

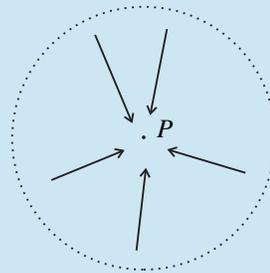
Problema 1.2.7. Comprobar que si $U_n = (0, \frac{n+1}{n}) \subset \mathbb{R}$, entonces $\cap_n U_n = (0, 1]$, que *no es abierto*.

Antes de seguir, quisiera que discutamos brevemente qué estamos haciendo.

Si recordamos la definición de límite en \mathbb{R} , la idea central es establecer con claridad qué quiere decir “estar cerca” de un punto x_0 . Para eso usamos el orden de \mathbb{R} , y alrededor de x_0 establecemos un intervalo $(x_0 - \varepsilon, x_0 + \varepsilon)$ para un $\varepsilon > 0$ dado,



Observemos que hay sólo dos direcciones desde donde acercarnos a x_0 (por la izquierda o por la derecha), que se corresponden con los límites laterales. Esto es así porque los números reales están ordenados, y por lo tanto sólo se puede ser mayor o menor que x_0 (o igual). Sin embargo, en el plano (o el espacio) no hay un orden entre puntos. Por lo tanto hay infinitas direcciones desde donde acercarse a un punto $P \in \mathbb{R}^n$.



De hecho, acercarnos por rectas no agota todas las posibilidades porque uno podría acercarse con una espiral u otra curva. Sin embargo, con las nociones de bola y abierto queda claro qué quiere decir estar cerca del punto P .

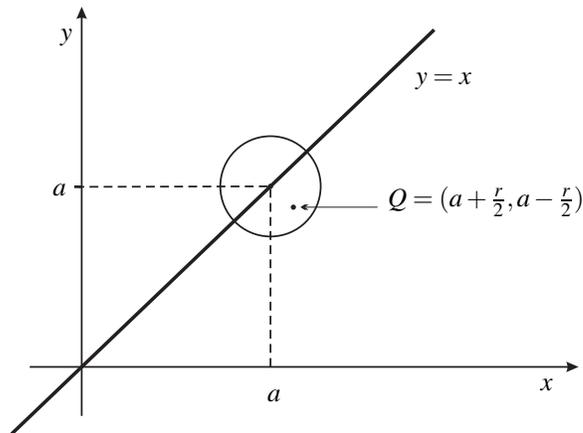
Otra noción útil que necesitamos es la de interior de un conjunto.

Definición 1.2.8. Si $C \subset \mathbb{R}^n$, un punto $P \in C$ es **interior** si existe una bola abierta centrada en P tal que $B_r(P) \subset C$. El conjunto de todos los puntos interiores de C se denomina **interior de C** y lo denotamos C° .

Un conjunto puede no tener ningún punto interior, es decir $C^\circ = \emptyset$. Por ejemplo si C es una recta en el plano. Aunque es evidente, esta afirmación requiere demostración. Seamos entonces concretos.

Ejemplo 1.2.9. Si $C = \{(x, y) : y = x\}$, entonces $C^\circ = \emptyset$.

Demostración. La manera más simple de probar lo enunciado es por el absurdo. Es decir, suponemos que hay un punto interior y llegamos a una contradicción. Supongamos entonces que $P \in C$ es interior. Debe ser $P = (a, a)$ para algún número real a . Si P es interior, existe $r > 0$ tal que $B_r(P) \subset C$. Nos guiamos por el dibujo siguiente



Si tomamos el punto $Q = (a + \frac{r}{2}, a - \frac{r}{2})$, evidentemente $Q \notin C$. Sin embargo, como

$$\|Q - P\| = \sqrt{\frac{r^2}{4} + \frac{r^2}{4}} = \frac{r}{\sqrt{2}} < r$$

se tiene $Q \in B_r(P)$, y como supusimos que $B_r(P) \subset C$, se tiene $Q \in C$, una contradicción. \square

En el otro extremo están los abiertos, donde todos los puntos son interiores. Primero una aclaración: el conjunto vacío, $C = \emptyset$, es abierto pues no hay puntos para verificar la condición. Se tiene

Proposición 1.2.10. 1. C° es un conjunto abierto de \mathbb{R}^n .

2. $C^\circ = C$ si y sólo si C es abierto.

Demostración. Veamos 1. Como dijimos, si $C = \emptyset$, $C^\circ = \emptyset$ con lo cual es abierto. Si es no vacío, y tomamos $P \in C^\circ$, por definición de interior se tiene que existe $r > 0$ tal que $B_r(P) \subset C$. Afirмо que todos los puntos de $B_r(P)$ son interiores, con lo cual en realidad $B_r(P) \subset C^\circ$, lo que probaría que C° es abierto. Para verlo tomemos $Q \in B_r(P)$; por el Lema 1.2.3 existe una bola $B_t(Q) \subset B_r(P)$. Como $B_r(P) \subset C$, tenemos $Q \in B_t(Q) \subset C$, o sea Q es interior, como queríamos ver.

Ahora veamos 2. Supongamos primero que $C^\circ = C$. Entonces por 1. C es abierto. Supongamos ahora que C es abierto. Tenemos por definición $C^\circ \subset C$. Veamos que vale la otra inclusión. Para ello tomamos $P \in C$, por ser C abierto existe $r > 0$ tal que $B_r(P) \subset C$. Pero esto nos dice exactamente (releamos la definición) que P es un punto interior, o sea $P \in C^\circ$. \square

El interior C° de un conjunto es, en algún sentido, el abierto más grande contenido en C , según nos enseña el ejercicio que dejamos para el lector.

Problema 1.2.11. Probar que C° se puede escribir como la unión de todos los conjuntos abiertos de C , es decir $C^\circ = \cup U_i$, donde $U_i \subset C$ es abierto en \mathbb{R}^n .

1.2.4. Límites en \mathbb{R}^n

Vamos a generalizar la noción de sucesión de números reales a sucesiones de vectores.

Definición 1.2.12. Una sucesión $\{P_k\}_{k \in \mathbb{N}}$ de puntos de \mathbb{R}^n es una tira ordenada de vectores

$$P_1, P_2, P_3, \dots, P_k, \dots$$

Equivalentemente, es una función $P: \mathbb{N} \rightarrow \mathbb{R}^n$.

Ejemplo 1.2.13. 1. Tomemos $P_k = (\frac{1}{k}, \cos \frac{1}{k})$. Entonces $\{P_k\}_{k \in \mathbb{N}}$ es una sucesión de \mathbb{R}^2 .

2. Si ponemos $Q_k = (e^{-k}, k^2, \frac{1}{k})$, entonces $\{Q_k\}_{k \in \mathbb{N}}$ es una sucesión de \mathbb{R}^3 .

3. Si $X_j = (e^{-j}, \frac{1}{j} \sin(2j), \frac{1}{j^2}, (1 + \frac{1}{j})^j)$ entonces $\{X_j\}_{j \in \mathbb{N}}$ es una sucesión de \mathbb{R}^4 .

Observemos que como un vector $P \in \mathbb{R}^n$ está determinado por sus n coordenadas (que son números reales), entonces una sucesión de \mathbb{R}^n se puede pensar como una tira ordenada de n sucesiones de números reales, y al k -ésimo término de la sucesión (que es un vector de \mathbb{R}^n) lo anotamos así

$$P_k = ((P_k)_1, (P_k)_2, \dots, (P_k)_n).$$

Es decir, usamos un paréntesis y un subíndice para señalar qué coordenada nos interesa. Así por ejemplo, si P_k es la sucesión de \mathbb{R}^2 dada por

$$P_k = (1 - 1/k, 2^k),$$

se tiene $(P_k)_1 = 1 - 1/k$ y $(P_k)_2 = 2^k$.

Ahora que sabemos qué es una sucesión, podemos dar la definición de límite de sucesiones usando la norma (que nos provee de los entornos).

Definición 1.2.14. Si $\{P_k\}_{k \in \mathbb{N}}$ es una sucesión de \mathbb{R}^n , y $P \in \mathbb{R}^n$, decimos que P_k tiende a P si para todo $\varepsilon > 0$ existe $k_0 \in \mathbb{N}$ tal que

$$\|P_k - P\| < \varepsilon \text{ para todo } k \geq k_0,$$

y lo escribimos $\lim_{k \rightarrow \infty} P_k = P$ o bien $P_k \xrightarrow{k \rightarrow \infty} P$.

Observemos que esta definición coincide con la definición de límite de una sucesión en el caso $n = 1$. Se puede definir también que quiere decir que una sucesión de vectores tienda a infinito, pero en el caso $n \geq 2$ lo que queremos decir es que la norma de los vectores se hace arbitrariamente grande, es decir

Definición 1.2.15. Si $\{P_k\}_{k \in \mathbb{N}}$ es una sucesión de \mathbb{R}^n , decimos que P_k tiende a infinito si para todo $M > 0$ existe $k_0 \in \mathbb{N}$ tal que

$$\|P_k\| > M \text{ para todo } k \geq k_0,$$

y lo escribimos $\lim_{k \rightarrow \infty} P_k = \infty$ o bien $P_k \xrightarrow{k \rightarrow \infty} \infty$. También es habitual decir que la sucesión $\{P_k\}$ **diverge**.

Lo que uno imagina, si piensa en las coordenadas del vector que forma la sucesión, es que si tiene límite cada coordenada entonces tiene límite la sucesión original. Esto no sólo es así sino que vale la recíproca. Para verlo, primero una observación útil sobre la norma.

Observación 1.2.16. Si $P = (p_1, p_2, \dots, p_j, \dots, p_n)$ es un vector de \mathbb{R}^n , entonces se tiene, para cualquier j desde 1 hasta n ,

$$|p_j| \leq \|P\| \leq \sqrt{n} \max_{j=1..n} |p_j|. \quad (1.1)$$

Para verlo basta recordar que

$$\|P\| = \sqrt{p_1^2 + p_2^2 + \dots + p_j^2 + \dots + p_n^2},$$

y en consecuencia se tiene

$$p_j^2 \leq p_1^2 + p_2^2 + \dots + p_j^2 + \dots + p_n^2 \leq n \max_{j=1..n} p_j^2.$$

Ahora sí, escribimos el resultado que establece que la convergencia en \mathbb{R}^n es simplemente la convergencia en todas y cada una de las coordenadas.

Proposición 1.2.17. Sea $\{P_k\}$ una sucesión de \mathbb{R}^n . Entonces $P = (p_1, p_2, \dots, p_n)$ es el límite de la sucesión P_k si y sólo si cada coordenada del vector sucesión tiene como límite la correspondiente coordenada de P . Es decir

$$\lim_{k \rightarrow \infty} P_k = P \iff \left(\lim_{k \rightarrow \infty} (P_k)_j = p_j \text{ para todo } j = 1 \dots n \right).$$

Demostración. Supongamos primero que todas las coordenadas convergen al correspondiente p_j . Esto quiere decir, que dado $\varepsilon > 0$,

$$1. \text{ existe } k_1 \in \mathbb{N} \text{ tal que } |(P_k)_1 - p_1| < \frac{\varepsilon}{\sqrt{n}} \text{ si } k \geq k_1.$$

2. existe $k_2 \in \mathbb{N}$ tal que $|(P_k)_2 - p_2| < \frac{\varepsilon}{\sqrt{n}}$ si $k \geq k_2$.
3. ...
4. existe $k_n \in \mathbb{N}$ tal que $|(P_k)_n - p_n| < \frac{\varepsilon}{\sqrt{n}}$ si $k \geq k_n$.

Si tomamos $k_0 = \max_{j=1 \dots n} k_j$, entonces

$$\|P_k - P\| = \sqrt{((P_k)_1 - p_1)^2 + \dots + (P_k)_n - p_n)^2} < \sqrt{\frac{\varepsilon^2}{n} + \dots + \frac{\varepsilon^2}{n}} = \sqrt{\varepsilon^2} = \varepsilon.$$

Esto prueba que $\lim_k P_k = P$.

Supongamos ahora que $\lim_k P_k = P$, queremos ver que todas las coordenadas convergen a la correspondiente coordenada de P . Pero por la definición, tenemos que dado $\varepsilon > 0$, existe $k_0 \in \mathbb{N}$ tal que $k \geq k_0$ implica $\|P_k - P\| < \varepsilon$. Pero si miramos la ecuación (1.1) de arriba, se deduce que

$$|(P_k)_j - p_j| \leq \|P_k - P\| < \varepsilon$$

para todo $j = 1 \dots n$, siempre que $k \geq k_0$. Esto nos dice que la sucesión de cada coordenada converge a la coordenada correspondiente de P . \square

Observación 1.2.18. *Atenti, que el resultado anterior no vale para sucesiones divergentes. Por ejemplo, si $P_k = (k, \frac{1}{k})$, entonces $\|P_k\| = \sqrt{k^2 + \frac{1}{k^2}} \rightarrow +\infty$, es decir P_k tiende a infinito o diverge de acuerdo a nuestra definición. Sin embargo, **no** es cierto que ambas coordenadas tiendan a infinito.*

Problema 1.2.19. *Calcular los límites de las sucesiones del Ejemplo 1.2.13.*

1.2.5. Puntos de acumulación y conjuntos cerrados

Ahora vamos a usar la noción de límite para pensar en otro tipo de conjuntos, que son los acompañantes naturales de los conjuntos abiertos que estuvimos charlando más arriba. Hagamos una lista de los objetos que nos interesan.

Definición 1.2.20. *Sea $C \subset \mathbb{R}^n$ un conjunto.*

*Decimos que P es un **punto de acumulación** de C si existe una sucesión $\{P_k\}$ de puntos de C tal que $P = \lim_{k \rightarrow \infty} P_k$.*

*La **clausura** del conjunto C es la unión de todos los puntos de acumulación de C . La denotamos \bar{C} .*

*Decimos que C es un conjunto **cerrado** si para toda sucesión convergente $\{P_k\}$, tal que todos sus términos son puntos de C , se tiene $P = \lim_{k \rightarrow \infty} P_k \in C$.*

Observemos que $C \subset \overline{C}$ puesto que dado un punto $P \in C$, la sucesión constante $P_k = P$ nos dice que P es punto de acumulación de C . A partir de estas definiciones, se tienen la siguiente propiedad fundamental, que relaciona los conjuntos abiertos con los cerrados:

Proposición 1.2.21. *Un conjunto $C \subset \mathbb{R}^n$ es cerrado si y sólo si su complemento C^c es abierto.*

Demostración. Supongamos primero que C es cerrado, pongamos $U = C^c$ y veamos que U es abierto. De acuerdo a la definición, basta probar que todo punto de U es interior. Tomemos $P \in U$, supongamos que no es interior y llegaremos a un absurdo. Si P no es interior, entonces para todo $k \in \mathbb{N}$ se tiene $B_{\frac{1}{k}}(P) \not\subset U$, es decir existe $P_k \in U^c = C$ tal que $P_k \in B_{\frac{1}{k}}(P)$. Observemos que $P_k \rightarrow P$, pues $\|P_k - P\| < \frac{1}{k}$. Hemos llegado a una contradicción, pues $P_k \in C$, $P \notin C$ y por hipótesis C es cerrado.

Supongamos ahora que $U = C^c$ es abierto, y veamos que C es cerrado. Tomemos $\{P_k\}$ una sucesión de puntos de C con límite $P \in \mathbb{R}^n$. Queremos ver que $P \in C$. Si esto no fuera así, sería $P \in U$ que por hipótesis es abierto. En consecuencia existiría $r > 0$ tal que $B_r(P) \subset U$. Tomemos $k \in \mathbb{N}$ tal que $\|P_k - P\| < r$ (recordemos que $P_k \rightarrow P$). Se tendría $P_k \in B_r(P) \subset U$, lo cual es imposible pues $P_k \in C = U^c$. Luego debe ser $P \in C$, y esto prueba que C es cerrado. \square

Observación 1.2.22. *Atención, que hay conjuntos que no son ni abiertos, ni cerrados. Por ejemplo, el intervalo $I = [0, 1)$ en \mathbb{R} . Que no es abierto se deduce de que $0 \in I$ no es un punto interior (¿por qué?). Que no es cerrado se deduce de que la sucesión $a_n = 1 - \frac{1}{n}$, de puntos de I , tiene límite 1, que no es un punto de I .*

Con esta dualidad entre abiertos y cerrados es fácil establecer los siguientes hechos, a partir de las demostraciones que ya hicimos para conjuntos abiertos.

Sea $C \subset \mathbb{R}^n$.

1. La clausura \overline{C} es un conjunto cerrado.
2. $\overline{C} = C$ si y sólo si C es cerrado.
3. Si $\{C_i\}_{i \in I}$ es una familia de conjuntos cerrados, entonces $\bigcap_{i \in I} C_i$ es un conjunto cerrado.
4. Si C_1 y C_2 son conjuntos cerrados, su unión $C_1 \cup C_2$ es un conjunto cerrado.

Demostración. Veamos 1. Si $A = \overline{C}$, y $P \in A$, entonces existe $r > 0$ tal que $B_r(P) \cap C = \emptyset$ (de lo contrario, uno podría fabricar una sucesión X_k de puntos de C tal que $X_k \rightarrow P$, lo que diría que $P \in \overline{C}$, contradiciendo que $P \in \overline{C}^c$). Afirmando que $B_r(P) \subset A$, lo que probaría que A es abierto, y entonces \overline{C} sería cerrado. Tomemos $Q \in B_r(P)$, $\|Q - P\| = d < r$. Si fuera $Q \notin A$, tendríamos $Q \in \overline{C}$. Existiría entonces una sucesión $Y_k \in C$ tal que $Y_k \rightarrow Q$. Pero con k suficientemente grande, tendríamos

$$\|Y_k - P\| \leq \|Y_k - Q\| + \|Q - P\| < (r - d) + d = r$$

lo que nos dice que $Y_k \in B_r(P)$, contradiciendo que $B_r(P) \cap C = \emptyset$.

Veamos 2. Supongamos primero que $\overline{C} = C$. Entonces por el ítem previo, C es cerrado. Recíprocamente, si C es cerrado, entonces dado un punto $P \in \overline{C}$, existe una sucesión de puntos P_k de C tal que $P_k \rightarrow P$, y como C es cerrado debe ser $P \in C$. Esto prueba que $\overline{C} \subset C$, y como la otra inclusión siempre vale, obtuvimos $\overline{C} = C$.

La propiedad 3. se verifica observando que, por las leyes de de Morgan, $(A \cap B)^c = A^c \cup B^c$. En consecuencia,

$$(\cap_{i \in I} C_i)^c = \cup_{i \in I} C_i^c.$$

Si cada C_i es cerrado, por la Proposición 1.2.21, cada C_i^c es abierto. Pero una unión arbitraria de abiertos es abierta por la Proposición 1.2.5. Así que el lado derecho de la ecuación es un conjunto abierto (y por lo tanto el lado izquierdo es un conjunto abierto). Si volvemos a tomar complementos, obtenemos que $\cap_i C_i$ es cerrado (recordar que $(A^c)^c = A$).

La propiedad 4. se verifica en forma idéntica, usando ahora que $(A \cup B)^c = A^c \cap B^c$ y el Lema 1.2.6. ¡Escribirlo! \square

Se deduce del último ítem que si C_1, \dots, C_m son **finitos** conjunto cerrados, su unión es un conjunto cerrado. Sin embargo una unión infinita de cerrados puede no ser cerrada, como muestra el siguiente problema.

Problema 1.2.23.

Consideremos los cerrados de \mathbb{R} dados por $C_j = [0, 1 - \frac{1}{j}]$, con $j \in \mathbb{N}$. Verificar que $\cup_j C_j = [0, 1)$, que **no** es un conjunto cerrado.

Observación 1.2.24. Si $B_r(P)$ es una bola abierta en \mathbb{R}^n , su clausura se denomina **bola cerrada** de radio r centrada en P , y es el conjunto

$$\overline{B}_r(P) = \{Q \in \mathbb{R}^n : \|Q - P\| \leq r\}.$$

Aunque bastante obvia, esta afirmación requiere alguna demostración

$$\overline{B_r(p)} = \overline{B}_r(P).$$

Queremos ver que el conjunto de puntos de acumulación de $B_r(P)$ es $\overline{B}_r(P)$. Sea Q un punto de acumulación de $B_r(P)$. Entonces existe $Q_k \in B_r(P)$ tal que $Q_k \rightarrow Q$. En consecuencia,

$$\|Q - P\| \leq \|Q - Q_k\| + \|Q_k - P\| < \|Q - Q_k\| + r.$$

Tomando límite para $k \rightarrow \infty$ se obtiene $\|Q - P\| \leq r$, es decir hemos probado que si Q es un punto de acumulación de $B_r(P)$, entonces $Q \in \overline{B}_r(P)$ (o sea hemos probado que $B_r(p) \subset \overline{B}_r(P)$). Veamos ahora que

vale la recíproca. Tomemos un punto $Q \in \overline{B}_r(P)$, es decir $\|Q - P\| \leq r$. Consideremos la sucesión $Q_k = P + (1 - \frac{1}{k})(Q - P)$. Ciertamente

$$\|Q_k - P\| = |1 - \frac{1}{k}| \|Q - P\| \leq |1 - \frac{1}{k}| r < r$$

para todo $k \in \mathbb{N}$, pues $|1 - \frac{1}{k}| = 1 - \frac{1}{k} < 1$, lo que prueba que $Q_k \in B_r(P)$. Y por otro lado, como $Q_k = Q - \frac{1}{k}(Q - P)$, se tiene

$$\|Q_k - Q\| = \frac{1}{k} \|Q - P\|$$

lo que prueba que $Q_k \rightarrow Q$. Entonces Q es punto de acumulación de $B_r(P)$, lo que prueba la inclusión $\overline{B}_r(P) \subset \overline{B}_r(p)$ que faltaba para tener la igualdad.

1.2.6. Conjuntos acotados

Antes de pasar a hablar de funciones, extendamos la idea de conjunto acotado a \mathbb{R}^n y veamos sus propiedades básicas.

Definición 1.2.25. Un conjunto $A \subset \mathbb{R}^n$ es **acotado** si existe $M > 0$ tal que

$$\|P\| \leq M \text{ para todo } P \in A.$$

Observación 1.2.26. Es decir, un conjunto es acotado si lo podemos poner dentro de una bola. Se deduce de la Observación 1.2.16 que un conjunto es acotado si y sólo si todas las coordenadas del conjunto están acotadas, ya que por un lado, cada coordenada está acotada por la norma, y por otro lado, la norma está acotada por n veces el máximo de las coordenadas.

Toda sucesión convergente es acotada, pues si $P_k \rightarrow P$ en \mathbb{R}^n , existe $k_0 \in \mathbb{N}$ tal que $k \geq k_0$ implica $\|P_k - P\| \leq 1$, con lo cual

$$\|P_k\| \leq \|P_k - P\| + \|P\| \leq 1 + \|P\|$$

para todo $k \geq k_0$. Y los primeros términos de la sucesión están acotados (pues es un conjunto finito). Es decir, si tomamos $M = \max_{k=1 \dots k_0} \{\|P_k\|, \|P\| + 1\}$, se tiene

$$\|P_k\| \leq M \text{ para todo } k \in \mathbb{N},$$

lo que nos dice que la sucesión está acotada por M .

Ahora podemos escribir una generalización del Teorema de Bolzano-Weierstrass que probamos para \mathbb{R} . La idea es simplemente que cada coordenada tiene una subsucesión convergente, pero hay que elegir en forma ordenada para no arruinar la convergencia de la anterior. Un hecho importante que vale en general, y que usamos aquí, es que una subsucesión de una sucesión convergente es siempre convergente.

Teorema 1.2.27. Si $\{P_k\} \subset \mathbb{R}^n$ es una sucesión acotada, entonces existe una subsucesión convergente.

Demostración. Si $\{P_k\}$ es una sucesión acotada, entonces todas las sucesiones $(P_k)_j$ que forman sus coordenadas (con $j = 1 \dots n$) son sucesiones acotadas de números reales. Por el Teorema de B-W (Corolario 1.1.29), como la sucesión $(P_k)_1$ de la primer coordenada está acotada, tiene una subsucesión $(P_{k_i})_1$ convergente. Ahora miramos la segunda coordenada, pero sólo los términos que corresponden a la subsucesión que ya elegimos en la primer coordenada. Es decir, miramos la subsucesión $(P_{k_i})_2$. Esta es de nuevo una sucesión acotada, por ser una subsucesión de una sucesión acotada. Entonces por el Teorema de B-W, existe una subsucesión $(P_{k_{i_2}})_2$ que es convergente. Seguimos así hasta llegar a la última coordenada del vector que forma la sucesión. Observemos que en cada paso nos estamos quedando con una subsucesión, por lo tanto esto no altera el hecho de que en la coordenada anterior nos quedamos con una subsucesión convergente (pues una subsucesión de una sucesión convergente es convergente). Obtenemos así una subsucesión convergente en cada coordenada, y por ende, tenemos una subsucesión convergente de la sucesión P_k original. \square

Definición 1.2.28. Sea $C \subset \mathbb{R}^n$. La **frontera de C** es el conjunto de puntos de acumulación de C que no son interiores. La denotamos

$$\partial C = \bar{C} - C^\circ.$$

Problema 1.2.29. Probar que la frontera de la bola (abierta o cerrada) es la esfera que forman los puntos del borde. Es decir

$$\partial B_r(P) = \{Q \in \mathbb{R}^n : \|Q - P\| = r\}.$$

Probar que si C es cerrado, con interior vacío (por ejemplo, un subespacio), entonces $\partial C = C$. ¿Se les ocurre algún ejemplo en el que $\partial C = \emptyset$?

1.3. Problemas

1.I. Probar usando los axiomas de cuerpo ordenado que $0 < 1$, y que si $a < b$ entonces el promedio $p = \frac{a+b}{2}$ verifica $a < p < b$.

1.II. Probar que $i \in \mathbb{R}$ es el ínfimo de $A \subset \mathbb{R}$ si $i \leq a$ para todo $a \in A$ y además, dado $\varepsilon > 0$, existe $a \in A$ tal que $i + \varepsilon > a$.

1.III. Probar que si $\{a_n\}_{n \in \mathbb{N}}$ es una sucesión de números reales tales que $a_n \leq l$ para todo n , entonces

$$[\forall \varepsilon > 0 \quad \exists n_0 \in \mathbb{N} \text{ tal que } l - a_n < \varepsilon \quad (\forall n \geq n_0)] \Rightarrow \lim_{n \rightarrow \infty} a_n = l.$$

- 1.IV. Probar que si $\inf A \notin A$, existe una sucesión estrictamente decreciente de elementos de A que converge al ínfimo.
- 1.V. Probar que si $\{a_n\}$ es una sucesión decreciente y acotada inferiormente, entonces $\lim_{n \rightarrow \infty} a_n = \inf\{a_n\}$.
- 1.VI. Supongamos que $\{a_n\}$ es una sucesión que no está acotada superiormente. Probar que $\{a_n\}$ tiene una subsucesión estrictamente creciente que tiende a $+\infty$.
- 1.VII. Sea U_n la colección de intervalos abiertos en \mathbb{R} dados por $U_n = (0, \frac{n+1}{n})$. Probar que $\bigcap_n U_n = (0, 1]$ y concluir que la intersección arbitraria de conjuntos abiertos no es necesariamente abierta.
- 1.VIII. Sea $C \subset \mathbb{R}^n$, y sea $\{U_i\}_{i \in I}$ la familia de todos los conjuntos abiertos de \mathbb{R}^n tales que $U_i \subset C$. Probar que $C^0 = \bigcup_i U_i$.
- 1.IX. Probar que $(A \cap B)^c = A^c \cap B^c$ para todo par de conjuntos. Concluir que la unión finita de conjuntos cerrados es un conjunto cerrado (asumiendo que la intersección finita de conjuntos abiertos es abierta).
- 1.X. Consideremos los cerrados de \mathbb{R} dados por $C_j = [0, 1 - \frac{1}{j}]$, con $j \in \mathbb{N}$. Probar que $\bigcup_j C_j = [0, 1)$, que **no** es un conjunto cerrado.
- 1.XI. Sea $C \subset \mathbb{R}^n$, y sea $\{C_i\}_{i \in I}$ la familia de todos los conjuntos cerrados de \mathbb{R}^n tales que $C \subset C_i$. Probar que $\overline{C} = \bigcap_i C_i$.
- 1.XII. Probar que la frontera de la bola de radio $r > 0$ alrededor de $P \in \mathbb{R}^n$ son los $X \in \mathbb{R}^n$ tales que $\|X - P\| = r$. A este conjunto se lo denomina *esfera centrada en P de radio r* .
- 1.XIII. Probar que si C es cerrado y tiene interior vacío entonces $\partial C = C$. Dar un ejemplo de un conjunto en \mathbb{R}^2 tal que $\partial C = \emptyset$.

Equality is not in regarding different things similarly, equality is in regarding different things differently.

TOM ROBBINS

2.1. Dominio, gráfico, imagen

Podemos empezar a pensar en funciones de varias variables; ya desarrollamos las herramientas para pensar en dominios, límites y demás nociones que involucren conjuntos en \mathbb{R}^n .

2.1.1. Funciones $f : \mathbb{R}^n \rightarrow \mathbb{R}$

Recordemos primero que el dominio natural de una función es el conjunto de puntos donde tiene sentido aplicar la función. Ya estuvimos trabajando sin decirlo con funciones de \mathbb{R}^n en \mathbb{R} .

Ejemplo 2.1.1. *Veamos primero algunos ejemplos donde $\text{Dom}(f) = \mathbb{R}^n$. Sea $X = (x_1, \dots, x_n) \in \mathbb{R}^n$.*

1. La norma $\|X\| = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$ es una función.
2. Si miramos una coordenada concreta, la asignación $X \mapsto x_j$ es una función. Estas se llaman **funciones coordenadas** o **proyecciones** a las coordenadas.
3. Si fijamos $P \in \mathbb{R}^n$ y movemos X en el producto escalar $\langle P, X \rangle$ lo que tenemos es una función $X \mapsto \langle P, X \rangle$.
4. Si $x, y \in \mathbb{R}$, entonces la $f(x, y) = xy$ es la función producto de las coordenadas, y por otro lado $g(x, y) = x + y$ es la función suma de las coordenadas (ambas funciones de \mathbb{R}^2 en \mathbb{R}).

Ahora consideremos ejemplos donde el dominio sea un subconjunto propio de \mathbb{R}^n .

1. $f(x, y) = \frac{y}{x}$. El dominio es el conjunto $\text{Dom}(f) = \{(x, y) \in \mathbb{R}^2 : x \neq 0\}$, es decir, todo el plano salvo el eje y .

2. $f(x, y, z) = \frac{1}{x^2+y^2+z^2} = \frac{1}{\|X\|^2}$. En este caso, para que se anule el denominador, debe ser $x = y = z = 0$. Luego $\text{Dom}(f) = \mathbb{R}^3 - \{\mathbf{0}\}$.

3. Sea

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

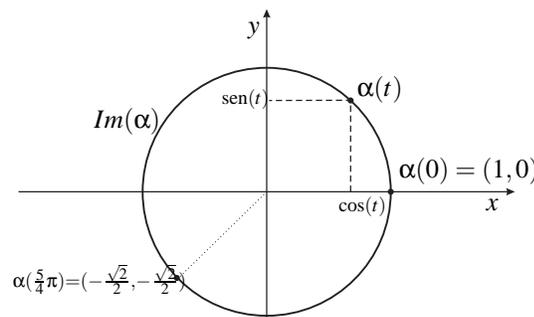
En este caso es $\text{Dom}(f) = \{(x, y) \in \mathbb{R}^2 : y \neq x + 1\}$, que es \mathbb{R}^2 menos una recta.

2.1.2. Curvas

Una curva es una función $\alpha : I \rightarrow \mathbb{R}^n$, donde I es algún subconjunto (en general un intervalo) de \mathbb{R} . Ejemplos de curvas son

1. $\alpha : \mathbb{R} \rightarrow \mathbb{R}^2$ dada por $\alpha(t) = (\cos(t), \text{sen}(t))$.
2. $\beta : \mathbb{R} \rightarrow \mathbb{R}^3$ dada por $\beta(x) = (x, x - 1, 3)$
3. $\gamma : \mathbb{R} \rightarrow \mathbb{R}^4$ dada por $\gamma(y) = (y, y^2, y^3, y^4)$.
4. Si $f : \text{Dom}(f) \subset \mathbb{R} \rightarrow \mathbb{R}$, $\delta(x) = (x, f(x))$ es una curva en el plano.

Observemos que en el último ejemplo la curva δ tiene como imagen lo que llamamos el *gráfico de f* , $\text{Gr}(f)$. En el caso de curvas a valores en \mathbb{R}^2 ó \mathbb{R}^3 , lo que interesa en general es la imagen de la misma (a diferencia de las funciones de más variables, donde lo que interesa es el gráfico como vimos). Así por ejemplo, la imagen de la curva α es una circunferencia de radio unitario centrada en el origen,



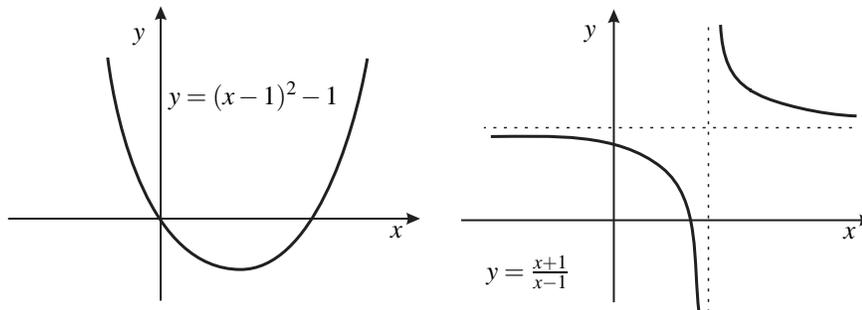
Mientras que la imagen de β es una recta en el espacio \mathbb{R}^3 como se deduce de $(x, x - 1, 3) = x(1, 1, 0) + (0, -1, 3)$. A las curvas también se las suele denominar *trayectorias* o *arcos*.

2.1.3. Gráfico, curvas y superficies de nivel

¿Se acuerdan del gráfico de una función $f : \mathbb{R} \rightarrow \mathbb{R}$? La siguiente definición es general, e incluye el caso que todos conocemos:

Definición 2.1.2. Sea $f : \text{Dom}(f) \subset \mathbb{R}^n \rightarrow \mathbb{R}$. El **gráfico de f** es el subconjunto de \mathbb{R}^{n+1} formado por los pares ordenados $(X, f(X))$, donde $X \in \text{Dom}(f)$. Lo denotamos $\text{Gr}(f) \subset \mathbb{R}^{n+1}$.

Así, por ejemplo, el gráfico de una función $f : \text{Dom}(f) \subset \mathbb{R} \rightarrow \mathbb{R}$ es un subconjunto del plano \mathbb{R}^2 , que muchas veces si f no es muy complicada podemos dibujar. Por ejemplo el gráfico de $f(x) = (x-1)^2 - 1$ es una parábola, mientras que el de $f(x) = \frac{x+1}{x-1}$ es una hipérbola.



Pensemos un poco en el caso $n = 2$, donde según la definición, el gráfico será un subconjunto de \mathbb{R}^3 . Por ejemplo, si $f(x, y) = 3$ (¡una función constante!), se tiene (ya que $\text{Dom}(f) = \mathbb{R}^2$)

$$\text{Gr}(f) = \{(x, y, 3) : x, y \in \mathbb{R}\}.$$

Como puntos del espacio, tienen la única restricción de tener $z = 3$, mientras que x e y pueden moverse libremente. Es decir, el gráfico se representa simplemente como un plano horizontal puesto a altura $z = 3$.

Si tomamos un ejemplo un poquito más complicado, $g(x, y) = x + y$, se tiene

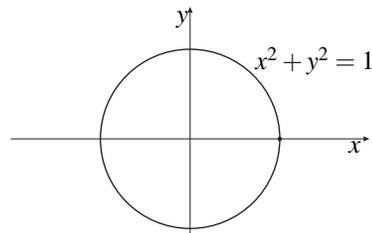
$$\text{Gr}(g) = \{(x, y, x + y) : x, y \in \mathbb{R}\}.$$

Los puntos del gráfico verifican entonces la relación $z = x + y$, lo que nos dice que el gráfico lo podemos representar en \mathbb{R}^3 como un plano por el origen, de normal $N = (1, 1, -1)$ (puesto que $x + y - z = 0$).

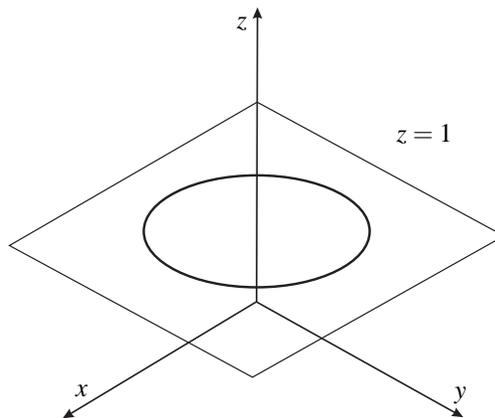
Veamos un par de ejemplos más interesantes. Consideremos $h(x, y) = x^2 + y^2$, y también $j(x, y) = \sqrt{x^2 + y^2}$. Ambas tienen como dominio \mathbb{R}^2 . Los puntos del gráfico de h verifican $z = x^2 + y^2$, mientras que los del gráfico de j verifican $z = \sqrt{x^2 + y^2}$. ¿Cómo podemos descubrir qué figuras representan en \mathbb{R}^3 los puntos que verifican estas relaciones?

Se recurre a las *curvas de nivel*. Esto quiere decir que tratamos de pensar qué figura queda si cortamos el gráfico con distintos planos. Por ejemplo, si consideramos planos horizontales, es decir $z = \text{cte}$, se observa que en ambos casos no puede ser z negativo. Dicho de otra forma, la curva de nivel que corresponde a z negativo es el conjunto vacío.

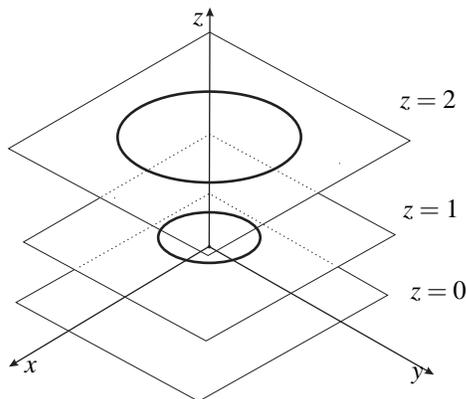
Esto nos dice que los gráficos de ambas funciones están por encima del plano xy de \mathbb{R}^3 . Por otro lado, si pensamos en valores positivos de z , como por ejemplo $z = 1$, obtenemos en ambos casos la relación $1 = x^2 + y^2$. Esta figura es una circunferencia. Estrictamente hablando, la curva de nivel es una circunferencia en el plano \mathbb{R}^2 .



Sin embargo, hemos descubierto que si cortamos cualquiera de los dos gráficos con un plano de altura 1, queda formada una circunferencia de radio 1, lo que nos permite dibujar lo siguiente:



Probemos ahora con $z = 4$. En el caso de h obtenemos $x^2 + y^2 = 4$, que es una circunferencia de radio 2, mientras que en el caso de j obtenemos $x^2 + y^2 = 16$, que es una circunferencia de radio 4.

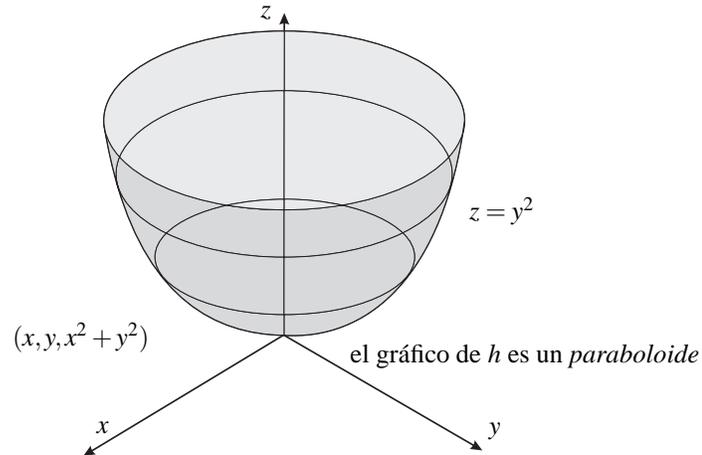


En general, si $z = z_0$ es un número positivo fijo, la ecuación $x^2 + y^2 = z_0$ es una circunferencia de radio $\sqrt{z_0}$, mientras que la ecuación $x^2 + y^2 = z_0^2$ es una circunferencia de radio z_0 . Pareciera que las dos figuras, salvo el ancho, son muy parecidas. Incluso el caso $z = 0$ en ambos casos da la ecuación $x^2 + y^2 = 0$, cuya única solución es el par $(0, 0)$.

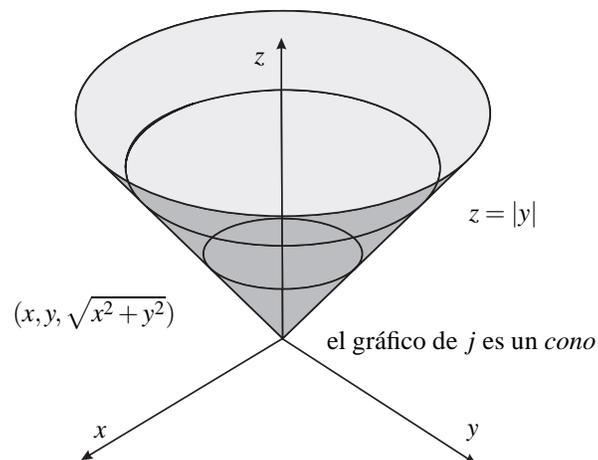
Sin embargo, la técnica no se agota aquí. No necesariamente tenemos que cortar con planos horizontales.

¡Podemos elegir cualquier plano!

Por ejemplo, si tomamos $x = 0$ estamos considerando la pared formada por el plano yz . En el caso de la función h se obtiene la ecuación $0^2 + y^2 = z$. Es decir, $z = y^2$, que es una parábola. Por lo tanto el gráfico de h es lo que se conoce como *paraboloide*



Por otro lado, en el caso de la función j se obtiene $0^2 + y^2 = z^2$. Esto es $z^2 - y^2 = 0$, que se factoriza como diferencia de cuadrados $(z - y)(z + y) = 0$. En consecuencia debe ser la recta $z = y$ o bien la recta $z = -y$. Recordemos que tenía que ser $z \geq 0$, con lo cual la figura que se obtiene es la siguiente,



que es lo que se conoce como *cono*. Miremos la diferencia de las dos figuras en el origen: ciertamente van a tener distintas propiedades cuando pensemos en las derivadas.

Superficies de nivel

En el caso de funciones $f : \mathbb{R}^3 \rightarrow \mathbb{R}$, el gráfico es un subconjunto de \mathbb{R}^4 . Para dibujarlo hay que recurrir a sustancias ilegales. Sin embargo, si recurrimos a la técnica de antes se pueden dibujar lo que se conocen como **superficies de nivel** de f , que nos dan una idea del comportamiento de f .

Definición 2.1.3. Dada $f : \text{Dom}(f) \subset \mathbb{R}^n \rightarrow \mathbb{R}$ y un número real c , se denomina **superficie de nivel de f** al subconjunto del dominio dado por

$$S_c(f) = \{(x_1, \dots, x_n) \in \text{Dom}(f) : f(x_1, \dots, x_n) = c\}.$$

Por ejemplo, consideremos $f(x, y, z) = x^2 + y^2 - z^2$. La superficie de nivel que corresponde a $c = 0$ es el conjunto $\{(x, y, z) \in \mathbb{R}^3 : z^2 = x^2 + y^2\}$, que como vimos antes, es un cono (aunque hay que destacar que en este caso el gráfico se extiende en forma simétrica a ambos lados del plano $z = 0$).

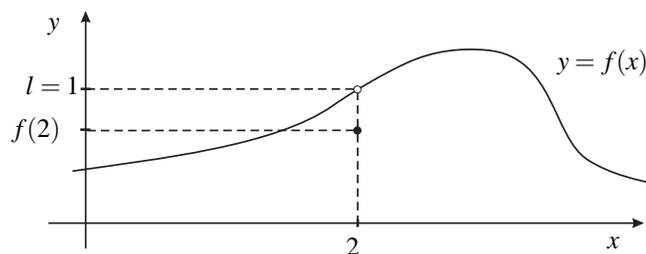
2.1.4. Límite

Pasemos ahora a hablar de límite de funciones. Escribimos la definición, que contiene al caso conocido cuando $n = 1$.

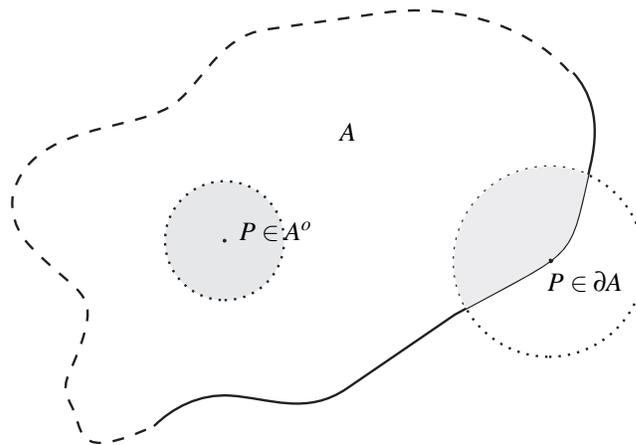
Definición 2.1.4. Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $P \in \mathbb{R}^n$. Decimos que $l \in \mathbb{R}$ es el límite de f cuando X tiende a P (y lo anotamos $\lim_{X \rightarrow P} f(X) = l$) si para todo $\varepsilon > 0$ existe $\delta > 0$ tal que

$$0 < \|X - P\| < \delta \text{ implica } |f(X) - l| < \varepsilon.$$

Notar que la definición de límite excluye al punto P en cuestión. Así por ejemplo, el límite de la función real de la figura de abajo, cuando $x \rightarrow 2$, es igual a 1, independientemente de cuánto valga f en $x_0 = 2$ (ni siquiera tiene por qué estar definida en $x_0 = 2$):



Observación 2.1.5. En el caso en el que el dominio de la función no sea todo \mathbb{R}^n (esto es usual), se puede razonar de la siguiente manera. Lo que queremos ver es a qué valor se aproxima $f(X)$ cuando X se aproxima a P , para una $f : A \rightarrow \mathbb{R}$ (con $A \subset \mathbb{R}^n$). Por lo tanto tiene sentido considerar como P no sólo a cualquier punto del conjunto A donde f esté definida, sino a **cualquier punto de acumulación** A . Es decir, es lícito tomar $P \in \bar{A}$. Hay que hacer la salvedad de que nos aproximemos con puntos donde se pueda evaluar f .



Con esto, se obtiene la definición de límite que cubre todos los casos relevantes:

Definición 2.1.6. Sea $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$. Sean $P \in \bar{A}$ y l un número real cualquiera. Decimos que $\lim_{X \rightarrow P} f(X) = l$ en A si para todo $\varepsilon > 0$ existe $\delta > 0$ tal que

$$0 < \|X - P\| < \delta \text{ y } X \in A \text{ implican } |f(X) - l| < \varepsilon.$$

Álgebra de límites, propiedades

En general probar a mano que un límite existe es muy engorroso (aunque a veces no queda otra). Sin embargo, si uno establece algunas propiedades que permitan desarmar el límite en límites más simples, esto facilita la tarea. Este es el objetivo del álgebra de límites.

Antes, una observación: si tenemos dos funciones distintas, pueden estar definidas en subconjuntos distintos A, B de \mathbb{R}^n . En consecuencia la suma, resta, producto y división de ellas sólo va a tener sentido en la intersección $A \cap B$ de estos conjuntos. Y los puntos donde tiene sentido calcular el límite tienen que ser entonces puntos de la clausura de $A \cap B$, de acuerdo a lo discutido en la Observación 2.1.5. Además el cociente sólo se puede hacer en aquellos puntos donde no se anula el denominador, que es la intersección de A con $B - C_0(g)$.

Vamos a suponer que ya hicimos la intersección de dominios, con lo cual f, g están definidas en el mismo conjunto $A \subset \mathbb{R}^n$.

Teorema 2.1.7. Sean $f, g : A \rightarrow \mathbb{R}$, dos funciones, con $A \subset \mathbb{R}^n$. Sea $P \in \bar{A}$, y supongamos que

$$\lim_{X \rightarrow P} f(X) = l_1 \in \mathbb{R} \quad \text{y} \quad \lim_{X \rightarrow P} g(X) = l_2 \in \mathbb{R}.$$

Entonces

1. $\lim_{X \rightarrow P} (f \pm g)(X) = l_1 \pm l_2$ en A .

2. $\lim_{X \rightarrow P}(f \cdot g)(X) = l_1 \cdot l_2$ en A .
3. Si $l_2 \neq 0$, $\lim_{X \rightarrow P}\left(\frac{f}{g}\right)(X) = \frac{l_1}{l_2}$ en $A - C_0(g)$.

Demostración. Veamos 1, la suma. Dado $\varepsilon > 0$, existe $\delta_1 > 0$ tal que $|f(X) - l_1| < \frac{\varepsilon}{2}$ siempre que $0 < \|X - P\| < \delta_1$ y $X \in A$ (pues $\lim_{X \rightarrow P} f(X) = l_1$). Análogamente, existe $\delta_2 > 0$ tal que $|g(X) - l_2| < \frac{\varepsilon}{2}$ siempre que $0 < \|X - P\| < \delta_2$ y $X \in A$ (pues $\lim_{X \rightarrow P} g(X) = l_2$). En consecuencia, si $X \in A$ y $0 < \|X - P\| < \delta = \min\{\delta_1, \delta_2\}$, entonces

$$|(f + g)(X) - (l_1 + l_2)| \leq |f(X) - l_1| + |g(X) - l_2| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Lo que prueba que $\lim_{X \rightarrow P}(f + g)(X) = l_1 + l_2$. La resta es análoga.

Vemos 2, el producto. Tengamos presente que podemos conseguir que $|f(X) - l_1|$ y $|g(X) - l_2|$ sean tan chicos como nosotros queramos, por hipótesis. Escribimos

$$|fg(X) - l_1 l_2| = |f(X)g(X) - f(X)l_2 + f(X)l_2 - l_1 l_2| \leq |f(X)||g(X) - l_2| + |f(X) - l_1||l_2|.$$

El único término que no es evidente que está acotado es $|f(X)|$. Pero de nuevo por la desigualdad triangular,

$$|f(X)| \leq |f(X) - l_1| + |l_1|,$$

con lo cual

$$|fg(X) - l_1 l_2| \leq |f(X) - l_1|(|g(X) - l_2| + |l_2|) + |l_1||g(X) - l_2|. \quad (2.1)$$

En consecuencia, como l_1 y l_2 están fijos, y las cantidades $|f(X) - l_1|$ y $|g(X) - l_2|$ se pueden hacer tan chicas como uno quiera (eligiendo convenientemente el entorno de P), se obtiene que $|fg(X) - l_1 l_2|$ es tan chico como uno quiera. Más precisamente, consideremos los dos casos $l_1 = 0$, $l_1 \neq 0$.

1. $l_1 = 0$. Entonces $l_1 l_2 = 0$, y por la cuenta (2.1) de arriba,

$$|fg(X) - l_1 l_2| = |f(x)g(X)| \leq |f(X)|(|g(X) - l_2| + |l_2|).$$

Dado $\varepsilon > 0$, elegimos $\delta_2 > 0$ tal que $0 < \|X - P\| < \delta_2$ implique $|g(X) - l_2| < 1$, y elegimos $\delta_1 > 0$ tal que $0 < \|X - P\| < \delta_1$ implique $|f(X)| < \frac{\varepsilon}{(1+|l_2|)}$. Si tomamos $\delta = \min\{\delta_1, \delta_2\}$, entonces

$$|fg(X) - l_1 l_2| < \frac{\varepsilon}{(1+|l_2|)}(1+|l_2|) = \varepsilon.$$

2. $l_1 \neq 0$. Dado $\varepsilon > 0$, elegimos $\delta_2 > 0$ tal que $0 < \|X - P\| < \delta_2$ implique $|g(X) - l_2| < \frac{\varepsilon}{2|l_1|}$, y elegimos $\delta_1 > 0$ tal que $0 < \|X - P\| < \delta_1$ implique $|f(X) - l_1| < \frac{\varepsilon}{2(\frac{\varepsilon}{2|l_1|} + |l_2|)}$. Si tomamos $\delta = \min\{\delta_1, \delta_2\}$, nuevamente por la cuenta (2.1) de arriba,

$$|fg(X) - l_1 l_2| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Veamos 3, el cociente. Observemos primero que f/g tiene sentido en el conjunto $A - C_0(g)$ pues removimos los ceros de g . Sacando denominadores comunes tenemos

$$|f/g(X) - l_1/l_2| = \frac{1}{|l_2||g(X)|} |f(X)l_2 - g(X)l_1|.$$

El numerador se puede hacer tan chico como uno quiera pues

$$|f(X)l_2 - g(X)l_1| = |f(X)l_2 - l_1l_2 + l_1l_2 - g(X)l_1| \leq |l_2||f(X) - l_1| + |l_1||l_2 - g(X)|.$$

Falta acotar el primer factor por alguna constante, es decir, falta ver que existe $M > 0$ tal que $\frac{1}{|g(X)|} < M$ si $0 < \|X - P\| < \delta$ y $X \in A - C_0(g)$. Para simplificar la demostración, supongamos que $l_2 > 0$ (el caso $l_2 < 0$ es análogo). Como $\lim_{X \rightarrow P} g(X) = l_2$, se tiene $|g(X) - l_2| < l_2/2$ para algún $\delta_0 > 0$ conveniente, es decir

$$-l_2/2 < g(X) - l_2 < l_2/2,$$

de donde se deduce que $l_2/2 < g(X)$ si $0 < \|X - P\| < \delta_0$ y $X \in A - C_0(g)$. En particular $g(X) > 0$ en este entorno de P , y además $\frac{1}{|g(X)|} = \frac{1}{g(X)} < \frac{2}{l_2}$, como queríamos. Con esto, llamando $M = \frac{2}{l_2}$ tenemos

$$\frac{1}{|g(X)||l_2|} < M$$

siempre que $0 < \|X - P\| < \delta_0$.

Ahora, como f tiende a l_1 y g tiende a l_2 , se tiene

$$|f(X) - l_1||l_2| \rightarrow 0 \text{ y } |g(X) - l_2||l_1| \rightarrow 0$$

por la propiedad del producto que ya probamos. Entonces existen $\delta_1, \delta_2 > 0$ tal que $0 < \|X - P\| < \min\{\delta_1, \delta_2\}$ implica

$$|f(X) - l_1||l_2| < \frac{\epsilon}{2M} \text{ y } |g(X) - l_2||l_1| < \frac{\epsilon}{2M}.$$

En consecuencia, si $\delta = \min\{\delta_0, \delta_1, \delta_2\}$, se tiene

$$|f/g(X) - l_1/l_2| \leq \frac{1}{|l_2||g(X)|} (|l_2||f(X) - l_1| + |l_1||g(X) - l_2|) < M \left(\frac{\epsilon}{2M} + \frac{\epsilon}{2M} \right) = \epsilon.$$

□

Observación 2.1.8. Una observación importante y muy sencilla: si $f_j : \mathbb{R}^n \rightarrow \mathbb{R}$ es una de las funciones coordenadas (la que se queda con la j -ésima coordenada), es decir

$$f_j(x_1, x_2, \dots, x_j, \dots, x_n) = x_j,$$

entonces si $P = (p_1, \dots, p_n)$, se tiene $\lim_{X \rightarrow P} f_j(X) = \lim_{X \rightarrow P} x_j = p_j$.

Por ejemplo, si $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ es la proyección a la segunda coordenada $f(x, y, z) = y$, entonces se tiene $\lim_{(x,y,z) \rightarrow (2,3,-1)} f(x, y, z) = \lim_{(x,y,z) \rightarrow (2,3,-1)} y = 3$.

Este hecho se demuestra a partir de la definición de límite observando que

$$|f_j(X) - p_j| = |x_j - p_j| \leq \sqrt{(x_1 - p_1)^2 + \dots + (x_n - p_n)^2} = \|X - P\|,$$

con lo cual si $\|X - P\|$ es pequeño, entonces $|f_j(X) - p_j|$ es pequeño.

Como comentario adicional (para confundir) observemos que de acuerdo a la notación que usamos para sucesiones, cuando queríamos señalar la j -ésima coordenada de una sucesión $\{P_k\}$ escribíamos $(P_k)_j$. Ahora podemos decir que $(P_k)_j = f_j(P_k)$ (¡para pensarlo, es una tontería de la notación nomás!).

Ejemplo 2.1.9. *Juntando el Teorema 2.1.7 con la Observación 2.1.8, se deduce que tomar límite en las funciones que involucran sumas, restas, productos y cocientes de las coordenadas, se traduce simplemente en evaluar las funciones en el punto dado (siempre y cuando el denominador no se anule). Así por ejemplo*

$$\lim_{(x,y,z) \rightarrow (1,3,-4)} \frac{xy + z^2 + y}{3x + z + y^2} = \frac{13 + 16 + 3}{3 - 4 + 9} = \frac{32}{8} = 4.$$

2.2. Funciones $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$

Una función de este tipo está dada por una tira de funciones $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ (con $i = 1 \dots m$), es decir

$$F(x_1, x_2, \dots, x_n) = (f_1(x_1, x_2, \dots, x_n), f_2(x_1, x_2, \dots, x_n), \dots, f_m(x_1, x_2, \dots, x_n)).$$

La definición de límite es enteramente análoga, simplemente hay que tener en cuenta que ahora los valores de F son vectores y no números. Consideremos $A \subset \mathbb{R}^n$ como siempre.

Definición 2.2.1. *Sea $F : A \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$. Sean $P \in \bar{A}$ y $L \in \mathbb{R}^m$. Decimos que $\lim_{X \rightarrow P} F(X) = L$ en A si para todo $\varepsilon > 0$ existe $\delta > 0$ tal que*

$$0 < \|X - P\| < \delta \text{ y } X \in A \text{ implican } \|F(X) - L\| < \varepsilon.$$

2.2.1. Composición

Podemos componer funciones de la manera obvia. Por ejemplo si por un lado $F(x, y) = (x + e^y, xy, \cos(xy))$ y por otro lado $G(x, y, z) = (x + y + z, xy \ln(y))$, se tiene

$$(G \circ F)(x, y) = (x + e^y + xy + \cos(xy), (x + e^y)xy \ln(xy)),$$

con las restricciones de dominio que corresponda en cada caso. Por ejemplo aquí

$$\text{Dom}(G \circ F) = \{(x, y) \in \mathbb{R}^2 : xy > 0\}.$$

En general hay que tener cuidado de que se pueda hacer la composición. Además de poder encadenar

$$\mathbb{R}^n \xrightarrow{F} \mathbb{R}^m \xrightarrow{G} \mathbb{R}^k$$

con el espacio \mathbb{R}^m intermedio, hay que recordar que el dominio de una composición, si $F : A \subset \mathbb{R}^m \rightarrow \mathbb{R}^m$ y $G : B \subset \mathbb{R}^m \rightarrow \mathbb{R}^k$ es

$$\text{Dom}(G \circ F) = \{X \in A : F(X) \in B\}.$$

Y tenemos la propiedad correspondiente del álgebra de límites:

Teorema 2.2.2. Sean $F : A \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ y $G : B \subset \mathbb{R}^m \rightarrow \mathbb{R}^k$.

Si $P \in \bar{A}$ es tal que $\lim_{X \rightarrow P} F(X) = L_1 \in B$, y $\lim_{Y \rightarrow L_1} G(Y) = L_2$, entonces existe el límite de la composición $\lim_{X \rightarrow P} (G \circ F)(X)$ en $\text{Dom}(G \circ F)$, y además $\lim_{X \rightarrow P} (G \circ F)(X) = L_2$, siempre y cuando $F(X) \neq L_1$ para $X \neq P$.

Demostración. Dado $\varepsilon > 0$, existe $\delta > 0$ tal que

$$\|(G \circ F)(X) - L_2\| = \|G(F(X)) - L_2\| < \varepsilon$$

siempre y cuando $0 < \|F(X) - L_1\| < \delta$ y $F(X) \in B$, por ser $\lim_{Y \rightarrow L_1} G(Y) = L_2$ (simplemente llamando $Y = F(X)$).

Pero dado $\delta > 0$ existe $\delta' > 0$ tal que $0 < \|F(X) - L_1\| < \delta$ siempre que $0 < \|X - P\| < \delta'$ y $X \in A$, pues $\lim_{X \rightarrow P} F(X) = L_1$ en A , y $F(X) \neq L_1$ si $X \neq P$.

Luego $0 < \|X - P\| < \delta'$ y $X \in \text{Dom}(G \circ F) \subset A$ garantizan $\|(G \circ F)(X) - L_2\| < \varepsilon$. \square

Observación 2.2.3. La condición $F(X) \neq L_1$ está puesta para que no ocurran situaciones ridículas como la siguiente: sea $f : \mathbb{R} \rightarrow \mathbb{R}$ la función nula (es decir $f(x) = 0$ para todo $x \in \mathbb{R}$) y sea

$$g(x) = \begin{cases} 1 & \text{si } x \neq 0 \\ 0 & \text{si } x = 0 \end{cases}$$

Entonces $g \circ f(x) = 0$ para todo $x \in \mathbb{R}$, con lo cual

$$\lim_{x \rightarrow 0} (g \circ f)(x) = 0.$$

Pero el límite de g no es cero, pues

$$\lim_{x \rightarrow 0} g(x) = \lim_{x \rightarrow 0} 1 = 1.$$

Con el teorema previo, se pueden calcular algunos límites de funciones $g : \mathbb{R}^n \rightarrow \mathbb{R}$ de manera muy sencilla, si uno descubre la composición.

Ejemplo 2.2.4. 1. Como $\lim_{(x,y,z) \rightarrow (0,0,0)} x^2 + y + z = 0$, entonces

$$\lim_{(x,y,z) \rightarrow (0,0,0)} \frac{\text{sen}(x^2 + y + z)}{x^2 + y + z} = 1.$$

2. Como $\lim_{(x,y,z) \rightarrow (3,2,1)} e^{xy} \cos(x^2) \ln(z) = 0$, se tiene

$$\lim_{(x,y,z) \rightarrow (3,2,1)} \left(1 + e^{xy} \cos(x^2) \ln(z)\right)^{1/e^{xy} \cos(x^2) \ln(z)} = e.$$

2.2.2. Curvas y el límite

Un caso particular (y bastante importante) de función $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ son las curvas, que es lo que se tiene cuando $n = 1$.

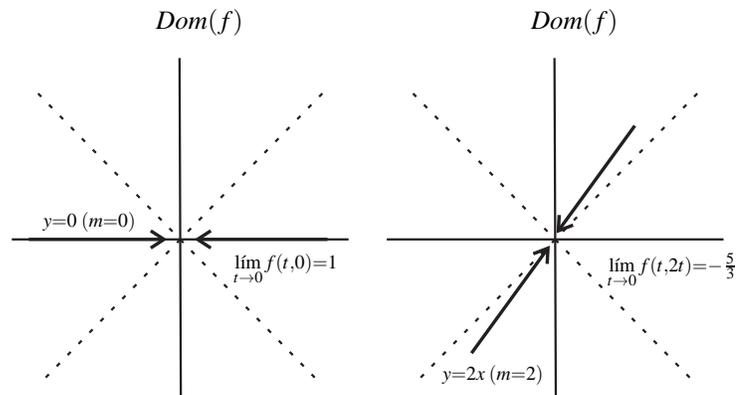
Que exista el límite de una función $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ a lo largo de una trayectoria, no garantiza que el límite vaya a existir. Pero, de acuerdo al Teorema 2.2.2, si el límite existe entonces el límite de la composición tiene que dar lo mismo, **independientemente** de con qué curva uno componga. Tenemos entonces la siguiente herramienta útil:

Observación 2.2.5. Sea $F : A \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$. Sean $\alpha : I \rightarrow A$ y $\beta : J \rightarrow A$ dos curvas en \mathbb{R}^n tales que $\lim_{t \rightarrow t_0} \alpha(t) = P$ y $\lim_{t \rightarrow t_1} \beta(t) = P$. Supongamos además que $\alpha(t) \neq P$ para $t \neq t_0$ y $\beta(t) \neq P$ para $t \neq t_2$. Entonces $\lim_{t \rightarrow t_0} F(\alpha(t)) \neq \lim_{t \rightarrow t_1} F(\beta(t))$ implica que **no** existe el límite $\lim_{X \rightarrow P} F(X)$.

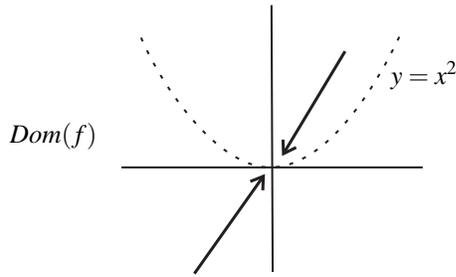
Ejemplo 2.2.6. Sea $f(x,y) = \frac{x^2+y^2}{x^2-y^2}$, y $P = (0,0)$. Consideremos la curva $\alpha(t) = (t, mt)$, que es una recta de pendiente $m \in \mathbb{R}$ por el origen. Observemos que el denominador de f se anula a lo largo de las rectas $y = x$, $y = -x$, por lo que hay que excluir las posibilidades $m = \pm 1$. Calculemos el límite de la composición, para cada $m \neq \pm 1$:

$$\lim_{t \rightarrow 0} f(\alpha(t)) = \lim_{t \rightarrow 0} \frac{t^2 + (mt)^2}{t^2 - (mt)^2} = \lim_{t \rightarrow 0} \frac{t^2}{t^2} \frac{1+m^2}{1-m^2} = \frac{1+m^2}{1-m^2}.$$

Esto prueba que el límite $\lim_{(x,y) \rightarrow (0,0)} f(x,y)$ no existe, pues si nos acercamos a lo largo del eje x (con $m = 0$), el límite de la composición da 1, mientras que si nos acercamos a lo largo de la recta $y = 2x$ (con $m = 2$), el límite de la composición da $-5/3$.



Ejemplo 2.2.7. Una función tal que $\text{Dom}(g) \neq \mathbb{R}^2$.



Sea $g(x,y) = \frac{x^2+y}{x^2-y}$, $P = (0,0)$. Consideremos $\alpha(t) = (t, mt)$ como antes. El denominador de f se anula en la parábola $y = x^2$, pero si t es chico ($y \neq 0$) la recta $y = mx$ no corta la parábola.

Para cada $m \in \mathbb{R}$:

$$\lim_{t \rightarrow 0} f(\alpha(t)) = \lim_{t \rightarrow 0} \frac{t^2 + mt}{t^2 - mt} = \lim_{t \rightarrow 0} \frac{t}{t} \left(\frac{t+m}{t-m} \right) = -1.$$

¡Esto no nos sirve! ¿Podría ser que el límite exista? Probemos con algunas trayectorias más. Nos faltó considerar, entre las rectas, el eje y . Para esto tomamos $\beta(t) = (0, t)$. Se tiene

$$\lim_{t \rightarrow 0} f(\beta(t)) = \lim_{t \rightarrow 0} \frac{0+t}{0-t} = -1$$

también; seguimos sin saber si el límite existe o no. Mirando la función, vemos que el numerador se anula a lo largo de la curva $y = -x^2$. Esto nos da una pista, tomamos $\gamma(t) = (t, -t^2)$. Entonces

$$\lim_{t \rightarrow 0} f(\gamma(t)) = \lim_{t \rightarrow 0} \frac{t^2 - t^2}{t^2 + t^2} = \lim_{t \rightarrow 0} 0 = 0.$$

Esto prueba que el límite $\lim_{(x,y) \rightarrow (0,0)} f(x,y)$ no existe.

2.2.3. Límite y sucesiones

¿Se pueden usar sucesiones para calcular límites?

Proposición 2.2.8. Sea $F : A \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$. Sean $P \in \bar{A}$ y $L \in \mathbb{R}^m$. Las siguientes afirmaciones son equivalentes:

1. $\lim_{X \rightarrow P} F(X) = L$
2. Para **toda** sucesión de puntos $P_k \in A$ tal que $P_k \neq P$ y $P_k \rightarrow P$, se tiene $\lim_k F(P_k) = L$.

Demostración. $1 \Rightarrow 2$) Supongamos que $\lim_{X \rightarrow P} F(X) = L$, y tomemos P_k una sucesión de puntos del A que tiende a P . Entonces por la definición de límite de función dado $\varepsilon > 0$ existe $\delta > 0$ tal que

$$0 < \|X - P\| < \delta \text{ y } X \in A \text{ implican } \|F(X) - L\| < \varepsilon.$$

Como $P_k \neq P$, claramente vale $0 < \|P_k - P\|$. Si tomamos $k_0 \in \mathbb{N}$ tal que $k \geq k_0$ implique $\|P_k - P\| < \delta$, el renglón de arriba nos garantiza que $\|F(P_k) - L\| < \varepsilon$. O sea $F(P_k) \rightarrow L$ de acuerdo a la definición de límite de sucesiones, aplicado a la sucesión $Q_k = F(P_k)$.

$2 \Rightarrow 1$) Razonemos por el absurdo. Negar que $\lim_{X \rightarrow P} F(X) = L$ es decir que existe $\varepsilon > 0$ tal que para todo $\delta > 0$ existe $X \in A$ con $0 < \|X - P\| < \delta$ y $\|F(X) - L\| \geq \varepsilon$. Tomemos $\delta = \frac{1}{k}$, con $k \in \mathbb{N}$, y sea P_k el punto correspondiente. Vemos que $P_k \rightarrow P$ y $P_k \neq P$, pero por otro lado $\|F(P_k) - L\| \geq \varepsilon$, con lo cual $F(P_k) \not\rightarrow L$, contradiciendo la hipótesis. \square

Es decir que "usar sucesiones" no es muy práctico porque no basta probar con una. Sin embargo, usar sucesiones sí sirve para ver que el límite **no** existe. La idea es la misma que la de las trayectorias.

Observación 2.2.9. Sean $\{P_k\}$ y $\{P'_k\}$ dos sucesiones de puntos de A , tales que ambas tienden a P , y las sucesiones $Q_k = F(P_k)$ y $Q'_k = F(P'_k)$ tienen límites distintos, entonces **no existe** el límite $\lim_{X \rightarrow P} F(X)$.

Aquí hay que hacer la salvedad, al igual que en el caso de las curvas, de que esto vale siempre que P_k y P'_k sean distintos de P , al menos para k suficientemente grande.

Ejemplo 2.2.10. Si $f(x, y) = \frac{x+y}{x-y}$, se tiene $\text{Dom}(f) = \{(x, y) : x \neq y\}$. Veamos que no existe

$$\lim_{(x,y) \rightarrow (0,0)} f(x, y).$$

Tomando $P_k = (\frac{1}{k}, 0)$ y $P'_k = (0, \frac{1}{k})$ se observa que ambas son sucesiones del dominio que tienden a cero. Pero $\lim_k f(P_k) = \lim_k \frac{1/k+0}{1/k-0} = 1$, y por otro lado $\lim_k f(P'_k) = \lim_k \frac{0+1/k}{0-1/k} = -1$.

Y de la Proposición 2.2.8 también se desprende el siguiente hecho más o menos obvio, que nos permite mirar cada coordenada por separado.

Corolario 2.2.11. Sea $F : A \rightarrow \mathbb{R}^m$, con $A \subset \mathbb{R}^n$. Sean $L \in \mathbb{R}^m$, $P \in \bar{A}$. Entonces F tiende a L cuando $X \rightarrow P$ si y sólo si cada coordenada de F converge a la coordenada correspondiente de L . Es decir, si $F = (f_1, f_2, \dots, f_m)$ y $L = (l_1, l_2, \dots, l_m)$, entonces

$$\lim_{X \rightarrow P} F(X) = L \iff \left(\lim_{X \rightarrow P} f_j(X) = l_j \text{ para cada } j = 1 \dots m \right).$$

Y desde este momento siempre que tengamos una función $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$, siempre podemos chequear cada coordenada por separado.

Límites infinitos

Se definen en forma análoga al caso real los límites "en infinito" y los límites "que dan infinito".

Definición 2.2.12. Límites infinitos. Sea $F : A \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$, $L \in \mathbb{R}^m$ y $P \in \bar{A}$.

1. Decimos que $\lim_{X \rightarrow \infty} F(X) = L$ en A si para **toda** sucesión $\{X_k\}$ de A que tiende a infinito se tiene $\lim_k F(X_k) = L$.

Equivalentemente, dado $\varepsilon > 0$ existe $M > 0$ tal que

$$\|X\| > M \text{ y } X \in A \text{ implican } \|F(X) - L\| < \varepsilon.$$

2. Decimos que $\lim_{X \rightarrow P} F(X) = \infty$ en A (o que F **diverge** cuando $X \rightarrow P$) si para cualquier sucesión $\{X_k\}$ de A tal que $X_k \rightarrow P$, se tiene $\lim_k F(X_k) = \infty$.

Equivalentemente, dado $M > 0$ existe $\delta > 0$ tal que

$$0 < \|X - P\| < \delta \text{ y } X \in A \text{ implican } \|F(X)\| \geq M.$$

2.3. Continuidad

La noción de continuidad es idéntica al caso conocido en una variable. Necesitamos que tomar límite coincida con evaluar.

Definición 2.3.1. Sean $F : A \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ y $P \in A$.

1. Decimos que F es **continua en P** si $\lim_{X \rightarrow P} F(X) = F(P)$.
2. Decimos que F es **continua en A** si F es continua en P para todo $P \in A$.

Probar que una función es continua en P , es asegurarse de que la diferencia $\|F(X) - F(P)\|$ se puede hacer tan chica como uno quiera, pidiendo que $\|X - P\|$ sea pequeño.

Observación 2.3.2. Es importante observar que, a diferencia de los límites en general, el punto P debe estar en el conjunto A donde F está definida, para poder calcular $F(P)$. Eso no quita que, dado $P \in \partial A$, podamos preguntarnos si la función se puede extender (redefinir) en el punto P de manera que quede continua. Esto lo podremos hacer si y sólo si el límite $\lim_{X \rightarrow P} F(X)$ existe en A . En ese caso diremos que la discontinuidad en P es **evitable**.

Observación 2.3.3. Puesto que el límite se calcula en cada coordenada por separado, se tiene que $F = (f_1, \dots, f_m)$ es continua en $P = (p_1, \dots, p_m)$ si y sólo si cada $f_j : A \rightarrow \mathbb{R}$ es continua en P , y también que F es continua en A si cada f_j es continua en A .

Dado que el límite de una composición es la composición de los límites (Teorema 2.2.2),

Teorema 2.3.4. Sean $F : A \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ y $G : B \subset \mathbb{R}^m \rightarrow \mathbb{R}^k$ funciones continuas. Entonces $G \circ F$ es una función continua en $\text{Dom}(G \circ F) = \{X \in A : F(X) \in B\}$.

2.3.1. Propiedades de las funciones continuas

En toda esta sección, $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$, y $P \in A$.

Probemos un par de propiedades sencillas que nos van a resultar útiles más adelante. La primera dice que si $f(X) \geq 0$, entonces $\lim f(X) \geq 0$. La segunda dice que si una función continua es no nula, hay un entorno donde no se anula, y se deduce de la primera.

Lema 2.3.5. Supongamos que f es continua en P . Entonces

1. Si $\{X_k\}$ es una sucesión tal que $X_k \rightarrow_k P$ en A , y $f(X_k) \leq 0$ para todo $k \in \mathbb{N}$, entonces $f(P) \leq 0$.
2. Si $\{X_k\}$ es una sucesión tal que $X_k \rightarrow_k P$ en A , y $f(X_k) \geq 0$ para todo $k \in \mathbb{N}$, entonces $f(P) \geq 0$.
3. Si $f(P) > 0$, entonces existe $r > 0$ tal que si $U = B_r(P) \cap A$, se tiene $f(X) > 0$ para todo $X \in U$.
4. Si $f(P) < 0$, entonces existe $r > 0$ tal que si $U = B_r(P) \cap A$, se tiene $f(X) < 0$ para todo $X \in U$.

Demostración. El ítem 1. lo probamos por el absurdo. Si fuera $f(P) > 0$, tomamos $\varepsilon = \frac{f(P)}{2}$, y existe entonces $k_0 \in \mathbb{N}$ tal que $k \geq k_0$ implica $|f(X_k) - f(P)| < \frac{f(P)}{2}$. Es decir,

$$-\frac{f(P)}{2} < f(X_k) - f(P) < \frac{f(P)}{2}.$$

Despejando del lado izquierdo se obtiene $f(X_k) > \frac{f(P)}{2} > 0$ si $k \geq k_0$ lo cual contradice la hipótesis.

El ítem 2. tiene una demostración análoga, es un buen ejercicio escribirlo.

El ítem 3. también lo probamos por el absurdo: supongamos que para todo $k \in \mathbb{N}$ existe $X_k \in B_{\frac{1}{k}}(P) \cap A$ tal que $f(X_k) \leq 0$. Entonces $X_k \rightarrow P$ en A y por ser f continua en P , $0 < f(P) = \lim_k f(X_k) \leq 0$, lo cual es imposible (en la última desigualdad usamos el ítem previo). La demostración de 4. es análoga. \square

Definición 2.3.6. 1. Un conjunto $A \subset \mathbb{R}^n$ es **acotado** si es acotado en norma, es decir, existe $M > 0$ tal que $\|X\| \leq M$ para todo $X \in A$.

2. Un conjunto $K \subset \mathbb{R}^n$ cerrado y acotado es un conjunto **compacto**.

3. Un conjunto $A \subset \mathbb{R}^n$ es **arcoconexo** si dados $P, Q \in A$ existe una curva continua $\alpha: [0, 1] \rightarrow A$ tal que $\alpha(0) = P$ y $\alpha(1) = Q$.

Observación 2.3.7. Por la desigualdad

$$|x_i| \leq \|X\| \leq \sqrt{n} \max_{i=1..n} |x_i|$$

se deduce que un conjunto A es acotado si y sólo si cada una de sus coordenadas está acotada.

Recordemos el teorema del valor medio en una variable:

Teorema 2.3.8. (Bolzano) Si $f: [a, b] \rightarrow \mathbb{R}$ es continua, y $f(a)f(b) < 0$, entonces existe $c \in (a, b)$ tal que $f(c) = 0$.

Demostración. Supongamos que $f(a) > 0$ y $f(b) < 0$. Sea $A = \{x \in [a, b] : f(x) > 0\}$, que es no vacío (pues $a \in A$). Como $A \subset [a, b]$, está acotado superiormente así que tiene supremo $s = \sup A$. Existe una sucesión creciente $\{a_n\}$ de puntos de A que tiende a $s \in [a, b]$. Como $f(a_n) > 0$ (pues $a_n \in A$), el ítem 1 del Lema 2.3.5 nos dice que debe ser $f(s) \geq 0$. Afirmo que $f(s) = 0$. Si no fuera así, sería $f(s) > 0$. Pero entonces el ítem 3 del mismo lema nos dice que hay un entorno de s de la pinta $(s-r, s+r) \cap [a, b]$, tal que $f(x) > 0$ allí.

Como $s < b$ pues $f(b) < 0$, podemos tomar $x_0 \in [a, b]$ tal que $x_0 \in (s, s+r)$. Este $x_0 \in A$ y es mayor que el supremo, lo que es imposible. Debe ser pues $f(s) = 0$. \square

Teorema 2.3.9. (Bolzano) Sea $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$. Supongamos que existen $P, Q \in A$ tales que $f(P) < 0$ y $f(Q) > 0$. Si A es arcoconexo y f es continua, entonces existe $R \in A$ tal que $f(R) = 0$.

Demostración. Por ser A arcoconexo existe $\alpha : [0, 1] \rightarrow A$ continua con $\alpha(0) = P$ y $\alpha(1) = Q$. Consideremos $g : [0, 1] \rightarrow \mathbb{R}$ la composición $g(t) = f \circ \alpha(t)$, que es una función continua por ser composición de funciones continuas. Se tiene $g(0) = f(P) < 0$ y $g(1) = f(Q) > 0$. Por el teorema del valor medio en una variable, existe $c \in (0, 1)$ tal que $g(c) = 0$. Si $R = \alpha(c)$, entonces $R \in A$ y además $f(R) = f(\alpha(c)) = g(c) = 0$. \square

Y se tienen los corolarios habituales,

Corolario 2.3.10. Sea $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$ continua. Supongamos que A es arcoconexo. Entonces

1. Dados $P, Q \in A$, f toma todos los valores intermedios entre $f(P)$ y $f(Q)$.
2. Sean $P \in A$ y $Q \in \bar{A}$. Si $\lim_{X \rightarrow Q} f(X) = l$ y $f(P) < l$, entonces $[f(P), l) \subset \text{Im}(f)$.

Demostración. Para probar 1, supongamos $f(P) < f(Q)$ (si son iguales no hay nada que probar). Tomemos $c \in [f(P), f(Q)]$ y consideramos $f_c : A \rightarrow \mathbb{R}$ dada por $f_c(X) = f(X) - c$. Entonces $f_c(P) = f(P) - c < 0$ y $f_c(Q) = f(Q) - c > 0$. Además f_c es continua, así que por el teorema anterior existe $R \in A$ tal que $f_c(R) = 0$. Esto es $f(R) - c = 0$, es decir $f(R) = c$.

Para probar 2, tomemos $c \in (f(P), l)$. Observemos que como $\lim_{X \rightarrow Q} f(X) = l$, existe una sucesión X_k de puntos de A tales que, dado $\varepsilon = l - c > 0$, si $k \geq k_0$ entonces $|f(X_k) - l| < \varepsilon = l - c$. Desarmando este módulo se tiene $-(l - c) < f(X_k) - l < l - c$, y considerando sólo el lado izquierdo se tiene $c < f(X_k)$. Tomemos $Q' = X_{k_0} \in A$. Entonces $f(Q') > c$, y como $f(P) < c$ por el ítem 1. se tiene que $c \in \text{Im}(f)$. \square

Teorema 2.3.11. (Weierstrass) Si A es un conjunto compacto de \mathbb{R}^n , y f es continua en A , entonces existen $m, M \in \mathbb{R}$ tales que $m \leq f(X) \leq M$ para todo $X \in A$. Además, existen $P_m, P_M \in A$ tales que $f(P_m) = \min\{f(X) : X \in A\}$ y $f(P_M) = \max\{f(X) : X \in A\}$.

Demostración. Supongamos primero que f no es acotada. Si no existe M tal que $f(X) \leq M$, entonces existe una sucesión de puntos $X_k \in A$ tal que $f(X_k) \geq k$ para todo $k \in \mathbb{N}$. Como A es cerrado y acotado, $\{X_k\}$ tiene una subsucesión convergente $\{X_{k_j}\}$, tal que $X_{k_j} \rightarrow_j P \in A$. Se tiene $f(X_{k_j}) \geq k_j$, y como f es continua en P , esto es imposible. De la misma manera se prueba que existe $m \in \mathbb{R}$ tal que $m \leq f(X)$.

Ahora veamos que f alcanza sus valores máximo y mínimo en A . Como $\text{Im}(f) \subset [m, M]$, $\text{Im}(f)$ es un conjunto acotado de \mathbb{R} . En particular es acotado superiormente, luego tiene supremo s . Veamos que en realidad es un máximo, es decir veamos que existe $P_M \in A$ tal que $f(P_M) = s$. Para ello, tomamos una sucesión creciente de puntos de $\text{Im}(f)$ que tienda al supremo. Tenemos entonces que existe una sucesión de puntos X_k en A tales que $\lim_k f(X_k) = s$. De la sucesión $\{X_k\}$ extraemos una subsucesión convergente $\{X_{k_j}\}$, y llamamos al límite P_M (que es un punto de A). Se tiene $f(P_M) = \lim_j f(X_{k_j}) = s$ por ser f continua. De forma análoga se prueba que f alcanza su valor mínimo, construyendo una sucesión de puntos que tienda al ínfimo de $\text{Im}(f)$. \square

Corolario 2.3.12. Sea $F : A \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ continua. Si A es compacto, entonces $F(A) \subset \mathbb{R}^m$ es compacto.

Demostración. Como $F = (f_1, \dots, f_m)$, y cada f_i es continua, por el teorema anterior cada coordenada del conjunto $F(A)$ está acotada, y en consecuencia $F(A)$ es acotado. Por otro lado, si $Q \in \overline{F(A)}$, existe $X_k \in A$ una sucesión de puntos tal que $F(X_k) \rightarrow Q$ en \mathbb{R}^m . Extraemos de X_k una subsucesión convergente $\{X_{k_j}\}$ a un punto $P \in A$. Se tiene, por la continuidad de F ,

$$Q = \lim_j F(X_{k_j}) = F(\lim_j X_{k_j}) = F(P).$$

Esto prueba que $Q \in F(A)$, y por lo tanto $F(A)$ es cerrado. \square

2.3.2. Caracterización de las funciones continuas

Se tiene la siguiente caracterización F es continua si y sólo si la preimagen de cualquier abierto es un abierto relativo, con más precisión:

Teorema 2.3.13. Sea $F : A \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$. Entonces F es continua si y sólo si para todo abierto $W \subset \mathbb{R}^m$, el conjunto

$$F^{-1}(W) = \{X \in A : F(X) \in W\}$$

se puede escribir como la intersección $A \cap U$ de un conjunto $U \subset \mathbb{R}^n$ abierto, con el conjunto A .

Demostración. Supongamos que F es continua, tomemos W abierto en \mathbb{R}^m . Tomemos P tal que $F(P) \in W$, es decir, $P \in F^{-1}(W)$. Como W es abierto, existe $r_P > 0$ tal que $B_{r_P}(F(P)) \subset W$. Como F es continua, dado este $r_P > 0$ existe $\delta_P > 0$ tal que $\|X - P\| < \delta_P$, $X \in A$, implican $\|F(X) - F(P)\| < r_P$. Es decir, existe $\delta_P > 0$ tal que $X \in A \cap B_{\delta_P}(P)$ implica $F(X) \in B_{r_P}(F(P)) \subset W$. Dicho de otra forma, $F(B_{\delta_P}(P) \cap A) \subset W$. Como esto se puede hacer para cualquier $P \in F^{-1}(W)$, podemos tomar $U = \cup_{P \in F^{-1}(W)} B_{\delta_P}(P)$ que por ser de unión de abiertos es abierto, y se tiene

$$F^{-1}(W) = U \cap A$$

como queríamos.

La otra implicación es similar, queda como ejercicio. \square

2.3.3. Continuidad uniforme

Para terminar, una definición y un teorema

Definición 2.3.14. Decimos que una función $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ es **uniformemente continua** si dado $\varepsilon > 0$, existe $\delta > 0$ tal que $X, Y \in A$ y $\|X - Y\| < \delta$ implican $\|f(X) - f(Y)\| < \varepsilon$.

La definición nos dice que pequeños movimientos en el dominio provocan pequeños movimientos en la imagen, independientemente de en qué punto hagamos estos movimientos.

Observación 2.3.15. Una función uniformemente continua es continua, puesto que dado $P \in A$, y $\varepsilon > 0$, se tiene por la definición que existe $\delta > 0$ tal que $\|X - P\| < \delta$ implica $\|f(X) - f(P)\| < \varepsilon$. Es decir, $\lim_{X \rightarrow P} f(X) = f(P)$.

Observemos sin embargo, que es habitual para funciones continuas el hecho siguiente: dado ε , el δ necesario para conseguir $\|f(X) - f(P)\| < \varepsilon$ **puede depender no sólo de ε , sino también del punto P en cuestión**. Un ejemplo sencillo de función continua que no es uniformemente continua es el siguiente.

Ejemplo 2.3.16. Tomemos $f(x) = x^2$ en el intervalo $(0, +\infty)$. Entonces si f fuera uniformemente continua, dado $\varepsilon = 2$ existiría $\delta > 0$ tal que si $x, y > 0$ y $|x - y| < \delta$, entonces $|x^2 - y^2| < 2$. Pero si tomamos $y = \frac{2}{\delta}$, $x = y + \frac{\delta}{2}$ (que verifican las hipótesis), entonces

$$|x^2 - y^2| = \left| \left(y + \frac{\delta}{2}\right)^2 - y^2 \right| = 2 + \frac{\delta^2}{4} \geq \varepsilon = 2.$$

Sin embargo, en dominios cerrados y acotados, no hay distinción, pues

Teorema 2.3.17. (Heine-Cantor) Si $F : A \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ es continua y A es compacto, entonces F es uniformemente continua.

Demostración. Si F no fuera uniformemente continua, existiría $\varepsilon > 0$ tal que para cada $k \in \mathbb{N}$ existirían X_k e Y_k en A tales que $\|X_k - Y_k\| < \frac{1}{k}$, pero

$$\|F(X_k) - F(Y_k)\| \geq \varepsilon. \quad (2.2)$$

Como A es compacto, se tiene una subsucesión $\{X_{k_j}\}$ convergente a un punto $X \in A$. Observemos que

$$\|Y_{k_j} - X\| \leq \|Y_{k_j} - X_{k_j}\| + \|X_{k_j} - X\| < \frac{1}{k_j} + \|X_{k_j} - X\|,$$

luego debe ser también $Y_{k_j} \rightarrow X$. Como F es continua, $\lim_j F(X_{k_j}) = \lim_j F(Y_{k_j}) = F(X)$. Pero esto contradice la desigualdad (2.2). \square

2.4. Problemas

2.I. Graficar las curvas de nivel de las siguiente funciones y hacer un dibujo aproximado de $Gr(f) \subset \mathbb{R}^3$.

- $f(x, y) = x^2 - y^2$
- $f(x, y) = 2x + 3y$

2.II. Sea $P \in A \subset \mathbb{R}^n$ y $f : A \rightarrow \mathbb{R}$ una función continua en P .

- Suponiendo que $\{X_k\}$ es una sucesión de puntos en A que tiende a P , y $f(X_k) \geq 0$ para todo $k \in \mathbb{N}$, probar que $f(P) \geq 0$.
- Suponiendo que $f(P) < 0$, probar que existe $r > 0$ tal que si $U = B_r(P) \cap A$, entonces se tiene $f(X) < 0$ para todo $X \in U$.

2.III. Sea $f : A \rightarrow \mathbb{R}^k$ con $A \subset \mathbb{R}^n$ arcoconexo. Probar que si f es continua, entonces $Im(f) = f(A) \subset \mathbb{R}^k$ es arcoconexo. En particular, si $k = 1$, la imagen de f debe ser un intervalo.

2.IV. Sea $K \subset \mathbb{R}^n$ un compacto conexo. Si $f : K \rightarrow \mathbb{R}$ es continua, probar que $Im(f)$ es un intervalo cerrado $[a, b]$. ¿Quiénes son a, b ?

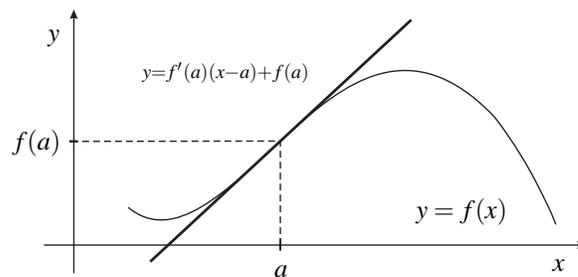
2.V. Sea $F : A \subset \mathbb{R}^n \rightarrow \mathbb{R}^k$. Supongamos que para todo abierto $W \subset \mathbb{R}^k$, existe $U \subset \mathbb{R}^n$ abierto tal que $F^{-1}(W) = A \cap U$. Probar que f es continua.

Considero aquí las magnitudes matemáticas no como compuestas de partes extraordinariamente pequeñas, sino como descritas por el movimiento continuo

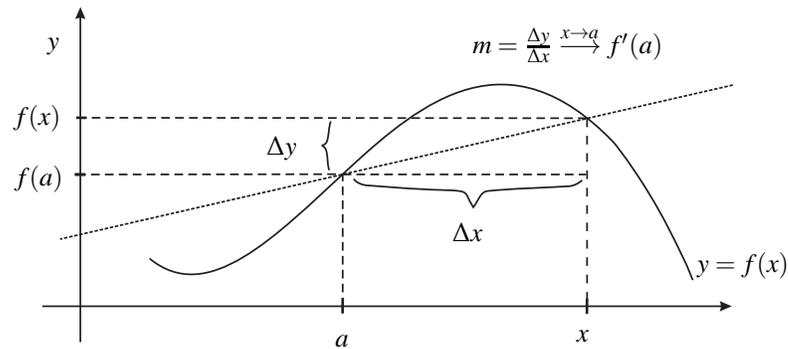
Sir Isacc Newton

3.1. Derivadas

Empecemos repasando la derivada de una función de una variable. Recordemos la idea de “recta que mejor aproxima”, que según la figura de abajo, es la recta tangente al gráfico de f .



Para armar una recta, hacen falta dos puntos distintos del plano. O un punto del plano y la pendiente. Entonces fijando $(a, f(a))$, hallamos $m = f'(a)$ de la siguiente manera. Tomamos otro punto $x \in \text{Dom}(f)$, construimos la recta secante



La pendiente de esta secante depende de x y a , es

$$\frac{\Delta y}{\Delta x} = \frac{f(x) - f(a)}{x - a}$$

Si hacemos tender x al punto a , se obtiene (si existe el límite) la pendiente de la recta deseada, y por lo tanto las pendientes de arriba tienden a la pendiente de la recta tangente, con lo cual

$$f'(a) = \lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a}.$$

O sea existe la derivada si y sólo si existe el límite de los cocientes incrementales. Y decimos que " f es derivable en a ". Con el cambio de variable $h = x - a$ se puede escribir

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

Recordemos también la ecuación de la recta tangente al gráfico de f en $x = a$, que es

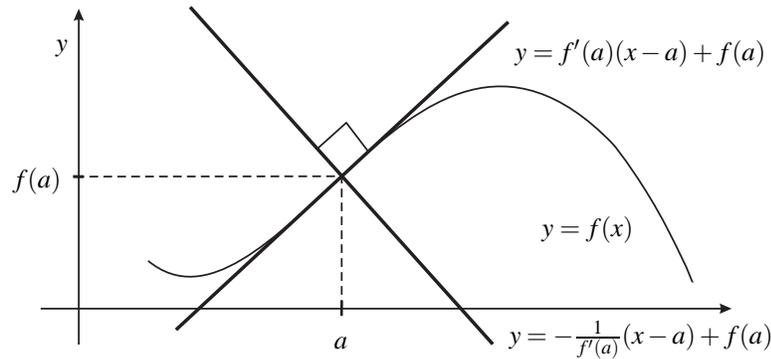
$$y = f'(a)(x - a) + f(a),$$

que en forma paramétrica es simplemente

$$L_a : (a, f(a)) + \lambda(1, f'(a)),$$

puesto que la pendiente de la recta tangente debe ser $\frac{\Delta y}{\Delta x} = \frac{f'(a)}{1} = f'(a)$. Esto nos dice que un incremento en una unidad en el dominio, nos produce un incremento de $f'(a)$ unidades en la imagen.

Por su parte, la recta normal pasa por el mismo punto según indica el gráfico



pero su vector director debe ser perpendicular al de la tangente, con lo cual

$$N_a : (a, f(a)) + \lambda(f'(a), -1).$$

3.1.1. Curvas

El siguiente caso es el de las curvas o trayectorias, es decir, funciones $\alpha : I \rightarrow \mathbb{R}^n$, donde $I \subset \mathbb{R}$ en general es algún intervalo. En este caso tenemos n -funciones reales, todas moviéndose al unísono con el mismo parámetro, y la derivada se define de la manera obvia, es decir, derivando cada coordenada. Solamente vamos a derivar en el interior del dominio, como es usual.

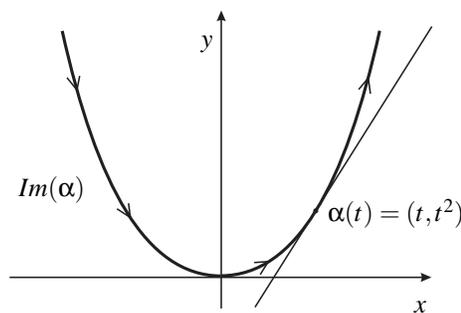
Atención que en este caso nos interesa la imagen de la curva, y no su gráfico.

Definición 3.1.1. Sea $\alpha : (a, b) \rightarrow \mathbb{R}^n$. La **derivada** o **velocidad** de α en t_0 es el vector que se obtiene derivando cada coordenada en $t = t_0$ (siempre que existan todas las derivadas).

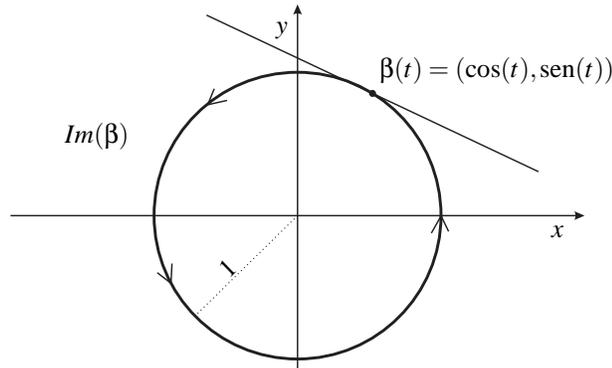
La **derivada** de α es la curva α' que se obtiene derivando cada coordenada de α .

Por ejemplo, si $\alpha(t) = (t, t^2)$, se tiene $\alpha'(0) = (1, 0)$ y en general $\alpha'(t) = (1, 2t)$ para todo $t \in \mathbb{R}$. En cambio si $\beta(t) = (\cos(t), \sin(t))$ se tiene $\beta'(t) = (-\sin(t), \cos(t))$ y en particular $\beta'(0) = (0, 1)$.

¿Qué representa la velocidad de una curva? Es un vector de \mathbb{R}^n , y representa la trayectoria rectilínea que seguiría una partícula si se soltara del alambre que representa α . Por ejemplo, como α describe una parábola $y = x^2$, se tiene



Y como β describe una circunferencia unitaria $x^2 + y^2 = 1$, se observa



Por supuesto que la derivada en realidad representa el vector director, que nosotros dibujamos pinchado en el punto correspondiente. La recta tangente a la curva tiene en general ecuación

$$L_{t_0} : \alpha(t_0) + \lambda \alpha'(t_0). \quad (3.1)$$

Hay que tener presente que el gráfico de una función $f : \mathbb{R} \rightarrow \mathbb{R}$, que es el conjunto

$$Gr(f) = \{(x, f(x)) : x \in Dom(f)\} \subset \mathbb{R}^2$$

se puede parametrizar siempre con la curva $\alpha(x) = (x, f(x))$, donde $\alpha : Dom(f) \rightarrow \mathbb{R}^2$. Esta curva será derivable en a si y sólo si f es derivable en a , y su recta tangente coincide por supuesto con la recta tangente al gráfico de f en el punto $(a, f(a))$.

Observación 3.1.2. Cambiemos el punto de vista. Supongamos que queremos dar la ecuación de la recta tangente al gráfico de $f : \mathbb{R} \rightarrow \mathbb{R}$, pero en forma paramétrica, es decir como en la ecuación (3.1). Entonces se tiene en cuenta que el gráfico de f se puede parametrizar con la curva $\alpha(x) = (x, f(x))$ con $x \in Dom(f)$ como mencionamos antes. Y por lo tanto, si f es derivable, la derivada es $\alpha'(x) = (1, f'(x))$. Con lo cual

$$L_a : \alpha(a) + \lambda \alpha'(a) = (a, f(a)) + \lambda(1, f'(a)).$$

Es fundamental observar que la primer coordenada del vector derivada es constantemente uno, mientras que la segunda coordenada expresa la razón de cambio, pues $\Delta y / \Delta x = f'(a) / 1 = f'(a)$. Es decir, que por cada unidad que avanzamos en el eje de las x , se avanza $f'(a)$ unidades en el eje de las y . Esto último, por supuesto, es una aproximación, y sólo es exacta esta afirmación cuando f es una recta, con lo cual coincide con su recta tangente.

Resumiendo, el vector derivada nos dice cuánto hay que moverse en y por cada unidad que nos movemos en x , si nos movemos desde el punto $(a, f(a))$, pero a lo largo de la recta tangente al gráfico de f en ese punto.

Se deduce de lo anterior que un vector normal al gráfico de f es $(f'(a), -1)$, pues como

$$\langle (1, f'(a)), (f'(a), -1) \rangle = f'(a) - f'(a) = 0,$$

se concluye que son ortogonales. Observemos que esta escritura es válida incluso en el caso $f'(a) = 0$, donde se obtiene el vector normal $(0, 1)$.

3.1.2. Derivadas direccionales

Supongamos que tenemos una función $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, como por ejemplo

$$f(x,y) = \sqrt{1 - x^2 - y^2}.$$

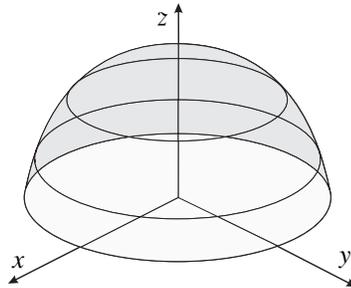
El dominio de esta función es la bola cerrada unitaria en el plano, centrada en el origen

$$\text{Dom}(f) = \overline{B}_1(\mathbf{0}) = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}.$$

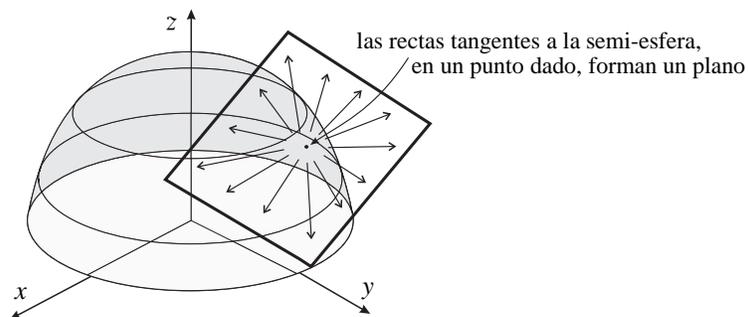
Si miramos las curvas de nivel, $z = cte$, debe ser $z \geq 0$ pues z está igualado a una raíz cuadrada. Es decir el gráfico está por sobre el piso. Además como $z_0^2 = 1 - x^2 - y^2$, se tiene $x^2 + y^2 = 1 - z_0^2$ lo que nos indica que debe ser $z_0 \leq 1$. Las curvas de nivel son circunferencias como se puede observar, pero podemos decir más: como $x^2 + y^2 + z_0^2 = 1$, esto nos dice que

$$\|(x, y, z_0)\| = 1.$$

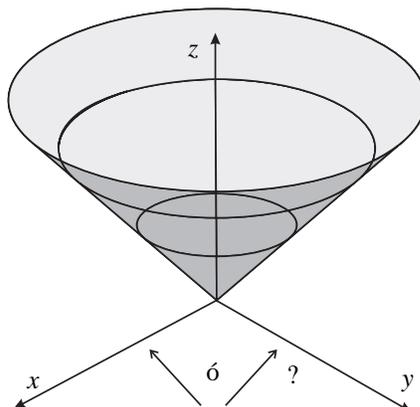
Entonces los puntos están sobre la cáscara de la esfera unitaria centrada en el origen. Pero es sólo el casquete superior, pues z debe ser positivo. El gráfico de f es entonces una media esfera



Pensemos qué pasa si nos movemos a lo largo de una trayectoria en la esfera. Si en algún momento la fuerza que nos sostiene atados a la superficie se rompe, lo que ocurriría es que saldríamos disparados en línea recta, en forma tangente a la esfera (¿es esto intuitivo?). Pero dado un punto P en la esfera, hay muchas direcciones tangentes. Se puede ver que forman un plano, que se denomina *plano tangente*



Ahora tomemos $g(x,y) = \sqrt{x^2 + y^2}$. Las curvas de nivel son nuevamente circunferencias, el gráfico sólo existe para $z \geq 0$. Cortando con los planos $x = 0$ e $y = 0$ se obtienen las ecuaciones $z = |y|$, $z = |x|$ respectivamente. Estamos en presencia de un cono, como ya discutimos con anterioridad.



Si nos movemos en esta superficie cerca de un punto que no sea el vértice, nuevamente direcciones tangentes forman (al menos eso imaginamos) un plano. Exceptuando el origen. Allí, si uno dibuja algunas rectas tangentes, lo que observa es nuevamente un cono. Es decir, no hay plano tangente en el origen. Es más, si uno mira cuidadosamente una dirección y ambos sentidos, obtiene distintas rectas como en el caso de la función real módulo $f(x) = |x|$. De hecho, en el vértice *no hay* rectas tangentes, ya que, como sabemos, la función $f(x) = |x|$ no tiene derivada en el origen. Dicho de otra manera, las derivadas laterales

$$\lim_{x \rightarrow 0^+} \frac{f(x) - f(0)}{x - 0} = 1 \quad \text{y} \quad \lim_{x \rightarrow 0^-} \frac{f(x) - f(0)}{x - 0} = -1$$

no coinciden, por más que cada una de ellas existe por separado.

Pongamos un poco de precisión a estas ideas gráficas.

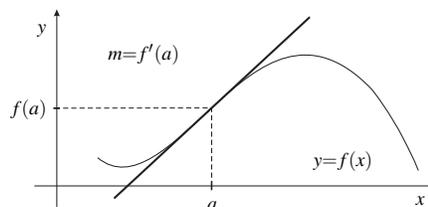
Definición 3.1.3. Sea $V \in \mathbb{R}^n$, $\|V\| = 1$. Sea $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$, y $P \in A^\circ$. El límite

$$\lim_{t \rightarrow 0} \frac{f(P + tV) - f(P)}{t}$$

(si existe) es la **derivada direccional de f en P en la dirección de V** . Lo anotamos

$$\frac{\partial f}{\partial V}(P) \text{ o bien } f_V(P).$$

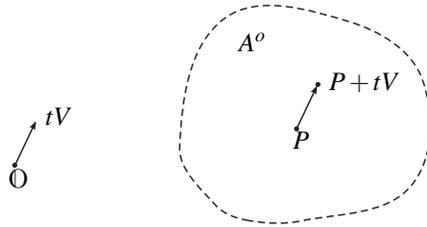
Observemos que $f_V(P)$ (si existe) es un número real. ¿Qué pasó con nuestra intuición gráfica? Es sencillo, basta recordar el caso $n = 1$. Allí, la definición nos da en el punto $x = a$ un número real que representa la pendiente de la recta tangente al gráfico de f en el punto $(a, f(a))$.



La recta tangente la armamos sabiendo la pendiente y el punto por el que pasa. La situación aquí es similar. El punto por el que pasa lo tenemos, lo que nos falta es relacionar la derivada direccional $\frac{\partial f}{\partial V}(P)$ con el vector director de la recta.

Conviene introducir la siguiente notación. Usaremos (X, x_{n+1}) para denotar al vector de \mathbb{R}^{n+1} formado por el vector $X = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ en las primeras n coordenadas, y el número x_{n+1} en la última.

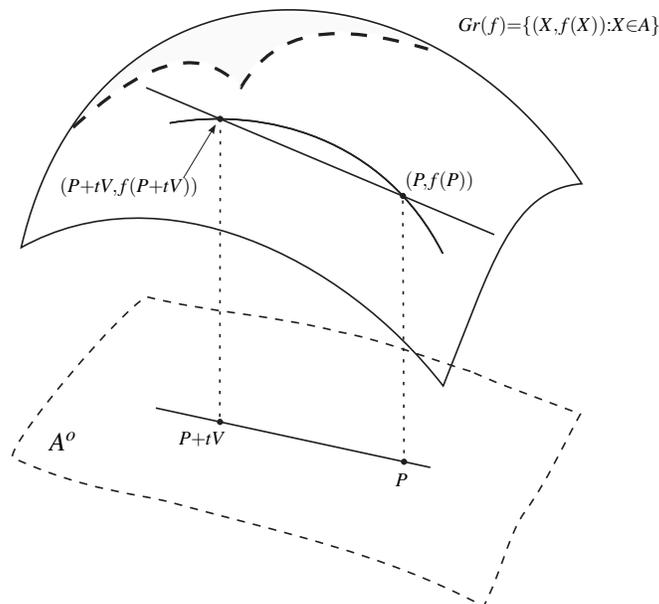
Si $P \in A^o$ y $\|V\| = 1$, en el dominio de f el vector $P + tV$ se representa así:



Y lo que que estamos haciendo cuando calculamos

$$f(P+tV) - f(P)$$

es calcular la diferencia de alturas de f según el siguiente gráfico:



Si dividimos por t y hacemos tender t a cero, tendremos la derivada direccional en la dirección de V . Antes de hacerlo, pensemos cual es la ecuación paramétrica de las rectas secantes. El punto por el que pasan todas es P . Y el vector director es, como se observa del gráfico, el vector que une los puntos $(P, f(P))$ con $(P+tV, f(P+tV))$. Entonces el vector que nos interesa (el vector director de la recta tangente) es

$$\lim_{t \rightarrow 0} \frac{1}{t} [(P+tV, f(P+tV)) - (P, f(P))] = \lim_{t \rightarrow 0} \left(\frac{P+tV - P}{t}, \frac{f(P+tV) - f(P)}{t} \right),$$

es decir

$$\left(V, \frac{\partial f}{\partial V}(P) \right).$$

Esta ecuación, aunque la dedujimos del gráfico de una función particular, es completamente general, y nos dice que la recta tangente se arma formando el vector de \mathbb{R}^{n+1} dado por V en las primeras n coordenadas y la derivada direccional (que es un número) en la última. Es decir su ecuación paramétrica es

$$L_{P,V} : (P, f(P)) + \lambda(V, f_V(P)).$$

Observación 3.1.4. Observemos que en el caso de una función $f : \mathbb{R} \rightarrow \mathbb{R}$, dado un punto del dominio $P = a \in \mathbb{R}$, hay sólo dos vectores de norma unitaria para considerar, que son el $V = 1$ y $V = -1$. Si usamos $V = 1$, obtenemos

$$\frac{\partial f}{\partial V}(P) = \lim_{t \rightarrow 0} \frac{f(a+t) - f(a)}{t} = f'(a),$$

con lo cual

$$\left(V, \frac{\partial f}{\partial V}(P) \right) = (1, f'(a)),$$

que es el vector que obtuvimos al escribir la recta tangente en forma paramétrica. Mientras que si usamos $V = -1$, obtenemos

$$\frac{\partial f}{\partial V}(P) = \lim_{t \rightarrow 0} \frac{f(a+t(-1)) - f(a)}{t} = \lim_{t \rightarrow 0} \frac{f(a-t) - f(a)}{t}.$$

En esta última cuenta podemos hacer el cambio de variable $t = -h$ con lo que se obtiene

$$\frac{\partial f}{\partial V}(P) = \lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{-h} = -f'(a).$$

Con lo cual

$$\left(V, \frac{\partial f}{\partial V}(P) \right) = (-1, -f'(a)),$$

pero la recta tangente es la misma por ser este último vector múltiplo del primero. Esto podríamos haberlo deducido sin hacer ninguna cuenta, del sólo hecho de que si f es derivable entonces hay una sola recta tangente.

3.1.3. Derivadas parciales

Vamos a denotar con E_i ($i = 1 \dots n$) a los vectores de la base canónica de \mathbb{R}^n , es decir $E_1 = (1, 0, 0, \dots, 0)$, $E_2 = (0, 1, 0, \dots, 0)$, \dots , $E_n = (0, 0, \dots, 1)$. Para abreviar, vamos a usar la siguiente notación.

Definición 3.1.5. Sea $E_i \in \mathbb{R}^n$, $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$, $P \in A^\circ$. La i -ésima derivada parcial de f en P es la derivada direccional de f en la dirección de E_i , y la denotamos $\frac{\partial f}{\partial x_i}(P)$ o bien f_{x_i} . Es decir

$$f_{x_i}(P) = \lim_{t \rightarrow 0} \frac{f(P + tE_i) - f(P)}{t}.$$

Es decir, las derivadas parciales son las derivadas en las direcciones de los ejes de coordenadas. Así, por ejemplo, si $f(x, y) = 3x^2 + 4y$ y $P = (0, 0)$

$$\frac{\partial f}{\partial x}(0, 0) = \lim_{t \rightarrow 0} \left[\frac{1}{t} 3(0 + t)^2 + 0 - 0 \right] = \lim_{t \rightarrow 0} 3t = 0,$$

mientras que

$$\frac{\partial f}{\partial y}(0, 0) = \lim_{t \rightarrow 0} \frac{1}{t} [3 \cdot 0^2 + 4t - 0] = \lim_{t \rightarrow 0} 4 = 4.$$

Si prestamos atención a la definición de derivada direccional, vemos que lo único que hacemos es poner un incremento en la coordenada i -ésima. Por ejemplo si $n = 3$ y tomamos $i = 2$, se tiene

$$\frac{\partial f}{\partial y}(p_1, p_2, p_3) = \lim_{t \rightarrow 0} \frac{f(p_1, p_2 + t, p_3) - f(p_1, p_2, p_3)}{t}.$$

Si llamamos $g(y) = f(x, y, z)$, lo que tenemos es una función de una sola variable, y de acuerdo a la definición de derivada para funciones $g: \mathbb{R} \rightarrow \mathbb{R}$,

$$g'(p_2) = \lim_{h \rightarrow 0} \frac{g(p_2 + h) - g(p_2)}{h} = \frac{\partial f}{\partial y}(p_1, p_2, p_3).$$

Es decir que una derivada parcial es simplemente una derivada respecto de una variable dada, que se calcula haciendo de cuenta que todas las demás variables son constantes. Y podemos calcularla en cualquier punto usando las reglas generales

Por ejemplo, si $f(x, y) = 3x^2 + 4y$, entonces $\frac{\partial f}{\partial x}(x, y) = 6x$ para cualquier $(x, y) \in \mathbb{R}^2$, y también $\frac{\partial f}{\partial y}(x, y) = 4$.

Hay que hacer la salvedad de que las derivadas existan, y por supuesto, de que los puntos estén en el dominio. El resultado exacto es el siguiente:

Proposición 3.1.6. Sea $f: A \subset \mathbb{R}^n \rightarrow \mathbb{R}$, $P = (p_1, \dots, p_n) \in A^\circ$. La función **derivada parcial de f respecto de x_i** es la función $\frac{\partial f}{\partial x_i}$ que se obtiene de f de la siguiente manera

$$\frac{\partial f}{\partial x_i}(p_1, \dots, p_n) = \lim_{t \rightarrow 0} \frac{f(p_1, \dots, p_i + t, \dots, p_n) - f(p_1, \dots, p_n)}{t}$$

siempre que el límite exista. Se tiene $\text{Dom}\left(\frac{\partial f}{\partial x_i}\right) \subset A^\circ$ con inclusión estricta si existe algún punto donde no exista el límite.

3.2. Plano tangente y Diferencial

En general, la existencia de una derivada parcial nos dice que en una dirección dada, el gráfico es suave. Pero acercándonos por otras direcciones podría tener saltos, y la función podría no ser continua.

Ejemplo 3.2.1. Un ejemplo a tener en cuenta es el siguiente:

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

Afirmo que existen todas las derivadas direccionales de f en el origen. Tomamos $V = (v_1, v_2) \in \mathbb{R}^2$ (con $\|V\| = 1$), y si $v_2 \neq 0$ calculamos

$$\lim_{t \rightarrow 0} \frac{f(tv_1, tv_2) - f(0, 0)}{t} = \lim_{t \rightarrow 0} \frac{1}{t} \frac{t^3 v_1^2 v_2}{t^2(t^2 v_1^4 + v_2^2)} = \frac{v_1^2 v_2}{v_2^2} = \frac{v_1^2}{v_2}.$$

En particular (con $V = (0, 1)$) se tiene $f_y(0, 0) = 0$. Falta calcular $f_x(0, 0)$, que también da cero puesto que

$$f_x(0, 0) = \lim_{t \rightarrow 0} \frac{f(t, 0) - f(0, 0)}{t} = \lim_{t \rightarrow 0} \frac{0}{t} = 0.$$

Sin embargo, f **no** es continua en $(0, 0)$ pues si consideramos $\alpha(t) = (t, t^2)$, entonces

$$\lim_{t \rightarrow 0} f \circ \alpha(t) = \lim_{t \rightarrow 0} \frac{t^2 t^2}{t^4 + t^4} = \frac{1}{2} \neq 0.$$

¿Cómo podemos recuperar la idea de que una función derivable es una función suave, en particular continua? El último ejemplo es desalentador ya que contradice la intuición, al existir en este caso todas las derivadas direccionales, pero no ser esto garantía de continuidad.

Lo que haremos será recurrir a la idea de aproximación lineal. Para ello, supongamos que tenemos una función $f: \mathbb{R}^n \rightarrow \mathbb{R}$ para la cual existen todas sus derivadas parciales. Lo que nos tenemos que preguntar es **si estas derivadas forman un hiperplano en \mathbb{R}^{n+1} , y si este plano aproxima a la función.**

El plano en cuestión que queremos considerar es, dado $P \in A^o$, el plano generado por todas las derivadas parciales. Es decir el plano en \mathbb{R}^{n+1} que pasa por el punto $(P, f(P))$ y está generado por los siguientes n vectores de \mathbb{R}^n :

$$\{(E_1, f_{x_1}(P)), \dots, (E_n, f_{x_n}(P))\}.$$

¿Cuál es la ecuación de este plano, que como dijimos es un hiperplano en \mathbb{R}^n ?

Recordemos que para dar un hiperplano de \mathbb{R}^{n+1} , basta con dar una ecuación

$$\langle N, Y - Q \rangle = 0$$

donde $N, Q \in \mathbb{R}^{n+1}$ están fijos. Esta ecuación nos dice que el plano tiene normal N y pasa por el punto Q .

Observemos que si ponemos $N_P = (f_{x_1}(P), f_{x_2}(P), \dots, f_{x_n}(P), -1) \in \mathbb{R}^{n+1}$ como normal del plano, esta es perpendicular a todos los vectores generados por las derivadas parciales, pues

$$\begin{aligned} \langle N_P, (E_i, f_{x_i}(P)) \rangle &= \langle (f_{x_1}(P), f_{x_2}(P), \dots, f_{x_n}(P), -1), (0, \dots, 0, 1, 0, \dots, 0, f_{x_i}(P)) \rangle \\ &= f_{x_i}(P) - f_{x_i}(P) = 0. \end{aligned}$$

Entonces la ecuación del plano debe ser

$$\Pi_P : \langle N_P, (X - P, x_{n+1} - f(P)) \rangle = 0$$

pues este plano tiene normal N_P y pasa por $(P, f(P)) = (p_1, \dots, p_n, f(P)) \in \mathbb{R}^{n+1}$.

Despejando se obtiene la ecuación

$$\Pi_P : x_{n+1} = f(P) + f_{x_1}(P)(x_1 - p_1) + f_{x_2}(P)(x_2 - p_2) + \cdots + f_{x_n}(P)(x_n - p_n).$$

Así por ejemplo si $f(x, y) = \sqrt{x^2 + y^2}$, como

$$f_x(x, y) = \frac{x}{\sqrt{x^2 + y^2}} \text{ y } f_y(x, y) = \frac{y}{\sqrt{x^2 + y^2}}, \quad (3.2)$$

dato $P = (x_0, y_0) \neq (0, 0)$ se tiene

$$\Pi_P : z = \sqrt{x_0^2 + y_0^2} + \frac{x_0}{\sqrt{x_0^2 + y_0^2}}(x - x_0) + \frac{y_0}{\sqrt{x_0^2 + y_0^2}}(y - y_0).$$

Por ejemplo si tomamos $P = (3, 4)$ entonces $f(P) = 5$ con lo cual

$$\Pi_P : z = 5 + \frac{3}{5}(x - 3) + \frac{4}{5}(y - 4),$$

es decir $z = 5 + \frac{3}{5}x + \frac{4}{5}y - \frac{9}{5} - \frac{16}{5} = \frac{3}{5}x + \frac{4}{5}y - 5$ o lo que es lo mismo

$$\Pi_P : 3x + 4y - 5z = 25.$$

¿Qué pasa en el vértice, es decir, en $P = (0, 0)$? Se tiene $f(P) = 0$, y las fórmulas (3.2) de las derivadas parciales allí no tienen sentido, pero cuidado que eso no quiere decir que no existan. Las intentamos calcular usando la definición

$$f_x(0, 0) = \lim_{t \rightarrow 0} \frac{1}{t} (\sqrt{t^2 + 0^2} - 0) = \lim_{t \rightarrow 0} \frac{|t|}{t}$$

que no existe y lo mismo ocurre con $f_y(0, 0)$. Por lo tanto en este caso **no** hay plano en el vértice del cono, lo cual nos devuelve un poco de la intuición que queremos desarrollar.

Veamos ahora lo que ocurre con el Ejemplo 3.2.1 de más arriba. Como vimos, f no es continua en el origen. Sin embargo, existían todas las derivadas direccionales y en particular

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

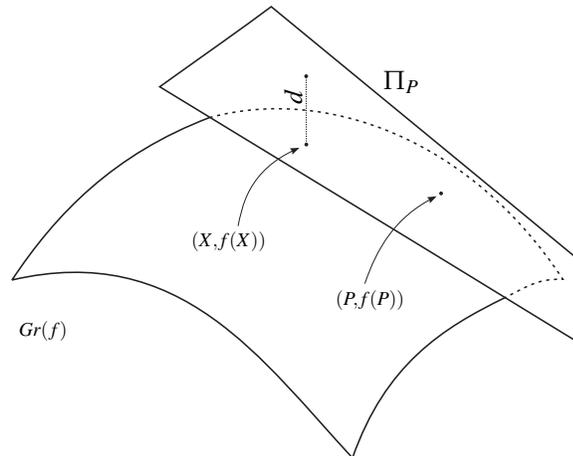
Con esto, el plano tangente en $(0, 0)$ tiene ecuación

$$z - 0 = 0(x - 0) + 0(y - 0)$$

pues $f(0, 0) = 0$. Es decir, hay un plano, y es el plano $z = 0$. ¿Cómo puede ser, si vimos que f no era continua?

Lo que veremos es que el problema aquí no es que no haya un plano en P , sino que este plano no aproxima suficientemente bien a la función cerca de P .

Lo que vamos a pedir es que la distancia entre este plano Π_P y la función f no solamente tienda a cero (lo cual siempre que haya plano ocurre, puesto que el plano y la función se tocan en el punto $(P, f(P))$ y allí la distancia es cero), sino que la distancia $d = d(X)$ entre ambos como se indica en la figura,



tienda a cero **más rápido** que la distancia entre X y P . Es decir, llamando Π_P a una parametrización del plano tangente, vamos a pedir que si $X \rightarrow P$, entonces

$$\frac{\text{dist}(\Pi_P(X), f(X))}{\text{dist}(X, P)} \rightarrow 0.$$

Cuando esto ocurra diremos que f es diferenciable en P . Recordemos que los puntos del plano tangente (si el plano existe) verifican la ecuación

$$x_{n+1} = f(P) + f_{x_1}(P)(x_1 - p_1) + \cdots + f_{x_n}(P)(x_n - p_n).$$

Veamos la definición precisa.

Definición 3.2.2. Sea $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$, $P \in A^\circ$. Si existen las derivadas parciales de f en P , decimos que f es **diferenciable en P** si

$$\lim_{X \rightarrow P} \frac{|f(X) - f(P) - f_{x_1}(P)(x_1 - p_1) - \cdots - f_{x_n}(P)(x_n - p_n)|}{\|X - P\|} = 0.$$

En ese caso a la función $(x_1, \dots, x_n) \mapsto f_{x_1}(P)x_1 + \cdots + f_{x_n}(P)x_n$ la denominamos **diferencial de f en P** y la anotamos Df_P .

Si A es abierto y f es diferenciable en todos los puntos de A , entonces decimos que f es **diferenciable en A** .

Veamos algunos ejemplos. Volvamos al último que nos desconcertaba. El plano tangente en $P = (0, 0)$ estaba dado por la ecuación $z = 0$. Entonces para que f sea diferenciable en P el siguiente límite debería dar cero.

$$\lim_{(x,y) \rightarrow (0,0)} \frac{|\frac{x^2y}{x^4+y^2} - 0 - 0|}{\sqrt{(x-0)^2 + (y-0)^2}} = \lim_{(x,y) \rightarrow (0,0)} \frac{\frac{x^2y}{x^4+y^2}}{\sqrt{x^2+y^2}} = \lim_{(x,y) \rightarrow (0,0)} \frac{x^2y}{(x^4+y^2)\sqrt{x^2+y^2}}.$$

Pero tomando $\alpha(t) = (t, t)$, con $t > 0$ (o lo que es lo mismo, tomando $y = x$, es decir aproximándonos al origen por la diagonal del primer cuadrante) se tiene

$$\lim_{t \rightarrow 0^+} \frac{t^2t}{(t^4+t^2)\sqrt{t^2+t^2}} = \lim_{t \rightarrow 0^+} \frac{t^3}{t^2(t^2+1)\sqrt{2}|t|} = \lim_{t \rightarrow 0^+} \frac{1}{(t^2+1)\sqrt{2}} = \frac{1}{\sqrt{2}} \neq 0.$$

Es decir, el límite no puede ser cero. Con lo cual, de acuerdo a nuestra definición, la función f **no** es diferenciable en el origen.

Veamos otro ejemplo. Si tomamos $f(x, y) = x^2 + y^2$ (el paraboloides), las derivadas parciales son $f_x(x, y) = 2x$, $f_y(x, y) = 2y$. Con lo cual las dos derivadas parciales en el origen existen y son nulas. Entonces el plano tangente en el origen es el plano $z = 0$. Calculamos

$$\lim_{(x,y) \rightarrow (0,0)} \frac{|x^2+y^2-0-0|}{\sqrt{(x-0)^2 + (y-0)^2}} = \lim_{(x,y) \rightarrow (0,0)} \frac{x^2+y^2}{\sqrt{x^2+y^2}} = \lim_{(x,y) \rightarrow (0,0)} \sqrt{x^2+y^2} = 0,$$

lo que confirma nuestra intuición, pues en este caso la función **si** es diferenciable en el origen.

Observación 3.2.3. *Algunas observaciones y definiciones útiles.*

1. Una condición necesaria para que f sea diferenciable en P es que existan las derivadas parciales de f en P . Sin embargo esto no es suficiente, como vimos.
2. Al vector formado por las n derivadas parciales de f en P lo llamamos **gradiente de f en P** , y lo denotamos ∇f_P (se lee "nabla" de f). Es decir

$$\nabla f_P = (f_{x_1}(P), \dots, f_{x_n}(P)).$$

3. La diferencial de f en P (si f es diferenciable) es una transformación lineal de \mathbb{R}^n en \mathbb{R} , pues según la definición

$$Df_P(Y) = \langle \nabla f_P, Y \rangle = \sum_{i=1 \dots n} f_{x_i}(P)y_i.$$

4. La ecuación del plano tangente a f en P es

$$x_{n+1} = f(P) + \langle \nabla f_P, X - P \rangle.$$

5. Se tiene, por la desigualdad de Cauchy-Schwarz,

$$|Df_P(X)| = |\langle \nabla f_P, X \rangle| \leq \|\nabla f_P\| \|X\|. \quad (3.3)$$

6. Asimismo, con esta notación, f es diferenciable en P si y sólo si

$$\lim_{X \rightarrow P} \frac{|f(X) - f(P) - \langle \nabla f_P, X - P \rangle|}{\|X - P\|} = 0.$$

7. Si f es diferenciable en P se puede escribir la condición equivalente

$$\lim_{X \rightarrow P} \frac{|f(X) - f(P) - Df_P(X - P)|}{\|X - P\|} = 0.$$

Con la definición correcta recuperamos la propiedad “si es derivable es continua”, como muestra la siguiente

Proposición 3.2.4. Sea $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$, $P \in A^\circ$. Si f es diferenciable en P entonces f es continua en P .

Demostración. Se tiene

$$\begin{aligned} |f(X) - f(P)| &\leq |f(X) - f(P) - Df_P(X - P)| + |Df_P(X - P)| \\ &\leq |f(X) - f(P) - Df_P(X - P)| + \|\nabla f_P\| \|X - P\|, \end{aligned}$$

donde en el último término usamos la desigualdad (3.3). Si en el primer sumando multiplicamos y dividimos por $\|X - P\|$, se obtiene

$$|f(X) - f(P)| \leq \frac{|f(X) - f(P) - Df_P(X - P)|}{\|X - P\|} \|X - P\| + \|\nabla f_P\| \|X - P\|.$$

Tanto el primero como el segundo sumando de la derecha tienden a cero cuando $X \rightarrow P$, con lo cual se obtiene $\lim_{X \rightarrow P} f(X) = f(P)$, es decir, f es continua en P . \square

3.2.1. Unicidad de la diferencial

El siguiente teorema tiene su utilidad para probar propiedades generales. Recordemos primero que una transformación lineal $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ es una función que verifica que si $\alpha, \beta \in \mathbb{R}$, $X, Y \in \mathbb{R}^n$ entonces

$$T(\alpha X + \beta Y) = \alpha T(X) + \beta T(Y).$$

En particular, toda transformación lineal $T : \mathbb{R}^n \rightarrow \mathbb{R}$ tiene la expresión

$$T(x_1, \dots, x_n) = v_1 x_1 + v_2 x_2 + \dots + v_n x_n$$

para alguna n -upla de números reales fija (v_1, \dots, v_n) . Equivalentemente, existe un vector $V \in \mathbb{R}^n$ tal que

$$T(X) = \langle V, X \rangle \text{ para todo } X \in \mathbb{R}^n,$$

y cada coordenada de V se puede recuperar aplicándole T al vector E_i de la base canónica, es decir

$$v_i = T(E_i).$$

Con estas herramientas estamos en condiciones de enunciar el teorema de unicidad.

Teorema 3.2.5. Sea $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$, $P \in A^\circ$. Si existe una transformación lineal $T_P : \mathbb{R}^n \rightarrow \mathbb{R}$ tal que

$$\lim_{X \rightarrow P} \frac{|f(X) - f(P) - T_P(X - P)|}{\|X - P\|} = 0,$$

entonces

1. Existen todas las derivadas direccionales de f en P , y vale

$$\frac{\partial f}{\partial V}(P) = T_P(V) \text{ para todo } V \in \mathbb{R}^n, \|V\| = 1.$$

2. En particular existen todas las derivadas parciales de f , se tiene $f_{x_i}(P) = T_P(E_i)$ y la transformación T_P es única.
3. Se tiene $T_P(X) = Df_P(X) = \langle \nabla f_P, X \rangle$ para todo $X \in \mathbb{R}^n$, y f es diferenciable en P . En particular,

$$\frac{\partial f}{\partial V}(P) = \langle \nabla f_P, V \rangle = Df_P(V) \text{ si } f \text{ es diferenciable en } P.$$

Demostración. Veamos que existen todas las derivadas direccionales, tomemos $V \in \mathbb{R}^n$ tal que $\|V\| = 1$. Si existe el límite del enunciado, existe en particular el límite componiendo con cualquier curva que tenga límite P . En particular, si ponemos $X = P + tV$ (con t suficientemente chico para que $X \in A$), se tiene $X - P = tV$ con lo cual $\|X - P\| = |t|$. Entonces

$$\lim_{t \rightarrow 0} \left| \frac{f(P + tV) - f(P)}{t} - T_P(V) \right| = \lim_{t \rightarrow 0} \frac{|f(P + tV) - f(P) - T_P(tV)|}{|t|} = 0.$$

Esto prueba que $f_V(P) = \lim_{t \rightarrow 0} \frac{f(P + tV) - f(P)}{t} = T_P(V)$. En particular existen todas las derivadas parciales.

Además, como $T_P(E_i) = f_{x_i}(P)$ vale para todos los E_i , y una transformación lineal queda determinada por su valor en una base de \mathbb{R}^n , se deduce que $T_P(X) = \langle \nabla f_P, X \rangle$.

La unicidad de T_P se desprende de que si hay otra transformación lineal S_P que verifique la condición del límite, entonces también va a cumplir $S_P(X) = \langle \nabla f_P, X \rangle$.

Lo que se afirma en 3. es, en primer lugar, que

$$\lim_{X \rightarrow P} \frac{|f(X) - f(P) - \langle \nabla f_P, X - P \rangle|}{\|X - P\|} = 0.$$

Pero esto es evidente pues $\langle \nabla f_P, X - P \rangle = T_P(X - P)$ por lo que dijimos recién, y el límite da cero por hipótesis del teorema. También es evidente entonces que $T_P = Df_P$. \square

Hay que observar que reemplazamos la idea geométrica de plano tangente por la idea de aproximación lineal, y vimos que ambas nociones son equivalentes. Es decir, una función es diferenciable si y sólo si se puede aproximar bien con un transformación lineal, y esta transformación representa, en términos de gráficos, al plano tangente que aproxima bien al gráfico de la función.

Observación 3.2.6. Si f es diferenciable en P , entonces tomando V de norma unitaria se tiene

$$f_V(P) = \langle \nabla f_P, V \rangle \leq |\langle \nabla f_P, V \rangle| \leq \|\nabla f_P\| \|V\| = \|\nabla f_P\|,$$

lo que nos muestra que la derivada direccional a lo sumo vale $\|\nabla f_P\|$. Y por otro lado, si $\nabla f_P \neq \mathbf{0}$, poniendo $V_P = \frac{\nabla f_P}{\|\nabla f_P\|}$ se tiene

$$f_{V_P} = \langle \nabla f_P, V_P \rangle = \langle \nabla f_P, \frac{\nabla f_P}{\|\nabla f_P\|} \rangle = \|\nabla f_P\|$$

lo que muestra que este máximo siempre se alcanza.

Esto nos dice que ∇f_P es la **dirección de mayor crecimiento de f en P** .

Ejemplo 3.2.7. Consideremos la función $f(x,y) = \sqrt{1-x^2-y^2}$ con dominio en el disco unitario

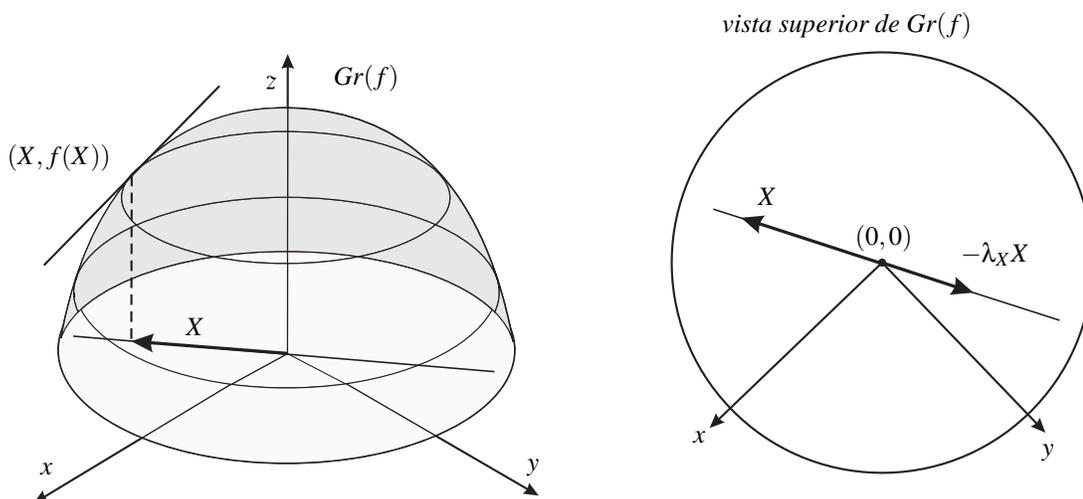
$$\text{Dom}(f) = \{(x,y) : x^2 + y^2 \leq 1\}.$$

En esta discusión vamos a excluir los bordes para hablar de derivadas. Es decir, vamos a considerar $X = (x,y) \in B_1(0,0)$. Recordemos que el gráfico de f es el casquete superior de la esfera unitaria, pues de la ecuación $z = f(x,y)$ se despeja $x^2 + y^2 + z^2 = 1$.

El gradiente de f se calcula fácilmente:

$$\nabla f_X = \left(\frac{-x}{\sqrt{1-x^2-y^2}}, \frac{-y}{\sqrt{1-x^2-y^2}} \right) = -\frac{1}{\sqrt{1-x^2-y^2}} (x,y) = -\lambda_{(x,y)} X.$$

Es decir, el gradiente de f en X es un múltiplo negativo del vector X . Si miramos el gráfico de f desde el costado y desde arriba,



observamos que, parados en el punto $X \in \text{Dom}(f)$, la dirección de mayor crecimiento es aquella que nos lleva al polo norte. En la vista superior del gráfico, podemos observar que lo que hay que hacer es caminar, justamente (empezando en X), en la dirección contraria del vector X que es la que nos lleva hacia el centro del dominio y por lo tanto al punto más alto si caminamos por la esfera.

3.2.2. Álgebra de funciones diferenciables

Se tienen las propiedades habituales de las derivadas. La primera es que la diferencial de una función constante (en cualquier punto) es la transformación lineal nula.

Observación 3.2.8. Sea $c \in \mathbb{R}$ y $f: \mathbb{R}^n \rightarrow \mathbb{R}$ la función constante $f(X) \equiv c$. Entonces f es diferenciable en \mathbb{R}^n y $Df_P(X) = 0$ para todo $P \in \mathbb{R}^n$ y todo $X \in \mathbb{R}^n$. Esto se deduce de

$$\lim_{X \rightarrow P} \frac{|f(X) - f(P) - 0|}{\|X - P\|} = \lim_{X \rightarrow P} \frac{0}{\|X - P\|} = 0.$$

y el Teorema 3.2.5.

Dadas dos funciones f, g , nos interesan los puntos del interior de la intersección de los dominios. Para simplificar consideramos que tienen el mismo dominio, y que es un conjunto abierto (lo que se consigue, como dijimos, tomando el interior de la intersección de los dominios correspondientes).

Teorema 3.2.9. Sean $f, g: A \subset \mathbb{R}^n \rightarrow \mathbb{R}$, con A abierto. Si $P \in A$, y f, g son diferenciables en P , entonces

1. La suma $f + g$ es diferenciable en A y se tiene la fórmula

$$D(f + g)_P(X) = Df_P(X) + Dg_P(X).$$

2. El producto fg es diferenciable en A , y se tiene

$$D(fg)_P(X) = Df_P(X)g(P) + f(P)Dg_P(X).$$

En particular si $\lambda \in \mathbb{R}$,

$$D(\lambda f)_P(X) = \lambda Df_P(X).$$

3. El conjunto $A - C_0(g)$ es abierto, allí el cociente f/g es diferenciable y se tiene

$$D(f/g)_P(X) = \frac{Df_P(X)g(P) - f(P)Dg_P(X)}{g(P)^2}.$$

Demostración. Veamos 1. Ciertamente si f, g son diferenciables existen Df_P y Dg_P . Como $T_P: X \mapsto Df_P(X) + Dg_P(X)$ es una transformación lineal de \mathbb{R}^n en \mathbb{R} , por el Teorema 3.2.5, sólo hay que ver que

$$\lim_{X \rightarrow P} \frac{|(f + g)(X) - (f + g)(P) - (Df_P + Dg_P)(X - P)|}{\|X - P\|} = 0,$$

pues esto probaría que la suma es diferenciable y su diferencial es la propuesta. Entonces

$$\frac{|f(X) - f(P) + g(X) - g(P) - Df_P(X - P) - Dg_P(X - P)|}{\|X - P\|} \leq$$

$$\leq \frac{|f(X) - f(P) - Df_P(X - P)|}{\|X - P\|} + \frac{|g(X) - g(P) - Dg_P(X - P)|}{\|X - P\|}.$$

Los últimos dos sumandos tienden a cero por hipótesis cuando $X \rightarrow P$, lo que prueba que el primer término de la desigualdad tiende a cero, como queríamos.

Ahora veamos 2. Usamos la misma idea (el Teorema 3.2.5). Observemos que, para cada $P \in A$,

$$T_P : X \mapsto Df_P(X)g(P) + f(P)Dg_P(X)$$

es una transformación lineal de \mathbb{R}^n en \mathbb{R} , pues $f(P), g(P) \in \mathbb{R}$ están fijos. Entonces, evaluando T_P en $X - P$, sumamos y restamos $f(X)g(P)$ para obtener

$$\begin{aligned} & |f(X)g(X) - f(X)g(P) - Df_P(X - P)g(P) - f(P)Dg_P(X - P)| \leq \\ & \leq |f(X)g(X) - f(X)g(P) - f(P)Dg_P(X - P)| \\ & \quad + |f(X)g(P) - f(P)g(P) - g(P)Df_P(X - P)|. \end{aligned}$$

Para terminar la demostración de 2., veamos que cada uno de estos dos sumandos (divididos por $\|X - P\|$) tiende a cero cuando $X \rightarrow P$. En el primer caso, sumando y restando $f(X)Dg_P(X - P)$, este término es menor o igual que

$$\begin{aligned} & \frac{|f(X)g(X) - f(X)g(P) - f(X)Dg_P(X - P)|}{\|X - P\|} + \frac{|f(X)Dg_P(X - P) - f(P)Dg_P(X - P)|}{\|X - P\|} = \\ & = \frac{|f(X)||g(X) - g(P) - Dg_P(X - P)|}{\|X - P\|} + \frac{|f(X) - f(P)||Dg_P(X - P)|}{\|X - P\|}. \end{aligned}$$

En el primer término, se tiene que existe $M > 0$ tal que $|f(X)| \leq M$ si $\|X - P\| \leq \delta$ pues f es continua por ser diferenciable (y toda función continua en un compacto alcanza máximo y mínimo). Y entonces

$$\frac{|f(X)||g(X) - g(P) - Dg_P(X - P)|}{\|X - P\|} \leq M \frac{|g(X) - g(P) - Dg_P(X - P)|}{\|X - P\|} \rightarrow_{X \rightarrow P} 0$$

por ser g diferenciable. En el segundo término, como g es diferenciable se tiene $|Dg_P(X - P)| \leq \|\nabla g_P\| \|X - P\|$ por la desigualdad de C-S (ver el ítem 5. de la Observación 3.2.3). En consecuencia,

$$\frac{|f(X) - f(P)||Dg_P(X - P)|}{\|X - P\|} \leq |f(X) - f(P)| \|\nabla g_P\| \rightarrow_{X \rightarrow P} 0$$

por ser f continua en P .

Por último, veamos la afirmación 3. sobre el cociente. Que $A - C_0(g) = \{X \in A : g(X) \neq 0\}$ es abierto se deduce de lo siguiente: si $X \in A - C_0(g)$, entonces $g(X) > 0$ ó $g(X) < 0$. Pero entonces existe una bola abierta B alrededor de X , contenida en A , donde g no se anula, por ser g continua. Es decir $B \subset C_0(g) \cap A$, que es lo que queríamos probar.

Ahora hay que observar una vez más que la transformación propuesta

$$T_P : X \mapsto \frac{Df_P(X)g(P) - f(P)Dg_P(X)}{g(P)^2}$$

es lineal en X . Usamos nuevamente el Teorema 3.2.5, tenemos que ver que la siguiente expresión tiende a cero cuando $X \rightarrow P$:

$$\frac{\left| \frac{f(X)}{g(X)} - \frac{f(P)}{g(P)} - T_P(X - P) \right|}{\|X - P\|}.$$

Observemos que esta expresión está bien definida en un entorno $0 < \|X - P\| < r$, pues g , como es diferenciable, es continua, y como $|g(P)| > 0$ existe un entorno digamos $B_r(P)$ donde $|g(X)| > c > 0$.

Sigamos con la prueba de que el cociente es diferenciable. Si reescribimos la expresión de arriba, obtenemos

$$\frac{1}{|g(X)||g(P)|} \frac{|f(X)g(P) - f(P)g(X) - g(X)g(P)T_P(X - P)|}{\|X - P\|}.$$

Según dijimos, $\frac{1}{|g(X)|} < \frac{1}{c}$ así que no es un problema si queremos ver que esta expresión tiende a cero. Si escribimos la expresión de T_P propuesta (evaluada ahora en $X - P$), se obtiene (omitiendo el término inicial $\frac{1}{|g(X)||g(P)|}$ que está acotado)

$$\begin{aligned} & \frac{\left| f(X)g(P) - f(P)g(X) - g(X)g(P) \frac{Df_P(X-P)g(P) - f(P)Dg_P(X-P)}{g(P)^2} \right|}{\|X - P\|} = \\ & = \frac{1}{|g(P)|^2} \frac{|f(X)g(P)^2 - f(P)g(P)g(X) - g(X)(Df_P(X-P)g(P) - f(P)Dg_P(X-P))|}{\|X - P\|}. \end{aligned}$$

Un vez más omitimos el primer término $\frac{1}{|g(P)|}$ pues está acotado y lo único que queremos ver es que esta expresión tiende a cero cuando $X \rightarrow P$. Si distribuimos $g(X)$ en los diferenciales, tenemos

$$\frac{|f(X)g(P)^2 - f(P)g(P)g(X) - g(X)g(P)Df_P(X-P) + g(X)f(P)Dg_P(X-P)|}{\|X - P\|}.$$

Sumamos y restamos tres cantidades: $f(P)g(P)^2$, $g(P)^2Df_P(X-P)$, $f(P)g(P)Dg_P(X-P)$. Usando la desigualdad triangular y agrupando convenientemente, queda la suma de las siguientes cuatro expresiones

1. $\frac{|g(P)|^2|f(X) - f(P) - Df_P(X-P)|}{\|X - P\|}.$
2. $\frac{|f(P)||g(P)||g(X) - g(P) - Dg_P(X-P)|}{\|X - P\|}.$
3. $\frac{|g(P)||g(X) - g(P)||Df_P(X-P)|}{\|X - P\|}.$
4. $\frac{|f(P)||g(X) - g(P)||Dg_P(X-P)|}{\|X - P\|}.$

Las dos primeras tienden a cero por ser f, g diferenciables en P . En las dos últimas, por la desigualdad de C-S, se tiene $|Df_P(X-P)| \leq \|\nabla f_P\| \|X - P\|$ y lo mismo para g . Con esto, cancelamos el denominador en ambas expresiones, y como g es continua en P (por ser diferenciable), también tienden a cero las dos últimas expresiones. \square

3.2.3. Repaso de los teoremas en una variable

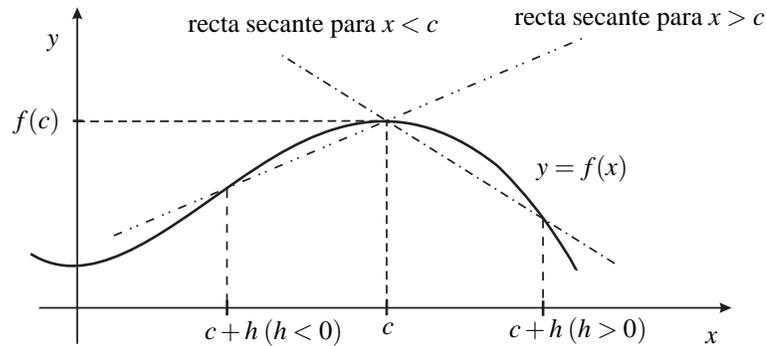
Recordemos aquí tres teoremas clásicos de funciones en una variable, que nos serán de gran utilidad.

Teorema 3.2.10. Sea $f : [a, b] \rightarrow \mathbb{R}$.

1. (Fermat). Si f es derivable en $c \in (a, b)$, y c es un extremo local de f , entonces $f'(c) = 0$.
2. (Rolle). Si f es continua en $[a, b]$, derivable en (a, b) , y además $f(a) = f(b)$, entonces existe $c \in (a, b)$ tal que $f'(c) = 0$.
3. (Lagrange). Si f es continua en $[a, b]$ y derivable en (a, b) entonces existe $c \in (a, b)$ tal que

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

Demostración. Fermat: supongamos que c es un máximo local de f . Entonces existe $\varepsilon > 0$ tal que si $x \in (c - \varepsilon, c + \varepsilon)$, se verifica $f(x) \leq f(c)$. La demostración se deduce del siguiente dibujo, mirando las rectas secantes que aproximan a la tangente:



Calculemos $f'(c)$ usando la definición, con los límites laterales. Por la izquierda se tiene

$$\lim_{h \rightarrow 0^-} \frac{f(c+h) - f(c)}{h} \geq 0$$

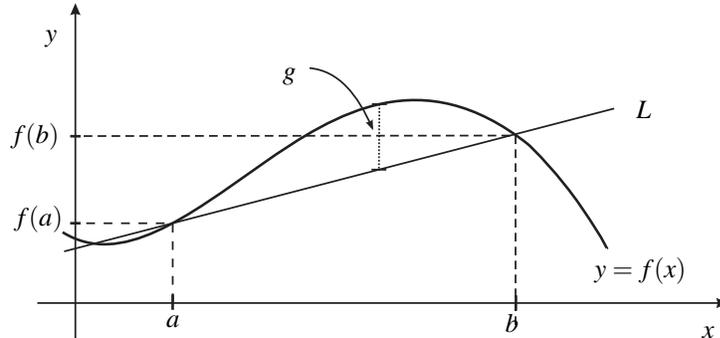
pues $f(c+h) - f(c) \leq 0$ y $h < 0$. Por la derecha,

$$\lim_{h \rightarrow 0^+} \frac{f(c+h) - f(c)}{h} \leq 0$$

pues el numerador sigue siendo negativo pero ahora el denominador es positivo. Como ambos laterales deben ser iguales a $f'(c)$, la única posibilidad es $f'(c) = 0$.

Rolle: Si f es continua en $[a, b]$, como este conjunto es compacto, f alcanza su máximo y su mínimo allí. Si el máximo y el mínimo son iguales, f es constante y por ende su derivada es nula en todo (a, b) . Si son distintos, como $f(a) = f(b)$, alguno de los dos se alcanza en el interior, digamos $c \in (a, b)$ es un extremo de f . Pero entonces, por Fermat, $f'(c) = 0$.

Lagrange: Consideramos la siguiente función auxiliar, que es la resta entre f y la recta L que une $(a, f(a))$ con $(b, f(b))$: $g(x) = f(x) - L(x)$.



Entonces como la recta y f se tocan en a y b , se tiene $g(a) = g(b) = 0$. Por otro lado, g verifica las otras hipótesis del teorema de Rolle pues f las verifica y L es una recta, y por ende existe $c \in (a, b)$ tal que $g'(c) = 0$. Pero $g'(x) = f'(x) - m$, donde m es la pendiente de la recta L . Evaluando en c , se obtiene $f'(c) = m$. Pero si la recta pasa por los puntos dados, su pendiente m se calcula como $m = \Delta y / \Delta x$, que es exactamente

$$m = \frac{f(b) - f(a)}{b - a}.$$

□

Por último, veamos cómo se puede deducir la regla de L'Hospital a partir del teorema del valor medio. Antes vamos a enunciar un resultado técnico con ese sólo propósito.

Proposición 3.2.11 (Cauchy). Si f, g son continuas en $[a, b]$ y derivables en (a, b) entonces existe $c \in (a, b)$ tal que

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

Demostración. Consideremos

$$h(x) = (f(x) - f(a))(g(b) - g(a)) - (g(x) - g(a))(f(b) - f(a)).$$

Entonces la función h verifica todas las hipótesis del teorema de Rolle en $[a, b]$. Por otra parte

$$h'(x) = f'(x)(g(b) - g(a)) - g'(x)(f(b) - f(a)).$$

Luego, existe $c \in (a, b)$ tal que $h'(c) = 0$. Reemplazando en la ecuación de h' y despejando se obtiene la conclusión. □

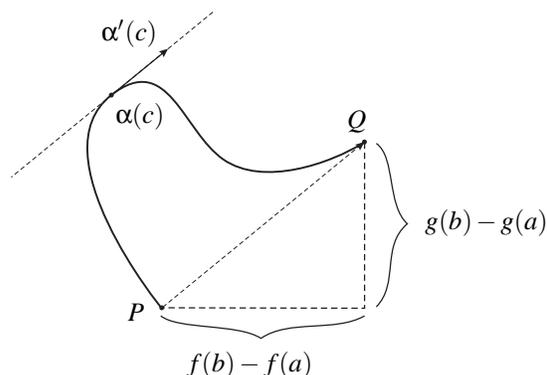
Es importante observar que las derivadas en el teorema de Cauchy están evaluadas en *el mismo punto* $c \in (a, b)$.

Cuando $g(b) \neq g(a)$, $f(b) \neq f(a)$, el resultado anterior tiene la siguiente interpretación geométrica: consideremos la curva en el plano dada por $\alpha(t) = (f(t), g(t))$, y sus extremos $P = (f(a), g(a))$, $Q = (f(b), g(b))$. Entonces $\alpha'(t) = (f'(t), g'(t))$ es el vector velocidad de la curva.

Supongamos que no hay ningún punto donde se anule esta velocidad en (a, b) . Entonces de la expresión

$$f'(c) = \frac{f(b) - f(a)}{g(b) - g(a)} g'(c)$$

se deduce que en el punto c también se tiene $g'(c) \neq 0$. Esto es porque en el caso contrario tendríamos $\alpha'(c) = 0$, lo cual es absurdo.



Luego podemos escribir

$$\frac{f'(c)}{g'(c)} = \frac{f(b) - f(a)}{g(b) - g(a)}$$

Esta expresión nos dice que hay algún punto $R = \alpha(c)$ en la curva, donde la recta tangente a la misma es paralela al vector que une los extremos P y Q , según indica la figura, puesto que la pendiente de la recta generada por el vector $\alpha'(c)$ es exactamente $\frac{f'(c)}{g'(c)}$.

Ahora enunciamos la regla de L'Hospital para calcular límites.

Proposición 3.2.12 (L'Hospital). Sean f, g funciones derivables en un entorno del punto $x = a$ (exceptuando tal vez el mismo punto $x = a$). Supongamos que $g'(x) \neq 0$ para todo $x \neq a$, y que

$$\lim_{x \rightarrow a} f(x) = \lim_{x \rightarrow a} g(x) = 0.$$

Si existe el límite

$$\lim_{x \rightarrow a} \frac{f'(x)}{g'(x)} = L$$

entonces existe el límite de f/g y coincide con el de las derivadas, es decir

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = L.$$

Demostración. Primero observemos que podemos extender a las funciones f, g al punto $x = a$ como $f(a) = g(a) = 0$, y quedan continuas por la hipótesis. Analicemos un límite lateral, el otro se calcula en forma idéntica. Como f, g eran derivables en un entorno de a , existe $\delta > 0$ tal que tanto f como g son continuas en $[a, a + \delta]$ y derivables en $(a, a + \delta)$. Afirmo que g no se anula en $(a, a + \delta)$: en efecto, si se anulara, por el teorema de Rolle aplicado a g en el intervalo $[a, a + \delta]$ tendría que anularse g' en $(a, a + \delta)$, contradiciendo la hipótesis $g'(x) \neq 0$. Tomemos $x \in (a, a + \delta)$, y apliquemos el teorema de Cauchy en $[a, x]$ para obtener que existe $c \in (a, x)$ donde

$$\frac{f'(c)}{g'(c)} = \frac{f(x) - f(a)}{g(x) - g(a)} = \frac{f(x)}{g(x)}.$$

Si hacemos tender $x \rightarrow a^+$, se deduce que $c \rightarrow a^+$ también con lo cual la expresión de la izquierda tiende a L por hipótesis. Luego la expresión de la derecha también tiende a L . \square

Se pueden deducir de aquí (pero no lo haremos) los corolarios habituales que nos permiten calcular límites cuando $x \rightarrow \infty$, y cuando $f, g \rightarrow \infty$.

3.2.4. Criterio de diferenciabilidad

Necesitamos algún criterio efectivo para ver si una función es diferenciable en un punto o no. Vamos a denotar un punto genérico de \mathbb{R}^2 como $P = (a, b)$, y otro punto cercano a P como $X = (a + h, b + k)$. Aquí h, k denotan los incrementos en x e y respectivamente, y $X - P = (h, k)$.

Dada $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, observemos que si las dos derivadas parciales de f existen en todas partes, entonces por el teorema de Lagrange en una variable, existen c entre a y $a + h$, d entre b y $b + k$, tales que

$$\begin{aligned} f(a + h, b + k) - f(a, b) &= f(a + h, b + k) - f(a, b + k) + f(a, b + k) - f(a, b) \\ &= \frac{\partial f}{\partial x}(c, b + k)h + \frac{\partial f}{\partial y}(a, d)k. \end{aligned}$$

En efecto, las funciones reales $x \mapsto f(x, b + k)$ e $y \mapsto f(a, y)$ son derivables, con lo cual se puede aplicar el teorema. Volviendo a la ecuación de arriba, tenemos

$$\begin{aligned} &f(a + h, b + k) - f(a, b) - \langle \nabla f_{(a,b)}, (a + h - a, b + k - b) \rangle = \\ &= \left(\frac{\partial f}{\partial x}(c, b + k) - \frac{\partial f}{\partial x}(a, b) \right) h + \left(\frac{\partial f}{\partial y}(a, d) - \frac{\partial f}{\partial y}(a, b) \right) k. \end{aligned}$$

Luego, como $|h| \leq \|(h, k)\| = \|(a + h, b + k) - (a, b)\|$ (y lo mismo ocurre con $|k|$), descubrimos que

$$\begin{aligned} \frac{|f(a + h, b + k) - f(a, b) - \langle \nabla f_{(a,b)}, (a + h - a, b + k - b) \rangle|}{\|(a + h, b + k) - (a, b)\|} &\leq \\ &\leq \left| \frac{\partial f}{\partial x}(c, b + k) - \frac{\partial f}{\partial x}(a, b) \right| + \left| \frac{\partial f}{\partial y}(a, d) - \frac{\partial f}{\partial y}(a, b) \right|. \end{aligned}$$

Si las derivadas parciales de f fueran continuas, haciendo tender $h, k \rightarrow 0$ tendríamos que el último término tiende a cero, con lo cual el primer término también. Pero esto último es exactamente lo que debe ocurrir para que f sea diferenciable en $P = (a, b)$.

Resumiendo, hemos probado el siguiente teorema, que resulta un criterio útil para ver si una función es diferenciable en un punto. Observemos que las condiciones son *suficientes* pero no *necesarias* (ver el Ejemplo 3.2.16).

Teorema 3.2.13. *Sea $f : A \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ con A abierto, y sea $P \in A$. Si las dos derivadas parciales de f existen en un entorno de P , y ambas son continuas allí, entonces f es diferenciable en P .*

El criterio en su forma general se enuncia de la siguiente manera, y tiene una prueba similar que omitimos:

Teorema 3.2.14. *Si $A \subset \mathbb{R}^n$ es abierto, todas las derivadas parciales de una función $f : A \rightarrow \mathbb{R}$ existen en A , y todas las derivadas parciales son continuas en A , entonces f es diferenciable en A .*

Es decir, basta chequear la continuidad de las n derivadas parciales de f . Se puede pedir un poquito menos, ver la NOTA I del final del capítulo.

Ejemplo 3.2.15. Tomemos $f(x, y, z) = \frac{y \cos(x) + z \ln(x)}{1 + z^2 e^y}$, con dominio

$$\text{Dom}(f) = \{(x, y, z) : x > 0\}.$$

Entonces es más o menos evidente que todas las derivadas parciales de f existen y son funciones continuas en todo el dominio, con lo cual se deduce que f es diferenciable en todo su dominio. ¿Se imaginan probando a mano, en cada punto del dominio, que f es diferenciable?

Otro ejemplo, para pensar:

Ejemplo 3.2.16. Tomemos $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ dada por

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

Entonces queda como ejercicio ver que

1. $f_x(0, 0) = f_y(0, 0) = 0$.
2. $\lim_{(x, y) \rightarrow (0, 0)} \frac{|f(x, y) - 0|}{\|(x, y)\|} = 0$ (o sea f es diferenciable en el origen).
3. Ninguna de las dos derivadas parciales es continua en el origen.

3.2.5. Funciones $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$

En general, dadas dos bases de \mathbb{R}^n y \mathbb{R}^m respectivamente, una transformación lineal $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ se representa como una matriz M de $n \times m$, donde las columnas de M se calculan haciendo $T(E_i)$ con $i = 1..n$. Esta matriz actúa sobre vectores columna de la manera usual.

Podemos extender la noción de diferenciabilidad a funciones a valores vectoriales así:

Definición 3.2.17. Sea $F : A \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$, sea $P \in A^\circ$. Si $T_P : \mathbb{R}^n \rightarrow \mathbb{R}^m$ es una transformación lineal que verifica

$$\lim_{X \rightarrow P} \frac{\|F(X) - F(P) - T_P(X - P)\|}{\|X - P\|} = 0$$

entonces decimos que F es **diferenciable en P** . A la transformación lineal la llamamos **diferencial de F en P** , y la anotamos DF_P .

Entonces extendimos la idea de diferenciabilidad a funciones a valores en \mathbb{R}^m usando la noción de aproximación lineal.

Observación 3.2.18. Una tal función se escribe como $F = (f_1, \dots, f_m)$ donde $f_i : A \rightarrow \mathbb{R}$. Además, la i -ésima coordenada del vector $T_P(X - P)$ se obtiene (si pensamos a la transformación lineal como matriz)

haciendo el producto escalar de la i -ésima fila de T_P con $X - P$.

$$\begin{pmatrix} (T_P)_{11} & (T_P)_{12} & \dots & (T_P)_{1n} \\ (T_P)_{21} & (T_P)_{22} & \dots & \dots \\ \dots & \dots & \dots & \dots \\ (T_P)_{i1} & (T_P)_{i2} & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & (T_P)_{mn} \end{pmatrix} \begin{pmatrix} x_1 - p_1 \\ \dots \\ \dots \\ \dots \\ \dots \\ x_m - p_m \end{pmatrix}$$

Como el límite existe y da cero si y sólo si existe el límite en cada coordenada y da cero, se deduce que F es diferenciable en P si y sólo si cada f_i es diferenciable en P . Además, si usamos la notación

$$\frac{\partial f_i}{\partial x_j}$$

para denotar la derivada j de la función i , se tiene que **la diferencial de F es la matriz que se obtiene poniendo como filas los gradientes de las f_i , es decir**

$$DF_P = \begin{pmatrix} \nabla f_1(P) \\ \nabla f_2(P) \\ \dots \\ \nabla f_m(P) \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(P) & \frac{\partial f_1}{\partial x_2}(P) & \dots & \frac{\partial f_1}{\partial x_n}(P) \\ \frac{\partial f_2}{\partial x_1}(P) & \frac{\partial f_2}{\partial x_2}(P) & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \frac{\partial f_m}{\partial x_n}(P) \end{pmatrix}$$

Observemos que se puede recuperar cada coordenada usando la base canónica:

$$\frac{\partial f_i}{\partial x_j}(P) = \langle DF_P(E_j), E_i \rangle.$$

También es importante observar que, si F es diferenciable en P y $V \in \mathbb{R}^n$, entonces poniendo $\alpha(t) = P + tV$ se tiene

$$0 = \lim_{t \rightarrow 0} \frac{\|F(P + tV) - F(P) - DF_P(tV)\|}{|t|} = \lim_{t \rightarrow 0} \left\| \frac{F(P + tV) - F(P)}{t} - DF_P(V) \right\|,$$

lo que prueba que

$$DF_P(V) = \lim_{t \rightarrow 0} \frac{F(P + tV) - F(P)}{t}.$$

para cualquier $V \in \mathbb{R}^n$.

De lo ya probado se deduce el criterio siguiente:

Teorema 3.2.19. Sea $F : A \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$, con A abierto y $P \in A$. Si existen todas las derivadas parciales de F en un entorno de P y son todas continuas en P , entonces F es diferenciable en P .

Observación 3.2.20. Por supuesto que la suma de funciones diferenciables es diferenciable como vimos. Para funciones a valores vectoriales, digamos $G, H : A \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$, tiene también sentido considerar su producto escalar

$$F(X) = \langle G(X), H(X) \rangle$$

que es una función diferenciable $F : A \rightarrow \mathbb{R}$ a valores reales, por ser suma y producto de funciones diferenciables. Además se tiene

$$\begin{aligned} \lim_{t \rightarrow 0} \frac{F(P+tX) - F(P)}{t} &= \lim_{t \rightarrow 0} \frac{1}{t} [\langle G(P+tX), H(P+tX) \rangle - \langle G(P), H(P) \rangle] \\ &= \lim_{t \rightarrow 0} \frac{1}{t} [\langle G(P+tX), H(P+tX) \rangle - \langle G(P), H(P+tX) \rangle + \\ &\quad + \langle G(P), H(P+tX) \rangle - \langle G(P), H(P) \rangle] \\ &= \lim_{t \rightarrow 0} \frac{1}{t} \langle G(P+tX) - G(P), H(P+tX) \rangle + \frac{1}{t} \langle G(P), H(P+tX) - H(P) \rangle. \end{aligned}$$

Esto prueba que

$$D\langle G(X), H(X) \rangle_P V = \langle DG_P V, H(P) \rangle + \langle G(P), DH_P V \rangle,$$

que es una regla fácil de recordar pues es idéntica a la regla de derivación de un producto de funciones reales.

Ya que la diferencial de una función la asimilamos a una matriz o transformación lineal, es bueno saber como controlar su tamaño. Para eso definiremos una norma en el espacio vectorial de las matrices, y veremos un par de propiedades útiles:

Lema 3.2.21. 1. Si $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ es una transformación lineal, existe una constante positiva C tal que $\|TX\| \leq C\|X\|$ para todo $X \in \mathbb{R}^n$. A la más chica de estas constantes la denotaremos $\|T\|_\infty$, y se tiene

$$\|T\|_\infty = \max\{\|TX\| : \|X\| \leq 1\}.$$

En particular $\|TX\| \leq \|T\|_\infty \|X\|$ para todo $X \in \mathbb{R}^n$.

2. La función $\|\cdot\|_\infty : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}$ es una norma, es decir, si $T, S \in \mathbb{R}^{n \times m}$ entonces

- $\|T\|_\infty \geq 0$ y es $\|T\|_\infty = 0$ si y sólo si $T = 0$.
- $\|\lambda T\|_\infty = |\lambda| \|T\|_\infty$ para todo $\lambda \in \mathbb{R}$.
- $\|T + S\|_\infty \leq \|T\|_\infty + \|S\|_\infty$.

3. Si $F : A \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ es diferenciable en $P \in A^\circ$, existe $r > 0$ y una constante positiva C_r tal que $X \in B_r(P)$ implica

$$\|F(X) - F(P)\| \leq C_r \|X - P\|.$$

Demostración. Veamos 1. Observemos que, como $T(X)$ es en cada coordenada una suma y producto de las coordenadas de X , se tiene que cada coordenada es continua, y en consecuencia T es continua. Como la función norma de \mathbb{R}^2 en \mathbb{R} , $Y \mapsto \|Y\|$, también es continua, pues

$$\| \|X\| - \|Y\| \| \leq \|X - Y\|,$$

se tiene que $X \mapsto \|TX\|$ es una función continua de \mathbb{R}^n en \mathbb{R} . Consideremos la bola cerrada unitaria,

$$\bar{B}_1 = \{X \in \mathbb{R}^n : \|X\| \leq 1\}.$$

Es un conjunto compacto pues es cerrado y acotado. Entonces, la función $X \mapsto \|TX\|$ es acotada en \bar{B}_1 , y además alcanza máximo y mínimo. Pongamos $\|T\|_\infty = \max_{X \in \bar{B}_1} \|TX\|$. Entonces, dado cualquier $X \neq \mathbb{O}$,

$$\|T\left(\frac{X}{\|X\|}\right)\| \leq \|T\|_\infty,$$

con lo cual como T es lineal se tiene $\|TX\| \leq \|T\|_\infty \|X\|$ como queríamos, además por como la elegimos es la mejor cota.

Veamos 2. Que $\|T\|_\infty$ es positivo es evidente por que es un máximo de cantidades positivas. Por otro lado, si $T = 0$, entonces $TX = \mathbb{O}$ para todo $X \in \mathbb{R}^n$, con lo cual $\|T\|_\infty = 0$. Recíprocamente, si $\|T\|_\infty = 0$ es porque el máximo es cero, y entonces debe ser $\|TX\| = 0$ para todo $X \in \bar{B}_1$, con lo cual $TX = 0$ para todo $X \in \bar{B}_1$. Pero entonces, por la linealidad de T , dado cualquier $X \neq \mathbb{O}$, se tiene

$$TX = \|X\|T\left(\frac{X}{\|X\|}\right) = \|X\|\mathbb{O} = \mathbb{O}$$

es decir, T es la transformación lineal nula. Que saca escalares con el módulo es también obvio de la definición.

Por último, veamos que verifica la desigualdad triangular. Si $S, T \in \mathbb{R}^{n \times m}$, entonces dado $X \in \bar{B}_1$,

$$\|(T+S)X\| = \|TX + SX\| \leq \|TX\| + \|SX\| \leq \|T\|_\infty \|X\| + \|S\|_\infty \|X\| \leq \|T\|_\infty + \|S\|_\infty.$$

Luego el máximo debe ser menor o igual que $\|T\|_\infty + \|S\|_\infty$.

Veamos 3. Como F es diferenciable en P , existe una bola $B_r(P)$ alrededor de P donde el cociente famoso es menor que uno. Con esto,

$$\frac{\|F(X) - F(P)\|}{\|X - P\|} \leq \frac{\|F(X) - F(P) - DF_P(X - P)\|}{\|X - P\|} + \frac{\|DF_P(X - P)\|}{\|X - P\|} \leq 1 + \|DF_P\|_\infty,$$

siempre que $X \in B_r(P)$, puesto que DF_P es una transformación lineal. Si ponemos $C_r = 1 + \|DF_P\|_\infty$, tenemos la cota deseada. \square

Podemos componer dos funciones, y se tiene la **regla de la cadena**

Teorema 3.2.22. Sea $F : A \subset \mathbb{R}^n \rightarrow B \subset \mathbb{R}^m$, y $G : B \rightarrow \mathbb{R}^k$, con A, B abiertos. Si F es diferenciable en $P \in A$, y G es diferenciable en $Q = F(P) \in B$, entonces $G \circ F$ es diferenciable en P y además

$$D(G \circ F)_P = DG_{F(P)}DF_P,$$

donde el producto denota composición de transformaciones lineales.

Demostración. Tenemos que probar que

$$\lim_{X \rightarrow P} \frac{\|G(F(X)) - G(F(P)) - DG_Q DF_P(X - P)\|}{\|X - P\|} = 0.$$

Si sumamos y restamos $DG_Q(F(X) - Q)$, se tiene que la expresión de la izquierda es menor o igual que

$$\frac{\|G(F(X)) - G(Q) - DG_Q(F(X) - Q)\|}{\|X - P\|} + \frac{\|DG_Q\|_\infty \|F(X) - F(P) - DF_P(X - P)\|}{\|X - P\|}.$$

El segundo término ciertamente tiende a cero pues F es diferenciable en P . Por otro lado, como F es diferenciable en P , en particular es continua en P pues cada coordenada de F es diferenciable y entonces cada coordenada es continua. Entonces $F(X) \rightarrow Q = F(P)$ y si en el primer sumando multiplicamos y dividimos por $\|F(X) - Q\|$, se tiene

$$\frac{\|G(F(X)) - G(Q) - DG_Q(F(X) - Q)\|}{\|F(X) - Q\|} \frac{\|F(X) - F(P)\|}{\|X - P\|}.$$

El último cociente está acotado por la parte 3. del lema previo, mientras que el primer cociente, como $F(X) \rightarrow Q$, y G es diferenciable en Q , tiende a cero. \square

Observación 3.2.23. Además de la propiedad general que establece la regla de la cadena, el teorema tiene la siguiente aplicación. Si $F = \mathbb{R}^k \rightarrow \mathbb{R}^n$ dada por $F = (f_1, \dots, f_n)$ la componemos con $g : \mathbb{R}^n \rightarrow \mathbb{R}$, entonces por la regla de la cadena (suponiendo que ambas son diferenciables)

$$\begin{aligned} D(g \circ F)_P &= \nabla_{g(F(P))} DF_P = \left(\frac{\partial g}{\partial x_1}, \dots, \frac{\partial g}{\partial x_n} \right)_{F(P)} \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \dots & \frac{\partial f_1}{\partial x_k} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \dots & \dots \\ \frac{\partial f_3}{\partial x_1} & \frac{\partial f_3}{\partial x_2} & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \frac{\partial f_n}{\partial x_1} & \dots & \dots & \frac{\partial f_n}{\partial x_k} \end{pmatrix}_P = \\ &= \left(\frac{\partial g}{\partial x_1} \frac{\partial f_1}{\partial x_1} + \frac{\partial g}{\partial x_2} \frac{\partial f_1}{\partial x_2} + \dots + \frac{\partial g}{\partial x_n} \frac{\partial f_1}{\partial x_1}, \dots, \frac{\partial g}{\partial x_1} \frac{\partial f_1}{\partial x_k} + \dots + \frac{\partial g}{\partial x_n} \frac{\partial f_n}{\partial x_k} \right), \end{aligned}$$

donde en cada producto el primer factor está evaluado en $F(P)$, mientras que el segundo está evaluado en P .

En particular, la coordenada i -ésima del gradiente de $g \circ F : \mathbb{R}^k \rightarrow \mathbb{R}$ (o lo que es lo mismo, la i -ésima derivada parcial de $g \circ F$), para cualquier $i = 1 \dots k$ está dada por

$$\frac{\partial(g \circ F)}{\partial x_i} = \frac{\partial g}{\partial x_1} \frac{\partial f_1}{\partial x_i} + \dots + \frac{\partial g}{\partial x_n} \frac{\partial f_n}{\partial x_i} = \sum_{j=1}^n \frac{\partial g}{\partial x_j} \frac{\partial f_j}{\partial x_i},$$

expresión en la que hay que recordar que las derivadas de F están evaluadas en P , mientras que las de g están evaluadas en $F(P)$.

Hagamos un ejemplo para descifrar la sopa de letras.

Ejemplo 3.2.24. Sea $F : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ dada por

$$F(u, v) = (f_1(u, v), f_2(u, v), f_3(u, v)) = (u^2 - 2v, \cos(u) + v + 3, uv).$$

Sea $g : \mathbb{R}^3 \rightarrow \mathbb{R}$ dada por $g(x, y, z) = x^2 y + \sin(z)$.

Entonces, si consideramos $g \circ F : \mathbb{R}^2 \rightarrow \mathbb{R}$, se tiene por ejemplo

$$\frac{\partial(g \circ F)}{\partial u} = \frac{\partial g}{\partial x} \frac{\partial f_1}{\partial u} + \frac{\partial g}{\partial y} \frac{\partial f_2}{\partial u} + \frac{\partial g}{\partial z} \frac{\partial f_3}{\partial u}.$$

Tenemos

$$\frac{\partial g}{\partial x} = 2xy, \quad \frac{\partial g}{\partial y} = x^2 + \cos(zy)z, \quad \frac{\partial g}{\partial z} = \cos(zy)y,$$

mientras que

$$\frac{\partial f_1}{\partial u} = 2u, \quad \frac{\partial f_2}{\partial u} = -\sin(u), \quad \frac{\partial f_3}{\partial u} = v.$$

Con esto, recordando que al componer $x = f_1(u, v)$, $y = f_2(u, v)$, $z = f_3(u, v)$, se tiene

$$\begin{aligned} \frac{\partial(g \circ F)}{\partial u} &= 2(u^2 - 2v)(\cos(u) + v + 3) \cdot 2u \\ &\quad + ((u^2 - 2v)^2 + \cos(uz(\cos(u) + v + 3))uv) \cdot (-\sin(u)) \\ &\quad + \cos(uv(\cos(u) + v + 3))(\cos(u) + v + 3) \cdot v. \end{aligned}$$

Observación 3.2.25. Es habitual usar la notación $x(u, v)$ en lugar de $f_1(u, v)$, así como $y(u, v)$ en lugar de $f_2(u, v)$, y también $z(u, v)$ en lugar de $f_3(u, v)$. Con lo cual se escribe $F(u, v) = (x(u, v), y(u, v), z(u, v))$ y la fórmula de la derivada parcial respecto de la primer coordenada de la composición queda escrita de manera sugestiva como

$$\frac{\partial(g \circ F)}{\partial u} = \frac{\partial g}{\partial x} \frac{\partial x}{\partial u} + \frac{\partial g}{\partial y} \frac{\partial y}{\partial u} + \frac{\partial g}{\partial z} \frac{\partial z}{\partial u},$$

lo que permite desarrollar una regla mnemotécnica sencilla para usar la regla de la cadena.

3.3. Teoremas de Lagrange y Fermat en \mathbb{R}^n

Supongamos que f es diferenciable. Entonces tiene sentido relacionar f con su diferencial, tal como hicimos en una variable. El siguiente resultado es la generalización del Teorema de Lagrange a varias variables.

Proposición 3.3.1. (Lagrange) Sea $f : B_r(P) \subset \mathbb{R}^n \rightarrow \mathbb{R}$ diferenciable. Entonces para todo $Q, R \in B_r(P)$ existe un punto P_0 en el segmento que une Q con R tal que

$$f(Q) - f(R) = \langle \nabla f_{P_0}, Q - R \rangle.$$

Demostración. Consideremos la parametrización del segmento que une Q con P , $g(t) = R + t(Q - R)$. Entonces $h(t) = f \circ g(t)$ está definida en $[0, 1]$, y es diferenciable en $(0, 1)$ por la regla de la cadena. Es más $h : [0, 1] \rightarrow \mathbb{R}$ es una función continua pues f es continua por ser diferenciable. Entonces, por el Teorema de Lagrange en una variable, existe $c \in (0, 1)$ tal que $h(1) - h(0) = h'(c)$. Pero por la regla de la cadena, $h'(c) = Df_{g(c)}g'(c) = Df_{g(c)}(Q - R)$. Si llamamos $P_0 = g(c)$, se tiene la afirmación. \square

Observación 3.3.2. Lo relevante del dominio de f en el teorema previo, no es tanto que sea una bola, sino que el segmento que une X con Y esté contenido en el dominio. El teorema **no** es válido en abiertos arbitrarios.

En funciones a valores vectoriales el teorema **es falso**, como muestra el siguiente contraejemplo.

Ejemplo 3.3.3. Sea $F : \mathbb{R} \rightarrow \mathbb{R}^2$ la curva dada por $F(t) = (t^2, t^3)$. Consideremos el intervalo $[0, 1]$ en el dominio de F . Entonces por un lado

$$F(1) - F(0) = (1, 1)$$

y por otro lado $F'(t) = (2t, 3t^2)$, con lo cual es evidente que no existe ningún $c \in (0, 1)$ que satisfaga

$$(1, 1) = (2c, 3c^2).$$

Sin embargo, se tiene el siguiente resultado útil:

Proposición 3.3.4. Si $G : B_r(P) \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ es diferenciable, y $\|DG_Q\|_\infty \leq M$ para todo $Q \in B_r(P)$, se tiene

$$\|G(X) - G(Y)\| \leq M\|X - Y\|$$

para todo $X, Y \in B_r(P)$.

Demostración. Si $G(X) = G(Y)$ no hay nada que probar. Supongamos que $G(X) \neq G(Y)$. Tomemos nuevamente $g(t) = Y + t(X - Y)$ el segmento que une X con Y en $B_r(P)$, pongamos ahora

$$h(t) = \langle G \circ g(t), G(X) - G(Y) \rangle.$$

Entonces $h'(t) = \langle DG_{g(t)}g'(t), G(X) - G(Y) \rangle$, luego por el teorema de Lagrange en \mathbb{R} (ítem 3. del Teorema 3.2.10), existe $c \in (0, 1)$ tal que

$$h(1) - h(0) = h'(c)(1 - 0),$$

lo que nos lleva a la siguiente expresión, escribiendo quienes son h y h' :

$$\langle G(X) - G(Y), G(X) - G(Y) \rangle = \langle DG_{Y+c(X-Y)}(X - Y), G(X) - G(Y) \rangle.$$

Es decir

$$\|G(X) - G(Y)\|^2 \leq \|DG_{Y+c(X-Y)}(X - Y)\|_\infty \|G(X) - G(Y)\| \leq M \|X - Y\| \|G(X) - G(Y)\|.$$

Como supusimos que $G(X) \neq G(Y)$, podemos dividir por $\|G(X) - G(Y)\|$ para terminar. \square

Y por último, la aplicación más concreta para extremos de una función, que desarrollaremos con cuidado más adelante. Si $f : \mathbb{R}^n \rightarrow \mathbb{R}$, un máximo local es un punto P tal que

$$f(P) \geq f(X)$$

para todo X en un entorno $B_r(P)$ de P . De la misma manera se define un mínimo local, y diremos que P es un **extremo local de f** si es máximo o mínimo local.

Teorema 3.3.5. (Fermat en \mathbb{R}^n) Sea $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$ diferenciable, con A abierto. Supongamos que $P \in A$ es un extremo local de f . Entonces $Df_P = \nabla f_P = \mathbf{0}$. Equivalentemente, todas las derivadas parciales de f se anulan en P .

Demostración. Supongamos que P es un máximo local de f . Como $P \in A$ es un punto interior, podemos suponer que existe $r > 0$ y una bola abierta $B_r(P) \subset A$ tal que $f(P) \geq f(X)$ para todo $X \in B_r(P)$. Consideremos la función auxiliar $g(t) = f(P + tE_i)$, donde E_i es un vector de la base canónica. Entonces $g : (-r, r) \rightarrow \mathbb{R}$ es una función derivable, que tiene un máximo local en $t = 0$ pues $g(0) = f(P) \geq f(P + tE_i) = g(t)$. En consecuencia, debe ser $g'(0) = 0$. Pero esta derivada en cero es exactamente (usando la definición) la derivada parcial i -ésima de f en P . Esto prueba que $f_{x_i}(P) = 0$, y como i es cualquiera entre 1 y n , se tiene que el gradiente de f es cero en P . \square

3.4. NOTAS

I. Vamos a mostrar aquí que el criterio de diferenciabilidad se puede mejorar. Enunciamos el resultado concreto, seguido de su prueba.

Sea $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$, $P \in A^\circ$. Supongamos que

- Existen todas las derivadas parciales de f en P .
- De las n derivadas parciales hay $n - 1$ que existen en un entorno $B_r(P)$ de P , y estas $n - 1$ funciones son continuas en P .

Entonces f es diferenciable en P .

Como existen todas las derivadas parciales en P , tenemos que probar que

$$\lim_{X \rightarrow P} \frac{|f(X) - f(P) - \langle \nabla f_P, X - P \rangle|}{\|X - P\|} = 0.$$

En esta prueba vamos a usar f_i para denotar la i -ésima derivada parcial. Para simplificar la prueba vamos a suponer que la única derivada parcial que no está definida en un entorno de P es la n -ésima (la última). Empecemos con un caso donde las cuentas son más fáciles, supongamos primero que $P = \mathbb{O}$ y $f(P) = 0$. Con esto, lo que hay que probar es que

$$\lim_{X \rightarrow \mathbb{O}} \frac{|f(X) - \langle \nabla f_{\mathbb{O}}, X \rangle|}{\|X\|} = 0.$$

Observemos que

$$f(X) = f(x_1, x_2, \dots, x_n) - f(0, x_2, x_3, \dots, x_n) + f(0, x_2, x_3, \dots, x_n). \quad (3.4)$$

Dado $X = (x_1, x_2, \dots, x_n)$ fijo, consideremos la función auxiliar $g_1(t) = f(t, x_2, \dots, x_n)$. Calculemos $g_1'(t)$. Se tiene

$$g_1'(t) = \lim_{h \rightarrow 0} \frac{f(t+h, x_2, \dots, x_n) - f(t, x_2, \dots, x_n)}{h}.$$

Pero este límite es exactamente la primera derivada parcial de f evaluada en (t, x_2, \dots, x_n) . Es decir, $g_1'(t) = f_1(t, x_2, \dots, x_n)$. Por el Teorema del valor medio de Lagrange, existe $c_1 \in (0, x_1)$ tal que

$$f(x_1, x_2, \dots, x_n) - f(0, x_2, \dots, x_n) = g_1(x_1) - g_1(0) = g_1'(c_1)(x_1 - 0) = f_1(c_1, x_2, \dots, x_n)x_1.$$

Volviendo a (3.4), nos quedó

$$f(X) = f_1(c_1, x_2, \dots, x_n)x_1 + f(0, x_2, x_3, \dots, x_n).$$

Ahora, con un argumento similar,

$$\begin{aligned} f(0, x_2, x_3, \dots, x_n) &= f(0, x_2, x_3, \dots, x_n) - f(0, 0, x_3, \dots, x_n) + f(0, 0, x_3, \dots, x_n) \\ &= f_2(0, c_2, x_3, \dots, x_n)x_2 + f(0, 0, x_3, \dots, x_n), \end{aligned}$$

donde $c_2 \in (0, x_2)$ como antes. Seguimos así hasta la penúltima variable, con lo cual nos queda

$$\begin{aligned} f(X) &= f_1(c_1, x_2, \dots, x_n)x_1 + f_2(0, c_2, x_3, \dots, x_n)x_2 + \dots \\ &\quad + \dots + f_{n-1}(0, 0, \dots, c_{n-1}, x_n)x_{n-1} + f(0, 0, 0, \dots, 0, x_n), \end{aligned}$$

con $c_i \in (0, x_i)$ para todo $i = 1 \dots n-1$. Ahora observemos que, como $P = \mathbb{O}$,

$$\langle \nabla f_{\mathbb{O}}, X \rangle = f_1(\mathbb{O})x_1 + f_2(\mathbb{O})x_2 + \dots + \dots + f_{n-1}(\mathbb{O})x_{n-1} + f_n(\mathbb{O})x_n.$$

Al restar y agrupar, usamos la desigualdad triangular para obtener

$$\begin{aligned} |f(X) - \langle \nabla f_{\mathbb{O}}, X \rangle| &\leq |f_1(c_1, x_2, \dots, x_n) - f_1(\mathbb{O})||x_1| + \dots \\ &\quad + |f_{n-1}(0, 0, \dots, c_{n-1}, x_n) - f_{n-1}(\mathbb{O})||x_{n-1}| + \\ &\quad + |f(0, 0, 0, \dots, 0, x_n) - f_n(\mathbb{O})x_n| \end{aligned}$$

Recordemos que $|x_i| \leq \|X\|$ para todo $i = 1 \dots n$. Luego al dividir por la norma de $\|X\|$, la expresión original nos queda acotada por

$$\begin{aligned} \frac{|f(X) - \langle \nabla f_{\mathbb{O}}, X \rangle|}{\|X\|} &\leq |f_1(c_1, x_2, \dots, x_n) - f_1(\mathbb{O})| + \dots \\ &\quad + |f_{n-1}(0, \dots, c_{n-1}, x_n) - f_{n-1}(\mathbb{O})| \\ &\quad + \frac{|f(0, 0, 0, \dots, 0, x_n) - f_n(\mathbb{O})x_n|}{\|X\|} \end{aligned}$$

Tomemos $\varepsilon > 0$. Como cada $c_i \in (0, x_i)$, y como las $n-1$ primeras derivadas parciales son continuas en $X = \mathbb{O}$ por hipótesis, todos estos términos tienden a cero cuando $X \rightarrow \mathbb{O}$, con lo cual se pueden hacer todos menores que ε/n si $0 < \|X\| < \delta_1$ para algún $\delta_1 > 0$.

Falta ver que el último término se puede hacer menor que ε/n (de esta manera la suma es menor o igual que $n \frac{\varepsilon}{n} = \varepsilon$). Si $x_n = 0$, ya terminamos. Si $x_n \neq 0$, observemos que, si $\|X\| \neq 0$, entonces $1/\|X\| \leq 1/|x_n|$. También recordemos que $f(\mathbb{O}) = 0$, con lo cual el último término es menor o igual que

$$\frac{|f(\mathbb{O} + x_n E_n) - f(\mathbb{O}) - f_n(\mathbb{O})x_n|}{|x_n|} = \left| \frac{f(\mathbb{O} + x_n E_n) - f(\mathbb{O})}{x_n} - f_n(\mathbb{O}) \right|.$$

Pero

$$\lim_{t \rightarrow 0} \frac{f(\mathbb{O} + t E_n) - f(\mathbb{O})}{t} = f_n(\mathbb{O})$$

según la definición de derivada direccional, lo que prueba que este último término también tiende a cero. Elegimos entonces $\delta_2 > 0$ tal que la última diferencia sea menor que ε/n siempre y cuando $|x_n| < \delta_2$.

Tomando $\delta = \min\{\delta_1, \delta_2\}$, se tiene que, si $0 < \|X\| < \delta$, entonces los primeros $n-1$ términos son menores que ε/n , y el último también pues $|x_n| \leq \|X\| < \delta \leq \delta_2$.

Hemos probado, en este caso particular donde $P = \mathbb{O}$, $f(\mathbb{O}) = 0$, que f es diferenciable en \mathbb{O} . Supongamos ahora que f cumple todas las hipótesis, pero $P \neq \mathbb{O}$, $f(P) \neq 0$. Si ponemos

$$\bar{f}(X) = f(X + P) - f(P),$$

entonces \bar{f} verifica $\bar{f}(\mathbb{O}) = 0$, y es fácil ver que las derivadas parciales de \bar{f} en un entorno del origen existen y coinciden con las derivadas parciales de f en un entorno de P , pues X está cerca de cero si y sólo si $X + P$ está cerca de P , con lo cual

$$\begin{aligned} \bar{f}_i(X) &= \lim_{t \rightarrow 0} \frac{\bar{f}(X + t E_i) - \bar{f}(X)}{t} = \lim_{t \rightarrow 0} \frac{f(X + t E_i + P) - f(P) - (f(X + P) - f(P))}{t} \\ &= \lim_{t \rightarrow 0} \frac{f(X + t E_i + P) - f(X + P)}{t} = f_i(P + X). \end{aligned}$$

En particular $\nabla f_P = \nabla \bar{f}_{\mathbb{O}}$, y además \bar{f} verifica todas las hipótesis que pusimos para hacer las cuentas, con lo cual es diferenciable en \mathbb{O} según acabamos de demostrar. Poniendo $Y = X - P$ se tiene

$$\begin{aligned} \lim_{X \rightarrow P} \frac{|f(X) - f(P) - \langle \nabla f_P, X - P \rangle|}{\|X - P\|} &= \lim_{Y \rightarrow \mathbb{O}} \frac{|f(Y + P) - f(P) - \langle \nabla f_P, Y \rangle|}{\|Y\|} \\ &= \lim_{Y \rightarrow \mathbb{O}} \frac{|\bar{f}(Y) - \langle \nabla \bar{f}_{\mathbb{O}}, Y \rangle|}{\|Y\|} = 0, \end{aligned}$$

luego f es diferenciable en P .

4.1. Función Inversa

Empezamos esta sección con un par de observaciones simples. Observemos que si F es diferenciable e inversible, y su inversa también es diferenciable, entonces por la regla de la cadena aplicada a $F^{-1}(F(X)) = X$ se tiene

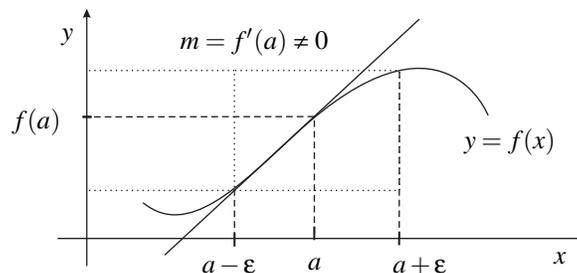
$$DF_{F(x)}^{-1} DF_x = I$$

donde Id es la transformación lineal identidad de \mathbb{R}^n . De aquí se deduce que DF^{-1} (la diferencial de la inversa) es la inversa de la diferencial de F , es decir

$$DF_{F(x)}^{-1} = (DF_x)^{-1}.$$

Supongamos ahora que F es diferenciable solamente ¿Será cierto que si su diferencial es una transformación lineal inversible, entonces F tiene una función inversa, diferenciable?

En el caso de una variable, el gráfico de f nos induce a pensar que sí, pues derivada no nula en un entorno de $x = a$ nos dice que por lo menos f es localmente inyectiva, es decir existe un entorno $(a - \varepsilon, a + \varepsilon)$ donde f es inyectiva y su inversa es derivable.



Atención que en general no podremos exhibir explícitamente esta inversa, sino solamente probar que existe y que verifica ciertas propiedades (y es para esto casos en los cuales no se pueda mostrar la fórmula de la inversa, donde el teorema que queremos enunciar más adelante tiene mayor sentido y utilidad).

Ejemplo 4.1.1. Consideramos $f(x) = x + e^x$. Esta función es derivable y estrictamente creciente pues $f'(x) = 1 + e^x > 0$, con lo cual tiene una inversa en todo \mathbb{R} , y la inversa es derivable. Sin embargo, despejar una fórmula para la inversa a partir de la ecuación $f(x) = y$ es imposible pues la ecuación $x + e^x = y$ no puede resolverse explícitamente.

Para enunciar y demostrar un resultado de estas características en \mathbb{R}^n , necesitamos pensar en todos los ingredientes necesarios, que son: propiedades de las transformaciones lineales, continuidad de las derivadas parciales.

4.1.1. Funciones de clase C^k

Queremos garantizar que si la derivada no se anula en un punto, no se anula en un entorno del punto. Para esto podemos pedir que la derivada sea una función continua. Eso motiva la siguiente definición.

Definición 4.1.2. Sea $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$, con A abierto.

Dado $k \in \mathbb{N}$, decimos que f es de clase C^k en A si para todo punto $P \in A$ existen todas las derivadas parciales hasta orden k y son funciones continuas en A . Al conjunto de todas las funciones de clase C^k en A lo anotamos $C^k(A)$.

Decimos que f es de clase C^∞ (o que $f \in C^\infty(A)$) si existen todas las derivadas parciales de cualquier orden y son todas continuas en A .

Decimos que f es C^0 en A (o que $f \in C^0(A)$) cuando f es continua en A .

Ejemplo 4.1.3.

Cualquier polinomio es una función C^∞ en \mathbb{R} .

Las funciones trigonométricas sen , cos y la exponencial son C^∞ en \mathbb{R} .

El logaritmo es una función C^∞ en $(0, +\infty)$.

Si $f(x) = x^{\frac{7}{3}}$, entonces como $f'(x) = \frac{7}{3}x^{\frac{4}{3}}$ y $f''(x) = \frac{28}{9}x^{\frac{1}{3}}$ se tiene $f \in C^2(\mathbb{R})$. Sin embargo $f \notin C^3(\mathbb{R})$ pues la derivada tercera de f no existe en cero.

Observación 4.1.4. Que la derivada de una función f exista no quiere decir que la derivada f' sea una función continua. Por ejemplo, si

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

entonces $f'(0) = \lim_{h \rightarrow 0} \frac{h^2 \text{sen}(1/h) - 0}{h} = \lim_{h \rightarrow 0} h \text{sen}(1/h) = 0$, mientras que

$$f'(x) = 2x \text{sen}(1/x) + x^2 \cos(1/x) \frac{-1}{x^2} = 2x \text{sen}(1/x) - \cos(1/x).$$

Esto prueba que f es derivable en todo \mathbb{R} . Sin embargo, $\lim_{x \rightarrow 0} f'(x)$ no existe, con lo cual la derivada no es una función continua pues no verifica

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

Es decir, $f \notin C^1(\mathbb{R})$ aunque f si es derivable en \mathbb{R} .

4.1.2. Transformaciones lineales

Algunas observaciones sobre transformaciones lineales $T : \mathbb{R}^n \rightarrow \mathbb{R}^k$.

1. Toda transformación lineal T es diferenciable, con $T_P = T$ para todo $P \in \mathbb{R}^n$, pues

$$T(X) - T(P) = T(X - P),$$

con lo cual el límite de la expresión $\frac{\|T(X) - T(P) - T(X - P)\|}{\|X - P\|}$ es trivialmente cero.

2. Si $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ es diferenciable (ó C^k) y T es una transformación lineal que sale de \mathbb{R}^n , entonces $T \circ F$ es diferenciable (ó C^k), y además

$$D(T \circ F)_P = DT_{F(P)}DF_P = T \circ DF_P.$$

3. Una transformación lineal T es inyectiva si y sólo si $Nu(T) = \{0\}$. En efecto, supongamos primero que T es inyectiva. Como $T0 = 0$ no puede haber otro $V \neq 0$ tal que $TV = 0$, y esto prueba que el núcleo es sólo el cero. Y recíprocamente, si $Nu(T) = 0$, entonces si $TV = TW$, se tiene $TV - TW = 0$, y como T es lineal, $T(V - W) = 0$. Por la hipótesis se concluye que $V - W = 0$, es decir $V = W$ lo que prueba que T es inyectiva.
4. Cuando $n = k$, se tiene que T es inversible (es decir biyectiva) si y sólo si $Nu(T) = \{0\}$, si y sólo si T es sobreyectiva. En efecto por el teorema de la dimensión (ver el libro de Lang [3]),

$$\dim(Nu(T)) + \dim(Im(T)) = n,$$

se deduce que el $Nu(T) = \{0\}$ si y sólo si T es sobreyectiva. Es decir, cuando $n = k$, ser sobreyectiva o ser inyectiva son equivalentes, y por ende cualquiera de las dos condiciones son equivalentes a ser biyectiva.

5. Indicaremos con $I : \mathbb{R}^n \rightarrow \mathbb{R}^n$ a la transformación lineal identidad, $IX = X$ para todo $X \in \mathbb{R}^n$.

Observación 4.1.5. Sea $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ transformación lineal, tal que $\|I - T\|_\infty < 1$. Entonces T es inversible.

Demostración. Supongamos que existe $V \neq 0$ tal que $TV = 0$. Es decir, supongamos que $\ker T$ (el núcleo de T) es no trivial. Entonces

$$\|V\| = \|V - 0\| = \|V - TV\| = \|(1 - T)V\| \leq \|I - T\|_\infty \|V\| < \|V\|,$$

una contradicción. Luego debe ser $Nu(T) = \{0\}$, y el teorema de la dimensión para transformaciones lineales nos dice que T es inversible. \square

Por último veamos cuál es la relación entre ser C^1 y las diferenciales, para ello necesitamos un lema previo de números reales.

Lema 4.1.6. Si a_i, b_i (con $i = 1 \dots n$) son números reales, entonces

$$\sum_{i=1}^n a_i b_i \leq \left(\sum_{i=1}^n a_i^2 \right)^{\frac{1}{2}} \left(\sum_{i=1}^n b_i^2 \right)^{\frac{1}{2}}$$

Demostración. Consideremos los vectores $A = (a_1, \dots, a_n)$ y $B = (b_1, \dots, b_n)$. Entonces por la desigualdad de C-S,

$$\sum_{i=1}^n a_i b_i = \langle A, B \rangle \leq |\langle A, B \rangle| \leq \|A\| \|B\| = \left(\sum_{i=1}^n a_i^2 \right)^{\frac{1}{2}} \left(\sum_{i=1}^n b_i^2 \right)^{\frac{1}{2}}.$$

□

Lema 4.1.7. Sea $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ una transformación lineal, $T = (T_{ij})$. Entonces

$$|T_{ij}| \leq \|T\|_{\infty} \leq n \max_{i,j=1..n} |T_{ij}|.$$

Demostración. La primer desigualdad se deduce de que, como $T(E_i)$ es la columna i -ésima de la matriz de T en la base canónica, el lugar ij lo obtenemos haciendo $\langle TE_i, E_j \rangle$. Como siempre con $\{E_i\}_{i=1..n}$ denotamos la base canónica de \mathbb{R}^n . Luego

$$|T_{ij}| = |\langle TE_i, E_j \rangle| \leq \|TE_i\| \|E_j\| = \|TE_j\| \leq \|T\|_{\infty} \|E_j\| = \|T\|_{\infty}.$$

Para probar la otra desigualdad, tomemos $X \in \mathbb{R}^n$ tal que $\|X\| \leq 1$. Entonces escribimos $X = \sum_{i=1}^n x_i E_i$ donde E_i son los vectores de la base canónica; observemos que $\sum_{i=1}^n |x_i|^2 = \|X\|^2 \leq 1$. Entonces

$$\|TX\| = \left\| \sum_{i=1}^n x_i TE_i \right\| \leq \sum_{i=1}^n |x_i| \|TE_i\|.$$

Esta última suma por el lema previo es menor o igual que el producto

$$\left(\sum_{i=1}^n |x_i|^2 \right)^{\frac{1}{2}} \left(\sum_{i=1}^n \|TE_i\|^2 \right)^{\frac{1}{2}} \leq \left(\sum_{i=1}^n \|TE_i\|^2 \right)^{\frac{1}{2}}.$$

Tomando máximo sobre los X de norma menor o igual a uno se tiene

$$\|T\|_{\infty} \leq \left(\sum_{i=1}^n \|TE_i\|^2 \right)^{\frac{1}{2}}.$$

Por otra parte, para cada $i \in \{1, \dots, n\}$, TE_i devuelve la columna i -ésima de la matriz T en la base canónica, con lo cual

$$\|TE_i\|^2 = \sum_{j=1}^n T_{ji}^2 \leq n \max_{j=1..n} T_{ji}^2.$$

Esto nos dice que

$$\sum_{i=1}^n \|TE_i\|^2 \leq n^2 \max_{i,j=1..n} T_{ji}^2.$$

Tomando raíz cuadrada se obtiene la desigualdad deseada. □

Lema 4.1.8. Sea $G : A \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ con A abierto. Supongamos que $G = (g_1, \dots, g_n)$ es diferenciable en A . Entonces, dados $P, Q \in A$ se tiene

$$\|DG_P - DG_Q\|_\infty \leq n \max_{i,j=1,\dots,n} \left| \frac{\partial g_i}{\partial x_j}(P) - \frac{\partial g_i}{\partial x_j}(Q) \right|.$$

En particular si G es C^1 en A , $\|DG_P - DG_Q\|_\infty \rightarrow 0$ si $Q \rightarrow P$.

Demostración. Le aplicamos el lema previo a $T = DG_P - DG_Q$. □

Enunciamos el teorema de la función inversa.

Teorema 4.1.9 (Teorema de la función inversa). Sean $F : A \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$, con A abierto y $P \in A$. Si F es C^1 en A y $DF_P : \mathbb{R}^n \rightarrow \mathbb{R}^n$ es una transformación lineal inversible, entonces existen entornos abiertos V, W de P y $F(P)$ respectivamente, tales que $F : V \rightarrow W$ es biyectiva, la inversa es diferenciable y además

$$DF_{F(Q)}^{-1} = (DF_Q)^{-1}$$

para todo $Q \in V$. En particular la diferencial de F es inversible en V .

Para demostrarlo conviene separar la demostración en una serie de enunciados. La demostración, y el ejemplo que le sigue, están extraídos del libro *Cálculo en variedades* de M. Spivak [9].

Teorema 4.1.10. Sean $F : A \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$, con A abierto y $P \in A$. Si F es C^1 en A y $DF_P : \mathbb{R}^n \rightarrow \mathbb{R}^n$ es una transformación lineal inversible, entonces

1. Existe $r > 0$ tal que $F : B_r(P) \rightarrow F(B_r(P))$ es biyectiva en $B_r(P)$, y la función inversa es continua en su dominio.
2. Existen abiertos $V \subset B_r(P)$ y $W \subset F(B_r(P))$ tales que $P \in V$ y $F^{-1} : W \rightarrow V$ es diferenciable, con diferencial

$$DF_{F(Q)}^{-1} = (DF_Q)^{-1} \text{ para todo } Q \in V.$$

3. La función inversa es C^1 en su dominio. Si además $F \in C^k(V)$, entonces $F^{-1} \in C^k(W)$. Esto no lo probaremos (hace falta ver que $T \mapsto T^{-1}$ es C^∞ en las matrices inversibles).

Demostración. Supongamos primero que $DF_P = I$.

Veamos 1. Elegimos $r_0 > 0$ tal que $B_{r_0}(P) \subset A$, de manera que $\|DF_Q - I\|_\infty \leq \frac{1}{2}$ siempre que $\|Q - P\| < r_0$ (usando el Lema previo). Tomamos cualquier r positivo tal que $r < r_0$ positivo, para asegurarnos que

$$\overline{B}_r(P) = \{Q \in \mathbb{R}^n : \|Q - P\| \leq r\} \subset A.$$

Si llamamos $G(X) = F(X) - X$ se tiene $DG = DF - I$, con lo cual por el Teorema del valor medio para campos se tiene

$$\|X - Y - F(X) + F(Y)\| = \|X - F(X) - (Y - F(Y))\| = \|G(X) - G(Y)\| \leq \frac{1}{2} \|X - Y\|,$$

siempre que $X, Y \in \overline{B}_r(P)$, pues $\overline{B}_r(P) \subset B_{r_0}(P)$. Por otra parte, por la desigualdad triangular, se tiene

$$\|X - Y\| \leq \|X - Y - F(X) + F(Y)\| + \|F(X) - F(Y)\|,$$

con lo cual

$$\|X - Y\| - \|F(X) - F(Y)\| \leq \|X - Y - F(X) + F(Y)\| \leq \frac{1}{2}\|X - Y\|,$$

es decir

$$\|X - Y\| \leq 2\|F(X) - F(Y)\| \quad (4.1)$$

siempre que $X, Y \in \overline{B}_r(P)$. Esto prueba que F es inyectiva en $\overline{B}_r(P)$, y en particular es inyectiva en su interior $B_r(P)$. Además, la desigualdad (4.1) se puede leer así:

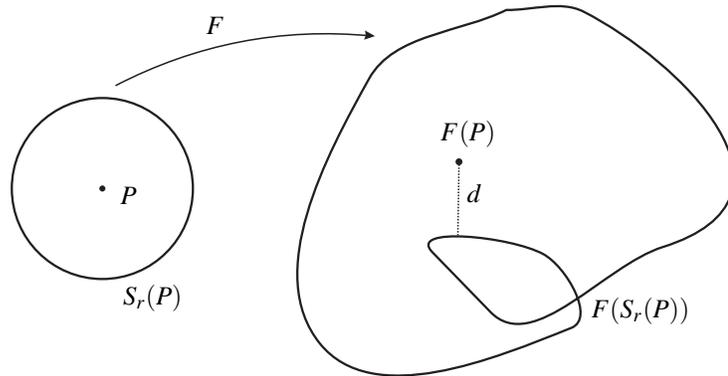
$$\|F^{-1}(Z) - F^{-1}(W)\| \leq 2\|Z - W\|$$

siempre que $Z, W \in F(B_r(P))$, y esto prueba que la función inversa es continua.

Veamos 2. Construyamos primero el abierto $W \subset F(B_r(P))$. Pongamos $S_r(P) = \{Q \in \mathbb{R}^n : \|Q - P\| = r\}$, que es un conjunto compacto de \mathbb{R}^n , con $S_r(P) \subset A$. Como F es inyectiva en $\overline{B}_r(P)$, se tiene $F(X) \neq F(P)$ para todo $X \in S_r(P)$. Como F es continua en A (por ser diferenciable), entonces la distancia $d_P : S_r(P) \rightarrow \mathbb{R}$ dada por $d_P(X) = \|F(X) - F(P)\|$ alcanza un mínimo en $S_r(P)$, digamos

$$d = \min_{X \in S_r(P)} \|F(X) - F(P)\| > 0,$$

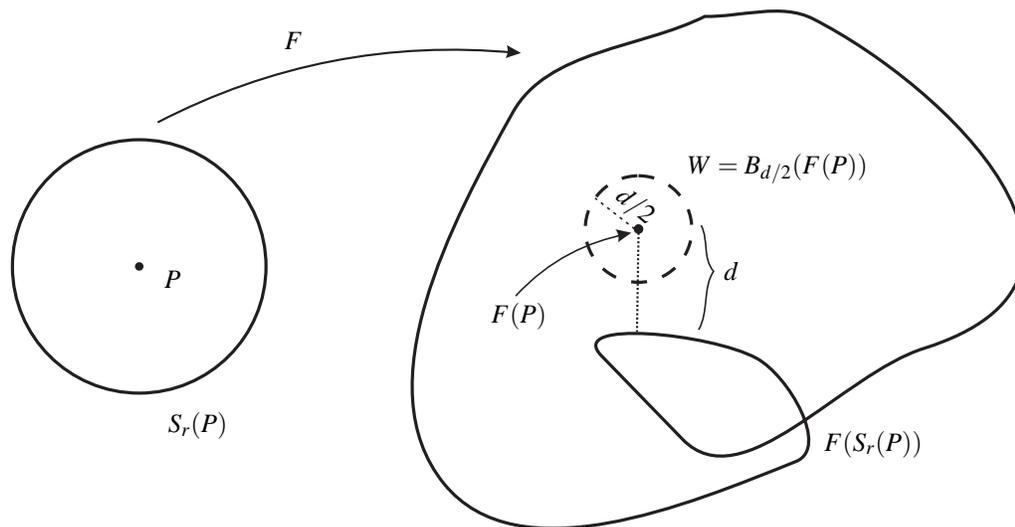
y se tiene en general $\|F(X) - F(P)\| \geq d$ para todo $X \in S_r(P)$.



Es decir, el número positivo d indica la distancia entre $F(S_r(P))$ y el punto $F(P)$.

Ponemos como W a la bola abierta centrada en $F(P)$ de radio $d/2$:

$$W = \{Y \in \mathbb{R}^n : \|Y - F(P)\| < d/2\}.$$



Tenemos que ver que $W \subset F(B_r(P))$, es decir, que dado $Y \in W$, existe $X \in B_r(P)$ tal que $F(X) = Y$. Consideremos (con $Y \in W$ fijo) la función $h : \bar{B}_r(P) \rightarrow \mathbb{R}$ dada por

$$h(X) = \|Y - F(X)\|^2 = \langle Y - F(X), Y - F(X) \rangle,$$

que es una función continua en el compacto $\bar{B}_r(P)$ y diferenciable en $B_r(P)$. Por ser continua tiene un mínimo, digamos $P_Y \in \bar{B}_r(P)$ es el mínimo de h . Afirmando que P_Y verifica $F(P_Y) = Y$, lo que probaría que $Y \in F(B_r(P))$.

Observemos que si X es un punto del borde de la bola, entonces

$$d \leq \|F(P) - F(X)\| \leq \|F(P) - Y\| + \|Y - F(X)\| < d/2 + \|Y - F(X)\|,$$

con lo cual $\|Y - F(X)\| > d/2$, es decir $h(X) > d^2/4$. Por otro lado, evaluando en P se tiene $h(P) = \|Y - F(P)\|^2 < d^2/4$ lo que nos dice que el mínimo de h se debe hallar en el interior de la bola, es decir $P_Y \in B_r(P)$. Como h es diferenciable en la bola abierta, y el punto P_Y es un extremo local, por el Teorema de Fermat en \mathbb{R}^n debe ser $Dh_{P_Y} = 0$. De acuerdo a la regla de derivación del producto escalar se tiene

$$0 = Dh_{P_Y}V = -2\langle DF_{P_Y}V, Y - F(P_Y) \rangle = -2\langle V, DF_{P_Y}^t(Y - F(P_Y)) \rangle$$

para todo $V \in \mathbb{R}^n$, donde $DF_{P_Y}^t$ denota la transformación lineal traspuesta de DF_{P_Y} . Si reemplazamos V por todos los vectores de la base canónica, se deduce que debe ser

$$DF_{P_Y}^t(Y - F(P_Y)) = 0.$$

Como $\|DF_Q - I\|_\infty \leq \frac{1}{2}$ si $\|Q - P\| < r$, entonces DF_Q es invertible por la Observación 4.1.5, y entonces $DF_{P_Y}^t$ es invertible. Debe ser entonces $Y - F(P_Y) = 0$, que es lo que queríamos probar, pues se tiene $P_Y \in B_r(P)$ y $F(P_Y) = Y$.

Observemos que $F^{-1}(W)$ es un conjunto abierto pues F es continua. Si ponemos $V = B_r(P) \cap F^{-1}(W) = \{X \in B_r(P) : F(X) \in W\}$, entonces claramente $P \in V$, $V \subset B_r(P)$, $F : V \rightarrow W$ es una biyección bicontinua y además V es abierto pues es la intersección de dos conjuntos abiertos.

Veamos 3. Tenemos una función inversa $F^{-1} : W \rightarrow V$, definida en un abierto, que es continua. Resta probar que es diferenciable. Para ello recordemos nuevamente que, como $\|DF_Q - I\|_\infty \leq \frac{1}{2}$ si $\|Q - P\| < r$, entonces DF_Q es inversible por la Observación 4.1.5. Resta ver que la diferencial propuesta funciona, es decir, tenemos que ver que

$$\lim_{Z \rightarrow F(Q)} \frac{\|F^{-1}(Z) - F^{-1}F(Q) - DF_Q^{-1}(Z - F(Q))\|}{\|Z - F(Q)\|} = 0.$$

Si hacemos el cambio de variable (válido por ser F biyectiva) $Z = F(Y)$, se tiene

$$\begin{aligned} \frac{\|F^{-1}(Z) - F^{-1}F(Q) - DF_Q^{-1}(Z - F(Q))\|}{\|Z - F(Q)\|} &= \frac{\|Y - Q - DF_Q^{-1}(F(Y) - F(Q))\|}{\|F(Y) - F(Q)\|} \\ &= \frac{\|DF_Q^{-1}(DF_Q(Y - Q) - F(Y) + F(Q))\|}{\|F(Y) - F(Q)\|} \leq \|DF_Q^{-1}\|_\infty \frac{\|F(Y) - F(Q) - DF_Q(Y - Q)\|}{\|F(Y) - F(Q)\|} \\ &\leq 2\|DF_Q^{-1}\|_\infty \frac{\|F(Y) - F(Q) - DF_Q(Y - Q)\|}{\|Y - Q\|} \end{aligned}$$

donde en el último paso hemos usado la desigualdad (4.1), y el hecho de que si $Z \neq F(Q)$, entonces $Y \neq Q$. El teorema queda entonces probado, pues $Y \rightarrow Q$ cada vez que $Z = F(Y) \rightarrow F(Q)$, puesto que la inversa F^{-1} es continua en su dominio, con lo cual el último término tiende a cero pues F es diferenciable en Q .

Nos queda ver que el teorema vale en general, sin suponer que $DF_P = I$. Pero si F es cualquier función que verifica las hipótesis del teorema, y ponemos

$$\bar{F}(X) = DF_P^{-1} \circ F(X),$$

entonces \bar{F} sigue verificando todas las hipótesis y además $D\bar{F}_P = DF_P^{-1}DF_P = I$, con lo cual por las cuentas que hicimos recién se tiene que \bar{F} verifica todos los ítems del enunciado. En consecuencia, $F = DF_P \circ \bar{F}$ también. \square

Observación 4.1.11. La hipótesis $F \in C^1(A)$ es esencial. El por qué, tiene que ver con el hecho siguiente: aunque F sea diferenciable, y su diferencial sea inversible en un punto P , puede no haber ningún entorno de P donde su diferencial sea también inversible. El más curioso puede ver un ejemplo concreto en las notas del final del capítulo (Nota II).

Ejemplo 4.1.12. Consideremos $F(r, \theta) = (r \cos(\theta), r \sin(\theta))$. Entonces se tiene

$$DF_{(r, \theta)} = \begin{pmatrix} \cos(\theta) & -r \sin(\theta) \\ \sin(\theta) & r \cos(\theta) \end{pmatrix}.$$

Pongamos $T = DF_{(r, \theta)}$. Entonces $\det(T) = r(\cos^2(\theta) + \sin^2(\theta)) = r$. Esto nos dice que, salvo en la recta $r = 0$, la diferencial de la función F es inversible.

Así que, para cada punto (r, θ) , con $r \neq 0$, existe un entorno donde F es inversible, y un entorno de $F(r, \theta)$ donde F^{-1} es diferenciable.

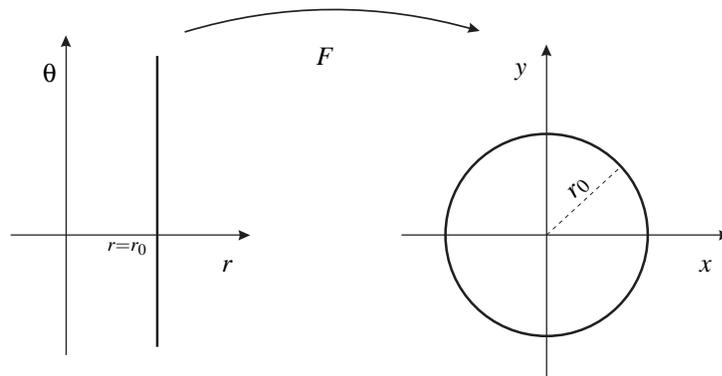
Sin embargo, observemos que F no es biyectiva, pues por ejemplo, $F(1,0) = (1,0) = F(1,2\pi)$. ¿Cómo puede ser? Lo que ocurre es que el teorema nos asegura que F es localmente inyectiva (y por lo tanto inversible) pero si los puntos están lejos, se pierde la inyectividad.

Sin embargo, si nos restringimos al conjunto $r > 0$, $0 \leq \theta < 2\pi$, la función F es inyectiva. Veamos por qué. Si $F(r_1, \theta_1) = F(r_2, \theta_2)$, entonces

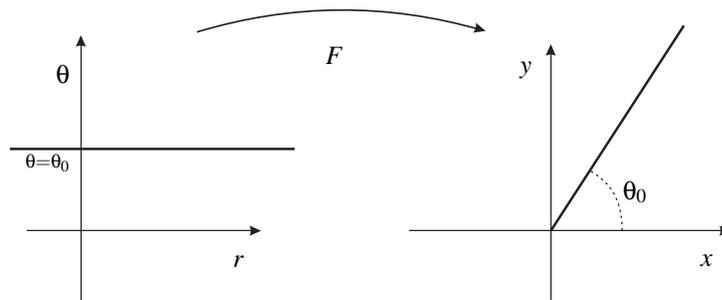
$$r_1 \cos(\theta_1) = r_2 \cos(\theta_2) \quad \text{y} \quad r_1 \sin(\theta_1) = r_2 \sin(\theta_2).$$

Elevando al cuadrado y sumando se tiene $r_1^2 = r_2^2$, con lo cual $r_1 = r_2$. Pero entonces de la primera ecuación se deduce que $\cos(\theta_1) = \cos(\theta_2)$, con lo cual, como el coseno es inyectivo en $[0, 2\pi)$, se tiene también $\theta_1 = \theta_2$.

Para tener en cuenta es que, fijado $r = r_0$, esta transformación manda la recta vertical $r = r_0$ en una circunferencia de radio r .



Asimismo, si fijamos $\theta = \theta_0$, esta recta horizontal tiene como imagen un rayo que parte del origen, formando un ángulo θ_0 con el eje x .



La función F con la restricción que hicimos para que sea biyectiva nos da un cambio de variable $(x,y) \mapsto (r \cos(\theta), r \sin(\theta))$, que se suelen denominar **coordenadas polares**.

4.2. Superficies de nivel y funciones implícitas

Recordemos que dada una función $f : \mathbb{R}^n \rightarrow \mathbb{R}$, las superficies de nivel eran los conjuntos

$$L_c = \{X \in \text{Dom}(f) : f(X) = c\},$$

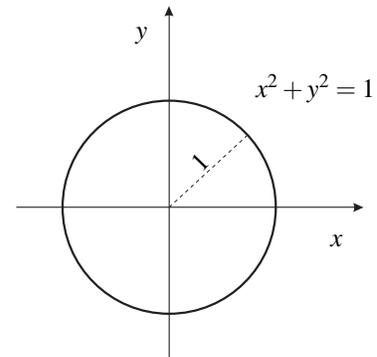
donde $c \in \mathbb{R}$ es una constante.

Observemos que este conjunto puede ser vacío, por ejemplo si $f(x,y) = x^2 + y^2$, y tomamos $c = -1$, entonces la ecuación $x^2 + y^2 = -1$ no tiene ninguna solución.

Estamos interesados en pensar a estas superficies como gráficos de funciones diferenciables. Para ello pensemos en un ejemplo elemental: la circunferencia unitaria. Se obtiene como curva de nivel de $f(x,y) = x^2 + y^2$ tomando $c = 1$, es decir es el conjunto

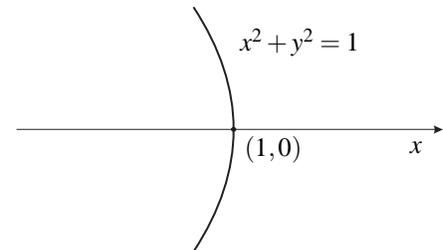
$$L_c(f) = \{(x,y) : x^2 + y^2 = 1\}.$$

Sabemos que podemos despejar y , obteniendo en este caso $y = \pm\sqrt{1-x^2}$. El problema es que esta expresión no está dada por una única función, para ello hay que elegir un signo.



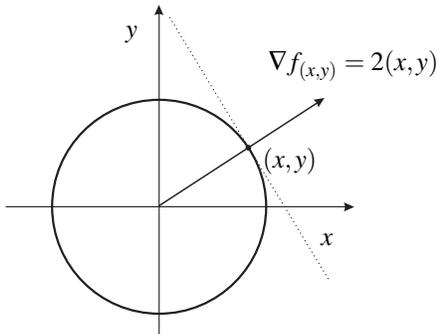
Si queremos la parte superior, consideramos $y = \sqrt{1-x^2}$, con lo cual si ponemos $\varphi(x) = \sqrt{1-x^2}$, la mitad superior de la circunferencia es el gráfico de la función φ , es decir, se escribe como los puntos de la pinta $(x, \varphi(x)) = (x, \sqrt{1-x^2})$. Si queremos la parte inferior, basta tomar $\varphi(x) = -\sqrt{1-x^2}$.

¿Qué pasa si queremos parametrizar algún pedazo alrededor del punto $(1,0)$? No podemos tomar funciones de x , pues cualquier entorno de la circunferencia en ese punto no es el gráfico de ninguna función.



Sin embargo, podemos despejar x para pensarla como función de y . Se obtiene entonces $x = \pm\sqrt{1-y^2}$, y como queremos el lado derecho de la circunferencia tomamos $\varphi(y) = \sqrt{1-y^2}$. Ahora esta porción de la circunferencia es el gráfico de la función φ pensada con dominio en el eje y , es decir son los puntos de la pinta $(\varphi(y), y) = (\sqrt{1-y^2}, y)$.

Observemos que $f(\varphi(y), y) = \varphi(y)^2 + y^2 = 1 - y^2 + y^2 = 1$ para todo y en el dominio de φ . Esto quiere decir que en efecto los puntos de la pinta $(\varphi(y), y)$ son puntos de la circunferencia.



Otro hecho que se observa en este ejemplo sencillo es que, en cada punto (x, y) de la circunferencia, el gradiente de la función $f(x, y) = x^2 + y^2$ está dado por

$$\nabla f_{(x,y)} = (2x, 2y) = 2(x, y).$$

Resulta evidente que el gradiente es perpendicular a la recta tangente de la curva, en cada punto.

Observación 4.2.1. Observemos que en general, si α parametriza la curva

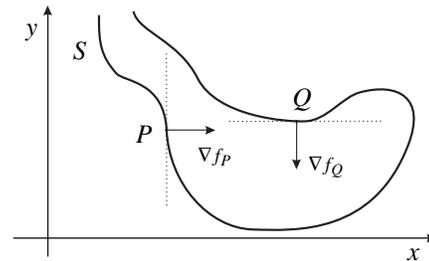
$$S = \{(x, y) : f(x, y) = 0\},$$

entonces $f \circ \alpha(t) = 0$ para todo t en el dominio de α . Entonces (asumiendo que todas las funciones son derivables), por la regla de la cadena se deduce que

$$\langle \nabla f_{\alpha(t)}, \alpha'(t) \rangle = 0 \text{ para todo } t.$$

Es decir, que $\nabla f_{(x,y)}$ es la dirección de la recta normal a la curva en el punto $(x, y) \in S$.

También observemos que, en general, la derivada parcial respecto de y será nula en $P \in S$ si y sólo si el vector gradiente es horizontal, con lo cual la curva de nivel S , en un entorno de ese punto, no puede pensarse como el gráfico de una función $\varphi(x)$. Asimismo, $\frac{\partial f}{\partial x}$ será nula en $Q \in S$ si y sólo si el vector gradiente es vertical, con lo cual la curva no puede parametrizarse en un entorno del punto como gráfico de una $\varphi(y)$.



Todo esto nos lleva a enunciar una versión sencilla del teorema de la función implícita:

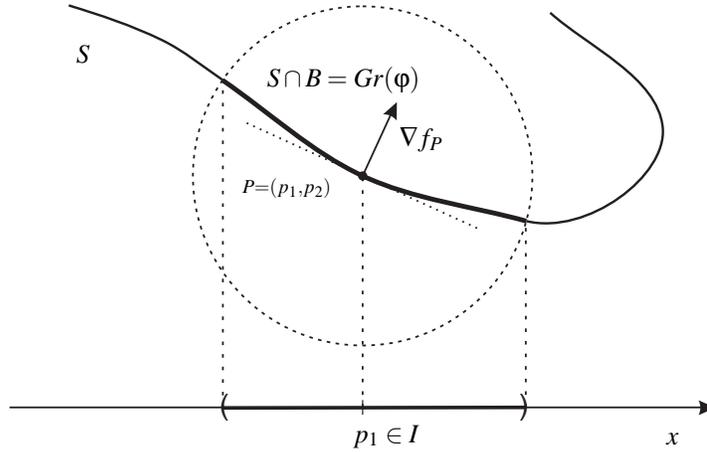
Teorema 4.2.2. (Función implícita en \mathbb{R}^2)

Sea $f : A \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ una función C^1 , y $S = \{(x, y) \in \mathbb{R}^2 : f(x, y) = 0\}$ una curva de nivel de f . Supongamos que $\frac{\partial f}{\partial y}(P) \neq 0$ para algún $P \in S$. Entonces

1. Existen un intervalo abierto $I \subset \mathbb{R}$, una función derivable $\varphi(x)$, $\varphi : I \rightarrow \mathbb{R}$, y una bola B alrededor de P tales que $S \cap B = Gr(\varphi)$.
2. $\frac{\partial f}{\partial y}(x, y) \neq 0$ para todo $(x, y) \in S \cap B$, y $\nabla f_{(x,y)}$ es perpendicular a S en $S \cap B$.
3. Para todo $x \in I$ se tiene

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

La figura a tener en cuenta es la siguiente:



Demostración. Consideramos $F(x,y) = (x, f(x,y))$ que es una función C^1 en A . Tenemos

$$DF = \begin{pmatrix} 1 & 0 \\ \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{pmatrix}.$$

En particular DF_P es inversible. Luego tiene (en una bola abierta B entorno de P) una inversa G definida y diferenciable en un entorno abierto W de $F(P) = (p_1, f(p_1, p_2)) = (p_1, 0)$, digamos $G = (g_1, g_2)$. Se tiene

$$(x,y) = F \circ G(x,y) = (g_1(x,y), f(g_1(x,y), g_2(x,y))),$$

lo que nos dice (mirando la primer coordenada) que $g_1(x,y) = x$, y que (mirando la segunda)

$$y = f(x, g_2(x,y)), \quad (4.2)$$

relación válida para todo (x,y) en un entorno abierto de $(p_1, 0)$. Es decir

$$G(x,y) = (x, g_2(x,y)).$$

En particular, volviendo a la ecuación (4.2), tomando $y = 0$ se tiene

$$0 = f(x, g_2(x, 0))$$

se verifica para todo x en un entorno I de p_1 , con lo cual si tomamos $\varphi(x) = g_2(x, 0)$ con dominio I , esta función es derivable y además cumple

$$f(x, \varphi(x)) = 0$$

para todo $x \in I$. Esto nos dice que $Gr(\varphi) \subset S$, pero como además φ se obtiene como una restricción de G , es decir $(x, \varphi(x)) = G(x, 0)$, también se tiene $Gr(\varphi) \subset B$. Esto prueba que $Gr(\varphi) \subset S \cap B$. Por otra parte dado $Q \in B$, existe $Z \in W$ tal que $G(Z) = Q$, con lo cual

$$(q_1, q_2) = (z_1, g_2(z_1, z_2))$$

y si además $Q \in S$, debe ser $f(q_1, q_2) = 0$ lo que nos dice que

$$z_2 = f(z_1, g_2(z_1, z_2)) = f(q_1, q_2) = 0$$

y entonces

$$(q_1, q_2) = (z_1, g_2(z_1, 0)) = (z_1, \varphi(z_1)),$$

lo que prueba la otra inclusión $S \cap B \subset Gr(\varphi)$.

Como DF_Q es inversible en B , debe ser $\frac{\partial f}{\partial y}(x, y) \neq 0$ para todo $(x, y) \in S \cap B$; ciertamente por la regla de la cadena aplicada a $f(x, \varphi(x)) = 0$ se tiene

$$\langle \nabla f_{(x,y)}, (1, \varphi'(x)) \rangle = 0$$

lo que prueba que el gradiente es perpendicular al gráfico, y además desarrollando queda

$$\frac{\partial f}{\partial x} \cdot 1 + \frac{\partial f}{\partial y} \cdot \varphi'(x) = 0,$$

de donde se deduce despejando la última fórmula del enunciado del teorema. \square

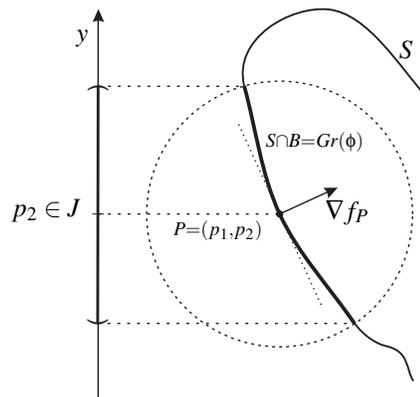
El teorema también vale si intercambiamos x con y , con la misma prueba.

Teorema 4.2.3. Sea $f : A \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ una función C^1 , y $S = \{(x, y) \in \mathbb{R}^2 : f(x, y) = c\}$ una curva de nivel de f . Supongamos que $\frac{\partial f}{\partial x}(P) \neq 0$ para algún $P \in S$. Entonces

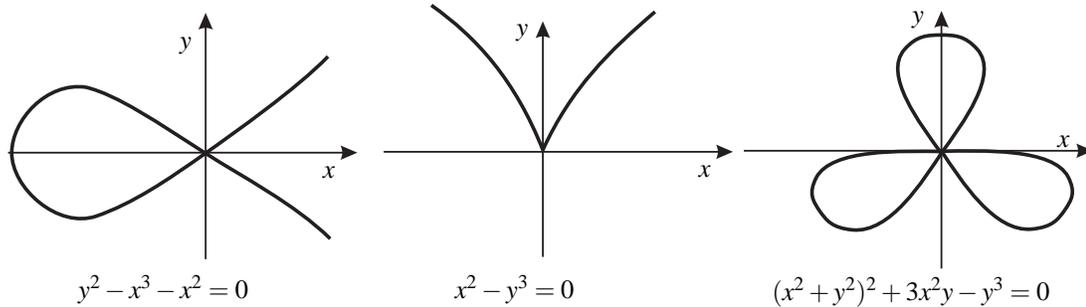
1. Existen un intervalo abierto $J \subset \mathbb{R}$, una función derivable $\phi(y)$, $\phi : J \rightarrow \mathbb{R}$, y una bola B alrededor de P tales que $S \cap B = Gr(\phi)$.
2. $\frac{\partial f}{\partial x}(x, y) \neq 0$ para todo $(x, y) \in S \cap B$, y $\nabla f_{(x,y)}$ es perpendicular a S en $S \cap B$.
3. Para todo $y \in J$ se tiene

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

En este caso la figura a tener en cuenta es la que sigue:



Observación 4.2.4. Algunos ejemplos sencillos, donde el teorema no se puede aplicar, se ilustran en la siguiente figura. En todos los casos el punto relevante es $P = (0,0)$.



En los tres ejemplos, la función f es C^1 (pues es un polinomio). En los tres ejemplos, el gradiente de f se anula en el punto $P = (0,0)$. En ninguno de los tres ejemplos se puede aplicar el teorema para hallar un entorno de cero y una función C^1 que parametrice la curva.

En el ejemplo del medio, sin embargo, podemos despejar $y = x^{2/3}$, luego la curva se puede describir como el gráfico de $\varphi(x) = x^{2/3}$. Lo que ocurre aquí es que esta φ **no** es C^1 en $x = 0$, y por lo tanto no se puede obtener usando el teorema general que enunciamos arriba.

Observación 4.2.5. La hipótesis de que f es C^1 es esencial para asegurar que si una derivada parcial no se anula en $P \in S$, entonces no se anula en un entorno de P . El lector curioso puede ver el ejemplo en la Nota III al final de este capítulo, para entender cómo puede fallar.

Sin embargo, puede probarse que si $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ es una función continua y la derivada parcial $\frac{\partial f}{\partial y}$ existe y no se anula **en un entorno de** $(x_0, y_0) \in S$, entonces existe una función $\varphi = \varphi(x)$ definida en un entorno de $x = x_0$ y continua en un x_0 , que parametriza S en un entorno de (x_0, y_0) . Y que si además f es diferenciable en el punto (x_0, y_0) , esta función φ resulta derivable en $x = x_0$. Ver el Vol. 2 de Rey Pastor [7], Sección 68.1 para una prueba.

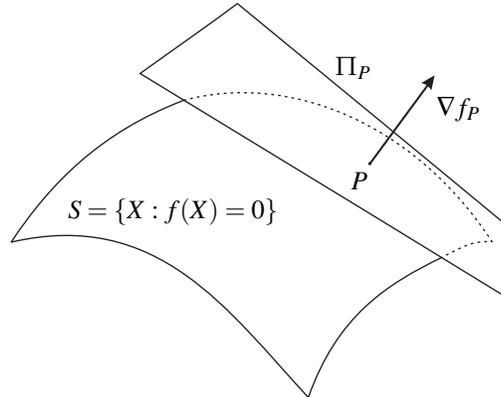
4.2.1. Función implícita en \mathbb{R}^n

El teorema de la función implícita que estudiamos en la sección anterior se generaliza a más variables. En ese caso lo que queremos encontrar son formas de parametrizar -con una función diferenciable φ - una superficie de nivel

$$S = \{X \in \mathbb{R}^n : f(X) = 0\},$$

(o un pedazo de ella) usando la cantidad adecuada de variables.

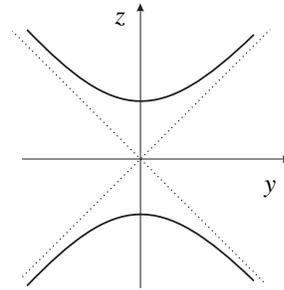
Si se trata de la superficie de nivel de una función $f: \mathbb{R}^n \rightarrow \mathbb{R}$, veremos que se necesitan $n - 1$ variables, y las llamaremos **hipersuperficies de nivel** en analogía con los hiperplanos de \mathbb{R}^n que tienen dimensión $n - 1$. Además, suponiendo que el gradiente de f no se anula en un punto P de la superficie, el vector ∇f_P resultará normal al hiperplano tangente a la superficie en el punto P . Por último podremos calcular cualquier derivada parcial de la función ϕ (que parametriza la superficie S) usando las derivadas parciales de f .



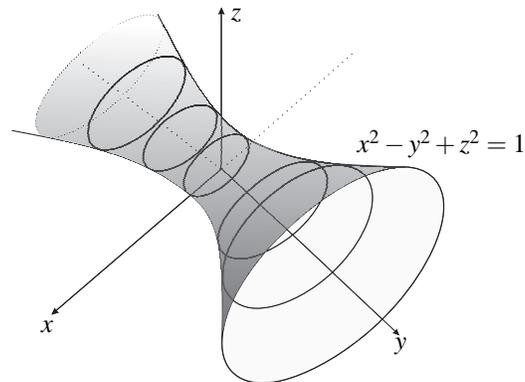
Veamos un ejemplo con tres variables:

Ejemplo 4.2.6. Consideremos la superficie $x^2 - y^2 + z^2 = 1$. Llamando $g(x, y, z) = x^2 - y^2 + z^2 - 1$, se tiene $S = \{(x, y, z) : g(x, y, z) = 0\}$. Esta superficie se conoce como **hiperboloides de una hoja**. Los cortes con los planos $y = cte$ nos dan las curvas $x^2 + z^2 = 1 + y^2 = cte \geq 1$ que son circunferencias centradas en el origen, en el plano xz (de radio ≥ 1).

Mientras que el corte con el plano $x = 0$ nos da la curva de nivel $-y^2 + z^2 = 1$, es decir $(z - y)(z + y) = 1$. El cambio de variable $z - y = u$, $z + y = v$ (que es una rotación de 45 grados) transforma esta ecuación en $uv = 1$, que es una hipérbola pues $v = 1/u$. Luego $z^2 - y^2 = 1$, que es una hipérbola en el plano yz con ejes en las rectas $y = \pm z$.



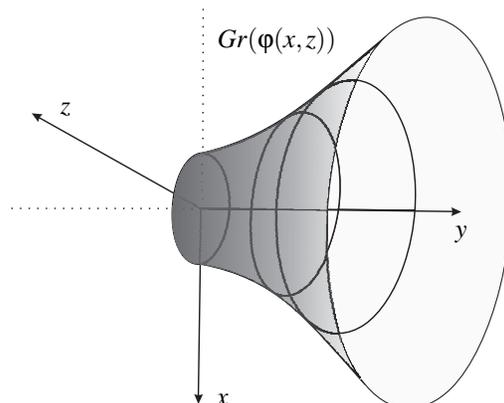
A continuación presentamos un dibujo de la superficie misma.



Si despejamos y en función de x, z obtenemos $y^2 = x^2 + z^2 - 1$. Supongamos que queremos obtener una parametrización en un entorno del punto $P = (1, 1, 1)$, que como se puede verificar fácilmente, es un punto de S . Entonces necesitamos que la coordenada y sea positiva, con lo cual se tiene $y = \phi(x, z) = \sqrt{x^2 + z^2 - 1}$.

Por supuesto que esta expresión tiene sentido solamente en el dominio $x^2 + z^2 - 1 \geq 0$, que como el lector puede verificar, es la parte externa de la circunferencia unitaria en el plano xz dada por $x^2 + z^2 = 1$. Entonces se obtiene una parametrización de la superficie S en un entorno del punto $P = (1, 1, 1)$, pensada como el gráfico de la función $\varphi(x, z)$, es decir

$$\{(x, \sqrt{x^2 + z^2 - 1}, z) : x^2 + z^2 \geq 1\}.$$



Observemos que $\nabla g_{(x,y,z)} = (2x, -2y, 2z)$ y en particular $\nabla g_P = (2, -2, 2)$ lo que nos dice que el plano tangente a la superficie S en el punto $P = (1, 1, 1)$ debe ser

$$\Pi_P : \langle N_P, X - P \rangle = 0 \text{ es decir } \langle (2, -2, 2), (x, y, z) - (1, 1, 1) \rangle = 0,$$

lo que se traduce en $2x - 2y + 2z = 2$. Por último, como

$$g(x, \varphi(x, z), z) = 0,$$

derivando respecto de x se obtiene

$$\frac{\partial g}{\partial x} \cdot 1 + \frac{\partial g}{\partial y} \cdot \frac{\partial \varphi}{\partial x} + \frac{\partial g}{\partial z} \cdot 0 = 0,$$

de donde se despeja

$$\frac{\partial \varphi}{\partial x} = - \frac{\frac{\partial g}{\partial x}}{\frac{\partial g}{\partial y}} \Big|_{(x, \varphi(x, z), z)}.$$

De forma análoga se puede calcular $\frac{\partial \varphi}{\partial z}$. Obviamente la fórmula tiene mayor utilidad cuando **no sabemos quien es φ** . En este caso particular, en cambio, podríamos derivar directamente la expresión que despejamos de φ ,

$$\varphi(x, z) = \sqrt{x^2 + z^2 - 1}.$$

Enunciamos el resultado con más precisión para demostrarlo:

Teorema 4.2.7. (Función Implícita en \mathbb{R}^n)

Sea $A \subset \mathbb{R}^{n+1}$ abierto y sea $f : A \rightarrow \mathbb{R}$ una función C^1 . Sea

$$S = \{X = (x_1, \dots, x_{n+1}) \in \mathbb{R}^{n+1} : f(X) = 0\}$$

una superficie de nivel de f . Si $P \in S$ es tal que alguna de las derivadas parciales de f no se anula en P , entonces

1. Existen una bola abierta $B \subset \mathbb{R}^n$, un abierto $V \subset \mathbb{R}^{n+1}$ entorno de P , y una función $\varphi : B \rightarrow \mathbb{R}$ tal que $S \cap V = Gr(\varphi)$.

2. Se tiene $\nabla f_Q \neq 0$ para todo $Q \in S \cap V$, y allí el plano tangente a S en Q está dado por la ecuación

$$\langle \nabla f_Q, X - Q \rangle = 0,$$

es decir, ∇f_Q es la normal del hiperplano tangente en Q .

3. Si la derivada que no se anula es la k -ésima (con $k \in \{1, \dots, n+1\}$), entonces no se anula en $S \cap V$, y la i -ésima derivada parcial de φ (con $i \in \{1, \dots, n, n+1\}$, $i \neq k$) se calcula de la siguiente manera:

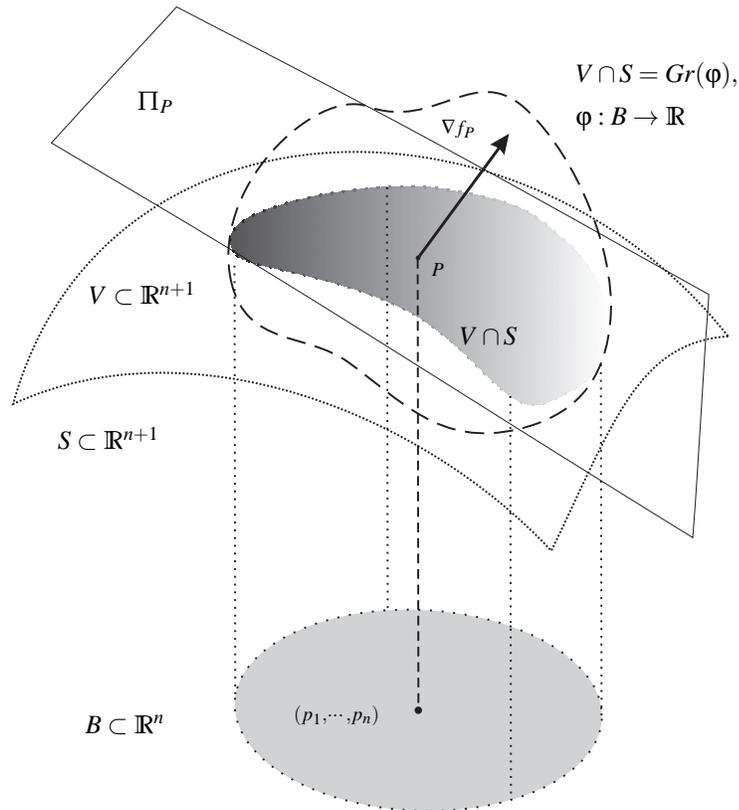
$$\frac{\partial \varphi}{\partial x_i}(Y) = -\frac{\frac{\partial f}{\partial x_i}(Y, \varphi(Y))}{\frac{\partial f}{\partial x_k}(Y, \varphi(Y))}, \text{ para } Y \in B.$$

Demostración. Para la demostración, suponemos que $\frac{\partial f}{\partial x_{n+1}}(P) \neq 0$. El caso general se deduce intercambiando el nombre de las variables. Observemos que esto quiere decir, en el caso de tres variables, que la normal a la superficie S no es horizontal, es decir, tiene componente vertical.

La figura de la derecha contiene los datos del teorema en ese caso, y nos da una idea de la situación general, el punto $P \in S \subset \mathbb{R}^{n+1}$, donde

$$P = (p_1, \dots, p_n, p_{n+1})$$

es donde se verifica que $\frac{\partial f}{\partial x_{n+1}}(P) \neq 0$.



Volviendo al caso de n variables, consideremos la función C^1 auxiliar $F : A \rightarrow \mathbb{R}^{n+1}$ dada por

$$F(x_1, \dots, x_n, x_{n+1}) = (x_1, \dots, x_n, f(x_1, \dots, x_{n+1})),$$

con lo cual la matriz de la diferencial de F en $P = (p_1, \dots, p_n, p_{n+1})$ está dada por la siguiente matriz de $(n+1) \times (n+1)$:

$$DF_P = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & 0 \\ \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \dots & \dots & \frac{\partial f}{\partial x_{n+1}} \end{pmatrix}_P$$

Como supusimos que la última derivada parcial es no nula, esta matriz resulta invertible, y por el teorema de la función inversa (Teorema 4.1.9) existe un entorno abierto V de P en \mathbb{R}^{n+1} y una bola abierta W de radio R centrada en

$$F(P) = (p_1, \dots, p_n, f(p_1, \dots, p_{n+1})) = (p_1, \dots, p_n, 0),$$

también en \mathbb{R}^{n+1} tales que $F^{-1} : W \rightarrow V$ es diferenciable.

Si escribimos $F^{-1}(Z) = F^{-1}(z_1, \dots, z_n, z_{n+1}) = (g_1(Z), \dots, g_n(Z), g_{n+1}(Z))$, la función inversa debe verificar, para todo $Z \in W$,

$$\begin{aligned} (g_1(Z), \dots, g_n(Z), f(g_1(Z), \dots, g_{n+1}(Z))) &= F(g_1(Z), \dots, g_n(Z), g_{n+1}(Z)) \\ &= F(F^{-1}(Z)) = Z. \end{aligned}$$

En particular debe ser, mirando cada una de las primeras n coordenadas,

$$z_i = g_i(Z).$$

Luego, mirando la última coordenada, también debe ser

$$z_{n+1} = f(z_1, \dots, z_n, g_{n+1}(Z))$$

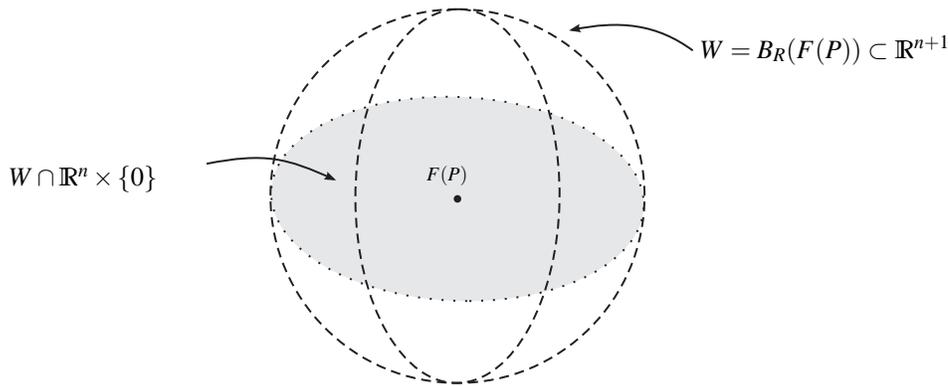
y en particular

$$0 = f(z_1, \dots, z_n, g_{n+1}(z_1, \dots, z_n, 0)). \quad (4.3)$$

Ahora tomamos el corte de la bola W con el plano del piso $\mathbb{R}^n \times \{0\}$, pero pensada en \mathbb{R}^n , es decir

$$B = \{(y_1, \dots, y_n) \in \mathbb{R}^n : \|(y_1 - p_1, \dots, y_n - p_n)\| < R\}$$

según indica la figura (recordemos que $F(P) = (p_1, \dots, p_n, 0)$):



y así definimos $\varphi : B \rightarrow \mathbb{R}$ como la restricción de g_{n+1} a B , es decir

$$\varphi(y_1, \dots, y_n) = g_{n+1}(y_1, \dots, y_n, 0),$$

donde $Y = (y_1, \dots, y_n) \in B \subset \mathbb{R}^n$, la que verifica $Gr(\varphi) \subset S \cap V$ pues la ecuación (4.3) dice que

$$0 = f(y_1, \dots, y_n, \varphi(y_1, \dots, y_n)) = f(Y, \varphi(Y)). \quad (4.4)$$

También se verifica que dado $Q \in V$, existe $Z = (z_1, \dots, z_n, z_{n+1}) \in W$ tal que $F^{-1}(Z) = Q$, y si además $Q \in S$, debe ser forzosamente

$$z_{n+1} = f(z_1, \dots, z_n, g_{n+1}(z_1, \dots, z_n, z_{n+1})) = 0,$$

lo que prueba que $S \cap V \subset Gr(\varphi)$.

Veamos que la normal del plano tangente en $Q \in S \cap V$ está dada por ∇f_Q . Observemos que este gradiente en primer lugar es no nulo, pues según indica el teorema de la función inversa, en este entorno de P se tiene DF_Q inversible, y esto sólo puede ocurrir mirando la matriz de arriba si $\frac{\partial f}{\partial x_{n+1}}(Q) \neq 0$. Por otro lado, el plano tangente en Q está generado por las n derivadas parciales de φ . Escribimos $Q = (Y, \varphi(Y))$, con $Y \in B$. Entonces sólo resta probar que

$$\langle \nabla f_Q, (E_i, \frac{\partial \varphi}{\partial y_i}(Y)) \rangle = 0$$

para todo $i = 1 \dots n$, es decir, que

$$\frac{\partial f}{\partial x_i}(Q) + \frac{\partial f}{\partial x_{n+1}}(Q) \frac{\partial \varphi}{\partial y_i}(Y) = 0.$$

Pero esto se deduce usando la regla de la cadena, derivando en la expresión (4.4) respecto de y_i en el punto Y , pues se tiene

$$0 = \frac{\partial f}{\partial x_i}(Y, \varphi(Y)) + \frac{\partial f}{\partial x_{n+1}}(Y, \varphi(Y)) \frac{\partial \varphi}{\partial x_i}(Y).$$

Por último, observemos que despejando de esta última ecuación se obtiene

$$\frac{\partial \varphi}{\partial x_i}(Y) = -\frac{\frac{\partial f}{\partial x_i}}{\frac{\partial f}{\partial x_{n+1}}}(Y, \varphi(Y))$$

como afirma el teorema. □

El teorema se generaliza para funciones a valores en \mathbb{R}^n .

Teorema 4.2.8. (Función Implícita para campos) Sea $A \subset \mathbb{R}^n$ abierto, y sea $F : A \rightarrow \mathbb{R}^k$ una función C^1 (con $k \leq n$). Sea

$$S = \{Z \in A : F(Z) = \mathbf{0}\}$$

una superficie de nivel de F . Si $P \in S$ es tal que $DF_P : \mathbb{R}^n \rightarrow \mathbb{R}^k$ es sobreyectiva, entonces existen abiertos $U \subset \mathbb{R}^{n-k}$, $V \subset \mathbb{R}^n$ entornos de $\mathbf{0}$ y P respectivamente, y una función $\varphi : U \rightarrow \mathbb{R}^n$ tal que $S \cap V = \varphi(U)$.

Demostración. Daremos sólo una idea de la demostración. Como la diferencial de F en P es sobreyectiva, esta matriz (de n columnas por k filas) tiene k columnas linealmente independientes (es porque la imagen de DF_P tiene que generar todo \mathbb{R}^k). Para simplificar vamos a suponer que son las primeras k columnas (el caso general se deduce intercambiando las variables de F para que esto ocurra). Escribimos $DF_P = \begin{pmatrix} DF_P^k & DF_P^{n-k} \end{pmatrix}$ para indicar con DF_P^k a la matriz inversible de $k \times k$ formada por las primeras k columnas.

Tenemos entonces, escribiendo \mathbb{R}^n como $\mathbb{R}^k \times \mathbb{R}^{n-k}$, y un punto genérico $Z \in \mathbb{R}^n$ como $Z = (X, Y)$ con $(X, Y) \in \mathbb{R}^k \times \mathbb{R}^{n-k}$, que la función $F : A \rightarrow \mathbb{R}^k$ se puede escribir como $F(Z) = F(X, Y)$. Consideramos la función auxiliar

$$H(X, Y) = (F(X, Y), Y),$$

que es una función de $\mathbb{R}^n = \mathbb{R}^k \times \mathbb{R}^{n-k}$ en sí mismo (y es C^1 pues F es C^1). La diferencial de H en $P = (P_1, P_2)$ es de la pinta

$$DH_P = \begin{pmatrix} DF_P^k & DF_P^{n-k} \\ 0 & I_{n-k} \end{pmatrix}_P$$

donde I_{n-k} indica la identidad de tamaño $n - k$. Esta matriz es inversible pues DF_P^k lo es, y por lo tanto el teorema de la función inversa nos da una función $H^{-1} : W \subset \mathbb{R}^k \times \mathbb{R}^{n-k} \rightarrow V \subset \mathbb{R}^k \times \mathbb{R}^{n-k}$ con W entorno de $H(P) = (F(P), P_2)$ y V entorno de $P = (P_1, P_2)$. Esta inversa es de la pinta

$$H^{-1}(X, Y) = (G(X, Y), Y),$$

con $G : \mathbb{R}^k \times \mathbb{R}^{n-k} \rightarrow \mathbb{R}^k$, lo cual se ve usando que es la inversa de H y H tiene esta pinta (ejercicio, ver el caso anterior). El entorno U en \mathbb{R}^{n-k} se construye como antes tomando un entorno producto $U' \times U \subset W$ en \mathbb{R}^n , y la función $\varphi : U \rightarrow \mathbb{R}^n$ se construye tomando, para $Y \in \mathbb{R}^{n-k}$,

$$\varphi(Y) = (G(\mathbf{0}, Y), Y).$$

Se tiene $(X, Y) = H(H^{-1}(X, Y)) = (F(G(X, Y), Y), Y)$ lo que prueba que

$$F(\varphi(Y)) = F(G(\mathbf{0}, Y), Y) = \mathbf{0}$$

lo que nos dice que φ parametriza la superficie de nivel de F , y el resto de las verificaciones quedan a cargo del lector. \square

4.3. NOTAS

- I. Que los ceros de la derivada de una función se pueden acumular en un punto $x = x_0$, aunque en ese punto $f'(x_0) \neq 0$, lo muestra el siguiente ejemplo. Observemos en primer lugar que la función

$$g(z) = \frac{1}{2}z + 2 \operatorname{sen}(z) - z \cos(z)$$

tiene infinitos ceros, puesto que si $n \in \mathbb{N}$, entonces

$$g(n\pi) = n\pi \left[\frac{1}{2} - (-1)^n \right]$$

que es un número positivo si n es impar, y negativo si n es par. Por el teorema de Bolzano, g tiene al menos un cero z_n en cada intervalo $[n\pi, (n+1)\pi]$. Cambiando la variable z por $\frac{1}{x}$, se deduce que la ecuación

$$\frac{1}{2x} + 2 \operatorname{sen}\left(\frac{1}{x}\right) - \frac{1}{x} \cos\left(\frac{1}{x}\right) = 0$$

tiene **infinitos ceros en cualquier entorno de** $x = 0$. Ahora tomemos la función $F : \mathbb{R} \rightarrow \mathbb{R}$ dada por

$$F(x) = \frac{x}{2} + x^2 \operatorname{sen}(1/x),$$

extendida como 0 en $x = 0$. Es fácil probar que F es derivable en \mathbb{R} , y además $F'(0) = \frac{1}{2} \neq 0$ (ver la Observación 4.1.4). Sin embargo, para $x \neq 0$ se tiene

$$F'(x) = \frac{1}{2} + 2x \operatorname{sen}\left(\frac{1}{x}\right) - \cos\left(\frac{1}{x}\right).$$

Y de acuerdo a lo recién señalado, esta derivada se anula arbitrariamente cerca de cero (basta igualar a cero y dividir por x).

- II. Si F es la función del ítem anterior, entonces F es derivable en todo \mathbb{R} , $F(0) = 0$, y su derivada en $x = 0$ es no nula. Sin embargo, no puede existir ninguna función *derivable* F^{-1} en un entorno de $y = 0$ que sea inversa de F . Esto es porque F^{-1} , de existir, tendría que verificar $F^{-1}(F(x)) = x$ para todo x en un entorno de cero, con lo cual derivando tendría que valer la relación

$$\left(F^{-1}\right)'(F(x)) \cdot F'(x) = 1$$

Pero para cada x_0 que anule F' (y en cualquier entorno de cero hay infinitos) se tendría

$$0 = \left(F^{-1}\right)'(F(x_0)) \cdot 0 = \left(F^{-1}\right)'(F(x_0)) \cdot F'(x_0) = 1,$$

un absurdo. Este ejemplo, aunque bastante natural, se lo debemos al libro de M. Spivak [9].

- III. El siguiente ejemplo muestra, en el teorema de la función implícita, no alcanza con que el gradiente no se anule, ni aún siendo f diferenciable. Sea $F(y) = \frac{1}{2}y + y^2 \operatorname{sen}\left(\frac{1}{y}\right)$ (extendida como cero en $y = 0$), la función de la nota anterior, que como vimos es derivable en todo \mathbb{R} , su derivada en cero es $F'(0) = \frac{1}{2}$, pero en cualquier entorno de cero existen infinitos puntos donde F' se anula. Pongamos $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ dada por

$$f(x, y) = x + F(y),$$

y la curva de nivel $S = \{(x, y) : f(x, y) = 0\}$. El punto $P = (0, 0)$ es un punto de S . Es fácil ver que f es diferenciable en todo \mathbb{R}^2 , y como

$$\nabla f(0, 0) = \left(1, \frac{1}{2}\right), \quad \nabla f(x, y) = (1, F'(y)) \text{ para } (x, y) \neq (0, 0)$$

entonces $\nabla f(x, y) \neq 0$ para todo $(x, y) \in \mathbb{R}^2$. Observemos que $\frac{\partial f}{\partial y}(0, 0) = \frac{1}{2}$. Sin embargo, no existe ninguna función derivable $\varphi = \varphi(x)$ definida en algún entorno de $x = 0$ tal que $Gr(\varphi) \subset S$. Para ver por qué, supongamos que **si** existe y llegaremos a un absurdo. Supongamos entonces que existe un intervalo $(-\delta, \delta)$ alrededor de $x = 0$ y una función derivable $\varphi : (-\delta, \delta) \rightarrow \mathbb{R}$ tal que $\varphi(0) = 0$ y además

$$f(x, \varphi(x)) = x + F(\varphi(x)) = 0$$

para todo $x \in (-\delta, \delta)$. Derivando con la regla de la cadena se tiene

$$1 + F'(\varphi(x))\varphi'(x) = 0 \tag{4.5}$$

para todo $x \in (-\delta, \delta)$. En particular, para $x = 0$ se tiene $1 + F'(0)\varphi'(0) = 0$, es decir $\varphi'(0) = -2$. Observemos ahora que φ no puede ser idénticamente nula, puesto que eso diría que su derivada es idénticamente nula. Toma entonces algún valor $b \neq 0$. Pero por ser derivable, φ es una función continua, y entonces toma (por el Teorema de Bolzano) todos los valores intermedios entre 0 y b . En particular φ tiene en su imagen alguno de los ceros $y_0 \neq 0$ de la derivada de la función F , que recordemos son infinitos en cualquier intervalo alrededor de cero. Sea $x_0 \neq 0$ en el dominio de φ tal que $\varphi(x_0) = y_0$. Entonces, por la ecuación (4.5) llegamos a una contradicción, ya que reemplazando x por x_0 tenemos

$$1 = 1 + 0 = 1 + F'(y_0)\varphi'(x_0) = 0.$$

Todos estuvimos de acuerdo en que tu teoría es loca. Lo que nos divide es si es suficientemente loca como para tener alguna chance de ser correcta. Mi impresión es que no es suficientemente loca.

Niels Bohr, en una carta a W. Pauli

5.1. Polinomio de Taylor

Dada una función derivable en un intervalo abierto $I \subset \mathbb{R}$, y un punto $a \in I$, podemos escribir para cualquier $x \in I$, mediante el teorema de Lagrange,

$$f(x) = f(a) + f'(c)(x - a)$$

donde c es un punto entre x y a . Esta fórmula vale para todo x en el intervalo, pero con precaución porque para cada x el c puede ser distinto.

La idea del polinomio de Taylor es aproximar a una función que sea n veces derivable con un polinomio de grado n . Vamos a usar la siguiente notación: $f^{(k)}$ denota la derivada k -ésima de una función y usamos el cero para incluir a la función original, es decir $f^{(0)} = f$.

Recordemos la definición con alguna precisión.

Proposición 5.1.1. *Sea I un intervalo abierto en \mathbb{R} , y sea $f \in C^n(I)$. El polinomio de Taylor de grado n de f en el punto $a \in I$ es el único polinomio $P(x)$ de grado n que verifica*

$$P^{(k)}(a) = f^{(k)}(a)$$

para todo $k \in \{0, \dots, n\}$. La expresión de P es la siguiente:

$$P(x) = f(a) + f'(a)(x - a) + \frac{1}{2}f''(a)(x - a)^2 + \dots + \frac{1}{n!}f^{(n)}(a)(x - a)^n.$$

Además para todo $x \in I$ se tiene

$$f(x) = P(x) + R(x)$$

donde $R(x) = f(x) - P(x)$ es el resto que verifica $\lim_{x \rightarrow a} \frac{R(x)}{(x-a)^n} = 0$.

Todo lo enunciado es conocido, lo único para aclarar es la validez de la última afirmación, que se deduce usando la regla de L'Hospital aplicada (n veces) a

$$\frac{r(x)}{(x-a)^n} = \frac{f(x) - P(x)}{(x-a)^n}.$$

Un resultado más refinado incluye una expresión concreta para el resto, y es el siguiente:

Proposición 5.1.2. (*Taylor con resto de Lagrange*)

Sea I un intervalo abierto en \mathbb{R} , y sea $f : I \rightarrow \mathbb{R}$ una función $n + 1$ veces derivable. Entonces dados $x, a \in I$, existe c estrictamente entre x y a tal que

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + \dots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}.$$

Es decir,

$$f(x) = P_n(x) + R_n(x).$$

Demostración. Fijados $x, a \in I$, consideramos la siguiente función auxiliar de variable t dada por $g : I \rightarrow \mathbb{R}$,

$$g(t) = f(x) - f(t) - f'(t)(x-t) - \frac{1}{2}f''(t)(x-t)^2 - \dots - \frac{1}{n!}f^{(n)}(t)(x-t)^n - \frac{K}{(n+1)!}(x-t)^{n+1}$$

donde $K \in \mathbb{R}$ es una constante adecuada elegida de manera que $g(a) = 0$. Observemos que se trata simplemente de la resta de $f(x)$ con el polinomio de f centrado en t , escrito en la variable x . Se tiene que g es una función derivable de la variable t , y continua en el intervalo cerrado entre x y a . Además se tiene $g(x) = g(a) = 0$. Con lo cual, por el teorema de Rolle existe una constante c entre x y a tal que $g'(c) = 0$. Pero si derivamos g se van cancelando términos y finalmente se tiene

$$g'(t) = \frac{K - f^{(n+1)}(t)}{n!}(x-t)^n,$$

luego reemplazando en $t = c$ se deduce que $K = f^{(n+1)}(c)$. La demostración del teorema concluye si evaluamos la expresión de g en $t = a$ y despejamos $f(x)$. \square

5.1.1. Varias variables

Las derivadas sucesivas de una función juegan un papel relevante en el enunciado del próximo resultado, tengamos antes una pequeña discusión sobre ellas. Dada una función $f : \mathbb{R}^n \rightarrow \mathbb{R}$, tiene sentido calcular (si existen) las derivadas sucesivas de la función, por ejemplo

$$\frac{\partial}{\partial x_i} \frac{\partial f}{\partial x_j},$$

que denotaremos para simplificar como $\frac{\partial^2 f}{\partial x_i \partial x_j}$. Hay que tener la precaución de respetar el orden, ya que en principio podría ser

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(P) \neq \frac{\partial^2 f}{\partial x_j \partial x_i}(P).$$

Por ejemplo: Sea $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ dada por

$$f(x, y) = \begin{cases} \frac{xy^3 - x^3y}{x^2 + y^2} & \text{si } (x, y) \neq (0, 0) \\ 0 & \text{si } (x, y) = (0, 0) \end{cases}$$

Entonces las derivadas parciales cruzadas no coinciden en $(0, 0)$. Esta verificación queda como ejercicio.

Sin embargo, bajo ciertas condiciones estas derivadas coinciden:

Teorema 5.1.3 (Clairaut-Schwarz). Sea $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$ con A abierto. Si $f \in C^2(A)$, entonces las derivadas cruzadas coinciden, es decir, para todo $i, j \in \{1, \dots, n\}$ se tiene

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

para todo $P \in A$.

Demostración. La demostración es más sencilla si consideramos una función de \mathbb{R}^2 en \mathbb{R} , y no se pierde generalidad ya que dada una función cualquiera, la restricción a las dos variables que nos interesan nos da una función de sólo dos variables. Consideremos entonces $f(x, y)$ y un punto $P = (a, b) \in A$, y pongamos

$$g(t) = f(a+t, b+t) - f(a+t, b) - f(a, b+t) + f(a, b).$$

Para recordarla sirve de guía el diagrama siguiente:

$$\begin{array}{ccc} f(a, b+t) & \boxed{} & f(a+t, b+t) \\ & & \\ f(a, b) & \boxed{} & f(a+t, b) \end{array}$$

Si ponemos $\varphi(x) = f(x, b+t) - f(x, b)$, entonces podemos escribir usando Taylor

$$g(t) = \varphi(a+t) - \varphi(a) = \varphi'(a)t + \varphi''(c)\frac{t^2}{2}$$

pues φ es una función dos veces derivable, donde c está entre a y t .

Se tiene

$$g(t) = \left(\frac{\partial f}{\partial x}(a, b+t) - \frac{\partial f}{\partial x}(a, b) \right) t + \left(\frac{\partial^2 f}{\partial x^2}(c, b+t) - \frac{\partial^2 f}{\partial x^2}(c, b) \right) \frac{t^2}{2}.$$

Si dividimos por t^2 y hacemos tender t a cero, se tiene

$$\lim_{t \rightarrow 0} \frac{g(t)}{t^2} = \frac{\partial^2 f}{\partial y \partial x}(a, b),$$

ya que, como $\frac{\partial f^2}{\partial x^2}$ es continua, el último término tiende a cero.

Ahora podemos repetir el argumento, considerando $\psi(y) = f(a+t, y) - f(a, y)$. Se tiene entonces

$$g(t) = \psi(b+t) - \psi(b) = \psi'(b)t + \psi''(c)\frac{t^2}{2},$$

para algún otro c entre b y $b+t$. Escribiendo las derivadas y razonando como antes,

$$\lim_{t \rightarrow 0} \frac{g(t)}{t^2} = \frac{\partial^2 f}{\partial x \partial y}(a, b)$$

lo que prueba la igualdad. □

Definición 5.1.4. Diremos que un conjunto $A \subset \mathbb{R}^n$ es **convexo** si dados $X, Y \in A$, el segmento que une X con Y está completamente contenido en A . Equivalentemente, la función $g(t) = tY + (1-t)X$ tiene su imagen contenida en A para todo $t \in [0, 1]$.

Para extender la idea del polinomio de Taylor a \mathbb{R}^n necesitamos hablar de derivadas segundas y terceras.

Si $A \subset \mathbb{R}^2$ es abierto y $g : A \rightarrow \mathbb{R}$, vamos a llamar **matriz Hessiana** o **Hessiano** de g en $P = (a, b) \in A$ a la siguiente matriz de las derivadas segundas de g (suponiendo que existen todas las derivadas segundas):

$$D^2 g_{(a,b)} = Hg_{(a,b)} = \begin{pmatrix} \frac{\partial^2 g}{\partial x^2} & \frac{\partial^2 g}{\partial y \partial x} \\ \frac{\partial^2 g}{\partial x \partial y} & \frac{\partial^2 g}{\partial y^2} \end{pmatrix}_P.$$

En general, se define para una función $g : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$ que tenga todas sus derivadas segundas, la matriz Hessiana de g en $P \in A$, como la matriz de $n \times n$ siguiente:

$$Hg_P = \begin{pmatrix} \dots & \nabla \frac{\partial g}{\partial x_1} & \dots \\ \dots & \nabla \frac{\partial g}{\partial x_1} & \dots \\ \vdots & \vdots & \vdots \\ \dots & \nabla \frac{\partial g}{\partial x_1} & \dots \end{pmatrix}_P.$$

Razonamos de la siguiente manera: si $g : \mathbb{R}^n \rightarrow \mathbb{R}$ es una función C^2 , entonces la función $\nabla g : P \mapsto \nabla g_P$ es una función $\nabla g : \mathbb{R}^n \rightarrow \mathbb{R}^n$, que resulta C^1 . Es decir ∇g es un campo C^1 . Recordemos que para un campo diferenciable cualquiera $H : \mathbb{R}^n \rightarrow \mathbb{R}^n$ dado por $H(P) = (h_1(P), \dots, h_n(P))$ su diferencial se calcula formando la matriz de $n \times n$ que tiene por filas los gradientes de las h_i , es decir

$$DH_P = \begin{pmatrix} \dots & \nabla h_1 & \dots \\ \dots & \nabla h_2 & \dots \\ \vdots & \vdots & \vdots \\ \dots & \nabla h_n & \dots \end{pmatrix}_P.$$

En el caso particular $H = \nabla g$, se deduce que

$$D^2 g_P = DDg_P = D\nabla g_P = Hg_P,$$

puesto que si reemplazamos h_i por $\frac{\partial g}{\partial x_i}$ en la matriz de arriba, obtenemos el Hessiano de g . Como corolario, por la regla de la cadena, si $\alpha : I \subset \mathbb{R} \rightarrow A \subset \mathbb{R}^n$ es una curva derivable,

$$(\nabla g_{\alpha(t)})' = Hg_{\alpha(t)} \cdot \alpha'(t).$$

La matriz Hessiana de una función $f \in C^2(A)$ es simétrica por el teorema de Clairaut. Así por ejemplo si $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ entonces

$$Hf = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial z \partial x} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} & \frac{\partial^2 f}{\partial z \partial y} \\ \frac{\partial^2 f}{\partial x \partial z} & \frac{\partial^2 f}{\partial y \partial z} & \frac{\partial^2 f}{\partial z^2} \end{pmatrix},$$

y la matriz es simétrica pues podemos intercambiar el orden de las derivadas segundas por ser f una función C^2 .

Notemos que, para una función $f(x, y)$ de dos variables, hay 8 derivadas de orden 3 que son

$$\begin{aligned} & \frac{\partial^3 f}{\partial x^3}, \frac{\partial^3 f}{\partial y^3} \\ & \frac{\partial^3 f}{\partial x^2 \partial y}, \frac{\partial^3 f}{\partial y \partial x^2}, \frac{\partial^3 f}{\partial x \partial y \partial x} \\ & \frac{\partial^3 f}{\partial y^2 \partial x}, \frac{\partial^3 f}{\partial x \partial y^2}, \frac{\partial^3 f}{\partial y \partial x \partial y}. \end{aligned}$$

Sin embargo, de las derivadas mixtas hay en realidad sólo dos distintas si f es C^3 . En efecto, se tiene por ejemplo

$$\frac{\partial^3 f}{\partial x \partial y \partial x} = \frac{\partial^2}{\partial x \partial y} \left(\frac{\partial f}{\partial x} \right) = \frac{\partial^2}{\partial y \partial x} \left(\frac{\partial f}{\partial x} \right) = \frac{\partial^3 f}{\partial y \partial x^2}$$

puesto que $\frac{\partial f}{\partial x}$ es C^2 si f es C^3 .

¿Qué ocurre en general con las derivadas de orden 3? De acuerdo a los cálculos de más arriba, dada una función $f : \mathbb{R}^n \rightarrow \mathbb{R}$ diferenciable, su Hessiano se obtiene como la matriz en la que cada fila aparece el gradiente de la respectiva derivada parcial de f , es decir

$$Hf_P = \begin{pmatrix} \nabla \frac{\partial f}{\partial x_1} \\ \nabla \frac{\partial f}{\partial x_2} \\ \vdots \\ \nabla \frac{\partial f}{\partial x_n} \end{pmatrix}_P.$$

Observemos que, visto como función de la variable P , el gradiente $Df_P = \nabla f_P$ toma valores en \mathbb{R}^n , mientras que el Hessiano toma valores en las matrices de $n \times n$, es decir $D^2 f = Hf : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$. Si derivamos una

vez más, sería razonable que nos quede $D^3 f : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n \times n}$. Esto no está mal, pero es poco práctico. Vamos a usar que $\mathbb{R}^{n \times n \times n}$ se puede identificar con las transformaciones lineales de \mathbb{R}^n en $\mathbb{R}^{n \times n}$, y para cada $P \in \mathbb{R}^n$, presentaremos $D^3 f_P$ como una transformación lineal de \mathbb{R}^n en $\mathbb{R}^{n \times n}$.

Para ver como es este término de orden 3, calculemos

$$(Hf_{\alpha(t)})' = DHf_{\alpha(t)} \cdot \alpha'(t) = DD^2 f_{\alpha(t)} \cdot \alpha'(t) = D^3 f_{\alpha(t)} \cdot \alpha'(t),$$

que debe ser (para cada t fijo) una matriz de $n \times n$.

Hay que derivar cada fila de $Hf_{\alpha(t)}$:

$$(Hf_{\alpha(t)})' = \begin{pmatrix} \left(\nabla \frac{\partial f}{\partial x_1} \alpha(t) \right)' \\ \left(\nabla \frac{\partial f}{\partial x_2} \alpha(t) \right)' \\ \vdots \\ \left(\nabla \frac{\partial f}{\partial x_n} \alpha(t) \right)' \end{pmatrix}.$$

Por la fórmula que ya probamos, $(\nabla g_{\alpha(t)})' = Hg_{\alpha(t)} \cdot \alpha'(t)$, se tiene en cada fila

$$(Hf_{\alpha(t)})' = \begin{pmatrix} H \frac{\partial f}{\partial x_1} \Big|_{\alpha(t)} \cdot \alpha'(t) \\ H \frac{\partial f}{\partial x_2} \Big|_{\alpha(t)} \cdot \alpha'(t) \\ \vdots \\ H \frac{\partial f}{\partial x_n} \Big|_{\alpha(t)} \cdot \alpha'(t) \end{pmatrix}. \quad (5.1)$$

Tomando $\alpha(t) = X + tV$ y evaluando en $t = 0$ se tiene que, como $\alpha(0) = X, \alpha'(0) = V$, entonces

$$D^3 f_X(V) = \begin{pmatrix} H \frac{\partial f}{\partial x_1} \Big|_X \cdot V \\ H \frac{\partial f}{\partial x_2} \Big|_X \cdot V \\ \vdots \\ H \frac{\partial f}{\partial x_n} \Big|_X \cdot V \end{pmatrix}.$$

para todo $X \in \mathbb{R}^n$ donde existan las derivadas terceras, para todo $V \in \mathbb{R}^n$. Ciertamente $D^3 f_X(V)$ es lineal en V (fijado X) y toma valores en $\mathbb{R}^{n \times n}$.

Para controlar el tamaño de esta expresión, que surgirá al escribir el resto cuando escribamos el polinomio de grado 2, usamos el siguiente lema:

Lema 5.1.5. Sean $A_i \in \mathbb{R}^{n \times n}$ una familia de n matrices cuadradas de $n \times n$, y sea $X \in \mathbb{R}^n$. Si los vector

$A_i \cdot X$ los ordenamos por filas, nos queda

$$M(X) = \begin{pmatrix} A_1 \cdot X \\ A_2 \cdot X \\ \vdots \\ A_n \cdot X \end{pmatrix},$$

que es una matriz cuadrada de $n \times n$. Entonces

$$\|M(X)\|_\infty \leq \left(\sum_{i=1}^n \|A_i\|_\infty^2 \right)^{\frac{1}{2}} \|X\|.$$

Demostración. Sea $Y \in \mathbb{R}^n$ tal que $\|Y\| \leq 1$. Entonces

$$\|M(X) \cdot Y\|^2 = \left\| \begin{pmatrix} \langle A_1 \cdot X, Y \rangle \\ \langle A_2 \cdot X, Y \rangle \\ \vdots \\ \langle A_n \cdot X, Y \rangle \end{pmatrix} \right\|^2 = \sum_{i=1}^n |\langle A_i \cdot X, Y \rangle|^2.$$

Por otro lado, para cada $i = 1 \dots n$, se tiene

$$|\langle A_i \cdot X, Y \rangle| \leq \|A_i \cdot X\| \|Y\| \leq \|A_i \cdot X\| \leq \|A_i\|_\infty \|X\|.$$

□

Estamos en condiciones de enunciar y demostrar el siguiente teorema sobre el polinomio de Taylor. Recordemos que la fórmula de Taylor para grado 2 nos dice que si g es una función C^3 en un intervalo abierto $I \subset \mathbb{R}$, entonces dados $x, a \in I$, se tiene

$$g(x) = g(a) + g'(a)(x-a) + \frac{1}{2}g''(a)(x-a)^2 + \frac{1}{3!}g'''(c)(x-a)^3,$$

donde c está entre x y a . En particular, si $x = 1$ y $a = 0$, se tiene

$$g(1) = g(0) + g'(0) + \frac{1}{2}g''(0) + \frac{1}{3!}g'''(c),$$

donde $c \in (0, 1)$.

Observemos también que si f es C^3 , en particular es C^2 y por el teorema de Clairaut se tiene que Hf_P es una matriz simétrica para todo $P \in A$.

Teorema 5.1.6 (Taylor de orden dos, en \mathbb{R}^n , con resto de Lagrange). *Sea $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$ con A abierto convexo. Supongamos f es C^3 en A . Entonces dado $P \in A$, para todo $X \in A$ se tiene*

$$f(X) = f(P) + \langle \nabla f_P, X - P \rangle + \frac{1}{2} \langle Hf_P \cdot (X - P), X - P \rangle + R_P(X - P),$$

donde

$$R_P(X-P) = \frac{1}{6} \langle D^3 f_C(X-P) \cdot (X-P), X-P \rangle$$

es el resto (C es algún punto en el segmento entre X y P). El resto verifica

$$\lim_{X \rightarrow P} \frac{R_P(X-P)}{\|X-P\|^2} = 0.$$

Demostración. Tomemos $P, X \in A$, y consideramos la función auxiliar

$$g(t) = f(P + t(X-P)),$$

que es C^3 en un entorno del intervalo $[0, 1]$. Derivando, obtenemos

$$g'(t) = \langle \nabla f_{P+t(X-P)}, X-P \rangle,$$

$$g''(t) = \langle (\nabla f_{P+t(X-P)})', X-P \rangle = \langle H f_{P+t(X-P)} \cdot (X-P), X-P \rangle,$$

y por último

$$g'''(t) = \langle D^3 f_{P+t(X-P)}(X-P) \cdot (X-P), X-P \rangle.$$

Por otro lado, por la fórmula de Taylor en una variable,

$$g(1) = g(0) + g'(0) + \frac{1}{2}g''(0) + \frac{1}{6}g'''(c),$$

donde c está entre 0 y 1. Entonces

$$\begin{aligned} f(X) &= f(P) + \langle \nabla f_P, X-P \rangle + \frac{1}{2} \langle H f_P \cdot (X-P), X-P \rangle \\ &\quad + \frac{1}{6} \langle D^3 f_{P+c(X-P)}(X-P) \cdot (X-P), X-P \rangle, \end{aligned}$$

donde $c \in (0, 1)$ con lo cual $C = P + c(X-P)$ está en el segmento entre P y X . Por último, observemos que

$$|\langle D^3 f_C(X-P) \cdot (X-P), X-P \rangle| \leq \|D^3 f_C(X-P)\|_\infty \|X-P\|^2,$$

y por el Lema previo,

$$\|D^3 f_C(X-P)\|_\infty \leq \sum_{i=1}^n \left(\|H \frac{\partial f}{\partial x_i} |_C\|_\infty^2 \right)^{\frac{1}{2}} \|X-P\|.$$

Notemos que $H \frac{\partial f}{\partial x_i}$ involucra, para cada i , las derivadas de orden 3 de f . Como cada una de ellas es una función continua por hipótesis, en particular es acotada en un entorno dado de P . Con lo cual

$$\|H \frac{\partial f}{\partial x_i} |_C\|_\infty \leq n \max \left| \frac{\partial^3 f}{\partial x_j \partial x_k \partial x_l} (C) \right| \leq M$$

si X está cerca de P . Luego, si X está cerca de P , se tiene

$$|R_P(X-P)| = \frac{1}{6} |\langle D^3 f_C(X-P) \cdot (X-P), X-P \rangle| \leq M \|X-P\|^3,$$

y de aquí se deduce inmediatamente la última afirmación del teorema, pues

$$\frac{|R_P(X-P)|}{\|X-P\|^2} \leq M\|X-P\|$$

para todo X suficientemente cerca de P . \square

5.1.2. Demostración alternativa de la fórmula de Taylor de orden 2 en \mathbb{R}^n

Otra forma de presentar la fórmula de Taylor del Teorema 5.1.6 es posible, y con otra demostración. Atención que se trata de la misma fórmula. La diferencia es que abandonamos la notación vectorial y pasamos a los índices y sumas. Así por ejemplo, si $X = (x_1, \dots, x_n)$ y $P = (p_1, \dots, p_n)$ entonces

$$\langle \nabla f_P, X-P \rangle = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(P)(x_i - p_i),$$

y lo mismo con los demás términos del Teorema 5.1.6. Usaremos reiteradas veces la siguiente identidad, que es una simple aplicación de la regla de la cadena: si g es una función diferenciable, entonces como $(p_i + t(x_i - p_i))' = (x_i - p_i)$, se tiene

$$g(P+t(X-P))' = \sum_{i=1}^n \frac{\partial g}{\partial x_i}(P+t(X-P))(x_i - p_i).$$

Teorema 5.1.7 (Taylor de orden dos, en \mathbb{R}^n , con resto de Lagrange). *Sea $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$ con A abierto convexo. Supongamos f es C^3 en A . Entonces dado $P = (p_1, \dots, p_n) \in A$, para todo $X = (x_1, \dots, x_n) \in A$ se tiene*

$$f(X) = f(P) + \sum_{i=1}^n \frac{\partial f}{\partial x_i}(P)(x_i - p_i) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f}{\partial x_j \partial x_i}(P)(x_j - p_j)(x_i - p_i) + R_P(X-P).$$

La expresión del resto es

$$R_P(X-P) = \frac{1}{6} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \frac{\partial^3 f}{\partial x_k \partial x_j \partial x_i}(C)(x_k - p_k)(x_j - p_j)(x_i - p_i).$$

con C algún punto en el segmento entre X y P , y verifica

$$\lim_{X \rightarrow P} \frac{R_P(X-P)}{\|X-P\|^2} = 0.$$

Demostración. Tomemos $P, X \in A$, y consideramos la función auxiliar

$$g(t) = f(P+t(X-P)),$$

que es C^3 en un entorno del intervalo $[0, 1]$. Derivando, obtenemos

$$g'(t) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(P+t(X-P))(x_i - p_i).$$

Derivando nuevamente, tenemos

$$\begin{aligned}
 g''(t) &= \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i}(P+t(X-P)) \right)' (x_i - p_i) \\
 &= \sum_{i=1}^n \left(\sum_{j=1}^n \frac{\partial}{\partial x_j} \frac{\partial f}{\partial x_i}(P+t(X-P))(x_j - p_j) \right) (x_i - p_i) \\
 &= \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f}{\partial x_j \partial x_i}(P+t(X-P))(x_j - p_j)(x_i - p_i).
 \end{aligned}$$

Derivamos una vez más para obtener

$$\begin{aligned}
 g'''(t) &= \sum_{i=1}^n \sum_{j=1}^n \left(\frac{\partial^2 f}{\partial x_j \partial x_i}(P+t(X-P)) \right)' (x_j - p_j)(x_i - p_i) \\
 &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \frac{\partial^3 f}{\partial x_k \partial x_j \partial x_i}(P+t(X-P))(x_k - p_k)(x_j - p_j)(x_i - p_i).
 \end{aligned}$$

Ahora invocamos la fórmula de Taylor de orden dos, en una variable para la función g en el intervalo $[0, 1]$, que nos dice que

$$g(1) = g(0) + g'(0) + \frac{1}{2}g''(0) + \frac{1}{6}g'''(c),$$

donde $c \in (0, 1)$. Observemos que $C = P + c(X - P)$ es en efecto un punto en el segmento que une X con P . Reemplazando en esta fórmula los valores de g y sus derivadas se tiene la fórmula del enunciado del teorema. Por último, para ver que el límite indicado da cero, observemos que las derivadas terceras son funciones continuas, con lo cual, tomando algún entorno compacto de P , son funciones acotadas, y con esto, si X está cerca de P se tiene

$$\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \left| \frac{\partial^3 f}{\partial x_k \partial x_j \partial x_i}(C) \right| \leq M.$$

Como

$$|x_k - p_k| |x_j - p_j| |x_i - p_i| \leq \|X - P\|^3,$$

se deduce que

$$|R_P(X - P)| \leq \frac{1}{6}M\|X - P\|^3$$

si X está cerca de P . Dividiendo por $\|X - P\|^2$ y tomando límite para $X \rightarrow P$ se tiene la conclusión. \square

¿Cómo quedan estas fórmulas en los casos concretos? Veamos para $n = 2$: ponemos $P = (a, b) \in \mathbb{R}^2$, $X = (x, y) \in \mathbb{R}^2$. Entonces

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

donde todas las derivadas parciales están evaluadas en $P = (a, b)$, y R es el resto dado por las derivadas de orden 3 de f ,

$$\frac{1}{6} \frac{\partial^3 f}{\partial x^3}(C)(x-a)^3 + \frac{1}{6} \frac{\partial^3 f}{\partial y^3}(C)(y-b)^3 + \frac{1}{2} \frac{\partial^3 f}{\partial x^2 \partial y}(C)(x-a)^2(y-b) + \frac{1}{2} \frac{\partial^3 f}{\partial x \partial y^2}(C)(x-a)(y-b)^2,$$

donde las derivadas parciales están evaluadas en un punto intermedio $C = (c_1, c_2)$ en el segmento que une P con X .

Para $n = 3$ el polinomio de grado dos está dado por (aquí $P = (a, b, c)$)

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

donde todas las derivadas parciales están evaluadas en $P = (a, b, c)$. No daremos la expresión explícita del resto, aunque se puede calcular desarrollando R_P en el teorema anterior, tal cual lo hicimos para $n = 2$.

5.2. Extremos

Recordemos que si una función $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}^k$ es diferenciable en un punto $P \in A^o$, y este punto es un extremo local de f , entonces la diferencial de f debe anularse en P . Equivalentemente, el vector gradiente es cero, es decir $\nabla f_P = \mathbf{0}$.

Recordemos un criterio sencillo para funciones en \mathbb{R} para determinar si el extremo es máximo o mínimo:

Proposición 5.2.1. (Criterio de la derivada segunda)

Si f es dos veces derivable en un intervalo abierto I , y se tiene $f'(a) = 0$ para algun $a \in I$, entonces

1. Si $f''(a) > 0$, el punto $x = a$ es un mínimo local de f .
2. Si $f''(a) < 0$, el punto $x = a$ es un máximo local de f .

Observación 5.2.2. Atención que si la derivada segunda también se anula, el criterio no nos dice nada. De hecho, el punto no tiene ni siquiera que ser un extremo: consideremos $f(x) = x^3$. Entonces si $a = 0$, se tiene $f'(a) = 0$ y $f''(a) = 0$, mientras que $x = 0$ no es un extremo local de f .

Por otro lado, si consideramos $g(x) = x^4$, se tiene $g'(0) = 0$ y $g''(0) = 0$, pero sin embargo $x = 0$ es un mínimo local (de hecho, absoluto) de g .

De manera análoga, si consideramos $h(x) = -x^4$, las dos primeras derivadas se anulan en cero, mientras que este punto es un máximo de h .

Vamos a pensar un poco como se generaliza este criterio a dos variables y luego lo demostramos. Recordemos que si f es C^2 , el Hessiano es la matriz simétrica de las derivadas segundas, evidentemente tiene que jugar un papel dominante en la formulación del mismo.

5.2.1. Formas cuadráticas

Dada una matriz cuadrada T de $n \times n$, consideramos la siguiente función

$$Q(X) = \langle TX, X \rangle,$$

denominada **forma cuadrática asociada a T** . Observemos que

$$Q(tX) = t^2 Q(X) \text{ para todo } t \in \mathbb{R},$$

de allí el nombre.

Supongamos que T está diagonalizada, con autovalores $\lambda_1, \dots, \lambda_n \in \mathbb{R}$. Entonces

$$TX = T(x_1, \dots, x_n) = (\lambda_1 x_1, \dots, \lambda_n x_n)$$

con lo cual

$$Q(x_1, \dots, x_n) = \lambda_1 x_1^2 + \lambda_2 x_2^2 + \dots + \lambda_n x_n^2. \quad (5.2)$$

Observemos que los λ_i pueden ser positivos, negativos o cero.

1. Si son todos no nulos, decimos que Q es **no degenerada**.
2. Si alguno (o varios) de los λ_i son nulos, decimos que Q es **degenerada**.
3. Q es **definida positiva** si $\lambda_i > 0$ para todo $i = 1 \dots n$.
4. Q es **definida negativa** si $\lambda_i < 0$ para todo $i = 1 \dots n$.
5. Q es **indefinida** si algunos λ_i son positivos y otros son negativos.

Observemos que una forma indefinida puede ser tanto degenerada como no degenerada. En el caso degenerado, pero no indefinido, decimos que Q es

1. **semidefinida positiva** si $\lambda_i \geq 0$ para todo $i = 1 \dots n$.
2. **semidefinida negativa** si $\lambda_i \leq 0$ para todo $i = 1 \dots n$.

Observemos que, inspeccionando la ecuación (5.2), se tiene

1. Q es **definida positiva** si y sólo si $Q(X) > 0$ para todo $X \neq 0$.
2. Q es **definida negativa** si y sólo si $Q(X) < 0$ para todo $X \neq 0$.
3. Q es **semidefinida positiva** si y sólo si $Q(X) \geq 0$ para todo X .
4. Q es **semidefinida negativa** si y sólo si $Q(X) \leq 0$ para todo X .
5. Q es **indefinida** si y sólo si existen X_1, X_2 tales que $Q(X_1) > 0$ y $Q(X_2) < 0$.

Pasemos ahora al caso general. Recordemos que si T es una matriz simétrica, es diagonalizable. Es decir, existe una base ortonormal $B = \{V_1, \dots, V_n\}$ de \mathbb{R}^n tal que

$$T = C_{BE} D C_{EB},$$

donde $D = M_{BB}(T)$ es una matriz diagonal que tiene a los autovalores de T . Recordemos también que $U = C_{BE}$ es la matriz de cambio de base, con los vectores de la base B escritos en la base canónica puestos como columnas, y se tiene la siguiente propiedad:

$$C_{BE}^t = C_{EB} = C_{BE}^{-1}, \text{ es decir } U^t = U^{-1},$$

con lo cual $T = UDU^t$. Luego

$$Q(X) = \langle TX, X \rangle = \langle UDU^t X, X \rangle = \langle D(U^t X), U^t X \rangle.$$

Ahora llamamos $Y = U^t X = C_{EB} X$; notar que Y es simplemente X_B , o sea $X_B = (y_1, \dots, y_n)$. Entonces

$$Q(X) = \langle DY, Y \rangle = \langle (\lambda_1 y_1, \dots, \lambda_n y_n), (y_1, \dots, y_n) \rangle = \lambda_1 y_1^2 + \dots + \lambda_n y_n^2. \quad (5.3)$$

Se observa que, salvo un cambio de base, la forma cuadrática Q se puede describir completamente con los autovalores.

Para los que se perdieron con la idea de la matriz de cambio de base, hacemos otra demostración: dado $X \in \mathbb{R}^n$, lo escribimos en la base B , es decir como combinación lineal de los vectores de la base B :

$$X = \sum_{i=1}^n y_i V_i = y_1 V_1 + y_2 V_2 + \dots + y_n V_n.$$

Entonces, por las propiedades del producto escalar se tiene

$$Q(X) = \langle TX, X \rangle = \sum_{i=1}^n \sum_{j=1}^n y_i y_j \langle TV_i, V_j \rangle.$$

Como $TV_i = \lambda_i V_i$ por ser los V_i autovectores de T , se tiene

$$Q(X) = \sum_{i=1}^n \sum_{j=1}^n y_i y_j \langle \lambda_i V_i, V_j \rangle = \sum_{i=1}^n \sum_{j=1}^n y_i y_j \lambda_i \langle V_i, V_j \rangle.$$

Como los V_i son una base ortonormal se tiene $\langle V_i, V_j \rangle = 0$ si $i \neq j$, con lo cual sólo quedan los términos en los que $i = j$,

$$Q(X) = \sum_{i=1}^n y_i^2 \lambda_i \langle V_i, V_i \rangle.$$

Por último, como $\langle V_i, V_i \rangle = \|V_i\|^2 = 1$, se tiene

$$Q(X) = \sum_{i=1}^n \lambda_i y_i^2 = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2,$$

que es la misma expresión que obtuvimos en (5.3), con otra prueba.

Si nos convencimos de que el signo de $Q(X)$ sólo depende de los autovalores de T , y no de X , entonces vamos a decir (para cualquier T simétrica), siguiendo la lógica de antes, que $Q(X) = \langle TX, X \rangle$ es

1. **degenerada** si $TX = \mathbf{0}$ para algún $X \neq \mathbf{0}$.
2. **no degenerada** si T es inversible.
3. **definida positiva** si $Q(X) > 0$ para todo $X \neq \mathbf{0}$.
4. **definida negativa** si $Q(X) < 0$ para todo $X \neq \mathbf{0}$.
5. **semidefinida positiva** si $Q(X) \geq 0$ para todo X .
6. **semidefinida negativa** si $Q(X) \leq 0$ para todo X .
7. **indefinida** si existen X_1, X_2 tales que $Q(X_1) > 0$ y $Q(X_2) < 0$.

El siguiente es un criterio útil que usa el determinante. Recordemos que los menores principales de una matriz cuadrada son las submatrices que se obtienen comenzando por la esquina superior izquierda. Por ejemplo, si

$$T = \begin{pmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{pmatrix},$$

entonces los menores principales de T son las siguientes tres matrices de 1×1 , 2×2 y 3×3 respectivamente:

$$T_{11}, \quad \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix}, \quad T.$$

Proposición 5.2.3. Sea T simétrica y $Q(X) = \langle TX, X \rangle$ la forma cuadrática asociada. Entonces Q es

1. **no degenerada** sii $\det(T) \neq 0$.

2. **definida positiva** sii todos los determinantes de los menores principales son estrictamente positivos.
3. **definida negativa** sii todos los determinantes de los menores principales tienen signos alternados, empezando por un número negativo.

Demostración. El ítem 1. es evidente. Demostraremos los ítems 2 y 3 solamente en el caso 2×2 . El caso general se deduce por inducción. Sea $B = \{V_1, V_2\}$ una base ortonormal de autovectores de T con autovalores λ_1, λ_2 respectivamente (que existe por ser T simétrica). Entonces vimos que, si $X_B = (y_1, y_2)$,

$$Q(X) = \lambda_1 y_1^2 + \lambda_2 y_2^2.$$

Veamos 2. Supongamos primero que los dos determinantes son positivos. Es decir $T_{11} > 0$ y $\det(T) > 0$. Entonces como $\det(T) = \lambda_1 \lambda_2$, los dos autovalores deben tener el mismo signo. Por otro lado, $0 < T_{11} = \langle TE_1, E_1 \rangle = Q(E_1)$ con lo cual no pueden ser los dos negativos pues sería $Q(X) < 0$ para todo $X \neq 0$ por la expresión de arriba. Entonces son los dos positivos, es decir T es definida positiva. Recíprocamente, si T es definida positiva, $T_{11} = Q(E_1) > 0$ y por otro lado los dos autovalores deben ser positivos con lo cual $\det(T) > 0$.

Supongamos ahora que $T_{11} < 0$ y $\det(T) > 0$. Nuevamente los dos autovalores tienen el mismo signo pero ahora $Q(E_1) = T_{11} < 0$ con lo cual tienen que ser los dos negativos así que T es definida negativa. Recíprocamente, si T es definida negativa, $\det(T) > 0$ pues los dos autovalores son negativos, y además $T_{11} = Q(E_1) < 0$. \square

En general uno se refiere indistintamente a T o a su forma cuadrática asociada Q . Así “ T es definida positiva” quiere decir que Q es definida positiva. Veamos los casos más relevantes de formas no degeneradas.

En $\mathbb{R}^{2 \times 2}$, tenemos

$$T = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

Entonces

1. T es definida positiva si y sólo si $\lambda_1 > 0$ y $\lambda_2 > 0$.
2. T es definida negativa si y sólo si $\lambda_1 < 0$ y $\lambda_2 < 0$.
3. T será indefinida si y sólo si $\lambda_1 \lambda_2 = \det(T) < 0$.

Conviene tener presentes los dos ejemplos más simples de formas definida positiva y negativa. En ambos casos debe ser $\det(T) > 0$. Se indican a un lado de la matriz los signos de los determinantes de los menores.

$$\begin{pmatrix} + & 0 \\ 0 & + \end{pmatrix} \begin{matrix} + \\ + \end{matrix} \qquad \begin{pmatrix} - & 0 \\ 0 & - \end{pmatrix} \begin{matrix} - \\ + \end{matrix}$$

Sólo con estos signos se consiguen formas respectivamente positivas y negativas, mientras que si $\det(T) < 0$ se tiene una forma indefinida como dijimos.

Veamos ahora que ocurre en $\mathbb{R}^{3 \times 3}$. Tenemos

$$T = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}.$$

Conviene tener presentes los dos ejemplos sencillos de forma definida positiva y negativa respectivamente, y recordar de allí los signos de los menores:

$$\begin{pmatrix} + & 0 & 0 \\ 0 & + & 0 \\ 0 & 0 & + \end{pmatrix} \begin{matrix} + \\ + \\ + \end{matrix} \qquad \begin{pmatrix} - & 0 & 0 \\ 0 & - & 0 \\ 0 & 0 & - \end{pmatrix} \begin{matrix} - \\ + \\ - \end{matrix}.$$

Observación 5.2.4. *Negando estos casos, se tiene que en matrices 3×3 , la forma será no degenerada e indefinida si y sólo si $\det(T) \neq 0$ y ocurre alguno de los dos casos siguientes:*

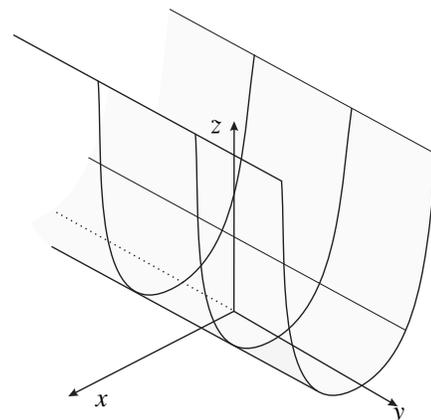
1. *El segundo menor tiene determinante menor a cero.*
2. *El segundo menor es estrictamente positivo y además $\det(T)T_{11} < 0$.*

5.2.2. El Hessiano y los extremos

Recordemos que para que un punto sea un extremo relativo de una función diferenciable f , se debe tener $\nabla f_P = \mathbf{0}$. Sin embargo, como en el caso de una variable esto sólo me da candidatos, resta ver si en efecto son extremos. A estos puntos donde el gradiente de f se anula los llamamos puntos críticos. Entran también en esta denominación aquellos puntos donde f no es diferenciable, pero por ahora nos concentraremos en el primer caso.

Toda la discusión de la sección previa fue para establecer un criterio efectivo para decidir si un punto crítico de una función $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es un extremo local o no, y si es un extremo, si es máximo o mínimo.

Una aclaración: todos los extremos considerados son **locales**, y un extremo es estricto si $f(P) > f(X)$ para todo X en un entorno de P , sin contar P . Un caso sencillo de mínimo (no estricto) es el dado por la parábola trasladada, $f(x, y) = x^2$ cuyo gráfico presentamos a la derecha. Aquí se observa que cualquier punto del eje y es un mínimo de f , pero no es estricto porque si nos movemos a lo largo de este eje la función es constante.



Definición 5.2.5. Diremos que P es **punto silla** de f si existen dos trayectorias α, β (no necesariamente rectas) que tienden a P (o sea son continuas y $\alpha(0) = \beta(0) = P$), y tales que $f \circ \alpha$ tiene un máximo en $t = 0$ y $f \circ \beta$ tiene un mínimo en $t = 0$. Es decir, si hay dos trayectorias continuas de manera que f tiene máximo y mínimo a lo largo de ellas en P . En este caso P no es ni máximo ni mínimo de f .

Vamos a referirnos indistintamente al Hessiano de f en P y a su forma cuadrática asociada

$$Q_P(V) = \frac{1}{2} \langle Hf_P V, V \rangle.$$

Lema 5.2.6. Sea $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$ una función C^3 , con A es abierto y $P \in A$. Supongamos que $\nabla f_P = \mathbf{0}$. Entonces

1. Si existe $V \in \mathbb{R}^n$ tal que $Q_P(V) < 0$, entonces a lo largo de la recta $P + tV$ (para t suficientemente pequeño) la función f tiene un máximo en P . Es decir $g(t) = f(P + tV)$ tiene un máximo local en $t = 0$.
2. Si existe $W \in \mathbb{R}^n$ tal que $Q_P(W) > 0$, entonces a lo largo de la recta $P + tW$ (para t suficientemente pequeño) la función f tiene un mínimo en P . Es decir $h(t) = f(P + tW)$ tiene un mínimo local en $t = 0$.

Demostración. Probamos la primera afirmación, la segunda se deduce de manera similar. Por la fórmula de Taylor, escribiendo $X = P + tV$, se tiene

$$f(P + tV) = f(P) + t^2 Q_P(V) + R_P(tV)$$

para t suficientemente pequeño. Sacando factor común $t^2 \|V\|^2$ se obtiene

$$f(P + tV) = f(P) + t^2 \|V\|^2 \left[\frac{Q_P(V)}{\|V\|^2} + \frac{R_P(tV)}{\|tV\|^2} \right]. \quad (5.4)$$

Si hacemos tender $t \rightarrow 0$, el cociente $\frac{R_P(tV)}{\|tV\|^2}$ tiende a cero. En particular, tomando $\varepsilon = -\frac{Q_P(V)}{\|V\|^2}$, que es un número positivo pues $Q_P(V) < 0$, existe $\delta > 0$ tal que

$$-\varepsilon < \frac{R_P(tV)}{\|tV\|^2} < \varepsilon$$

si $|t| < \delta$. Se deduce de la desigualdad de la derecha, recordando quién es ε , que

$$\frac{Q_P(Z_1)}{\|Z_1\|^2} + \frac{R(tZ_1)}{\|tZ_1\|^2} < 0.$$

Esto es lo mismo, observando la ecuación (5.4), que decir que $f(P + tV) < f(P)$ para t suficientemente pequeño. \square

Observación 5.2.7. Dada $T \in \mathbb{R}^{n \times n}$, la función $Q : \mathbb{R}^n \rightarrow \mathbb{R}$ dada por la forma cuadrática $Q(X) = \langle TX, X \rangle$ es una función continua. De hecho, es diferenciable. Para simplificar, supongamos que, como en toda esta sección, la matriz T es simétrica. Entonces afirmamos que para cualquier $X \in \mathbb{R}^n$, se tiene

$$DQ_X(V) = 2\langle TX, V \rangle.$$

Para verlo, fijado $X \in \mathbb{R}^n$, basta probar que $\frac{Q(Y) - Q(X) - DQ_X(Y - X)}{\|Y - X\|}$ tiende a cero cuando $Y \rightarrow X$. Pero, usando las propiedades del producto escalar,

$$\begin{aligned} Q(Y) - Q(X) - 2\langle TX, Y - X \rangle &= \langle TY, Y \rangle - \langle TX, X \rangle - 2\langle TX, Y \rangle + 2\langle TX, X \rangle \\ &= \langle TY, Y \rangle + \langle TX, X \rangle - 2\langle TX, Y \rangle \\ &= \langle TY, Y \rangle + \langle TX, X \rangle - 2\langle TX, Y \rangle \\ &= \langle TY, Y \rangle - \langle TX, Y \rangle + \langle TX, X \rangle - \langle TX, Y \rangle. \end{aligned}$$

Ahora, como T es simétrica (y el producto escalar también) se tiene $\langle TX, Y \rangle = \langle X, TY \rangle = \langle TY, X \rangle$, con lo cual podemos agrupar de la siguiente manera

$$\begin{aligned} Q(Y) - Q(X) - DQ_X(Y - X) &= \langle TY, Y - X \rangle + \langle TX, X - Y \rangle \\ &= \langle TY, Y - X \rangle + \langle -TX, Y - X \rangle \\ &= \langle TY - TX, Y - X \rangle = \langle T(Y - X), Y - X \rangle. \end{aligned}$$

Con esto, por la desigualdad de C-S, se tiene

$$|Q(Y) - Q(X) - DQ_X(Y - X)| \leq \|T(Y - X)\| \|Y - X\| \leq \|T\|_\infty \|Y - X\|^2.$$

Luego

$$\frac{|Q(Y) - Q(X) - DQ_X(Y - X)|}{\|Y - X\|} \leq \|T\|_\infty \|Y - X\|,$$

lo que prueba que este cociente tiende a cero cuando $Y \rightarrow X$.

Teorema 5.2.8. Sea $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$ una función C^3 , con A es abierto y $P \in A$. Supongamos que $\nabla f_P = \mathbf{0}$. Entonces

1. Si Hf_P es definido negativo, P es un máximo estricto de f .
2. Si Hf_P es definido positivo, P es un mínimo estricto de f .
3. Si Hf_P es indefinida, P es un punto silla de f .

Demostración. Supongamos primero que P es un punto donde Hf_P es definido positivo. Entonces por el Teorema de Taylor (Teorema 5.1.6), para X suficientemente cerca de P se tiene

$$\begin{aligned} f(X) &= f(P) + \frac{1}{2} \langle Hf_P(X - P), X - P \rangle + R(X - P) \\ &= f(P) + \|X - P\|^2 \left[\frac{1}{2} \langle Hf_P \frac{X - P}{\|X - P\|}, \frac{X - P}{\|X - P\|} \rangle + \frac{R(X - P)}{\|X - P\|^2} \right] \end{aligned} \quad (5.5)$$

pues $\nabla f_P = \mathbf{0}$. Llamando $Q_P(V) = \langle \frac{1}{2} Hf_P V, V \rangle$, esta expresión se reescribe así:

$$f(X) = f(P) + \|X - P\|^2 \left[Q_P \left(\frac{X - P}{\|X - P\|} \right) + \frac{R(X - P)}{\|X - P\|^2} \right]. \quad (5.6)$$

Observemos que para cualquier $X \neq P$, el vector en el cuál está evaluado Q_P tiene norma unitaria, es decir

$$\left\| \frac{X-P}{\|X-P\|} \right\| = \frac{1}{\|X-P\|} \|X-P\| = 1.$$

Como $S^{n-1} = \{V \in \mathbb{R}^n : \|V\| = 1\}$ es un conjunto compacto y $Q_P : \mathbb{R}^n \rightarrow \mathbb{R}$ es una función continua, esta función tiene un mínimo m_P en la esfera, es decir existe V_P de norma unitaria tal que $Q(V_P) = m_P$. Como $V_P \neq 0$, este mínimo m_P debe ser mayor a cero pues Hf_P es definido positivo por hipótesis. Entonces, como $\frac{R(X-P)}{\|X-P\|^2} \rightarrow 0$ cuando $X \rightarrow P$, tomando $\varepsilon = m_P$, existe $\delta > 0$ tal que $X \in B_\delta(P)$ implica

$$-m_P < \frac{R(X-P)}{\|X-P\|^2} < m_P.$$

En consecuencia, usando que m_P es el mínimo de Q_P en la esfera, y usando el lado izquierdo de esta última desigualdad, se tiene

$$Q_P \left(\frac{X-P}{\|X-P\|} \right) + \frac{R(X-P)}{\|X-P\|^2} \geq m_P + \frac{R(X-P)}{\|X-P\|^2} > 0$$

siempre que $\|X-P\| < \delta$. Luego, lo que le sigue a $f(P)$ en la ecuación (5.6) es positivo siempre que $\|X-P\| < \delta$, y esto prueba que $f(X) > f(P)$ si $X \in B_\delta(P)$. Es decir, P es un mínimo local de f .

La demostración para el caso definido negativo es similar.

Si Q_P es indefinida, existen dos vectores $Z_1, Z_2 \in \mathbb{R}^n$ tales que $Q_P(Z_1) > 0$ y $Q_P(Z_2) < 0$. Por el lema previo, a lo largo de las trayectorias $P + tZ_1, P + tZ_2$, la función f es respectivamente mayor y menor que $f(P)$, con lo cual P no puede ser ni máximo ni mínimo, y es un punto silla por definición. \square

Ejemplo 5.2.9. Veamos algunos ejemplos.

1. Sea $f(x, y) = 4xy - x^4 - y^4$. Entonces $\nabla f = (4y - 4x^3, 4x - 4y^3) = (0, 0)$ si y sólo si $y = x^3$, $x = y^3$. Entonces $y = y^9$, es decir $y(y^8 - 1) = 0$, de donde se deduce que $y = 0$ o bien $y = \pm 1$. Entonces los puntos críticos son $(0, 0)$, $(1, 1)$, $(-1, -1)$. Calculamos el Hessiano de f :

$$Hf = \begin{pmatrix} -12x^2 & 4 \\ 4 & -12y^2 \end{pmatrix},$$

con lo cual

$$Hf_{(0,0)} = \begin{pmatrix} 0 & 4 \\ 4 & 0 \end{pmatrix}, \quad Hf_{(1,1)} = \begin{pmatrix} -12 & 4 \\ 4 & -12 \end{pmatrix} = Hf_{(-1,-1)}.$$

Como $\det(Hf_{(0,0)}) = -16 < 0$ se deduce que el origen es un punto silla. Los otros dos puntos verifican que el lugar $1, 1$ es estrictamente negativo, mientras que el determinante es igual a $(-12)^2 - 16 = 144 - 16 > 0$, con lo cual los dos autovalores son negativos así que se trata en ambos casos de máximos.

2. Si $f(x, y, z) = x^4 - 2x^2 + y^2 + yz + z^2$, entonces $\nabla f = (4x^3 - 4x, 2y + z, 2z + y) = (0, 0, 0)$ si y sólo si $x^3 - x = 2y + z = 2z + y = 0$. Se deduce que $x = 0$ o bien $x = \pm 1$, y por otro lado que $y = z = 0$. Entonces los puntos críticos son $(0, 0, 0)$, $(\pm 1, 0, 0)$. La matriz Hessiana de f es

$$Hf = \begin{pmatrix} 12x^2 - 4 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix},$$

luego en el origen se tiene

$$Hf_{(0,0,0)} = \begin{pmatrix} -4 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}.$$

Los determinantes de los menores correspondientes son -4 , $-4 \cdot 2 = -8$ y $-4 \cdot 3 = -12$. Como son todos negativos, el origen es un punto silla de f . Por otro lado,

$$Hf_{(\pm 1,0,0)} = \begin{pmatrix} 8 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}.$$

Ahora los determinantes son 8 , $8 \cdot 2 = 16$ y $8 \cdot 3 = 24$. Como son todos positivos, los puntos $(\pm 1, 0, 0)$ son mínimos de f .

Observación 5.2.10. Notemos que en el caso degenerado no podemos asegurar que P sea un extremo de f . Por ejemplo, si $f(x, y) = x^2 + y^3$, entonces

$$\nabla f = (2x, 3y^2),$$

luego $\nabla f_{(0,0)} = (0, 0)$ con lo cual el origen es un punto crítico de f . También

$$Hf = \begin{pmatrix} 2 & 0 \\ 0 & 6y \end{pmatrix}, \text{ con lo cual } Hf_{(0,0)} = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix},$$

que es ciertamente semidefinida positiva (un autovalor positivo y el otro nulo). Sin embargo, si nos acercamos al origen a lo largo del eje y se observa que

$$f(0, y) = y^3,$$

con lo cual el origen no es ni máximo ni mínimo de f .

Sin embargo, hay una relación, dada por el siguiente teorema

Teorema 5.2.11. Sea $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$ una función C^3 (A es abierto). Entonces

1. Si P es un máximo local de f , entonces Hf_P es semidefinido negativo.
2. Si P es un mínimo local de f , entonces Hf_P es semidefinido positivo.

Demostración. En el caso del máximo, sabemos que existe $r > 0$ tal que $f(X) \leq f(P)$ para todo $X \in B_r(P)$. Si Hf_P no fuera semidefinido negativo, existiría $V \in \mathbb{R}^n$ tal que $Q_P(V) > 0$. Por el Lema 5.2.6, a lo largo de la trayectoria $X(t) = P + tV$ se tendría

$$f(P + tV) > f(P)$$

para t suficientemente pequeño, lo cual es una contradicción. La demostración para el caso de un mínimo es análoga. \square

El criterio del Hessiano nos asegura que si el mismo es indefinido, entonces el punto P es punto silla, y las trayectorias son dos rectas como se desprende de la demostración. Sin embargo, si el Hessiano es degenerado podría ocurrir que P fuera un punto silla pero que a lo largo de cualquier recta se encuentre siempre máximo (o siempre mínimo). Un ejemplo de esta situación bastante anti-intuitiva es el siguiente:

Ejemplo 5.2.12. Consideremos $f(x,y) = 2x^4 + y^2 - 3yx^2$. Entonces es fácil ver que $P = (0,0)$ es un punto crítico de f , y que

$$Hf_{(0,0)} = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}.$$

Esto nos dice que a lo largo del eje de las y , f tiene un mínimo en $(0,0)$. También se verifica que a lo largo de cualquier recta la función f tiene un mínimo en el origen. Sin embargo, considerando la trayectoria $y = \frac{3}{2}x^2$ (es decir $\alpha(x) = (x, \frac{3}{2}x^2)$, con x alrededor del cero), se obtiene

$$(f \circ \alpha)(x) = -\frac{1}{4}x^4,$$

con lo cual a lo largo de esta trayectoria f tiene un máximo en el origen. Luego el origen es un punto silla aunque a lo largo de cualquier recta se consiga un mínimo.

5.3. Extremos con restricciones, multiplicadores de Lagrange

5.3.1. Extremos en una región

Dada una función continua $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$ con A un conjunto cualquiera, el problema que nos interesa es el de hallar extremos (tanto relativos como absolutos si los hubiera) de f en A .

Para los puntos del interior A° , primero buscamos los puntos críticos de f con el gradiente. Después hay dos situaciones:

1. La frontera de A la podemos parametrizar con una función ϕ (o varias de ellas), de manera que $Im(\phi) = \partial A$.
2. La frontera de A está dada en forma implícita por una ecuación, de manera que no resulta conveniente (o es directamente imposible) parametrizarla explícitamente.

En el primer caso, lo que hacemos es estudiar la función f restringida al borde componiendo con la parametrización, es decir, hallamos los punto críticos de $g = f \circ \phi$. Recordemos que si A es compacto y f es continua, debe haber tanto máximo como mínimo absoluto de f en A . En el caso general, puede no haber extremos absolutos.

También nos interesan los casos en los que el conjunto A no tiene interior. Típicamente, cuando A es una curva o una superficie de nivel. Estos los tratamos como en el ítem 2 recién mencionado.

5.3.2. Extremos en regiones con borde que se puede parametrizar

Hagamos algunos ejemplos de este caso:

Ejemplo 5.3.1. Sea $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ dada por $f(x,y) = x^2 + y^2 + x$. Hallar los extremos de f restringidos a la bola unitaria cerrada, es decir

$$\bar{B} = \{(x,y) : x^2 + y^2 \leq 1\}.$$

El interior es la bola abierta, allí calculamos

$$\nabla f = (2x+1, 2y), \quad Hf = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

y el único punto crítico es $P_1 = (-\frac{1}{2}, 0)$. Como Hf_P es definido positivo, se trata de un mínimo local de f .

Por otra parte, la frontera se parametriza mediante $\alpha(t) = (\cos t, \sin t)$, para $t \in [0, 2\pi)$. La composición es la función real $g(t) = \cos^2 t + \sin^2 t + \cos t = 1 + \cos t$, cuya derivada es $g'(t) = -\sin t$, que se anula en $t = 0, \pi$. Entonces f tiene extremos en $P_2 = (1, 0)$ y en $P_3 = (-1, 0)$. Evaluando se tiene

$$f(P_1) = \frac{1}{4} - \frac{1}{2} = -\frac{1}{4}, \quad f(P_2) = f(P_3) = 2,$$

lo que nos dice que en P_1 se alcanza el mínimo absoluto de f , mientras que en P_2 y P_3 se alcanza el máximo absoluto de f .

Ejemplo 5.3.2. Sea $f(x,y,z) = x^2 - \frac{1}{2}x + e^{xy}$ en el cuadrado $[0, 1] \times [0, 1]$. En el interior el gradiente y el Hessiano de f son

$$\nabla f = (2x + e^{xy}y - \frac{1}{2}, e^{xy}x), \quad Hf = \begin{pmatrix} 2 + e^{xy}y^2 & e^{xy}xy \\ e^{xy}xy & e^{xy}x^2 \end{pmatrix}.$$

El gradiente se anula únicamente en $P_1 = (0, \frac{1}{2})$. Allí el Hessiano es

$$Hf = \begin{pmatrix} \frac{9}{4} & 0 \\ 0 & 0 \end{pmatrix}$$

que es semidefinido positivo, luego el criterio no se puede usar.

Por otro lado, a lo largo de los cuatro lados se tiene:

1. En $x = 0$, hay que buscar los extremos de $g(y) = f(0,y) = 1$ en el intervalo $[0, 1]$. Como es constante no hay nada para hacer (f vale constantemente uno en el lado izquierdo del cuadrado).
2. En $x = 1$, hay que buscar los extremos de $g(y) = f(1,y) = \frac{1}{2} + e^y$ en el intervalo $[0, 1]$. Como $g'(y) = e^y$ no se anula nunca, lo único relevante son los extremos, $g(0) = f(1,0) = \frac{3}{2}$ y $g(1) = f(1,1) = \frac{1}{2} + e$.
3. En $y = 0$, hay que buscar los extremos de $g(x) = f(x,0) = x^2 - \frac{1}{2}x + 1$ en el intervalo $[0, 1]$. Como $g'(x) = 2x - \frac{1}{2} = 0$ únicamente en $x = \frac{1}{4}$, los puntos relevantes son los bordes (que ya los consideramos porque son los vértices del cuadrado) y el punto $P_2 = (\frac{1}{4}, 0)$.
4. Por último, en $y = 1$, hay que buscar los extremos de $g(x) = f(x,1) = x^2 - \frac{1}{2}x + e^x$ en el intervalo $[0, 1]$. Como $g'(x) = 2x - \frac{1}{2} + e^x$ no se anula nunca (pues $e^x \geq 1$ para $x \geq 0$), lo único relevante son los extremos, que son dos vértices del cuadrado que ya consideramos.

En síntesis,

$$f(0, \frac{1}{2}) = 1, \quad f(\frac{1}{4}, 0) = \frac{15}{16},$$

$f = 1$ en todo el lado izquierdo del cuadrado (incluyendo los vértices), y por último

$$f(1, 0) = \frac{3}{2}, \quad f(1, 1) = \frac{1}{2} + e.$$

De esta lista se deduce que el vértice $(1, 1)$ es el máximo absoluto de f , mientras que el mínimo absoluto se alcanza en el punto $(\frac{1}{4}, 0)$ de la base.

5.3.3. Multiplicadores de Lagrange

No siempre se puede parametrizar la región de manera sencilla. Consideremos el siguiente ejemplo:

Ejemplo 5.3.3. Hallar los máximos y mínimos de $f(x, y) = x^2 + 2x + y^2$ restringida a la curva de nivel $g(x, y) = (x - 1)^2 + 4y^2 - 1 = 0$ de la función g .

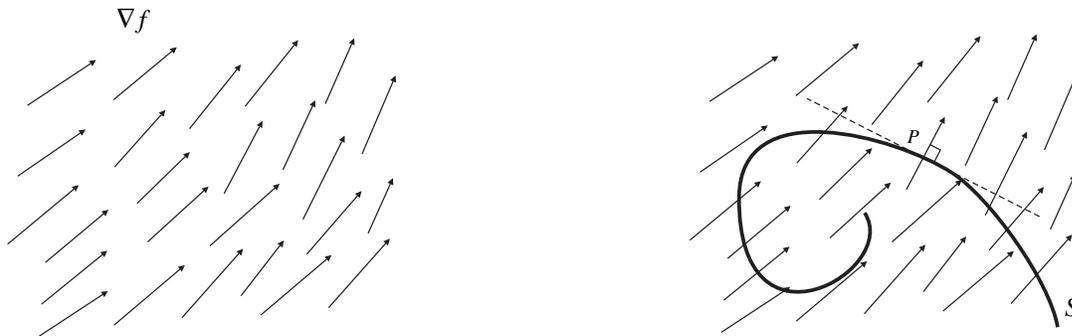
Aquí resulta conveniente otra estrategia, ya que la región $A = \{(x, y) \in \mathbb{R}^2 : g(x, y) = 0\}$ está dada en forma implícita.

Recordemos que para una función escalar f , dado un punto P donde f es diferenciable, la dirección de mayor crecimiento está dada por ∇f_P (y la dirección opuesta es hacia donde f decrece más rápido). También se deduce de

$$\frac{\partial f}{\partial V}(P) = \langle \nabla f_P, V \rangle$$

que si miramos las direcciones perpendiculares al gradiente, las derivadas direccionales son nulas, ya que allí el producto escalar da cero.

Obsevemos la siguiente figura. A la izquierda están representados algunos valores de ∇f , para algunos puntos del plano. A la derecha, superponemos sobre estos valores de ∇f una curva dada S .



En la figura de la derecha, hay que observar que en algunos puntos de la curva, como en el que indicamos como P , el gradiente de f es perpendicular a la curva S . En esos puntos, la dirección de mayor crecimiento es imposible de seguir (habría que salirse de la curva), y por otro lado si nos movemos a lo largo de la curva,

nos estaremos moviendo en forma ortogonal a ∇f_P , con lo cual la derivada direccional de f allí será nula por lo antes dicho. Estos puntos son los candidatos naturales a extremos, si recordamos la idea del teorema de Fermat que dice que si f tiene un extremo entonces su derivada se anula.

Dicho de otra manera, si estamos buscando puntos críticos de f restringidos a una curva, los candidatos naturales son aquellos puntos de la curva donde esta tiene una dirección ortogonal al gradiente de f . Si la curva S está dada por el conjunto de ceros de una función $g : \mathbb{R}^2 \rightarrow \mathbb{R}$, entonces la normal de la curva está dada por ∇g (en el caso en que g sea C^1 y su gradiente no se anule).

Luego, buscamos los puntos P tales que $g(P) = c$, donde además

$$\nabla f_P = \lambda \nabla g_P$$

para algún número real λ . Este método se conoce como método de los **multiplicadores de Lagrange**.

Volvamos al ejemplo concreto: calculamos

$$\nabla f = (2x + 2, 2y), \quad \nabla g = (2(x - 1), 8y)$$

y planteamos la igualdad

$$(2x + 2, 2y) = \lambda(2(x - 1), 8y).$$

De aquí se deduce que debe ser

$$2x + 2 = \lambda 2(x - 1), \quad 2y = \lambda 8y.$$

De la segunda ecuación, se deduce que hay dos posibilidades:

1. $\lambda = \frac{1}{4}$, con lo cual (reemplazando en la primera), se tiene $4x + 4 = x - 1$ es decir $x = -\frac{5}{3}$. Como el punto tiene que estar en la curva de nivel de g , debe verificar la ecuación implícita. Entonces obtenemos

$$\frac{64}{9} + 4y^2 = 1, \text{ es decir } y^2 = -\frac{55}{4 \cdot 9},$$

de donde se deduce que no hay ninguna solución con $x = -\frac{5}{3}$, o equivalentemente, que no hay ninguna solución con $\lambda = \frac{1}{4}$.

2. $y = 0$, con lo cual reemplazando nuevamente en la ecuación se tiene $(x - 1)^2 = 1$, es decir $x = 0$ y $x = 2$. En este caso hay dos puntos que son el $(0, 0)$ y el $(2, 0)$ que son solución.

¿Cómo sabemos si son máximos o mínimos? Observemos que la ecuación de g

$$(x - 1)^2 + 4y^2 = 1$$

define una elipse en el plano \mathbb{R}^2 , que es un conjunto cerrado y acotado (es decir compacto). Como la función f es continua, f debe alcanzar máximo y mínimo absoluto en la elipse. Calculamos

$$f(0, 0) = 0 \quad f(2, 0) = 8.$$

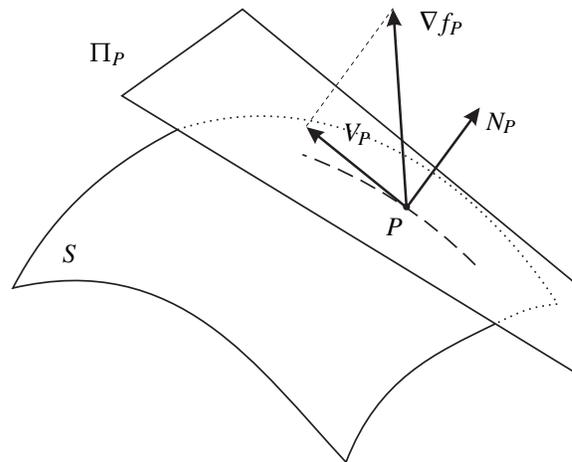
Se concluye que $(0, 0)$ es el mínimo absoluto de f restringida a la curva S , y que $(2, 0)$ es el máximo absoluto de f allí.

5.3.4. Multiplicadores en \mathbb{R}

¿Cómo se generaliza el método al espacio? Dada una función f y una superficie

$$S = \{(x, y, z) \in \mathbb{R}^3 : g(x, y, z) = c\},$$

nuevamente el gradiente ∇f_P indica la dirección de mayor crecimiento (partiendo del punto P), y lo que queremos identificar son aquellos puntos donde las direcciones tangentes a la superficie hacen que las derivadas direccionales de f se anulen. Observemos el siguiente gráfico:



Si el gradiente de f en P tiene alguna componente V_P en el plano tangente Π_P a S en P , entonces moviéndonos en esa dirección, sobre la superficie, la función crecería (pues como dijimos el gradiente es la dirección de mayor crecimiento). La manera de conseguir que no haya ninguna dirección (en la superficie, o equivalentemente en su plano tangente), donde la función crezca, es pidiendo que el gradiente ∇f_P sea **perpendicular** al plano tangente. Equivalentemente, que el gradiente de f sea paralelo a la normal N_P al plano tangente en P , lo que se traduce en la condición

$$\nabla f_P = \lambda \nabla g_P$$

para algún número $\lambda \in \mathbb{R}$. No hay que olvidar la condición $g(P) = c$ para que el punto esté en la superficie. Estas dos condiciones se pueden resumir en el siguiente enunciado, que generaliza lo que ya discutimos en el plano y en el espacio:

Proposición 5.3.4. (*Multiplicadores de Lagrange*)

Para hallar los puntos críticos de una función $f : \mathbb{R}^n \rightarrow \mathbb{R}$ restringida a la superficie de nivel $g(X) = c$ de una función $g : \mathbb{R}^n \rightarrow \mathbb{R}$, basta estudiar los puntos críticos de la función de $n + 1$ variables dada por

$$f(X) - \lambda(g(X) - c).$$

Más precisamente: si $P \in \mathbb{R}^n$ es un extremo de f restringida a la superficie de nivel $g(X) = c$, y $\nabla g_P \neq \mathbf{0}$, entonces existe $\lambda \in \mathbb{R}$ tal que

$$\nabla f_P = \lambda \nabla g_P.$$

Demostración. Si el gradiente de g no se anula en P , entonces por el teorema de la función implícita, existen una bola $B \subset \mathbb{R}^{n-1}$ y una parametrización $\varphi : B \rightarrow \mathbb{R}$ de la superficie S en un entorno de P . Llamemos $Z_0 \in \mathbb{R}^{n-1}$ al centro de la bola B , y para simplificar supongamos que la derivada que no se anula de g es la última, de manera que $(Z_0, \varphi(Z_0)) = P$, y además

$$g(Z, \varphi(Z)) = c \text{ para todo } Z \in B.$$

Es importante recordar que los vectores $(E_i, \frac{\partial \varphi}{\partial z_i}(Z_0))$ (con $i = 1 \dots n$) son generadores del plano tangente a la superficie S en P .

Ahora si P es (por ejemplo) un máximo de f restringida a S , debe ser $f(P) \geq f(X)$ para todo $X \in S$ en un entorno de P . Esto es

$$f(Z_0, \varphi(Z_0)) \geq f(Z, \varphi(Z))$$

para todo $Z \in U$ suficientemente cerca de Z_0 . Entonces la función $h(Z) = f(Z, \varphi(Z))$ tiene un extremo local en Z_0 , con lo cual su gradiente se anula en ese punto, o equivalentemente todas sus derivadas parciales son nulas en el punto. Pero por la regla de la cadena,

$$\frac{\partial h(Z)}{\partial z_i} \Big|_{Z_0} = \frac{\partial}{\partial z_i} f(Z, \varphi(Z)) \Big|_{Z_0} = \langle \nabla f_P, (E_i, \frac{\partial \varphi}{\partial z_i}(Z_0)) \rangle$$

o sea el gradiente de f en P es perpendicular a todos los generadores del plano tangente a la superficie, y en consecuencia tiene que ser paralelo a la normal. Como la normal es el gradiente de g en el punto, se tiene la conclusión. \square

Atención que nada garantiza que los puntos críticos hallados sean extremos.

5.3.5. Un ejemplo elemental

Veamos ahora un ejemplo del uso del método con tres variables. Con este ejemplo empezamos estas notas.

Ejemplo 5.3.5. Hallar las dimensiones de la caja rectangular, de lados a, b, l y de volumen máximo, sujeta a la restricción $a + b + l \leq 300$.

La función a maximizar es $V(x, y, z) = xyz$, y la región que nos interesa estudiar es la comprendida por $x + y + z \leq 300$, $x, y, z > 0$ (observemos que si x, y ó z son cero entonces el volumen es cero). Si calculamos el gradiente de V se tiene

$$\nabla V = (yz, xz, xy)$$

que se anula sólo en los caso que no nos interesan. Resta ver que ocurre en la superficie $x + y + z = 300$, que es un plano (sujeta por supuesto a las restricciones $x, y, z > 0$).

Nuevamente los bordes no son interesantes por dar cero allí el volumen, y lo que nos resta ver es si f tiene extremos restringida al plano. Para ello planteamos $\nabla f = \lambda \nabla g$, obteniendo

$$yz = \lambda, \quad xz = \lambda, \quad xy = \lambda$$

pues $\nabla g = (1, 1, 1)$. Como ninguna de las variables puede ser nula, se deduce fácilmente que $x = y = z$, con lo cual reemplazando en la ecuación del plano se obtiene $3x = 300$, es decir $x = y = z = 100$. Entonces el punto en cuestión debe ser $P = (100, 100, 100)$ (o sea la caja es un cubo de lado 100), que es un máximo pues la región es compacta y los otros son mínimos de f . El volumen máximo entonces es $V = 100^3 = 1000000$.

5.3.6. Varias ligaduras

Por último, una breve discusión sobre el caso en el que uno quiere hallara extremos de una f con restricciones dadas por más de una superficie de nivel. Por ejemplo, hallar los extremos de $f(x, y, z)$ sujeta a las condiciones $g_1(x, y, z) = c_1$, $g_2(x, y, z) = c_2$. En este caso, si los gradientes de g_1 , g_2 son no nulos, ambas ecuaciones definen superficies en el espacio. Si hay algún menor no nulo en la matriz que se obtiene apilando los gradientes, esto indica que la intersección es una curva. En ese caso, los extremos de f en esta curva se hallan planteando

$$\nabla f(P) = \alpha \nabla g_1(P) + \beta \nabla g_2(P)$$

para valores genéricos de $\alpha, \beta \in \mathbb{R}$. No hay que olvidar que P también debe verificar

$$g_1(P) = c_1 \text{ y } g_2(P) = c_2$$

para estar en la curva. La explicación de porqué el gradiente de f en un extremo es combinación lineal de los gradientes de g_1 y g_2 está en que un vector genérico, ortogonal a la intersección de los planos tangentes de ambas superficies, se puede escribir como combinación lineal de las normales a los planos.

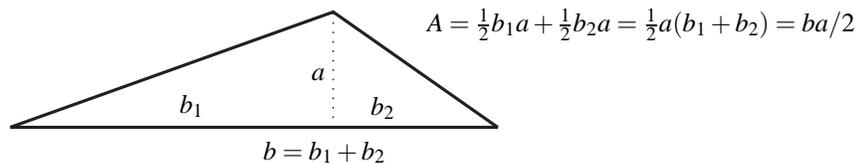
¡Si sólo tuviera los teoremas! Entonces podría encontrar las demostraciones fácilmente.

B. Riemann

6.1. Integrales

Vamos a empezar este capítulo recordando la idea de área en el plano. Todos tenemos presente que dado un rectángulo cualquiera R en el plano, su área o superficie se calcula multiplicando su base por su altura, $A = ba$. También que el área es independiente de la posición del rectángulo, es decir que si lo trasladamos o rotamos, su área permanece inalterada.

Si tenemos una figura más complicada, a veces el área la podemos calcular a partir de este hecho: el primer ejemplo es el de un triángulo rectángulo, donde $A = ba/2$. Esta fórmula se extiende a todos los triángulos, donde ahora la altura está dada por una perpendicular a la base que pase por el vértice opuesto:



¿Qué pasa si la curva que delimita el área es más complicada, como por ejemplo en la región delimitada por el eje de las x y la función $y = \sqrt{x}$? Al área comprendida entre $x = 0$, $x = 1$, el eje x y la gráfica de $f(x) = \sqrt{x}$ podemos aproximarla con rectángulos como indica la figura:



Entonces el área A es aproximadamente la suma de las áreas de los rectángulos de altura a_i y base b_i , es

decir

$$A \simeq \sum a_i \cdot b_i = \sum f(x_i)(t_{i+1} - t_i)$$

donde x_i es algún punto en el intervalo $[t_i, t_{i+1}]$. Como se ve, mientras más rectángulos tomemos (es decir mientras más subdividamos el intervalo), mejor será la aproximación. A estas aproximaciones se las denomina **sumas parciales** o **sumas de Riemann**. Es usual denotar $\Delta_i = [t_i, t_{i+1}]$, aunque con un abuso de notación también usaremos Δ_i para denotar el tamaño del intervalo, es decir $\Delta_i = t_{i+1} - t_i$.

También se ve en el caso particular de arriba que, como tomamos siempre como altura al punto más bajo de la curva, el área que queremos calcular es siempre mayor que nuestra aproximación. Surge naturalmente el concepto de suma inferior, que es una suma que nos da un área inferior a la buscada (pero próxima). Definimos a continuación formalmente las suma inferior y superior asociadas a una partición, para cualquier función acotada (no necesariamente positiva):

Definición 6.1.1. Sean $f : [a, b] \rightarrow \mathbb{R}$ acotada, y P una partición

$$P = \cup_{i=0 \dots n-1} [t_i, t_{i+1}] = \cup_{i=0 \dots n-1} \Delta_i$$

del intervalo $[a, b]$ en n pedazos

$$a = t_0 < t_1 < \dots < t_i < t_{i+1} < \dots < t_n = b.$$

La **suma inferior de f en la partición P** es el número

$$I(f, P) = \sum_{i=1}^n m_i(t_{i+1} - t_i)$$

donde m_i es el ínfimo de $f(x)$ en el intervalo $[t_i, t_{i+1}]$.

Si tomamos siempre el supremo M_i de $f(x)$ en Δ_i , se obtienen la **suma superior de f en la partición P** :

$$S(f, P) = \sum_{i=1}^n M_i(t_{i+1} - t_i).$$

Siguiendo con el ejemplo anterior, una suma superior sería la siguiente:



Como se puede ver, en el caso de una función positiva, las sumas superiores son todas aproximaciones del área bajo la curva (aparentemente cada vez mejores) con la propiedad adicional de ser todas mayores o iguales al área buscada.

Diremos que la partición P' del intervalo $[a, b]$ es un **refinamiento** de la partición P si P' se puede obtener de P subdividiendo los intervalos de P . Asimismo diremos que la norma de $\|P\|$ es el tamaño del intervalo más grande de la partición.

Evidentemente, dada una partición cualquiera P del intervalo, como $m_i \leq M_i$ para todo i , se tiene siempre

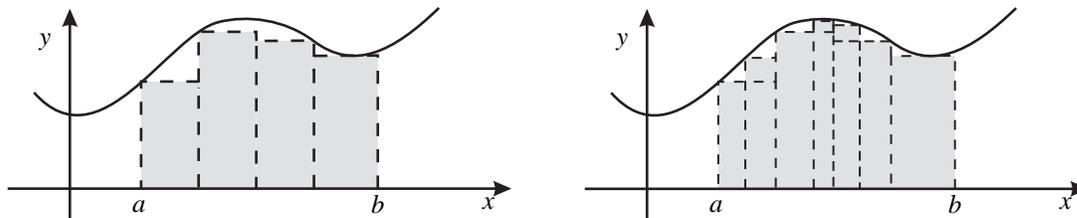
$$I(f, P) \leq S(f, P).$$

Algunas propiedades sencillas que se deducen de este hecho:

Proposición 6.1.2. Sea $f : [a, b] \rightarrow \mathbb{R}$ acotada. Entonces

1. Si P' es un refinamiento de P , entonces $I(f, P') \geq I(f, P)$ y también $S(f, P') \leq S(f, P)$.
2. $I(f, P) \leq S(f, Q)$ **para todo par de particiones P, Q .**

Demostración. Observemos la siguiente figura, donde puede observarse que al refinar la partición, el área cubierta por la suma inferior es cada vez mayor:



Si P' refina P , entonces dado un intervalo cualquiera Δ_i de P este se descompone como una unión de intervalos de P' ,

$$[t_i, t_{i+1}] = \Delta_i = \cup_{j=1 \dots k_i} \Delta'_j = \cup_{j=1 \dots k_i} [t'_j, t'_{j+1}].$$

Entonces como $m_i \leq m_j$ para todo $j = 1 \dots k_i$ (pues el ínfimo en un conjunto es menor o igual que el ínfimo en un subconjunto), se tiene

$$m_i(t_{i+1} - t_i) = m_i \left(\sum_{j=1}^{k_i} t'_{j+1} - t'_j \right) = \sum_{j=1}^{k_i} m_i(t'_{j+1} - t'_j) \leq \sum_{j=1}^{k_i} m_j(t'_{j+1} - t'_j).$$

Sumando sobre i se tiene

$$I(f, P) \leq I(f, P').$$

Esto prueba que al refinar la partición las sumas inferiores aumentan. Con un argumento análogo se deduce que al refinar la partición las sumas superiores disminuyen.

Para probar el segundo ítem, tomemos P' un refinamiento en común de las particiones P, Q . Este siempre se puede conseguir subdividiendo el intervalo en partes más pequeñas que incluyan los puntos de corte de P y Q . Entonces por el ítem anterior,

$$I(f, P) \leq I(f, P') \leq S(f, P') \leq S(f, Q),$$

lo que prueba que $I(f, P) \leq S(f, Q)$. □

Este resultado nos dice algo importante: que las sumas superiores descienden (esperamos que hasta el valor del área buscada, en el caso de una función positiva) y que las sumas inferiores ascienden. Tomemos los siguientes conjuntos de números reales:

$$I(f) = \{I(f, P) : P \text{ es una partición de } [a, b]\}$$

$$S(f) = \{S(f, P) : P \text{ es una partición de } [a, b]\}.$$

Estos conjuntos pueden ser no acotados, pero si f es una función acotada entonces son ambos acotados, pues en cada intervalo

$$\inf(f) \leq m_i(f) \leq M_i(f) \leq \sup(f),$$

donde $\inf(f)$ y $\sup(f)$ denotan respectivamente al ínfimo de f en $[a, b]$ y al supremo de f en $[a, b]$. Multiplicando por Δ_i y sumando sobre i se tiene

$$\inf(f) \sum_i \Delta_i \leq \sum_i m_i(f) \Delta_i \leq \sum_i M_i(f) \Delta_i \leq \sup(f) \sum_i \Delta_i.$$

Pero

$$\sum \Delta_i = t_n - t_{n-1} + t_{n-1} - t_{n-2} + \cdots + t_2 - t_1 + t_1 - t_0 = t_n - t_0 = b - a,$$

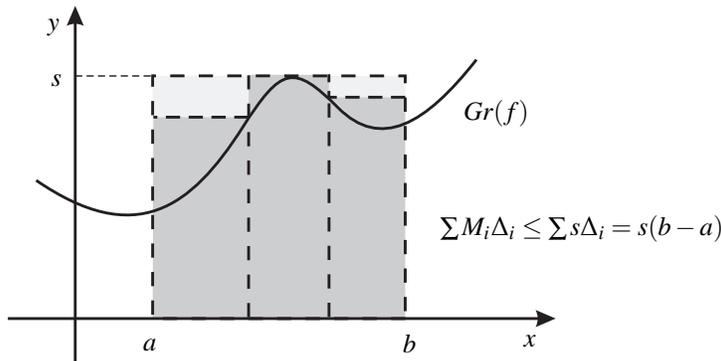
luego

$$\inf(f)(b - a) \leq I(f, P) \leq S(f, P) \leq \sup(f)(b - a).$$

De hecho, como probamos que $I(f, P) \leq S(f, Q)$ **para todo par de particiones P, Q** del intervalo, entonces

$$\inf(f)(b - a) \leq I(f, P) \leq S(f, Q) \leq \sup(f)(b - a).$$

Si denotamos $s = \sup(f)$ en $[a, b]$, en la figura que sigue se representa la desigualdad $S(f, P) \leq s(b - a)$ para el caso particular de una f positiva:



Se define la **integral superior de f** como el ínfimo de las sumas superiores (sobre cualquier partición), y la **integral inferior de f** como el supremo de las sumas inferiores (sobre cualquier partición), es decir

$$I_*(f) = \sup I(f), \text{ mientras que } I^*(f) = \inf S(f).$$

Si f es acotada, como

$$\inf(f)(b-a) \leq I(f, P) \leq S(f, Q) \leq \sup(f)(b-a)$$

para cualquier par de particiones, si tomamos ínfimo sobre particiones Q (fijando P) se deduce que

$$\inf(f)(b-a) \leq I(f, P) \leq I^*(f) \leq \sup(f)(b-a).$$

Pero como P también era cualquiera, tomando supremo sobre particiones P se deduce que

$$\inf(f)(b-a) \leq I_*(f) \leq I^*(f) \leq \sup(f)(b-a).$$

Es decir que en general, $I_*(f) \leq I^*(f)$. Diremos que f es **Riemann integrable** en $[a, b]$ si $I_*(f) = I^*(f)$, y denotaremos a este número con el símbolo integral

$$\int_a^b f = \int_a^b f(x) dx.$$

Un criterio para decidir si una función es integrable es el siguiente:

Proposición 6.1.3. Si $f : [a, b] \rightarrow \mathbb{R}$ es acotada, entonces f es integrable en $[a, b]$ si y sólo si para todo $\varepsilon > 0$ existe una partición P de $[a, b]$ tal que

$$S(f, P) - I(f, P) < \varepsilon.$$

Demostración. Supongamos primero que vale la condición. Entonces para todo $\varepsilon > 0$ existe una partición P tal que

$$I^*(f) - I_*(f) \leq S(f, P) - I(f, P) < \varepsilon,$$

con lo cual $I^*(f) \leq I_*(f)$. Como la otra desigualdad vale siempre, f es integrable.

Supongamos ahora que f es integrable, entonces dado $\varepsilon > 0$ existe por las propiedades de ínfimo y supremo una partición P tal que

$$\int_a^b f - I(f, P) < \varepsilon/2 \text{ y también } S(f, P) - \int_a^b f < \varepsilon/2.$$

Juntando estas dos desigualdades se tiene

$$S(f, P) - I(f, P) < \varepsilon$$

como queríamos. □

En particular toda función constante es integrable, y

$$\int_a^b c \, dx = c(b-a) = c \int_a^b 1 \, dx.$$

También se deduce que toda función monótona y acotada es integrable.

Proposición 6.1.4. *Sea $f : [a, b] \rightarrow \mathbb{R}$ una función monótona y acotada. Entonces f es integrable en $[a, b]$.*

Demostración. Supongamos que f es creciente. Entonces dada una partición, en cada intervalo de la partición se tiene $f(t_i) = m_i$, $f(t_{i+1}) = M_i$ con lo cual

$$\begin{aligned} S(f, P) - I(f, P) &= \sum_i (M_i - m_i)(t_{i+1} - t_i) = \sum_i [f(t_{i+1}) - f(t_i)](t_{i+1} - t_i) \\ &\leq \sum_i [f(t_{i+1}) - f(t_i)] \text{máx}(t_{i+1} - t_i) = [f(b) - f(a)] \|P\|, \end{aligned}$$

puesto que $t_0 = a$, $t_n = b$ y además

$$f(t_1) - f(t_0) + f(t_2) - f(t_1) + f(t_3) - f(t_2) + \cdots + f(t_n) - f(t_{n-1}) = f(b) - f(a).$$

Con lo cual si refinamos la partición, haciendo tender $\|P\| \rightarrow 0$, por la Proposición 6.1.3 deducimos que f es integrable. El caso f decreciente es análogo. □

Teorema 6.1.5. *Sea $f : [a, b] \rightarrow \mathbb{R}$ una función continua. Entonces f es integrable en $[a, b]$.*

La demostración la dejamos para más adelante, en el Corolario 6.1.10.

Ejemplo 6.1.6. *Un caso de función no integrable Riemann es la función de Dirichlet, definida en el intervalo $[0, 1]$ como:*

$$f(x) = \begin{cases} 1 & \text{si } x \in \mathbb{Q} \\ 0 & \text{si } x \notin \mathbb{Q} \end{cases}$$

Si tomamos una partición cualquiera P del $[0, 1]$, en cada intervalo Δ_i de la partición tendremos por lo menos un número racional y un número irracional. Por lo tanto el supremo de f en cada Δ_i es 1, y el ínfimo es 0. Con esto, las sumas inferiores para cualquier partición dan cero, mientras que las superiores dan siempre 1. En consecuencia, f no es integrable porque la integral superior $I^(f)$ da 1 y la inferior $I_*(f)$ da 0.*

6.1.1. Propiedades

Puede ser útil, dada f , considerar las siguientes funciones positivas en $[a, b]$ que se obtienen a partir de f :

$$f_+(x) = \begin{cases} f(x) & \text{si } f(x) > 0 \\ 0 & \text{en cualquier otro caso} \end{cases}$$

$$f_-(x) = \begin{cases} -f(x) & \text{si } f(x) < 0 \\ 0 & \text{en cualquier otro caso} \end{cases}$$

Entonces podemos escribir, para todo $x \in [a, b]$,

$$f(x) = f_+(x) - f_-(x).$$

Una de las propiedades más útiles que tienen f_+, f_- , es que nos permiten calcular el módulo a partir de una suma:

$$|f(x)| = f_+(x) + f_-(x) \text{ para todo } x \in [a, b].$$

Algunas propiedades sencillas de la integral:

Proposición 6.1.8. Sean $f, g : [a, b] \rightarrow \mathbb{R}$ funciones integrables. Entonces

1. Si $\alpha \in \mathbb{R}$ entonces αf es integrable, $\int_a^b \alpha f = \alpha \int_a^b f$.
2. $f + g$ es integrable, $\int_a^b f + g = \int_a^b f + \int_a^b g$.
3. Si $a \leq c \leq b$ entonces f es integrable en $[a, c]$ y en $[c, b]$ y además $\int_a^b f = \int_a^c f + \int_c^b f$.
4. Si $f \leq g$, entonces $\int_a^b f \leq \int_a^b g$.
5. Si f es integrable entonces $|f|$ es integrable y además $|\int_a^b f| \leq \int_a^b |f|$.

Demostración. Las primeras tres propiedades se deducen de escribir la definición como límite de sumas superiores.

La cuarta propiedad se deduce de lo siguiente: si $h \geq 0$ y h es integrable, sus sumas parciales son positivas y entonces $\int_a^b h \geq 0$. Luego si $f \leq g$, se tiene que $g - f \geq 0$ es una función integrable por los ítems 1 y 2, y por lo recién dicho $\int_a^b (g - f) \geq 0$. Nuevamente por los ítems 1 y 2, se deduce que

$$\int_a^b g - \int_a^b f \geq 0.$$

Problema 6.1.7. Probar que, si f es acotada, entonces f es integrable en $[a, b]$ si y sólo si f_+, f_- son ambas funciones integrables en $[a, b]$, y además

$$\int_a^b f = \int_a^b f_+ - \int_a^b f_-.$$

Por último, si f es integrable, entonces f_+ y f_- son integrables. Luego $|f| = f_+ + f_-$ es integrable, y además

$$\left| \int_a^b f \right| = \left| \int_a^b f_+ - \int_a^b f_- \right| \leq \left| \int_a^b f_+ \right| + \left| \int_a^b f_- \right| = \int_a^b f_+ + \int_a^b f_- = \int_a^b f_+ + f_- = \int_a^b |f|.$$

□

También es usual definir (para funciones integrables en $[a, b]$)

1. $\int_a^b f$ como $-\int_b^a f$ cuando $a > b$, es decir

$$\int_x^y f = -\int_y^x f \text{ para todo } x, y \in [a, b].$$

2. $\int_x^x f = 0$ para todo $x \in [a, b]$.

Recordemos que toda función continua en un conjunto compacto es uniformemente continua.

Proposición 6.1.9. Si $h : [a, b] \rightarrow [c, d]$ es integrable, y $\varphi : [c, d] \rightarrow \mathbb{R}$ es una función continua, entonces $\varphi \circ h : [a, b] \rightarrow \mathbb{R}$ es integrable.

Demostración. Como h es integrable, dado $\varepsilon_1 > 0$ existe una partición P de $[a, b]$ tal que

$$S(h, P) - I(h, P) = \sum (M_i - m_i) \Delta_i < \varepsilon_1. \quad (6.1)$$

Observemos que, como M_i es el supremo de h en Δ_i , y m_i es el ínfimo, entonces

$$M_i - m_i = \sup\{h(x) : x \in \Delta_i\} - \inf\{h(y) : y \in \Delta_i\} = \sup\{h(x) - h(y) : x, y \in \Delta_i\},$$

puesto que la diferencia entre el valor más grande y más chico de h coincide con la diferencia más grande de valores de h . Como φ es uniformemente continua, dado $\varepsilon_2 > 0$, existe $\delta > 0$ tal que $|t - s| < \delta$ en $[c, d]$ implica $|\varphi(s) - \varphi(t)| < \varepsilon_2$.

Queremos probar que la siguiente cantidad es arbitrariamente pequeña:

$$S(\varphi \circ h, P) - I(\varphi \circ h, P) = \sum (M_i^* - m_i^*) \Delta_i,$$

donde M_i^* , m_i^* son respectivamente, el supremo y el ínfimo de $\varphi \circ h$ en Δ_i . Notemos que, como antes

$$M_i^* - m_i^* = \sup\{\varphi(h(x)) - \varphi(h(y)) : x, y \in \Delta_i\}.$$

Consideremos los dos conjuntos de índices de la partición

$$A = \{i : M_i - m_i < \delta\}, \quad B = \{i : M_i - m_i \geq \delta\}$$

Si $i \in A$, entonces

$$M_i^* - m_i^* < \varepsilon_2$$

por la uniforme continuidad de φ . En consecuencia,

$$\sum_{i \in A} (M_i^* - m_i^*) \Delta_i < \varepsilon_2 \sum_{i \in A} \Delta_i \leq \varepsilon_2 (b - a).$$

Es decir, eligiendo ε_2 pequeño podemos asegurar que las sumas superiores e inferiores están tan cerca como querramos, siempre restringiéndonos al subconjunto de $[a, b]$ donde $M_i - m_i < \delta$.

Veamos que pasa en el resto del conjunto. Como φ es una función continua en un compacto, alcanza máximo, es decir existe $M > 0$ tal que $\varphi(x) \leq M$ para todo $x \in [c, d]$. En general, podemos asegurar que

$$M_i^* - m_i^* = \sup\{\varphi(h(x)) - \varphi(h(y)) : x, y \in \Delta_i\} \leq 2M.$$

Luego

$$\sum_{i \in B} (M_i^* - m_i^*) \Delta_i \leq 2M \sum_{i \in B} \Delta_i = \frac{2M}{\delta} \sum_{i \in B} \delta \Delta_i \leq \frac{2M}{\delta} \sum_{i \in B} (M_i - m_i) \Delta_i < \frac{2M}{\delta} \varepsilon_1$$

por la ecuación (6.1). Así que eligiendo ε_1 pequeño (o sea eligiendo una partición P suficientemente fina) podemos asegurar que la diferencia entre la suma superior y la inferior es tan pequeña como uno quiera, también en este caso. Luego podemos hacer que las sumas superiores estén tan próximas a las inferiores como queramos, lo que prueba que $\varphi \circ h$ es integrable. \square

Corolario 6.1.10. Sean $f, g : [a, b] \rightarrow \mathbb{R}$. Entonces

1. Si f es continua entonces f es integrable.
2. Si f es integrable entonces $f^n = f \cdot f \cdots f$ es integrable para todo $n \in \mathbb{N}$.
3. Si f, g son integrables entonces fg es integrable.

Demostración. Para ver 1. tomamos $h(x) = x$, $\varphi(x) = f(x)$ y usamos el teorema anterior, pues $\varphi \circ h(x) = f(x)$. Que h es integrable se deduce del hecho de que es una función monótona y acotada en $[a, b]$.

Para ver 2. tomamos $h = f$, $\varphi(x) = x^n$.

Para ver 3. escribimos

$$fg = \frac{1}{2}(f+g)^2 - \frac{1}{2}f^2 - \frac{1}{2}g^2$$

y usamos que $f+g$ es integrable y el ítem previo con $n = 2$. \square

Observación 6.1.11. En general es *falso* que la composición de dos funciones integrables resulta integrable. Un ejemplo está dado por las funciones siguientes, ambas definidas en el intervalo $[0, 1]$:

$$f(x) = \begin{cases} 1/q & \text{si } x = p/q \text{ es racional (y está simplificado todo lo posible)} \\ 0 & \text{si } x \notin \mathbb{Q} \end{cases}$$

$$g(x) = \begin{cases} 0 & \text{si } x = 0 \\ 1 & \text{si } x \neq 0 \end{cases}$$

Se define $f(0) = 1$ para evitar ambigüedades, aunque no es muy relevante desde el punto de vista de la integral. Entonces f es integrable (difícil, ver la Nota 1 al final de este capítulo), g es integrable (obvio), pero $g \circ f$ es la función de Dirichlet, que como vimos no es integrable.

Dada una función continua en $[a, b]$, su restricción al intervalo $[a, x]$ para cualquier $x \in [a, b]$ sigue siendo continua. Luego es integrable, y llamamos a

$$F(x) = \int_a^x f(t) dt$$

función primitiva de f . Esta primitiva es una función derivable, y su derivada coincide con f , resultado conocido TFCl:

Teorema 6.1.12 (Teorema Fundamental del Cálculo Integral). *Sea $f : [a, b] \rightarrow \mathbb{R}$ una función continua. Dado $x \in [a, b]$, sea $F : [a, b] \rightarrow \mathbb{R}$ definida como*

$$F(x) = \int_a^x f = \int_a^x f(t) dt.$$

Entonces F es continua en $[a, b]$, derivable en (a, b) y además, para todo $x \in (a, b)$ se tiene

$$F'(x) = f(x).$$

Demostración. Sea $s = \max\{|f(t)| : t \in [a, b]\}$ (que es finito porque f es continua). Veamos que F es continua en los bordes. Observemos que $F(a) = 0$.

$$|F(x) - F(a)| = \left| \int_a^x f \right| \leq \int_a^x |f| = \inf S(|f|, P)$$

donde P es una partición del intervalo $[a, x]$, es decir (si M_i es el supremo de $|f(t)|$ en el intervalo Δ_i)

$$|F(x) - F(a)| \leq S(|f|, P) = \sum_i M_i \Delta_i \leq s \sum_i \Delta_i = s(x - a).$$

Con esto es evidente que $F(x) \rightarrow F(a)$ si $x \rightarrow a^+$. Por otro lado,

$$|F(b) - F(x)| = \left| \int_a^b f - \int_a^x f \right| = \left| \int_x^b f \right| \leq \int_x^b |f| = \inf S(|f|, P) \leq s(b - x)$$

donde ahora la partición P es del intervalo $[x, b]$. Esto prueba que $F(x) \rightarrow F(b)$ cuando $x \rightarrow b^-$.

Tomemos ahora un punto x_0 interior del intervalo $[a, b]$. Si x es otro punto cualquiera en (a, b) , calculamos

$$\frac{F(x) - F(x_0)}{x - x_0} = \frac{\int_a^x f - \int_a^{x_0} f}{x - x_0} = \frac{\int_{x_0}^x f}{x - x_0}$$

Queremos ver que esta expresión tiende a $f(x_0)$ cuando $x \rightarrow x_0$. Esto probaría que F es derivable, que $F' = f$, y en particular F es continua en el interior.

Como f es continua entre x y x_0 , existen dos números reales (el máximo y el mínimo de f allí) tales que

$$m_x \leq f(t) \leq M_x$$

para todo t entre x_0 y x . Supongamos primero que $x > x_0$. Entonces integrando se tiene

$$m_x(x - x_0) \leq \int_{x_0}^x f \leq M_x(x - x_0).$$

Luego, como $x - x_0 > 0$,

$$m_x \leq \frac{1}{x - x_0} \int_{x_0}^x f \leq M_x.$$

Cuando $x \rightarrow x_0^+$, tanto m_x como M_x tienden a $f(x_0)$, por ser f continua. Esto prueba que

$$\lim_{x \rightarrow 0^+} \frac{F(x) - F(x_0)}{x - x_0} = f(x_0).$$

Si $x < x_0$, con un razonamiento análogo se obtiene que el otro límite lateral también da $f(x_0)$. Hay que tener cuidado solamente en el siguiente hecho: si $x < x_0$, entonces $x - x_0 < 0$, luego las desigualdades se invierten. \square

Entonces, dada una función continua f , existe por lo menos una función derivable tal que $F' = f$. ¿Cuántas puede haber? El siguiente lema muestra que no son muchas, en el sentido de que no son muy distintas todas las que hay:

Lema 6.1.13. Si F, G son funciones continuas en $[a, b]$, derivables en el interior, tales que $F'(x) = G'(x)$ allí, entonces existe una constante $c \in \mathbb{R}$ tal que

$$F(x) = G(x) + c.$$

Demostración. Si $H(x) = F(x) - G(x)$, entonces

$$H'(x) = (F - G)'(x) = F'(x) - G'(x) = 0$$

en el interior, entonces por el teorema de Lagrange H es constante allí pues

$$H(x) - H(y) = H'(c)(x - y) = 0.$$

\square

Es decir, dos primitivas de f difieren a lo sumo en una constante, o lo que es lo mismo, si uno conoce una primitiva F , entonces todas las primitivas de f están dadas por la familia

$$\int f = \{F + c : c \in \mathbb{R}\}.$$

Para calcular una integral definida basta encontrar entonces alguna primitiva de f y se tiene

$$\int_a^b f = F(b) - F(a).$$

En efecto, esto es obvio si F es la primitiva hallada en el TFC, es decir, si $F(x) = \int_a^x f$. Pero si G es otra primitiva, entonces $G = F + c$ y en consecuencia

$$G(b) - G(a) = F(b) + c - (F(a) + c) = F(b) - F(a) = \int_a^b f.$$

Este resultado se conoce como *Regla de Barrow*.

Del teorema de arriba también se deduce el **teorema del valor medio para integrales** (TVMI),

Proposición 6.1.14. Si $f : [a, b] \rightarrow \mathbb{R}$ es continua, entonces existe $c \in (a, b)$ tal que

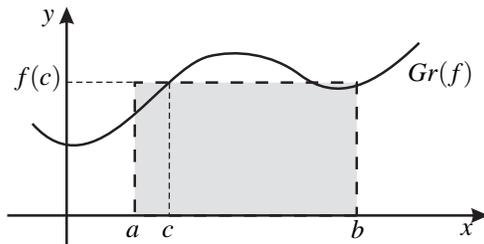
$$\int_a^b f = f(c)(b - a).$$

Demostración. Basta aplicar el teorema de Lagrange a la función $F(x) = \int_a^x f$:

$$\int_a^b f = F(b) - F(a) = F'(c)(b - a) = f(c)(b - a)$$

para algún c entre a y b . □

Este teorema, cuando f es positiva, tiene contenido geométrico, pues nos dice que el área bajo el gráfico de f (entre a y b) está dada por el área de algún rectángulo de base $b - a$, cuya altura $f(c)$ es alguna altura intermedia entre el mínimo y el máximo de f en $[a, b]$.



La idea es que si tomamos un rectángulo de base $[a, b]$, y hacemos variar la altura entre el máximo de f y el mínimo de f , en algún momento el área obtenida con el rectángulo coincide con el área bajo el gráfico de f en $[a, b]$.

6.2. Integrales impropias

Dada una función continua en un intervalo $[a, b)$, podemos calcular su integral en cualquier intervalo más pequeño. Surge la pregunta de si existirá el límite de estas integrales cuando el intervalo tiende al intervalo total. Es decir, dado $a \leq x < b$, pongamos

$$F(x) = \int_a^x f(t) dt.$$

Entonces podemos definir a la integral impropia como un límite, es decir

$$\int_a^b f = \lim_{x \rightarrow b^+} F(x),$$

en el supuesto de que este límite exista. En este caso diremos que $\int_a^b f$ **converge**.

Las mismas consideraciones se aplican para el otro extremo.

Ejemplo 6.2.1. Consideremos $f(t) = t^{-\frac{1}{2}}$, que es una función con una discontinuidad en $t = 0$. Entonces

$$\int_0^1 f = \lim_{x \rightarrow 0^+} \int_x^1 \frac{1}{\sqrt{t}} dt = \lim_{x \rightarrow 0^+} 2\sqrt{t} \Big|_x^1 = \lim_{x \rightarrow 0^+} 2 - 2\sqrt{x} = 2.$$

Esto es, el área encerrada por el gráfico de f , los ejes, y la recta $x = 1$ es **finita**.

¿Qué pasa si la discontinuidad está dentro del intervalo y no en el borde? Por ejemplo, tomemos $f(t) = t^{-\frac{2}{3}}$ en el intervalo $[-1, 1] - \{0\}$. Tenemos, por un lado

$$\int_{-1}^x f(t) dt = 3t^{\frac{1}{3}} \Big|_{-1}^x = 3(x^{\frac{1}{3}} + 1) \text{ si } x < 0,$$

y por otro lado

$$\int_y^1 f(t) dt = 3t^{\frac{1}{3}} \Big|_y^1 = 3(1 - y^{\frac{1}{3}}) \text{ si } y > 0.$$

En el caso particular en el que tomamos $x = -y$, se tiene

$$\lim_{y \rightarrow 0^+} \int_{-1}^{-y} f(t) dt + \int_y^1 f(t) dt = \lim_{y \rightarrow 0^+} 3(-y^{\frac{1}{3}} + 1) + 3(1 - y^{\frac{1}{3}}) = 3 + 3 = 6.$$

Esta manera de calcular una integral impropia (tomando un intervalo simétrico alrededor de la discontinuidad) se conoce como **valor principal** de la integral, es decir

$$\text{v.p.} \int_{-1}^1 t^{-\frac{2}{3}} dt = 6;$$

en general, dada $f : [a, b] - \{x_0\} \rightarrow \mathbb{R}$ definiremos

$$\text{v.p.} \int_a^b f(t) dt = \lim_{\varepsilon \rightarrow 0^+} \left\{ \int_a^{x_0 - \varepsilon} f(t) dt + \int_{x_0 + \varepsilon}^b f(t) dt \right\}.$$

Hay que tener presente que en muchos casos modificando el intervalo, se puede modificar este límite.

Ejemplo 6.2.2. Tomemos $f(t) = \frac{1}{t}$ en el intervalo $[-1, 1]$. Se tiene

$$\text{v.p.} \int_{-1}^1 \frac{1}{t} = \lim_{x \rightarrow 0^+} \int_{-1}^{-x} \frac{1}{t} dt + \int_x^1 \frac{1}{t} dt = \lim_{x \rightarrow 0^+} \ln|-x| - \ln|x| = 0,$$

mientras que si no miramos el valor principal, podemos obtener otro resultado:

$$\lim_{x \rightarrow 0^+} \int_{-1}^{-2x} \frac{1}{t} dt + \int_x^1 \frac{1}{t} dt = \lim_{x \rightarrow 0^+} \ln|-2x| - \ln|x| = \ln 2.$$

Estos ejemplos nos muestran que tenemos que tener claro lo que queremos calcular cuando integramos una función no acotada; en casos concretos puede desprenderse del problema que estamos estudiando cuál es la interpretación correcta que tenemos que hacer.

6.2.1. Intervalos infinitos

El otro caso que nos interesa es el de un intervalo infinito, y se piensa de la misma manera, es decir, como un límite.

Ejemplo 6.2.3. Si $f(t) = \frac{1}{1+t^2}$, entonces

$$\int_0^{+\infty} f(t) dt = \lim_{x \rightarrow +\infty} \int_0^x \frac{1}{1+t^2} dt = \lim_{x \rightarrow +\infty} \operatorname{arctg}(x) - \operatorname{arctg}(0) = \frac{\pi}{2}.$$

Nuevamente surgen las mismas consideraciones si hay dos infinitos; si tomamos un intervalo simétrico alrededor del cero diremos que es el valor principal.

Ejemplo 6.2.4. Consideremos $f(t) = \frac{2t}{1+t^2}$. Entonces v.p. $\int_{-\infty}^{+\infty} f(t) dt$ es el límite

$$\lim_{x \rightarrow +\infty} \int_{-x}^x f(t) dt = \lim_{x \rightarrow +\infty} \ln(1+t^2)|_{-x}^x = \lim_{x \rightarrow +\infty} \ln(1+x^2) - \ln(1+x^2) = 0$$

Por otro lado, si no tomamos un intervalo simétrico, el límite puede dar distinto:

$$\lim_{x \rightarrow +\infty} \int_{-x}^{2x} f(t) dt = \lim_{x \rightarrow +\infty} \ln(1+t^2)|_{-x}^{2x} = \lim_{x \rightarrow +\infty} \ln(1+4x^2) - \ln(1+x^2) = \ln 4.$$

Ejemplo 6.2.5. Un caso relevante es el de la integral $\int_1^{+\infty} \frac{1}{t^p} dt$, para $p > 0$. Como

$$\int_1^x t^{-p} dt = \frac{t^{-p+1}}{-p+1} \Big|_1^x = \frac{x^{1-p}}{1-p} - \frac{1}{1-p}$$

para cualquier $p > 0$, $p \neq 1$, la expresión de la derecha tiene límite cuando $x \rightarrow +\infty$ si y sólo si $1-p < 0$. ¿Qué pasa si $p = 1$? En este caso la primitiva es el logaritmo y por lo tanto la integral diverge. En resumen,

$$\int_1^{\infty} \frac{1}{t^p} dt \quad \begin{cases} \text{converge si } p > 1 \\ \text{diverge si } 0 < p \leq 1 \end{cases}.$$

Como se pueden imaginar, si uno reemplaza el intervalo $[1, +\infty)$ por el intervalo $(0, 1]$, el comportamiento de $f(t) = 1/t^p$ no es similar (es bueno en este punto, graficar $1/\sqrt{t}$, $1/t$ y $1/t^2$ en $[0, +\infty)$ en el mismo sistema de ejes para tener alguna intuición de lo que está ocurriendo). Dejamos como ejercicio ver que vale un resultado análogo al enunciado arriba, con similar demostración, en el intervalo $(0, 1]$.

Problema 6.2.6. Probar que

$$\int_0^1 \frac{1}{t^p} dt \quad \begin{cases} \text{converge si } 0 < p < 1 \\ \text{diverge si } p \geq 1 \end{cases}.$$

6.2.2. Convergencia condicional y absoluta

Diremos que $\int_a^b f$ **converge absolutamente** si $\int_a^b |f|$ converge. Si no es así, pero la integral de f converge, diremos que la integral de f **converge condicionalmente**.

El siguiente lema generaliza una idea que ya desarrollamos para sucesiones. Nos va a resultar útil para las integrales de funciones positivas.

Lema 6.2.7. 1. Sea $G : [a, b) \rightarrow \mathbb{R}$ una función creciente (b puede ser $+\infty$). Entonces existe el límite $\lim_{x \rightarrow b^+} G(x)$. Este límite es finito si y sólo si G es acotada superiormente.

2. Las mismas consideraciones valen para $G : (a, b] \rightarrow \mathbb{R}$ y el límite $\lim_{x \rightarrow a^-} G(x)$.

Demostración. Veamos 1. Supongamos que G es creciente, sea $l = \sup_{x \in [a, b)} \{G(x)\}$. Afirmamos que $l = \lim_{x \rightarrow b^+} G(x)$. En efecto, por definición de supremo, dado $\varepsilon > 0$ existe $x_0 \in [a, b)$ tal que $l - G(x_0) < \varepsilon$. Si $|b - x| = b - x < \delta = b - x_0$, entonces como $x \geq x_0$ y G es creciente, $G(x) \geq G(x_0)$. Entonces

$$l - G(x) \leq l - G(x_0) < \varepsilon,$$

lo que prueba que el límite existe y coincide con el supremo. En particular el límite es finito si y sólo si G es acotada superiormente.

La demostración de 2. es análoga y queda como ejercicio. □

Observación 6.2.8. Un hecho importante es que, en el caso de funciones integrables y positivas $f \geq 0$, el Lema 6.2.7 nos dice que el problema de convergencia se reduce a probar que la integral $\int_a^x f(t)dt$ está acotada, ya que es una función creciente de x .

Teorema 6.2.9. Si f es continua en $[a, b)$, y $\int_a^b f$ converge absolutamente, entonces converge. Es decir, si existe $\lim_{x \rightarrow b^+} \int_a^x |f(t)|dt$ y es finito entonces existe $\lim_{x \rightarrow b^+} \int_a^x f(t)dt$ y es finito.

Demostración. Si $\lim_{x \rightarrow b^+} \int_a^x |f(t)|dt < +\infty$, entonces cada una de las funciones positivas f_+, f_- deben verificar la misma condición. En efecto, en primer lugar como f es continua, es integrable, y entonces ambas funciones f_+, f_- son integrables en $[a, x]$. Pongamos

$$F_+(x) = \int_a^x f_+(t)dt, \quad F_-(x) = \int_a^x f_-(t)dt, \quad F(x) = F_+(x) + F_-(x).$$

Como $|f(t)| = f_+(t) + f_-(t)$, se tiene $F(x) = \int_a^x |f(t)|dt$. Las tres funciones F_+, F_-, F son crecientes y positivas. Además como $\lim_{x \rightarrow b^+} F(x)$ es finito, F está acotada superiormente por este límite. Entonces

$$F_+(x) \leq F_+(x) + F_-(x) = F(x) \leq \int_a^b |f(t)|dt = \lim_{x \rightarrow b^+} F(x).$$

Esto prueba que F_+ (y con un argumento idéntico F_-) es una función acotada superiormente. Por el lema previo, existen los límites

$$l_1 = \lim_{x \rightarrow b^+} F_+(x) = \lim_{x \rightarrow b^+} \int_a^x f_+(t) dt,$$

$$l_2 = \lim_{x \rightarrow b^+} F_-(x) = \lim_{x \rightarrow b^+} \int_a^x f_-(t) dt.$$

En consecuencia, como $f = f_+ - f_-$,

$$\lim_{x \rightarrow b^+} \int_a^x f(t) dt = \lim_{x \rightarrow b^+} \int_a^x f_+(t) dt - \lim_{x \rightarrow b^+} \int_a^x f_-(t) dt = l_1 - l_2.$$

□

6.2.3. Criterios de comparación

Dadas dos funciones positivas, $f, g : [a, b) \rightarrow \mathbb{R}$ (b puede ser $+\infty$) si las suponemos integrables en $[a, b)$, entonces la convergencia de g implica la convergencia de f siempre que $f \leq g$. Más precisamente:

Proposición 6.2.10. Sean $f, g : [a, b) \rightarrow \mathbb{R}$ funciones integrables (b puede ser $+\infty$) y no negativas.

1. Si $f(t) \leq g(t)$ para todo $t \in [a, b)$, entonces si $\int_a^b g$ converge, también converge $\int_a^b f$.
2. Si $f(t) \leq g(t)$ para todo $t \in [a, b)$, entonces si $\int_a^b f$ diverge, también diverge $\int_a^b g$.

Demostración. Tomemos $F(x) = \int_a^x f(t) dt$, y $G(x) = \int_a^x g(t) dt$. Ambas son crecientes en $[a, b)$, por ser $f, g \geq 0$ allí. Además

$$F(x) \leq G(x)$$

para todo $x \in [a, b)$ por ser $f \leq g$. Como $\int_a^b g$ converge, se tiene además que $G(x) \leq \int_a^b g$. Luego F es creciente y acotada, con lo cual tiene límite finito cuando $x \rightarrow b^+$ por el Lema 6.2.7.

El segundo ítem es consecuencia del primero. □

Corolario 6.2.11. Sean $f, g : [a, b) \rightarrow \mathbb{R}$ funciones integrables (b puede ser $+\infty$).

Si $|f(t)| \leq |g(t)|$ para todo $t \in [a, b)$, y la integral de g converge absolutamente, entonces también converge absolutamente la integral de f , y si la integral de f no converge absolutamente entonces tampoco converge absolutamente la integral de g .

Teorema 6.2.12. Sean f, g son positivas en $[a, b)$, donde b puede ser $+\infty$. Si $\lim_{x \rightarrow b^-} \frac{f(x)}{g(x)} = L$ existe, entonces hay sólo tres posibilidades, que se detallan a continuación:

1. L es finito y no nulo. Entonces $\int_a^b f$ converge si y sólo si $\int_a^b g$ converge.
2. $L = 0$. Entonces $\int_a^b g < +\infty \Rightarrow \int_a^b f < +\infty$.

3. $L = +\infty$. Entonces $\int_a^b g$ diverge $\Rightarrow \int_a^b f$ diverge.

Demostración. Dado $\varepsilon > 0$, si L es finito, se tiene

$$(L - \varepsilon)g(t) < f(t) < (L + \varepsilon)g(t)$$

para todo t tal que $x_0 < t < b$. Observemos que $L \geq 0$ siempre que exista, por ser f, g positivas.

1. Si L no es cero entonces podemos elegir ε positivo de manera que $L - \varepsilon$ siga siendo positivo. Entonces por el criterio de comparación, se tiene que una es convergente si y sólo si la otra también.

2. Si $L = 0$, el lado izquierdo es negativo por lo tanto sólo podemos asegurar que la convergencia de g garantiza la de f , y no la recíproca.

3. Si $L = +\infty$, entonces dado M positivo existe x_0 tal que

$$f(t) \geq Mg(t)$$

para todo $x_0 < t < b$. Si la integral de f converge se deduce que la de g también, sin poder decir nada sobre la recíproca. \square

Observación 6.2.13. *Los mismos resultados valen si tomamos intervalos $(a, b]$ donde ahora el extremo izquierdo es abierto.*

Criterio integral de Cauchy

En general, dada una integral impropia definida en un intervalo, es posible decidir la convergencia de la misma comparándola con una serie adecuada. Recordemos primero algunos hechos elementales de series. Sea a_k una sucesión de números reales y

$$S_n = \sum_{k=0}^n a_k$$

la sucesión de sumas parciales. Entonces

1. $a_k \rightarrow 0$ **no** implica que la serie converge.
2. Sí es cierto que si la serie converge entonces $a_k \rightarrow 0$.
3. Una serie de términos positivos $a_k \geq 0$ es convergente si y sólo si sus sumas parciales están acotadas, pues en este caso S_n es una sucesión creciente de números reales.

Teorema 6.2.14 (Criterio de comparación de Cauchy). *Sea $f : [N, +\infty) \rightarrow \mathbb{R}$ una función decreciente y no negativa. Entonces la integral $\int_N^{+\infty} f$ converge si y sólo si la sucesión de sumas parciales*

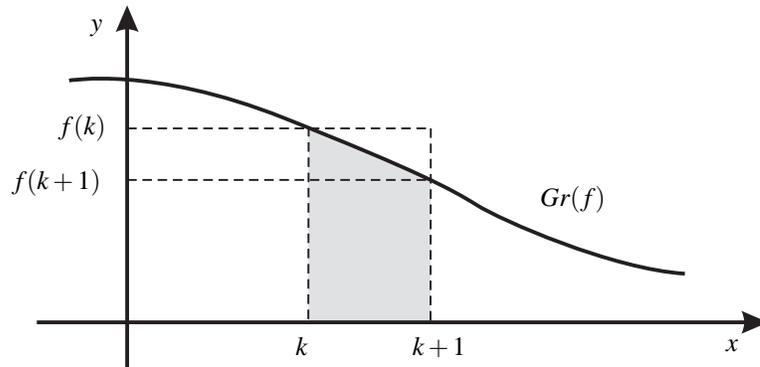
$$S_n = \sum_{k=N}^n f(k)$$

es convergente. Es decir, si llamamos $a_k = f(k)$, la integral converge si y sólo si la serie $\sum a_k$ converge.

Demostración. Observemos que por ser f decreciente y positiva, para todo $k \geq N$ se tiene

$$f(k+1) \leq \int_k^{k+1} f(t) dt \leq f(k),$$

como se puede apreciar en la figura, pues la base del rectángulo mide exactamente 1:



Sumando desde $k = N$ hasta $k = n$ se tiene

$$\sum_{k=N}^n a_k - a_N = \sum_{k=N+1}^n a_k \leq \int_N^{n+1} f(t) dt \leq \sum_{k=N}^n a_k.$$

Como $a_k \geq 0$, la serie es convergente si y sólo si las sumas parciales están acotadas, y esto ocurre si y sólo si la integral de f está acotada, y como $f \geq 0$ esto ocurre si y sólo si la integral es convergente. \square

Corolario 6.2.15. *Combinando el teorema previo con el Ejemplo 6.2.5, se deduce que para $p > 0$, la serie*

$$\sum_{k \geq 1} \frac{1}{k^p}$$

converge si y sólo si $p > 1$.

Ejemplo 6.2.16. *Observemos que,*

$$\int_1^x \left| \frac{\text{sen}(t^2)}{t^2} \right| dt \leq \int_1^x \frac{1}{t^2} = -\frac{1}{t} \Big|_1^x = 1 - \frac{1}{x} \rightarrow 1$$

cuando $x \rightarrow +\infty$. Luego la integral

$$\int_1^{+\infty} \frac{\text{sen}(t^2)}{t^2} dt$$

converge absolutamente, y en particular converge.

Por otro lado, integrando por partes,

$$\int_1^x \cos(t^2) dt = \frac{\operatorname{sen}(t^2)}{2t} \Big|_1^x + \int_1^x \frac{\operatorname{sen}(t^2)}{2t^2} dt = \frac{\operatorname{sen}(x^2)}{2x} - \frac{\operatorname{sen}(1)}{2} + \frac{1}{2} \int_1^x \frac{\operatorname{sen}(t^2)}{t^2} dt.$$

Como $\lim_{x \rightarrow +\infty} \frac{\operatorname{sen}(x^2)}{2x} = 0$, se deduce que $\int_1^{+\infty} \cos(t^2) dt$ es convergente.

Sin embargo, como $|\cos(x)|$ es periódica, continua, y se mueve entre 0 y 1, existe $\delta > 0$ tal que para todo $k \in \mathbb{N}$,

$$|\cos(x)| \geq \frac{1}{2}$$

siempre que $x \in [k\pi - \delta, k\pi + \delta]$. Luego

$$|\cos(t^2)| \geq \frac{1}{2}$$

siempre que $\sqrt{k\pi - \delta} \leq t \leq \sqrt{k\pi + \delta}$, para cualquier $k \in \mathbb{N}$. Entonces como $|\cos(t^2)|$ es positiva,

$$\begin{aligned} \int_1^{n\pi + \frac{\pi}{2}} |\cos(t^2)| dt &\geq \sum_{k=1}^n \int_{\sqrt{k\pi - \delta}}^{\sqrt{k\pi + \delta}} |\cos(t^2)| dt \geq \sum_{k=1}^n \frac{1}{2} (\sqrt{k\pi + \delta} - \sqrt{k\pi - \delta}) \\ &= \delta \sum_{k=1}^n \frac{1}{\sqrt{k\pi + \delta} + \sqrt{k\pi - \delta}}. \end{aligned}$$

Como $\sqrt{k\pi + \delta} + \sqrt{k\pi - \delta} \leq 2\sqrt{k\pi + \delta} \leq 2\sqrt{\delta + \pi} \sqrt{k}$ para todo $k \in \mathbb{N}$, se deduce que

$$\int_1^{n\pi + \frac{\pi}{2}} |\cos(t^2)| dt \geq C \sum_{k=1}^n \frac{1}{\sqrt{k}}.$$

Esta última serie es divergente por el criterio de la integral, luego la integral de $\cos(t^2)$ **no converge absolutamente**. Como vimos antes, converge sin el módulo, luego la convergencia es condicional.

Criterios de convergencia de series

En general, si podemos comparar dos sucesiones de términos positivos podemos razonar como en el caso de funciones. Se tiene el siguiente resultado, de demostración sencilla que queda como ejercicio:

Proposición 6.2.17. Sean $\{a_k\}, \{b_k\}$ dos sucesiones de términos positivos. Si $a_k \leq b_k$ a partir de un k_0 , entonces

1. Si $\sum b_k < \infty$ entonces $\sum a_k < \infty$.
2. Si $\sum a_k$ diverge, entonces $\sum b_k$ diverge.

Ejemplo 6.2.18. Un ejemplo importante y simple de serie, contra la que es fácil comparar, es la llamada **serie geométrica**: si $c > 0$, entonces la serie

$$\sum_{k \geq 0} c^k = 1 + c + c^2 + c^3 + \dots$$

es convergente si $c < 1$ y divergente si $c \geq 1$. En efecto, si $S_n = \sum_{k=0}^n c^k$, se tiene

$$S_n = 1 + c + c^2 + c^3 + \dots + c^n.$$

Si $c = 1$, esta serie es evidentemente divergente pues $S_n = n$. Supongamos que $c \neq 1$, entonces multiplicando por c ,

$$cS_n = c + c^2 + c^3 + \dots + c^n + c^{n+1} = S_n + c^{n+1} - 1,$$

de donde se deduce despejando que

$$S_n = 1 + c + c^2 + c^3 + \dots + c^n = \frac{1 - c^{n+1}}{1 - c}.$$

En esta expresión es evidente la afirmación sobre la convergencia.

Diremos que una serie $\sum a_n$ converge absolutamente si $\sum |a_n|$ es convergente. Dejamos para más adelante la demostración del hecho siguiente:

$$\sum |a_n| < \infty \Rightarrow \sum a_n < \infty.$$

Repasemos un par de criterios útiles y conocidos por todos:

Proposición 6.2.19. Sea $\{a_n\}$ una sucesión de números reales.

1. Criterio de D'Alembert (cociente). Si

$$L = \lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right|$$

existe, entonces $L > 1 \Rightarrow \sum a_k$ diverge, mientras que $L < 1 \Rightarrow \sum a_k$ converge absolutamente.

2. Criterio de Cauchy (raíz enésima). Si

$$C = \lim_{n \rightarrow \infty} \sqrt[n]{|a_n|}$$

existe, entonces $C > 1 \Rightarrow \sum a_k$ diverge, mientras que $C < 1 \Rightarrow \sum a_k$ converge absolutamente.

Demostración. 1. Supongamos que $L < 1$. Elegimos $\varepsilon > 0$ chico tal que $c = L + \varepsilon < 1$. Entonces tomamos $n_0 \in \mathbb{N}$ tal que si $n \geq n_0$,

$$-\varepsilon < \left| \frac{a_{n+1}}{a_n} \right| - L < \varepsilon. \quad (6.2)$$

Despejando del lado derecho se deduce que

$$|a_{n+1}| < (L + \varepsilon)|a_n| = c|a_n|.$$

Como esto vale para cualquier $n \geq n_0$, se tiene para cualquier $p \in \mathbb{N}$

$$|a_{n_0+p}| < c|a_{n_0+p-1}| < c^2|a_{n_0+p-2}| < \dots < c^p|a_{n_0}| = c^{n_0+p} \frac{|a_{n_0}|}{c^{n_0}} = Kc^{n_0+p}.$$

Esto prueba que $|a_n| < Kc^n$ para todo $n \geq n_0$, y por el criterio de comparación con la serie geométrica, como $c < 1$, se deduce que $\sum a_n$ converge absolutamente. La demostración de que diverge si $L > 1$ es análoga:

ahora hay que elegir $\varepsilon > 0$ tal que $L - \varepsilon > 1$, y despejar del lado izquierdo de la desigualdad (6.2). Iterando nuevamente, se compara (ahora por debajo) contra la serie geométrica en $c = L - \varepsilon$, que ahora diverge.

2. La idea de la demostración es similar. Ahora hay que observar que podemos conseguir n_0 tal que $n \geq n_0$ implica

$$C - \varepsilon < \sqrt[n]{|a_n|} < C + \varepsilon,$$

y con esto

$$(C - \varepsilon)^n < |a_n| < (C + \varepsilon)^n.$$

Si ε es suficientemente pequeño, se consigue $C - \varepsilon > 0$. Queda como ejercicio terminar de escribir la demostración en este caso, comparando la serie $|a_n|$ con las series geométricas de $C - \varepsilon$ y $C + \varepsilon$ respectivamente. \square

Observación 6.2.20. En ambos casos, si el límite da 1 no podemos usar el criterio. Para convencernos, basta considerar $a_n = 1/n$, que nos da una serie divergente, y donde $\sqrt[n]{1/n} \rightarrow 1 = L$, mientras que si consideramos $a_n = 1/n^2$, nos da una serie convergente, y también $\sqrt[n]{n^2} \rightarrow 1 = L$.

Necesitamos, antes de seguir, un resultado fundamental sobre sucesiones (que enunciaremos en particular para series).

Teorema 6.2.21. Sea $\{a_k\}$ una sucesión de números reales, $S_n = \sum_{k=0}^n a_k$. Entonces existe el límite $S = \lim_n S_n = \sum_{k=0}^{\infty} a_k$ si y sólo si la cola de la serie tiende a cero. Es decir, si y sólo si dado $\varepsilon > 0$, existe $n_0 \in \mathbb{N}$ tal que, para cualquier $p \in \mathbb{N}$,

$$|S_{n_0+p} - S_{n_0}| = \left| \sum_{k=n_0+1}^{n_0+p} a_k \right| < \varepsilon.$$

Demostración. Supongamos primero que la serie es convergente, es decir $S_n \rightarrow S$. Entonces, dado $\varepsilon > 0$, existe n_0 tal que $|S_n - S| < \varepsilon/2$ para todo $n \geq n_0$, con lo cual

$$|S_{n_0+p} - S_{n_0}| \leq |S_{n_0+p} - S| + |S - S_{n_0}| < \varepsilon/2 + \varepsilon/2 = \varepsilon,$$

y esto vale para cualquier $p \in \mathbb{N}$.

Supongamos ahora que se verifica la condición del teorema, es decir, la cola tiende a cero. En particular, se tiene, dado $\varepsilon = 1$, que si $n \geq n_0$ entonces

$$|S_n| \leq |S_n - S_{n_0}| + |S_{n_0}| < 1 + |S_{n_0}|$$

para todo $n \geq n_0$. Es decir, todos los términos de la sucesión S_n , de n_0 en adelante, están acotados por $1 + |S_{n_0}|$. Entonces toda la sucesión $\{S_n\}$ es acotada pues los primeros términos, que son finitos, también están acotados. Se deduce que existe una subsucesión convergente, $S_{n_k} \rightarrow L$ cuando $k \rightarrow \infty$. Afirmando que este número L es el límite de la serie. Para verlo, dado $\varepsilon > 0$, existe por la condición que es hipótesis un $n_0 \in \mathbb{N}$ tal

que $|S_n - S_{n_0}| < \varepsilon/2$ para todo $n > n_0$. Como S_{n_k} tiende a L , también existe un $k_0 \in \mathbb{N}$ tal que $|S_{n_k} - L| < \varepsilon/2$ si $k \geq k_0$. Podemos suponer, agrandando k_0 , que $n_{k_0} \geq n_0$. Entonces, si $n > n_0$

$$|S_n - L| \leq |S_n - S_{n_{k_0}}| + |S_{n_{k_0}} - L| < \varepsilon/2 + \varepsilon/2 = \varepsilon$$

lo que prueba que $S_n \rightarrow L$ cuando $n \rightarrow \infty$. \square

Observación 6.2.22. Se deduce fácilmente ahora que la serie $\sum a_k$ converge cuando $\sum |a_k|$ es convergente, pues

$$\left| \sum_{k=n+1}^{n+p} a_k \right| \leq \sum_{k=n+1}^{n+p} |a_k|.$$

Esta última cantidad tiende a cero cuando $n \rightarrow \infty$ si la serie de los módulos converge, por el teorema. Se deduce que la cola de la serie tiende a cero, y nuevamente invocando el teorema, se deduce que la serie $\sum a_k$ es convergente. Este resultado se suele mencionar diciendo que **la convergencia absoluta implica la convergencia de una serie**, al igual que con las integrales.

Antes de pasar a una aplicación veamos un criterio para series alternadas.

Proposición 6.2.23. *Criterio de Leibniz (series alternadas).* Si $a_n \geq 0, a_n \rightarrow 0$ y $a_{n+1} < a_n$ para todo n , entonces $S = \sum (-1)^k a_k < +\infty$. Además $|S - S_n| < a_{n+1}$, donde S_n es la suma parcial $\sum_{k=0}^n (-1)^k a_k$.

Demostración. Consideremos la cola, y supongamos que n, p son pares para simplificar. Todos los otros casos tienen idéntica demostración. Observemos que, como la sucesión a_k es estrictamente decreciente, cada uno de los términos entre paréntesis es estrictamente positivo:

$$\begin{aligned} |S_{n+p} - S_n| &= \left| \sum_{k=n+1}^{n+p} (-1)^k a_k \right| = |a_{n+p} - a_{n+p-1} + \cdots - a_{n+3} + a_{n+2} - a_{n+1}| \\ &= |(a_{n+1} - a_{n+2}) + (a_{n+3} - a_{n+4}) + \cdots + (a_{n+p-1} - a_{n+p})| \\ &= a_{n+1} - a_{n+2} + a_{n+3} - a_{n+4} + \cdots + a_{n+p-1} - a_{n+p} \\ &= a_{n+1} - (a_{n+2} - a_{n+3}) - \cdots - (a_{n+p} - a_{n+p-1}). \end{aligned}$$

El último término entonces siempre es menor que a_{n+1} . Esto prueba que $|S_{n+p} - S_n| < a_{n+1}$, y como $a_k \rightarrow 0$, se deduce que la cola tiende a cero. Por el Teorema 6.2.21, la serie es convergente. Haciendo tender p a infinito se deduce que $|S - S_n| < a_{n+1}$. \square

Ejemplo 6.2.24. Sabemos que la serie $\sum_{n \geq 1} \frac{1}{n}$ es divergente por el criterio integral. Sin embargo, por el criterio de Leibniz la serie

$$\sum_{n \geq 1} (-1)^n \frac{1}{n} = -1 + \frac{1}{2} - \frac{1}{3} + \frac{1}{4} - \cdots$$

es convergente y su suma S difiere de $S_2 = -\frac{1}{2}$ en menos que $a_3 = \frac{1}{3}$. Este ejemplo también nos dice que la recíproca de lo enunciado en la Observación 6.2.22 es falsa, pues la serie converge pero no converge absolutamente.

6.3. NOTAS

I. La función f definida en $[0, 1]$, con $f(0) = 1$ y dada por

$$f(x) = \begin{cases} \frac{1}{q} & \text{si } x = \frac{p}{q} \in \mathbb{Q} \text{ y } p, q \text{ no tienen divisores comunes} \\ 0 & \text{si } x \notin \mathbb{Q}. \end{cases}$$

es integrable en $[0, 1]$ y además $\int_0^1 f = 0$. Para probarlo, contemos un poco:

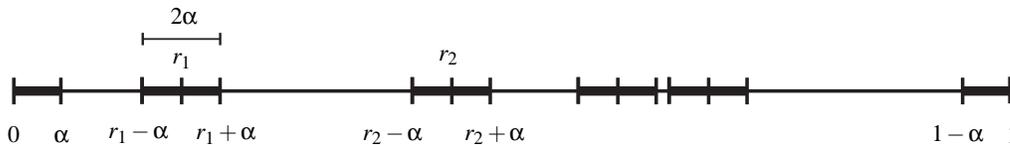
- ¿Cuántos números racionales p/q hay en $(0, 1)$ con la propiedad de tener un número 2 en el denominador? Únicamente uno, el número $\frac{1}{2}$, pues si ponemos en el numerador $\frac{p}{2}$ un número mayor o igual a 2, obtenemos el número $\frac{2}{2}$, que tiene factores comunes, o números mayores que 1.
- ¿Cuántos números racionales $\frac{p}{3}$ hay en $(0, 1]$? Hay dos, que son $\frac{1}{3}, \frac{2}{3}$, por los mismos motivos.
- ¿Cuántos números racionales $\frac{p}{4}$ hay en $(0, 1]$? Hay dos, que son $\frac{1}{4}, \frac{3}{4}$, por los mismos motivos. Observemos que el $\frac{2}{4}$ no hay que contarlo pues no verifica que p y q no tienen factores comunes.
- En general, habrá a lo sumo $j - 1$ racionales $\frac{p}{j}$ en el intervalo $(0, 1]$.

Luego, los que tienen denominador menor que K , que se obtienen juntando todos los de la lista de arriba, son finitos.

Dado $\varepsilon > 0$, usemos el criterio de integrabilidad: alcanza con encontrar una partición del $[0, 1]$ donde $S(f, P) - I(f, P) < \varepsilon$. Como en cualquier intervalo de cualquier partición, habrá siempre un número irracional, está claro que las sumas inferiores dan todas cero. Basta probar entonces que, dado $\varepsilon > 0$, hay una partición tal que la suma $S(f, P) < \varepsilon$. Tomemos $K \in \mathbb{N}$ tal que $\frac{1}{K} < \frac{\varepsilon}{2}$. Por lo que observamos recién, hay finitos números racionales no nulos $r = \frac{p}{q}$ en el intervalo $[0, 1]$ tales que $q < K$, contando $r = 1$. Digamos que son $N = N(K)$, los ordenamos de menor a mayor y los nombramos $\{r_i\}_{i=1 \dots N}$, con $r_N = 1$. Ahora tomamos α suficientemente pequeño de manera que la siguiente sucesión resulte bien ordenada

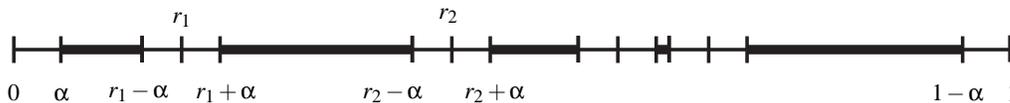
$$0, r_1 - \alpha, r_1, r_1 + \alpha, r_2 - \alpha, r_2, r_2 + \alpha, r_3 - \alpha, \dots, r_N - \alpha, r_N.$$

Podemos definir una partición P del intervalo $[0, 1]$ de la siguiente manera: separamos en intervalos que tienen racionales con $q < K$ y en intervalos que no tienen racionales con $q < K$, como indica la figura:



Es decir, tomamos primero los intervalos Δ'_k remarcados en la figura, que son aquellos de ancho 2α , centrados en los puntos r_i (como por ejemplo $[r_1 - \alpha, r_1 + \alpha]$). De estos hay $N - 1$. Incluimos también los intervalos $[0, \alpha]$ y $[1 - \alpha, 1]$.

Si a los intervalos restantes los denominamos Δ''_k , se observa que son aquellos intervalos que quedan entre medio (como por ejemplo $[r_1 + \alpha, r_2 - \alpha]$), donde no hay ningún número racional con $q < K$:



Esto define una partición $P = \{\Delta_k\}$ del intervalo $[0, 1]$. Calculemos $S(f, P) = \sum M_k \Delta_k$, pero separando en dos términos:

$$S(f, P) = \sum M'_k \Delta'_k + \sum M''_k \Delta''_k,$$

donde M'_k es el supremo de f en Δ'_k , mientras que M''_k es el supremo de f en Δ''_k . Observemos que, como $f(x) \leq 1$ para todo $x \in [0, 1]$, se tiene en particular que $M'_k \leq 1$ para todo k . Por otro lado, como señalamos antes, en cualquiera de los intervalos Δ''_k , cualquier racional es de la pinta $r = \frac{p}{q}$ con $q \geq K$, con lo cual: $f(x) = 0$ si $x \notin \mathbb{Q}$, mientras que $f(p/q) = 1/q \leq 1/K$. Esto nos dice que $M''_k < \frac{\varepsilon}{2}$ para todo k . Con esto,

$$S(f, P) \leq 1 \cdot \sum \Delta'_k + \frac{\varepsilon}{2} \sum \Delta''_k.$$

Ciertamente $\sum \Delta''_k \leq 1$ pues estos intervalos son sólo una parte del $[0, 1]$. Por otro lado, cada intervalo Δ'_k tiene ancho 2α , y hay N de ellos (en realidad son $N - 1$ y después hay que contar las dos mitades de los extremos), con lo cual

$$S(f, P) \leq 2\alpha N + \frac{\varepsilon}{2}.$$

Si elegimos (achicando si es necesario) α de manera que $\alpha < \frac{\varepsilon}{4N}$, se tiene $S(f, P) < \varepsilon$, que es lo que queríamos probar.

II. La función f definida en la nota anterior verifica la siguiente propiedad: para cualquier $a \in [0, 1]$, se tiene $\lim_{x \rightarrow a} f(x) = 0$.

Para probarlo, sean $a \in [0, 1]$, $\varepsilon > 0$, y tomemos cualquier número natural K tal que $\frac{1}{K} < \varepsilon$. Recordemos que por lo observado en el ítem previo, los números que tienen denominador menor que K , que se obtienen juntando todos los de la lista de arriba, son finitos. Esto quiere decir que tiene que haber algún entorno $(a - \delta, a + \delta)$ del punto a en el intervalo $[0, 1]$ donde no hay ningún racional p/q con $q < K$ (con la posible excepción de $p/q = a$, pero no importa porque estamos mirando el límite $\lim_{x \rightarrow a} f(x)$, y entonces el punto $x = a$ no nos interesa). Para convencernos de esta última afirmación sobre el entorno, pensemos en la afirmación opuesta: sería decir que para todo $\delta > 0$ hay algún racional $p/q \neq a$ con $q < K$ y tal que $|p/q - a| < \delta$; pero esto nos permitiría fabrica infinitos de estos puntos contradiciendo lo que contamos anteriormente. Ahora veamos que el límite es en efecto 0. Tomemos x con $|x - a| < \delta$, $x \neq a$. Entonces si $x \notin \mathbb{Q}$, se tiene $f(x) = 0$ por definición y ciertamente $|f(x)| < \varepsilon$. Por otra parte si $x \in \mathbb{Q}$, por como elegimos δ , debe ser $x = p/q$ con $q \geq K$, luego

$$|f(x)| = |f(p/q)| = \frac{1}{q} \leq \frac{1}{K} < \varepsilon,$$

que es lo que queríamos probar.

En particular, observando la definición de f , resulta que f es continua en los irracionales, y discontinua en los racionales. Esto permitiría otra prueba de la integrabilidad de f , pero para ello necesitaríamos probar (y no vamos a hacerlo aquí) que una función acotada, que es continua salvo un conjunto numerable de puntos, es integrable.

Los matemáticos son como los franceses; lo que sea que les digas ellos lo trasladan a su propio lenguaje e inmediatamente es algo completamente distinto.

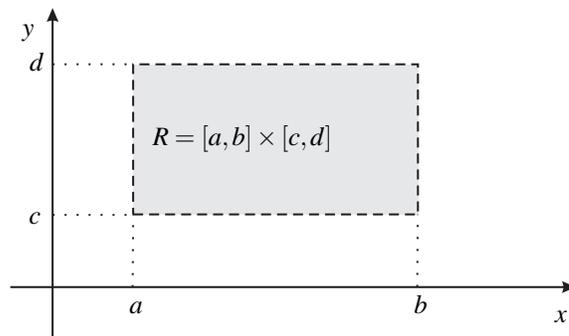
Goethe

7.1. Integrales en el plano

Comencemos por definir integrales para funciones con dominio en algún subconjunto $D \subset \mathbb{R}^2$. Nos vamos a encontrar con dificultades propias de los dominios de este tipo. Para comenzar es bueno entender el caso más simple, en el que el dominio es un rectángulo.

7.1.1. Rectángulos y particiones

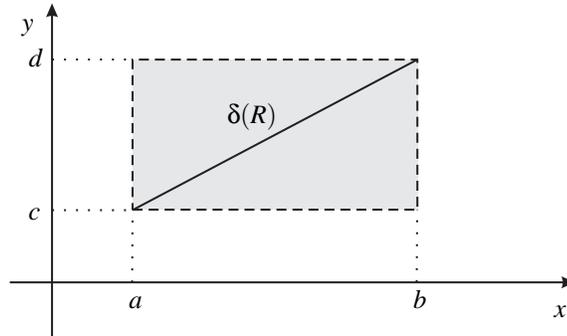
El análogo de un intervalo en el plano es un rectángulo, que se puede escribir siempre como el producto cartesiano de dos intervalos:



En este caso $R = [a, b] \times [c, d]$ que es lo mismo que decir que $X = (x, y) \in R$ si y sólo si $x \in [a, b]$ e $y \in [c, d]$. Su medida o área es el producto de los lados, que denotaremos μ . Es decir, el número positivo.

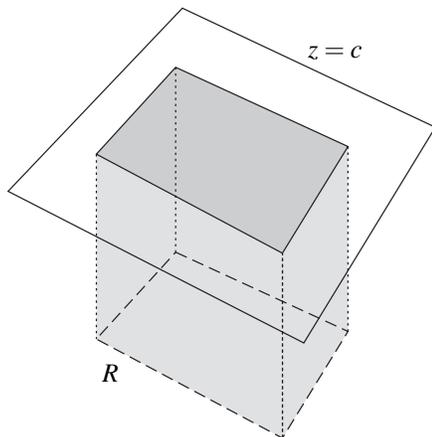
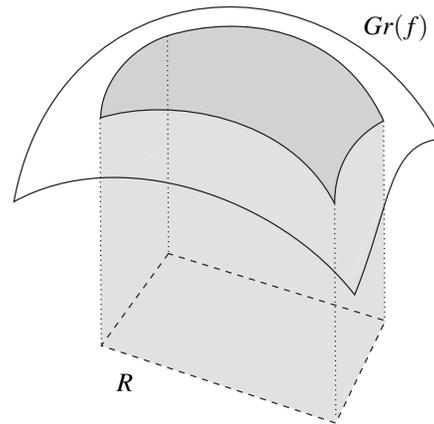
$$\mu(R) = (b - a)(d - c).$$

El **diámetro** $\delta(R)$ del rectángulo R es la mayor distancia entre dos de los puntos del rectángulo, que como se puede observar, es simplemente el largo de la diagonal del rectángulo:



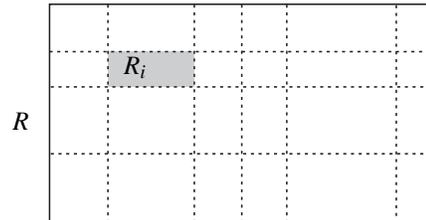
El hecho a rescatar de esta definición es $X, Y \in R \Rightarrow \|X - Y\| \leq \delta(R)$.

Si tenemos una función $f(x, y)$ definida en R , podemos preguntarnos cuál es el sentido de calcular la integral de f en R . La respuesta es simple si pensamos en el gráfico de f , que es un subconjunto de \mathbb{R}^3 , y por ahora, hacemos la simplificación de suponer que $f \geq 0$. Entonces la integral de f queremos que represente el volumen de la figura formada por el rectángulo con piso el plano $z = 0$ y techo el gráfico de f , como en la figura de la derecha.

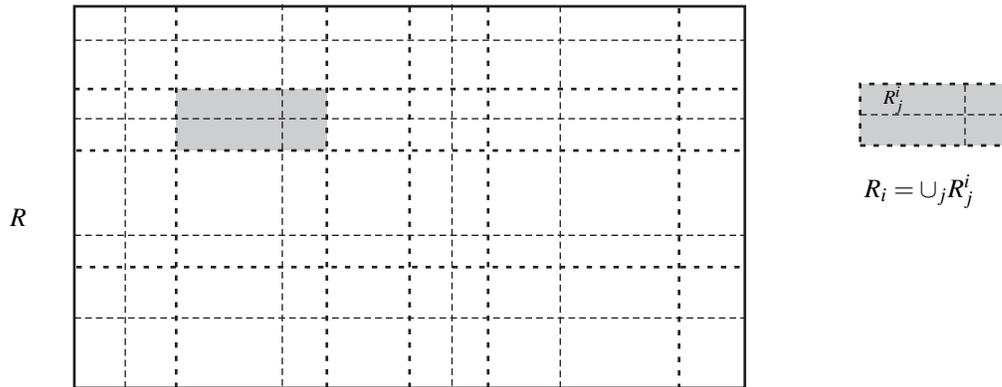


Observemos lo que pasa en un ejemplo sencillo: si $f(x, y) = 1$, obtenemos un “paralelepípedo” (ladrillo) de base R y altura 1, cuyo volumen es $1(b - a)(d - c)$. En general, si $f(x, y) = c$ con $c \geq 0$, obtenemos que la figura es un ladrillo de base R y altura c , por lo que su volumen es $c(b - a)(d - c)$.

Vamos a definir las sumas superiores e inferiores de una función en forma análoga a como hicimos en la recta. Dado un rectángulo R , podemos subdividirlo siempre en una cantidad finita de rectángulos más pequeños, es decir $R = \cup R_i$ donde los R_i son rectángulos (disjuntos salvo los bordes) del plano, como se ve en la figura:



Esto es lo que denominamos una partición P de R , y en general lo anotamos como $P = \{R_i\}$. Un **refinamiento** P' de P es otra partición que descomponga cada uno de los rectángulos R_i en una unión (disjunta salvo los bordes) de $R_i = \cup_j R_j^i$ de nuevos rectángulos, como indica la figura:



de manera que ahora $R = \cup_{i,j} R_j^i$.

El **diámetro** $\delta(P)$ de una partición (también llamado **norma** de la partición y denotado $\|P\|$) es el máximo de los diámetros de todos los rectángulos que forman la partición. Indica que tan pequeños son los rectángulos: a menor diámetro, más chicos los rectángulos de la partición. Observemos que dada una partición P cualquiera del rectángulo original R , y un número positivo ε , siempre podemos obtener un refinamiento P' de P de manera tal que $\delta(P') < \varepsilon$.

Definición 7.1.1. Dada una función positiva $f : R \rightarrow \mathbb{R}$ y acotada, y una partición $P = \cup R_i$ de R ,

1. La suma superior de f en la partición P es

$$S(f, P) = \sum_i M_i \mu(R_i)$$

donde M_i es el supremo de f en el rectángulo R_i .

2. La suma inferior es

$$I(f, P) = \sum_i m_i \mu(R_i)$$

donde ahora m_i indica el ínfimo de f en R_i .

7.1.2. Integral de Riemann

Claramente, si $M = \sup(f)$ en R y $m = \inf(f)$ en R , entonces para cualquier partición $P = \cup R_i$ de R se tiene

$$m \leq m_i \leq M_i \leq M$$

para todo i , donde m_i, M_i indican el supremo e ínfimo de f en R_i . Entonces se tiene

$$I(f, P) \leq S(f, P)$$

para cualquier partición. Como en el caso de una variable, es fácil ver que las sumas superiores decrecen a medida que la partición se refina (pues el supremo en un conjunto más pequeño es menor o igual que en el conjunto más grande), y también es fácil ver que las sumas inferiores crecen. También se deduce entonces que dado un par arbitrario P, Q de particiones,

$$I(f, P) \leq S(f, Q).$$

Se observa que las sumas inferiores están acotadas superiormente. Luego estas cantidades tienen un supremo que denotaremos $I_*(f)$, la **integral inferior** de Riemann. Análogamente, al ínfimo de las sumas superiores lo denotaremos $I^*(f)$, y evidentemente

$$I_*(f) \leq I^*(f).$$

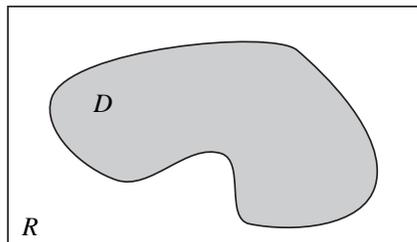
Cuando ambas cantidades coinciden decimos que f es **Riemann integrable en R** . A esta cantidad la denotamos con $\int_R f$, y la llamamos **integral doble de f** , y lo usual es usar la notación

$$\iint_R f$$

para aclarar que se trata de una función de dos variables.

Dominios generales

Cuando una función está definida en un dominio $D \subset \mathbb{R}^2$ acotado, pero que no es necesariamente un rectángulo, se considera la siguiente función auxiliar \bar{f} . Se toma un rectángulo R cualquiera tal que $D \subset R$



y se define

$$\bar{f}(X) = \begin{cases} f(X) & \text{si } X \in D \\ 0 & \text{si } X \in R - D \end{cases}$$

Se definen las sumas superiores e inferiores de f utilizando \bar{f} , y decimos que f es integrable en D si \bar{f} es integrable en R .

Esta integral no depende del rectángulo elegido, puesto que si cambiamos el rectángulo por otro R' , en la diferencia de ambos rectángulo la función \bar{f} es nula.

Se define la integral de f en D como la integral de \bar{f} en R , es decir

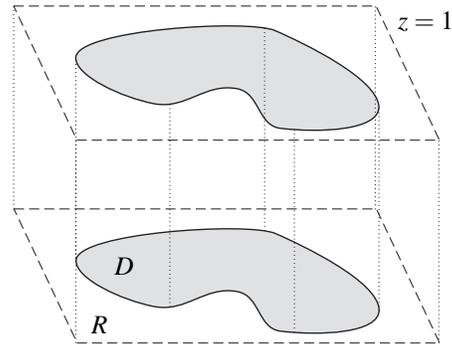
$$\int_D f := \int_R \bar{f}.$$

Una pregunta clave, a la que volveremos más adelante, es como saber si esta función extendida es integrable o no. Dicho de otra manera, ¿cómo sabemos si dada una función definida en un dominio que no es un rectángulo, si esta función es integrable o no?

Definición 7.1.2. Dado un conjunto acotado $D \subset \mathbb{R}^2$, decimos que D es **medible** si la función $f = 1$ es integrable en D , y en ese caso la medida de D se calcula como

$$\mu(D) = \int_D 1.$$

Este número es el **área** del conjunto D , como puede observarse en la figura, pues la altura es constantemente uno.



Dada cualquier función acotada f definida en un dominio acotado D , consideramos sus partes positiva y negativa dadas por

$$f_+(X) = \begin{cases} f(X) & \text{si } f(X) > 0 \\ 0 & \text{si } f(X) \leq 0 \end{cases},$$

$$f_-(X) = \begin{cases} -f(X) & \text{si } f(X) < 0 \\ 0 & \text{si } f(X) \geq 0 \end{cases}$$

Ambas son funciones positivas, y f es integrable Riemann en D si f_+ y f_- lo son, y como $f = f_+ - f_-$,

$$\int_D f = \int_D f_+ - \int_D f_-$$

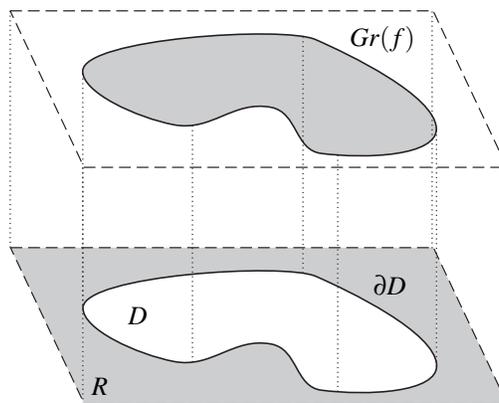
Observemos que, como en el caso de una variable, $|f| = f_+ + f_-$.

El criterio para ver si una función integrable es análogo al de una variable, con la misma demostración que por lo tanto omitimos.

Teorema 7.1.3. Sea $D \subset \mathbb{R}^2$ un conjunto acotado. Entonces $f : D \rightarrow \mathbb{R}$ es integrable en D si y sólo si para todo $\varepsilon > 0$ existe una partición P de un rectángulo que contenga a D tal que

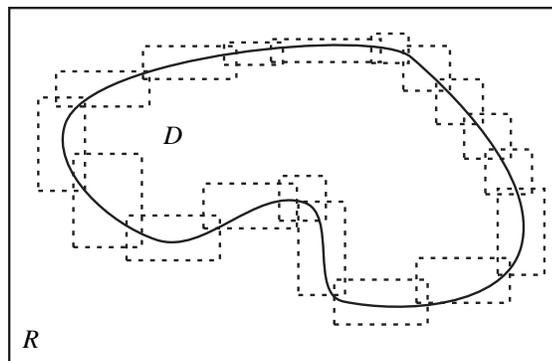
$$S(f, P) - I(f, P) < \varepsilon.$$

De acuerdo a lo que discutimos, si el dominio no es un rectángulo, hay que integrar la función extendida por cero en algún rectángulo que contenga al dominio. Pero surge el problema de si esta nueva función es integrable en el rectángulo.



Supongamos que tenemos un dominio D como en el dibujo, y lo extendemos a un rectángulo R que lo contenga, haciendo que f sea cero fuera del dominio D , como discutimos antes. El problema, es que esta función no es más continua en R , aunque lo sea en D , pues en el borde de D tenemos un salto, aún en el caso en el que f sea constante, como se observa en la figura.

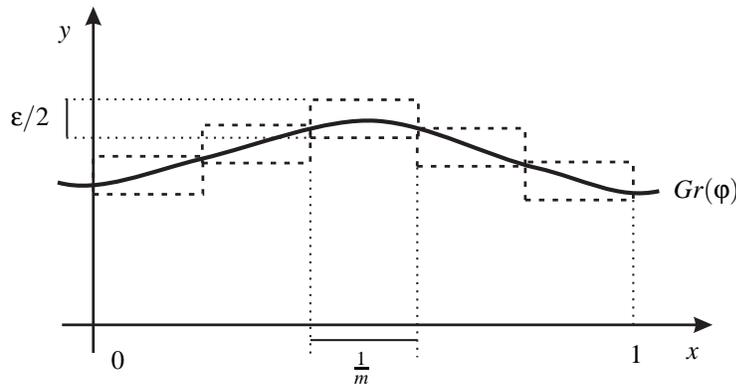
La observación importante es que allí, en ∂D , es en el único lugar donde no es continua, pues fuera es continua por ser constante. Nos alcanzaría con ver que esta región, la curva ∂D , no aporta nada a la integral en el rectángulo R . **En el caso general, esto no es necesariamente cierto.** Sin embargo, bajo ciertas condiciones razonables, como por ejemplo si la curva ∂D se obtiene como unión finita de gráficos de funciones continuas, se puede integrar con tranquilidad.



La idea es cubrir el borde con rectángulos de manera que la suma de las áreas de estos rectángulos sea arbitrariamente pequeña. Veamos que esto es posible.

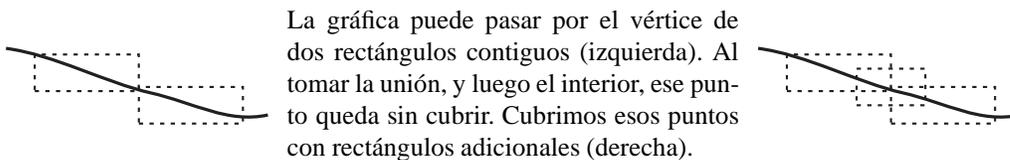
Proposición 7.1.4. Sea $\varphi : [a, b] \rightarrow \mathbb{R}$ una función continua. Entonces dado $\varepsilon > 0$, existe un conjunto finito de rectángulos $R_i \subset \mathbb{R}^2$ de manera que $Gr(\varphi) \subset (\cup R_i)^o$ y además $\sum \mu(R_i) = \varepsilon$.

Demostración. Para simplificar las cuentas, supongamos que $I = [0, 1]$. Por ser continua, φ es uniformemente continua en I . Entonces, dado $\varepsilon > 0$, existe $\delta > 0$ tal que $|x - x'| < \delta$ implica $|\varphi(x) - \varphi(x')| < \varepsilon/2$. Tomamos $m \in \mathbb{N}$ tal que $\frac{1}{m} < \delta$, y particionamos el intervalo $[0, 1]$ en m pedazos iguales de longitud $1/m$. En cada intervalo las imágenes de φ están contenidas en una banda de altura $\varepsilon/2$, como indica la figura:



Formamos los rectángulos de base $1/m$ y altura $\varepsilon/2$, de manera que $Gr(\varphi) \subset \cup R_i$. Como hay m rectángulos, y cada uno de ellos tiene área $= b.h = \frac{1}{m} \frac{\varepsilon}{2}$, se tiene

$$\sum \mu(R_i) = m \frac{1}{m} \frac{\varepsilon}{2} = \varepsilon/2.$$



La gráfica puede pasar por el vértice de dos rectángulos contiguos (izquierda). Al tomar la unión, y luego el interior, ese punto queda sin cubrir. Cubrimos esos puntos con rectángulos adicionales (derecha).

De estos puntos conflictivos, hay a lo sumo $m + 1$. Cubrimos cada uno de ellos con un rectángulo de base $\frac{1}{(m+1)}$ y altura $\varepsilon/2$. Entonces ahora se tiene en efecto $Gr(\varphi) \subset (\cup R_i)^o$, y por otro lado la suma de las áreas de los primeros rectángulos da $\varepsilon/2$, mientras que la de los nuevos a lo sumo

$$(m+1)b.h = (m+1) \frac{1}{(m+1)} \frac{\varepsilon}{2} = \frac{\varepsilon}{2}.$$

□

Se tiene en consecuencia el siguiente teorema. Recordemos que f integrable en un dominio D quiere decir que si metemos el dominio dentro de un rectángulo, y extendemos a f como 0 fuera de D , entonces f es integrable en el rectángulo.

Teorema 7.1.5. Sea $D \subset \mathbb{R}^2$ un compacto, y $f : D \rightarrow \mathbb{R}$ una función continua. Si ∂D está formado por una unión finita de gráficas de funciones continuas, entonces f es integrable en D .

Demostración. Tomamos un rectángulo cualquiera R tal que $D \subset R$, y extendemos a f por cero en $R - D$. Sea $\varepsilon > 0$. Queremos ver que existe una partición P de R tal que

$$S(f, P) - I(f, P) < \varepsilon.$$

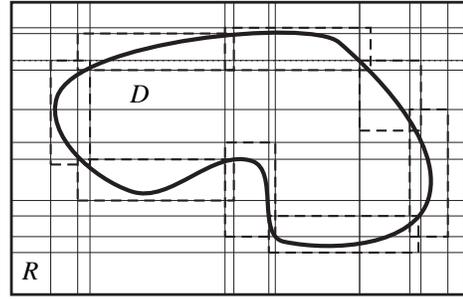
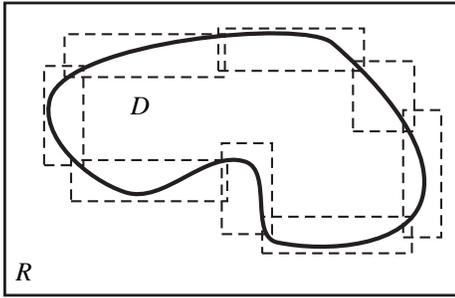
Sea $s = \max(f) - \min(f)$ en R ($s > 0$ si $f \neq 0$ en R). Cubrimos primero el borde de D con finitos rectángulos $\{R_i\}_{i \in A}$, de manera que

$$\sum \mu(R_i) = \frac{\varepsilon}{2s}.$$

Como $R - (\cup R_i)^o$ es compacto, y f es continua allí, es uniformemente continua (teorema de Heine-Borel). Es decir, dado $\varepsilon' = \frac{\varepsilon}{2\mu(R)} > 0$, existe $\delta > 0$ tal que $\|X - Y\| \leq \delta$ ($X, Y \in R - (\cup R_i)^o$) implica

$$|f(X) - f(Y)| \leq \varepsilon'.$$

Tomamos ahora una partición $P = \{R_i\}_{i \in F}$ de R que contenga a los rectángulos $\{R_i\}_{i \in A}$ que cubren al borde de D , como indica la figura (estamos subdividiendo aquellos rectángulos que estaban superpuestos para pensarlos como rectángulos disjuntos):



Es decir $P = (\cup_{i \in A} R_i) \cup (\cup_{i \in F-A} R_i)$. Refinando, podemos suponer que la partición tiene diámetro $\delta(P) < \delta$. Entonces

$$\begin{aligned} S(f, P) - I(f, P) &= \sum_{i \in F} (M_i - m_i) \mu(R_i) = \sum_{i \in A} (M_i - m_i) \mu(R_i) + \sum_{i \in F-A} (M_i - m_i) \mu(R_i) \\ &< \sum_{i \in A} s \mu(R_i) + \sum_{i \in F-A} \varepsilon' \mu(R_i) \\ &\leq s \frac{\varepsilon}{2s} + \varepsilon' \mu(R) = \varepsilon. \end{aligned}$$

□

7.1.3. Integrales iteradas y el Teorema de Fubini

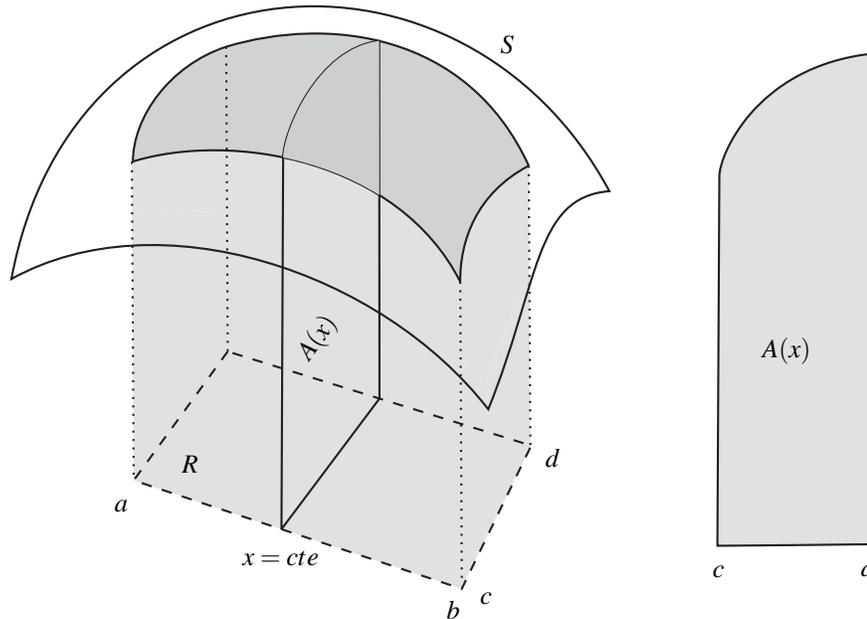
En general puede ser muy complicado calcular integrales múltiples usando la definición. Sin embargo, el siguiente teorema nos dice que, bajo ciertas condiciones bastante generales, el cálculo se reduce al cálculo sucesivo de integrales en una variable, donde podemos aplicar los métodos que conocemos.

Estas integrales se conocen como **integrales iteradas**, un ejemplo elemental es el siguiente:

$$\int_0^1 \left(\int_2^3 x^2 y dy \right) dx = \int_0^1 x^2 \frac{y^2}{2} \Big|_2^3 dx = \frac{1}{2} \int_0^1 (9-4)x^2 dx = \frac{5}{2} \int_0^1 x^2 dx = \frac{5}{2} \frac{1}{3} x^3 \Big|_0^1 = \frac{5}{6}.$$

En los próximos párrafos veremos que lo que acabamos de calcular puede interpretarse como la integral de $f(x,y) = x^2 y$ en el rectángulo $R = [0, 1] \times [2, 3]$, la cual también puede calcularse integrando en el otro orden (primero x y luego y).

La idea del siguiente teorema es que, si queremos calcular la integral de f , pensándola como el volumen indicado en la figura,



podemos cortar con planos $x = cte.$, y luego calcular el área $A(x)$ de la figura que se obtiene integrando la función $f_x(y) = f(x,y)$ respecto de la variable y de la manera usual. Finalmente integramos estas áreas para x moviéndose entre a y b . Esto se conoce como **principio de Cavalieri**. Para $x \in [a, b]$, sea $f_x : [c, d] \rightarrow \mathbb{R}$ la función que se obtiene al fijar $x \in [a, b]$, es decir $f_x(y) = f(x,y)$. Entonces

$$A(x) = \int_c^d f_x(y) dy,$$

y el principio de Cavalieri establece que bajo ciertas condiciones

$$vol = \int_a^b A(x) dx.$$

Vamos a ver bajo que condiciones se puede aplicar este principio. Daremos una versión simplificada del teorema, que será la que usaremos en las aplicaciones. La versión general del teorema puede verse en las notas del final del capítulo (Nota I).

Teorema 7.1.6 (Fubini simplificado). *Sea $R = [a, b] \times [c, d]$, un rectángulo en \mathbb{R}^2 , y $f : R \rightarrow \mathbb{R}$ una función integrable en R .*

1. Si para todo $x \in [a, b]$, $f_x : [c, d] \rightarrow \mathbb{R}$ es integrable, entonces

$$\int \int_R f = \int_a^b \int_c^d f(x, y) dy dx.$$

2. Si para todo $y \in [c, d]$, $f_y : [a, b] \rightarrow \mathbb{R}$ es integrable, entonces

$$\int \int_R f = \int_c^d \int_a^b f(x, y) dx dy.$$

Demostración. Consideremos una partición P del rectángulo, donde $\Delta_i = [x_i, x_{i+1}]$ es una partición de $[a, b]$ y $\Delta'_j = [y_j, y_{j+1}]$ es una partición de $[c, d]$, de manera que los rectángulos $R_{ij} = \Delta_i \times \Delta'_j$ de área $(x_{i+1} - x_i)(y_{i+1} - y_i)$ forman la partición P del rectángulo original. Como antes, abusamos un poquito de la notación y usamos Δ para denotar al intervalo y a su longitud. Entonces

$$I(f, P) = \sum_{i,j} m_{ij}(f) \Delta_i \Delta'_j = \sum_i (\sum_j m_{ij}(f) \Delta'_j) \Delta_i,$$

donde $m_{ij}(f)$ indica el ínfimo de f en R_{ij} .

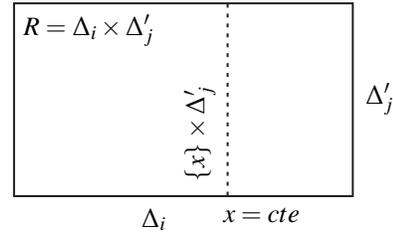
Vamos a demostrar sólo 1., la demostración de 2. es análoga. Supongamos entonces que f_x es integrable para todo $x \in [a, b]$, y bautizemos $I : [a, b] \rightarrow \mathbb{R}$ a la integral de f_x respecto de y :

$$I(x) = \int_c^d f_x(y) dy = \int_c^d f(x, y) dy.$$

Si $x \in R_{ij}$, entonces $m_j(f_x)$ (que es el ínfimo de la función f_x en Δ'_j) es mayor o igual que el ínfimo de f en R_{ij} , pues

$$m_j(f_x) = \inf\{f(x, y) : x \text{ fijo}, y \in \Delta'_j\},$$

y el ínfimo en un subconjunto -en este caso $\{x\} \times \Delta'_j$ - siempre es mayor o igual que el ínfimo en el conjunto original -en este caso $\Delta_i \times \Delta'_j$ -, como indica la figura.



Se deduce que, cualquiera sea $x \in \Delta_i$,

$$\sum_j m_{ij}(f) \Delta'_j \leq \sum_j m_j(f_x) \Delta'_j \leq I_*(f_x) = \int_c^d f_x(y) dy = I(x).$$

Si en el extremo derecho tomamos el ínfimo para $x \in \Delta_i$ - que anotamos $m_i(I)$ -, multiplicamos por Δ_i y sumamos, se deduce que

$$I(f, P) = \sum_i (\sum_j m_{ij}(f) \Delta'_j) \Delta_i \leq \sum_i m_i(I) \Delta_i \leq I_*(I),$$

es decir

$$I(f, P) \leq I_*(I).$$

Con un razonamiento análogo se deduce que

$$I^*(I) \leq S(f, P).$$

Luego, como siempre vale $I_*(I) \leq I^*(I)$,

$$I(f, P) \leq I_*(I) \leq I^*(I) \leq S(f, P).$$

Como f es integrable, los extremos se pueden hacer tan próximos como queramos refinando la partición P , y en consecuencia debe ser $I(x)$ integrable, y además

$$\int \int_R f = I_*(I) = I^*(I) = \int_a^b I = \int_a^b \int_c^d f(x, y) dy.$$

□

Corolario 7.1.7. Si $f : R \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ es continua, entonces f es integrable y además

$$\int \int_R f = \int_a^b \left(\int_c^d f dy \right) dx = \int_c^d \left(\int_a^b f dx \right) dy.$$

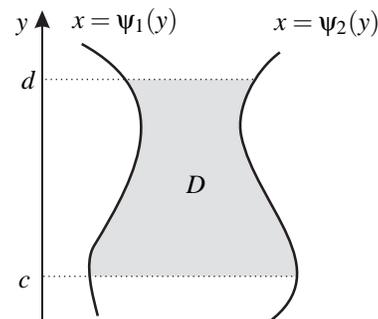
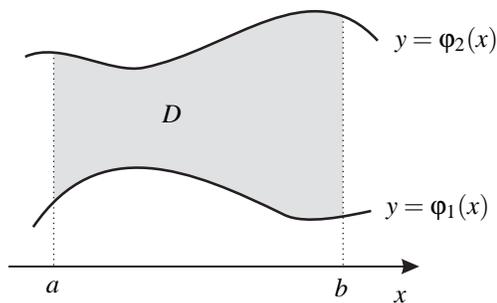
Demostración. Si f es continua en R es integrable por el teorema 7.1.5. Además f_x, f_y son continuas con lo cual son integrables, así que se puede usar el teorema de Fubini. □

Ejemplo 7.1.8. Sea $f(x, y) = e^{xy}x$. Se quiere calcular $\int_R f$ donde $R = [0, 1] \times [0, 1]$. Como f es continua, calculamos una integral iterada

$$\begin{aligned} \int \int_R f &= \int_0^1 \int_0^1 e^{xy} x dy dx = \int_0^1 e^{xy} \Big|_{y=0}^{y=1} dx = \int_0^1 (e^x - 1) dx \\ &= e^x - x \Big|_{x=0}^{x=1} = (e - 1) - (1 - 0) = e - 2. \end{aligned}$$

¿Qué pasa si integramos en el otro orden? El teorema nos asegura que debe dar lo mismo. Sin embargo, la cuenta en el otro orden es más larga ya que requiere calcular primero $\int e^{xy} x dx$, para lo cual es necesario recurrir al método de partes.

Una aplicación mucho más útil la obtenemos combinando el Teorema 7.1.5 con el Teorema de Fubini, observando las siguientes figuras en el plano:



Corolario 7.1.9. 1. Sea $D \subset \mathbb{R}^2$ un compacto que se puede escribir como

$$\{(x, y) \in \mathbb{R}^2 : \varphi_1(x) \leq y \leq \varphi_2(x), x \in [a, b]\}.$$

Supongamos que $f : D \rightarrow \mathbb{R}$ es continua, y las φ_i son continuas. Entonces

$$\iint_D f = \int_a^b \left(\int_{y=\varphi_1(x)}^{y=\varphi_2(x)} f(x, y) dy \right) dx.$$

2. Sea $D \subset \mathbb{R}^2$ un compacto que se puede escribir como

$$\{(x, y) \in \mathbb{R}^2 : \psi_1(y) \leq x \leq \psi_2(y), y \in [c, d]\}.$$

Supongamos que $f : D \rightarrow \mathbb{R}$ es continua, y las ψ_i son continuas. Entonces

$$\iint_D f = \int_c^d \left(\int_{x=\psi_1(y)}^{x=\psi_2(y)} f(x, y) dx \right) dy.$$

Demostración. Consideramos la función f extendida como cero a cualquier rectángulo

$$R = [a, b] \times [c, d]$$

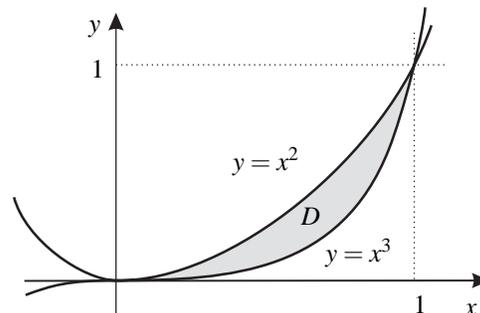
que contenga a D . Por el teorema previo, la función obtenida es integrable en el rectángulo. No es difícil ver que las f_x son funciones integrables para cada $x \in [a, b]$ fijo, puesto que f_x vale cero para aquellos y tales que $c \leq y \leq \varphi_1(x)$ o bien $\varphi_2(x) \leq y \leq d$, y es una función continua para aquellos y tales que $\varphi_1(x) \leq y \leq \varphi_2(x)$. Por el mismo motivo,

$$\int_c^d f(x, y) dy = \int_{y=\varphi_1(x)}^{y=\varphi_2(x)} f(x, y) dy.$$

Se deduce la conclusión usando el Teorema de Fubini. Con un razonamiento análogo se deduce el ítem 2. \square

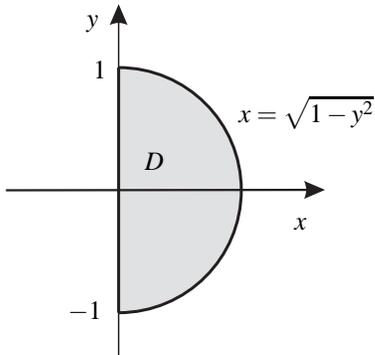
Ejemplo 7.1.10. Algunos cálculos de integrales dobles.

1. Se quiere calcular la integral $\iint_D f$, donde $f(x, y) = x^2 y$, mientras que D es la región del plano encerrado entre $y = x^3$, $y = x^2$, con $x \in [0, 1]$ según indica la figura.



Luego

$$\begin{aligned}\int_D f &= \int_0^1 \left(\int_{y=x^3}^{y=x^2} x^2 y dy \right) dx = \int_0^1 x^2 \frac{1}{2} y^2 \Big|_{x^3}^{x^2} \\ &= \frac{1}{2} \int_0^1 x^2 (x^6 - x^4) = \frac{1}{2} \left(\frac{1}{9} x^9 - \frac{1}{7} x^7 \right) \Big|_0^1 = \frac{1}{2} \left(\frac{1}{9} - \frac{1}{7} \right).\end{aligned}$$



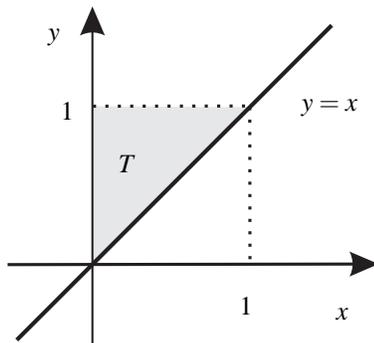
2. Si D es un semicírculo, de radio unitario y centrado el origen se quiere calcular $\iint_D \frac{y}{x+1}$.

Entonces

$$\begin{aligned}\int_D f &= \int_{-1}^1 \left(\int_{x=0}^{x=\sqrt{1-y^2}} y \frac{1}{x+1} dx \right) dy = \int_{-1}^1 y \ln(x+1) \Big|_0^{\sqrt{1-y^2}} dy \\ &= \int_{-1}^1 y (\ln(1 + \sqrt{1-y^2}) - 0) dy = \int_{-1}^1 y \ln(1 + \sqrt{1-y^2}) dy.\end{aligned}$$

Esta integral da cero pues la función a integrar es impar.

3. Se quiere calcular $\int_0^1 \int_y^1 e^{x^2} dx dy$. El problema en este caso es que no conocemos una primitiva elemental de e^{x^2} ¿Qué hacer? Observemos que las condiciones de integración definen un triángulo, pues $1 \leq x \leq y$ mientras que $0 \leq y \leq 1$:

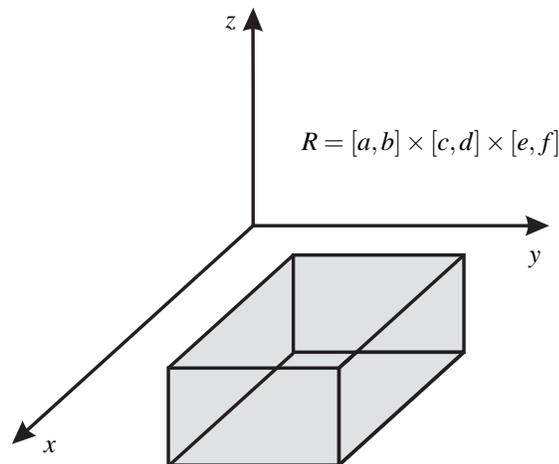


Entonces, usamos Fubini para cambiar el orden de integración. Se tiene

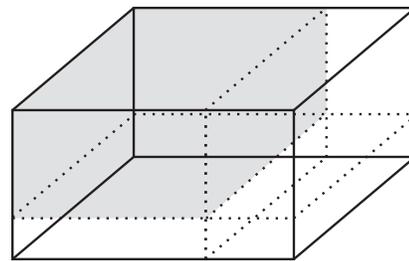
$$\begin{aligned} \int_0^1 \int_y^1 e^{x^2} dx dy &= \int_T e^{x^2} = \int_0^1 \int_0^x e^{x^2} dy dx \\ &= \int_0^1 e^{x^2} y|_0^x dx = \int_0^1 e^{x^2} x dx = \frac{1}{2} e^{x^2} \Big|_0^1 = \frac{1}{2}(e-1). \end{aligned}$$

7.2. Integrales en \mathbb{R}^3 y \mathbb{R}^n

Observemos ahora el caso de integrales de funciones con dominio en \mathbb{R}^3 . Aquí el análogo de un intervalo es un ladrillo, que seguiremos llamando R por comodidad,



Su **medida** es el volumen $\mu(R) = (b-a)(d-c)(f-e)$, y su diámetro $\delta(R)$ el largo de la diagonal mayor del ladrillo. Una partición P de R es una subdivisión de R en ladrillos disjuntos P_i como en la figura, y un refinamiento P' de P una subdivisión de cada uno de los ladrillos $P_i = \cup P'_j$. El diámetro $\delta(P)$ de la partición es el diámetro del ladrillo más grande que la conforma.



¿Qué querría decir aquí la integral de una función? En principio, esperamos que la integral de la función $f(x,y,z) = 1$ nos devuelva la medida del cubo

$$\text{volumen} = 1(b-a)(d-c)(f-e),$$

aunque no podamos visualizar el gráfico de f . La integral se define usando particiones de manera análoga al

caso $n = 2$, y cuando existe la integral de f se denomina **integral triple**, y se denota

$$\iiint_R f.$$

En general, para cualquier $n \geq 1$, un **rectángulo** R es el producto cartesiano de n intervalo

$$R = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n],$$

su **medida** es el producto

$$\mu(R) = (b_1 - a_1)(b_2 - a_2) \cdots (b_n - a_n),$$

y su diámetro $\delta(R)$ el largo de la diagonal mayor, es decir, la mayor distancia posible entre dos puntos de R . La integral se define de la manera obvia.

El siguiente es un corolario inmediato del Teorema de Fubini:

Corolario 7.2.1. Si $R = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$ es un rectángulo en \mathbb{R}^n y $f : R \rightarrow \mathbb{R}$ es una función continua en R , entonces

$$\int_R f = \int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_n}^{b_n} f(x_1, x_2, \dots, x_n) dx_1 dx_2 \cdots dx_n,$$

donde la integración se puede efectuar en cualquier orden.

Demostración. Hay que observar que si f es continua, entonces es continua en cada una de sus variables por separado, y entonces cada

$$\int_{a_i}^{b_i} f(x_1, x_2, \dots, x_n) dx_i$$

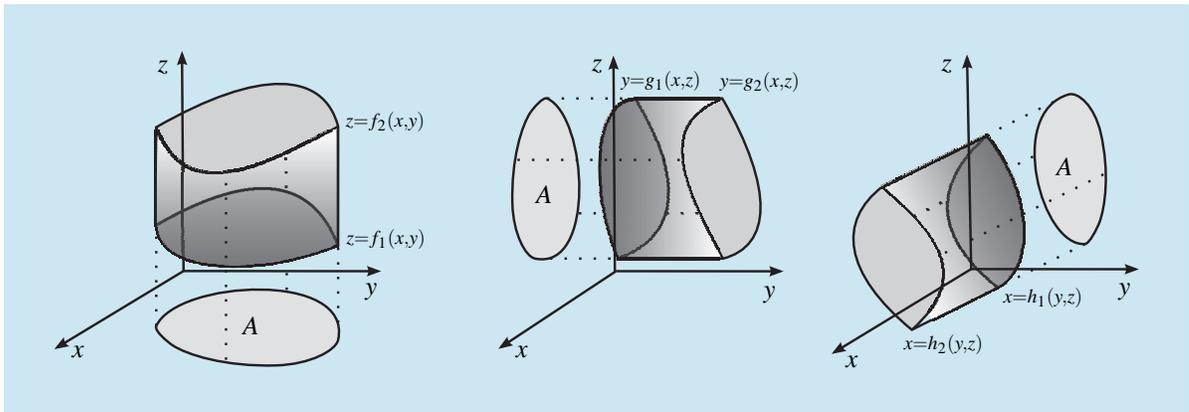
existe y es una función continua en $[a_i, b_i]$, en particular integrable. Usando el teorema de Fubini repetidas veces se obtiene el corolario. \square

Lo que es más interesante es pensar qué ocurre cuando el dominio no es un ladrillo en \mathbb{R}^3 . Con razonamientos similares a los del plano, se puede probar lo siguiente:

Teorema 7.2.2. Sea $D \subset \mathbb{R}^3$ un compacto, y $f : D \rightarrow \mathbb{R}$ una función continua. Si ∂D está formado por una unión finita de gráficos de funciones continuas definidas en conjuntos compactos del plano, entonces f es integrable en D .

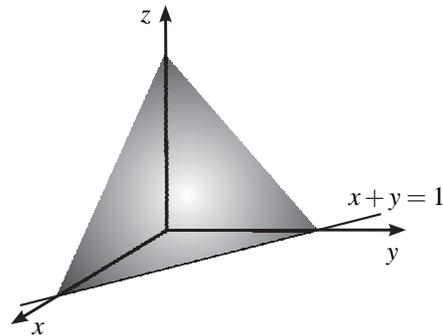
Aquí las superficies reemplazan a las curvas como borde de la región, y la idea central es que una superficie en \mathbb{R}^3 , que es gráfica de una función continua en un compacto, tiene volumen nulo, y por lo tanto el borde de la región no influye en la existencia de la integral.

Respecto del cálculo, aquí hay más posibilidades, simplemente por ser tres las variables. Observemos las siguientes figuras. En el primer caso, se trata de de la región encerrada por los gráficos de funciones $f_i(x, y)$, cuyo dominio común A se indica como una sombra en el plano xy . En los otros casos se tienen las variantes obvias.



Hay que poder escribir a las superficies que delimitan el dominio $D \subset \mathbb{R}^3$ como gráficos de funciones, con respecto a un par de variables. La mejor manera de entender cómo funciona la teoría es haciendo ejemplos.

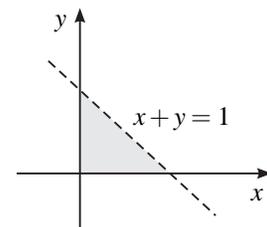
Ejemplo 7.2.3. 1. Queremos calcular el volumen de la siguiente figura, que es la región encerrada por el plano $x + y + z = 1$ en el primer octante:



En consecuencia, podemos integrar la función 1 en el volumen determinado por la figura. Las condiciones que definen la figura son, en primer lugar

$$0 \leq z \leq 1 - x - y.$$

Luego hay que observar la sombra de la figura en el plano xy , que es un triángulo, de donde se deduce que $0 \leq y \leq 1 - x$, y que $0 \leq x \leq 1$.

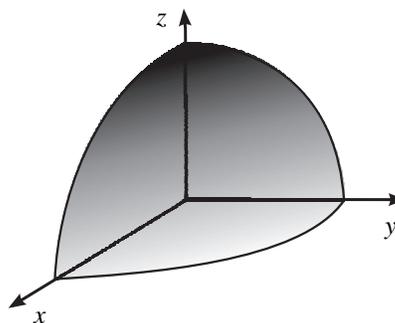


Luego

$$\begin{aligned}
 \text{vol}(D) &= \mu(D) = \int \int \int_D 1 = \int_0^1 \int_0^{1-x} \int_0^{1-x-y} 1 \, dz dy dx \\
 &= \int_0^1 \int_0^{1-x} (1-x-y) dy dx = \int_0^1 (y - xy - \frac{1}{2}y^2) \Big|_0^{1-x} dx \\
 &= \int_0^1 (1-x-x(1-x) - \frac{1}{2}(1-x)^2) dx \\
 &= \int_0^1 (1-2x+x^2 - \frac{1}{2}(1-x)^2) dx \\
 &= x - x^2 + \frac{1}{3}x^3 + \frac{1}{6}(1-x)^3 \Big|_0^1 = 1 - 1 + \frac{1}{3} - (0 + \frac{1}{6}) = \frac{1}{6}.
 \end{aligned}$$

2. Masa, volumen y densidad.

Se quiere calcular la masa del octavo de esfera que se halla en el primer octante. De acuerdo al principio físico que establece que $\rho = m/v$, la masa se puede calcular como el producto $m = \rho v$, en el caso en que la densidad ρ es constante. Se sabe que la densidad está dada por la función $\rho(x, y, z) = z$, es decir, aumenta a medida que ascendemos.



Si recordamos que una integral triple se calcula como límite del volumen de pequeñas cajitas C_i multiplicadas por la función $\rho(x, y, z)$

$$\int \int \int_D \rho = \lim \sum \rho(x, y, z) \text{vol}(C_i),$$

donde (x, y, z) es algún punto en la caja C_i , entonces lo que estamos haciendo al integrar, es sumar las masas aproximadas de las cajas. Al refinar la partición, las cajas son más pequeñas pero el valor $\rho(x, y, z) \text{vol}(C_i)$ es cada vez más parecido a la masa real de la caja C_i . Idealmente, si la densidad no varía bruscamente de un punto a otro cercano, se tiene la fórmula para la masa total del objeto dada por el límite

$$m = \int \int \int_D \rho.$$

En este caso se tiene $0 \leq z \leq \sqrt{1-x^2-y^2}$, mientras que mirando la sombra en el piso se observa que

$0 \leq y \leq \sqrt{1-x^2}$, $0 \leq x \leq 1$. Entonces

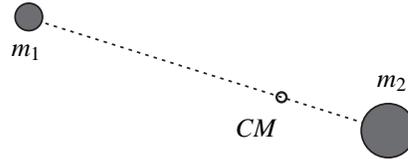
$$\begin{aligned}
 m(D) &= \int_0^1 \int_0^{\sqrt{1-x^2}} \int_0^{\sqrt{1-x^2-y^2}} z \, dz \, dy \, dx = \int_0^1 \int_0^{\sqrt{1-x^2}} \frac{1}{2}(1-x^2-y^2) \, dy \, dx \\
 &= \frac{1}{2} \int_0^1 \left((y-x^2y - \frac{1}{3}y^3) \Big|_0^{\sqrt{1-x^2}} \right) dx \\
 &= \frac{1}{2} \int_0^1 \sqrt{1-x^2}(1-x^2) - \frac{1}{3}(1-x^2)^{\frac{3}{2}} dx \\
 &= \frac{1}{2} \int_0^1 \frac{2}{3}(1-x^2)^{\frac{3}{2}} dx = \frac{1}{3} \int_0^{\frac{\pi}{2}} \cos^3(u) \cos(u) du = \frac{1}{3} \int_0^{\frac{\pi}{2}} \cos^4(u) du \\
 &= \frac{1}{3} \left(\frac{1}{4} \cos(u)^3 \sin(u) + \frac{3}{8} \cos(u) \sin(u) + \frac{3}{8} u \right) \Big|_0^{\frac{\pi}{2}} = \frac{1}{3} \frac{3\pi}{16} = \frac{\pi}{16}.
 \end{aligned}$$

donde hicimos la sustitución $x = \sin(u)$, y luego integramos por partes (¡o mejor, usamos una tabla!).

3. Dadas dos partículas puntuales (ideales) de masas m_1 y m_2 , ubicadas respectivamente en los puntos V_1 y V_2 del espacio, se define su **centro de masa** o **baricentro** como la ubicación promediada de las posiciones, teniendo en cuenta sus respectivas masas. Esto es, si $m_1 + m_2 = m_T$ la masa total, entonces el centro de masa CM se define de la siguiente manera:

$$CM = \frac{m_1 V_1 + m_2 V_2}{m_1 + m_2} = \frac{m_1 V_1 + m_2 V_2}{m_T}.$$

Si las partículas tienen la misma masa, el centro de masa es simplemente el punto medio del segmento que une la posición de ambas. En el caso general, el centro de masa es algún punto de este segmento, más cercano siempre a la partícula de mayor masa:



Dada una nube de partículas m_i de posiciones $V_i \in \mathbb{R}^3$, se define en forma análoga el centro de masa de la nube como

$$CM = \frac{\sum m_i V_i}{\sum m_i}.$$

Es fácil ver que si todas las partículas tienen la misma masa este es el centro geométrico de sus ubicaciones.

En el caso de objetos sólidos $D \subset \mathbb{R}^3$, de densidad ρ , si pensamos a la masa infinitesimal de una caja C_i como dada por $\rho \text{vol}(C_i)$, entonces el centro de masa puede pensarse como la suma

$$CM = \frac{\sum \rho(P_i) \text{vol}(C_i) P_i}{\sum \rho(P_i) \text{vol}(C_i)}.$$

donde $P_i = (x_i, y_i, z_i)$ es un punto genérico en la caja C_i . Es decir, las coordenadas del centro de masa se piensan como

$$(CM)_x \simeq \frac{\sum x_i \rho(P_i) \mu(C_i)}{\sum \rho(P_i) \mu(C_i)},$$

$$(CM)_y \simeq \frac{\sum y_i \rho(P_i) \mu(C_i)}{\sum \rho(P_i) \mu(C_i)},$$

$$(CM)_z \simeq \frac{\sum z_i \rho(P_i) \mu(C_i)}{\sum \rho(P_i) \mu(C_i)}.$$

Estas expresiones, pasando al límite de cajas cada vez más pequeñas nos dan las expresiones de las coordenadas del baricentro del sólido D :

$$(CM)_x = \frac{\int \int \int_D x \rho(x, y, z)}{m(D)},$$

$$(CM)_y = \frac{\int \int \int_D y \rho(x, y, z)}{m(D)},$$

$$(CM)_z = \frac{\int \int \int_D z \rho(x, y, z)}{m(D)},$$

donde $m(D)$ indica la masa total de D como en el ejemplo anterior. Esto está fundamentado por el Teorema 7.3.2 que enunciamos y demostramos más abajo, que nos dice que para calcular la integral de Riemann de una función integrable, se puede tomar cualquier punto en el rectángulo y evaluar f allí (no es necesario tomar el ínfimo o el supremo).

7.3. Propiedades

La próxima proposición establece una serie de propiedades de la integral, tanto en el plano como el espacio, cuya demostración es análoga al caso $n = 1$ y por tanto es omitida.

Proposición 7.3.1. Sea $D \subset \mathbb{R}^n$ un conjunto acotado, sean $f, g : D \rightarrow \mathbb{R}$ funciones integrables en D . Entonces

1. Si $\alpha \in \mathbb{R}$, αf es integrable en D y además

$$\int_D \alpha f = \alpha \int_D f.$$

2. Si $\varphi : \text{Im}(f) \rightarrow \mathbb{R}$ es continua, entonces $\varphi \circ f$ es integrable en D .
3. Las funciones $|f|$, $f^k = f \cdot f \cdots f$ ($k \in \mathbb{N}$), $f + g$, fg son integrables en D . Además

$$\left| \int_D f \right| \leq \int_D |f|, \quad \int_D (f + g) = \int_D f + \int_D g.$$

4. Si $D' \subset \mathbb{R}^n$ es otro conjunto acotado tal que $\mu(D \cap D') = 0$, entonces si f es integrable en $D \cup D'$, se tiene

$$\int_{D \cup D'} f = \int_D f + \int_{D'} f.$$

Por último una propiedad que usaremos de aquí en adelante: no es necesario tomar el ínfimo o el supremo de f para calcular su integral de Riemann:

Teorema 7.3.2. Sea $f : R \subset \mathbb{R}^n \rightarrow \mathbb{R}$ es integrable, $\{P_i\}$ una partición de R y $R_i \in P_i$ un punto cualquiera en el rectángulo R_i . Entonces $\int_R f$ se puede calcular como límite de sumas

$$\sum f(P_i)\mu(R_i),$$

en el sentido siguiente: al refinar la partición estas sumas tienden a la integral de f en R .

Demostración. Simplemente hay que observar que, para cualquier partición P del rectángulo R , y cualquier rectángulo R_i en la partición P , dado un punto $P_i \in R_i$ se tiene

$$m_i \leq f(P_i) \leq M_i.$$

Entonces

$$I(f, P) \leq \sum f(P_i)\mu(R_i) \leq S(f, P),$$

y como las cantidades de los extremos tienden a la integral de f en R , la cantidad del medio también. \square

7.3.1. Medida de una región y Teorema del valor medio

Recordemos que definimos la medida de una región $D \subset \mathbb{R}^n$ como la integral (si existe) de la función 1. Es decir

$$\mu(D) = \int_D 1.$$

Si existe, en el plano esta cuenta nos devuelve la superficie de la región D , mientras que en el espacio nos devuelve el volumen de la región D . ¿Podemos dar condiciones para que este número exista? Sí, es sencillo a partir de lo ya hecho, pues la función 1 es continua, y por lo tanto se tiene el siguiente resultado:

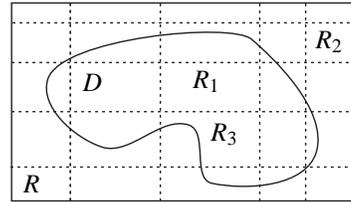
Teorema 7.3.3. Sea $D \subset \mathbb{R}^n$ compacto, $f : D \rightarrow \mathbb{R}$ continua, y supongamos que ∂D está formado por una unión finita de gráficos de funciones $\varphi : K_i \subset \mathbb{R}^{n-1} \rightarrow \mathbb{R}$ continuas, con K_i compactos. Entonces f es integrable en D , y además $\int_D f$ se puede calcular con n integrales iteradas.

En particular D es medible y

$$\mu(D) = \int_D 1 = \sup \sum_{U_i \subset D} \mu(R_i) = \inf \sum_{D \subset \cup R_j^*} \mu(R_j^*).$$

Demostración. Lo único nuevo son las últimas igualdades que dan $\mu(D)$. Son consecuencia del siguiente hecho: para calcular $\int_D 1$ tomamos un rectángulo R que contenga a D , extendemos a la función como cero en $R - D$. Es decir, $f = 1$ en D y $f = 0$ en $R - D$.

Ahora particionamos $R = \cup R_i$. Si R_i es un rectángulo íntegramente contenido en D (como el que indicamos con R_1 en la figura), entonces allí la integral inferior de f vale siempre 1, por lo tanto estos rectángulos aportan algo a la integral. Mientras que si el rectángulo toca el exterior de D (o sea no está contenido en D , como los que señalamos como R_2 y R_3 en la figura), entonces allí el ínfimo de f es cero y estos no aportan nada a la integral inferior.



En cambio para la integral superior aportan todos los rectángulos que tocan D pues en ellos el supremo de f es 1. \square

Observación 7.3.4. El teorema previo generaliza una idea que para intervalos $I \subset \mathbb{R}$ es mucho más sencilla. Por ejemplo si $I = [0, 1]$, uno puede obtener la medida del intervalo aproximándolo por dentro con intervalos de la pinta $[\frac{1}{n}, 1 - \frac{1}{n}]$. O bien por fuera con intervalos de la pinta $[-\frac{1}{n}, 1 + \frac{1}{n}]$.

Podemos relacionar la integral de cualquier función continua con $\mu(D)$ mediante el siguiente Teorema del Valor Medio Integral:

Teorema 7.3.5. Sea $D \subset \mathbb{R}^n$ compacto, $f : D \rightarrow \mathbb{R}$ continua, y supongamos que ∂D está formado por una unión finita de gráficos de funciones $\phi : K_i \subset \mathbb{R}^{n-1} \rightarrow \mathbb{R}$ continuas, con K_i compactos. Entonces

1. Si $M = \max(f)$ en D y $m = \min(f)$ en D , entonces

$$m\mu(D) \leq \int_D f \leq M\mu(D).$$

2. Si D es arcoconexo entonces existe $X_0 \in D$ tal que

$$f(X_0)\mu(D) = \int_D f.$$

Demostración. 1. Se deduce de la definición de integral, pues para cualquier partición P suficientemente fina, si consideramos a los rectángulos que están dentro de D se tiene

$$m \sum_{R_i \subset D} \mu(R_i) \leq \sum_{R_i \subset D} m_i(f)\mu(R_i) \leq I(f, P) \leq I_*(f) = \int_D f.$$

Refinando la partición se tiene que $\sum_{R_i \subset D} \mu(R_i) \rightarrow \mu(D)$ y entonces

$$m\mu(D) \leq \int_D f.$$

Con un argumento similar, pero ahora usando los rectángulos dentro de D más los que cubren el borde de D se tiene

$$M\mu(D) \geq I^*(f) = \int_D f.$$

Se deduce lo enunciado juntando las últimas dos desigualdades.

2. Si $\mu(D) = 0$ no hay nada que probar. Si $\mu(D) \neq 0$, entonces dividiendo la desigualdad del ítem previo por $\mu(D)$ se deduce que

$$m \leq \frac{1}{\mu(D)} \int_D f \leq M.$$

Como f es continua en un arcoconexo toma todos los valores intermedios entre m y M , en particular existe X_0 donde

$$f(X_0) = \frac{1}{\mu(D)} \int_D f.$$

□

Observación 7.3.6. Atención que el último enunciado es falso si el dominio no es arcoconexo. Por ejemplo si D es la unión de dos cuadrados disjuntos de lado 1, y f vale -1 en uno de ellos y 1 en el otro:



Entonces

$$\int \int_D f = \int \int_C f + \int \int_{C'} f = 1 \cdot 1 + (-1) \cdot 1 = 0,$$

mientras que f no se anula en D .

7.4. NOTAS

- I. El Teorema de Fubini que enunciamos es una versión más débil del que sigue en esta nota. La diferencia está en que en realidad, no es necesario suponer que cada f_x, f_y es integrable. Si $R = [a, b] \times [c, d]$ es un rectángulo en \mathbb{R}^2 , y $f : R \rightarrow \mathbb{R}$ es una función integrable en R , usamos la notación $f_x(y) = f(x, y) = f_y(x)$ como antes. Tomamos

$$I(x) = I_*(f_x) = \sup\{I(f_x, P) : P \text{ partición de } [c, d]\},$$

$$S(x) = I^*(f_x) = \inf\{S(f_x, P) : P \text{ partición de } [c, d]\}.$$

Entonces (sin ninguna hipótesis sobre las f_x o las f_y), vale que I, S son funciones integrables en $[a, b]$ y además

$$\int_R f = \int_a^b S(x) dx = \int_a^b I(x) dx.$$

Vale el resultado análogo para f_y .

Para la demostración, consideremos una partición P del rectángulo, donde $\Delta_i = [x_i, x_{i+1}]$ es una partición de $[a, b]$ y $\Delta'_j = [y_j, y_{j+1}]$ es una partición de $[c, d]$, de manera que los rectángulos $R_{ij} = \Delta_i \times \Delta'_j$ de área $(x_{i+1} - x_i)(y_{i+1} - y_i)$ forman la partición P del rectángulo original. Como antes, abusamos un poquito de la notación y usamos Δ para denotar al intervalo y a su longitud. Entonces

$$I(f, P) = \sum_{i,j} m_{ij}(f) \Delta_i \Delta'_j = \sum_i \left(\sum_j m_{ij}(f) \Delta'_j \right) \Delta_i,$$

donde $m_{ij}(f)$ indica el ínfimo de f en R_{ij} . Si $x \in R_{ij}$, entonces $m_j(f_x)$ (el ínfimo de f_x en Δ'_j) es mayor o igual que el ínfimo de f en R_{ij} , pues

$$m_j(f_x) = \inf\{f(x,y) : x \text{ fijo}, y \in \Delta'_j\},$$

y el ínfimo en un subconjunto -en este caso $\{x\} \times \Delta'_j$ - siempre es mayor o igual que el ínfimo en el conjunto original -en este caso $\Delta_i \times \Delta'_j$. Se deduce que, cualquiera sea $x \in \Delta_i$,

$$\sum_j m_{ij}(f)\Delta'_j \leq \sum_j m_j(f_x)\Delta'_j \leq I_*(f_x) = I(x).$$

Si en el extremo derecho tomamos el ínfimo para $x \in \Delta_i$ - que anotamos $m_i(I)$ -, multiplicamos por Δ_i y sumamos, se deduce que

$$I(f,P) = \sum_i \left(\sum_j m_{ij}(f)\Delta'_j \right) \Delta_i \leq \sum_i m_i(I)\Delta_i \leq I_*(I),$$

es decir $I(f,P) \leq I_*(I)$. Con un razonamiento análogo se deduce que $I^*(S) \leq S(f,P)$. Luego, como $I(x) \leq S(x)$ para todo $x \in [a,b]$, también se tiene la desigualdad $I^*(I) \leq I^*(S)$. Juntando estas desigualdades se obtiene la cadena

$$I(f,P) \leq I_*(I) \leq I^*(I) \leq I^*(S) \leq S(f,P).$$

Como f es integrable, los extremos se pueden hacer tan próximos como queramos refinando la partición P , y en consecuencia debe ser

$$\int_R f = I_*(I) = I^*(I) = \int_a^b I,$$

lo que prueba que I es integrable en $[a,b]$, con integral igual a la integral doble de f .

La prueba para la integral de S es idéntica y la omitimos, lo mismo vale para las afirmaciones respecto de f_y .

Si yo estuviera comenzando mis estudios nuevamente, seguiría el consejo de Platón y comenzaría por las matemáticas.

Galileo Galilei

8.1. El método de sustitución

Nuestro objetivo para terminar con estas notas es entender como funciona el teorema de cambio de variables en dimensiones 2 y 3, teorema que generaliza la idea del método de sustitución para integrales en \mathbb{R} .

Recordemos que si f es una función continua en un intervalo cerrado $I \subset \mathbb{R}$, y g es una función C^1 en $[a, b]$ tal que se puede hacer la composición $f \circ g$ en $[a, b]$, entonces $h(x) = f(g(x))g'(x)$ es una función integrable y además

$$\int_a^b f(g(x))g'(x)dx = \int_{g(a)}^{g(b)} f(u)du. \quad (8.1)$$

En efecto, si F es una primitiva de f , es decir $F'(t) = f(t)$ para todo t en el intervalo I , entonces se chequea fácilmente (derivando) que $H(x) = F(g(x))$ es una primitiva de $f(g(x))g'(x)$. Si aplicamos el Teorema Fundamental del Cálculo Integral a la función f en el intervalo con extremos $g(a), g(b)$, obtenemos

$$F(g(b)) - F(g(a)) = \int_{g(a)}^{g(b)} f(u)du$$

puesto que F es primitiva de f . Por otro lado, aplicando el mismo teorema a la función H' en el intervalo $[a, b]$ obtenemos

$$F(g(b)) - F(g(a)) = H(b) - H(a) = \int_a^b f(g(x))g'(x)dx,$$

lo que prueba la igualdad (8.1).

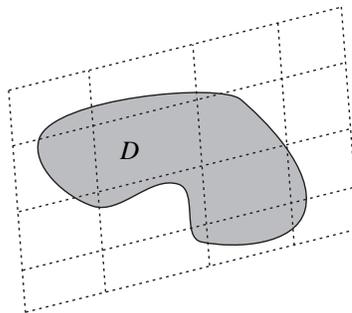
8.2. Particiones generales

Empecemos por una consideración general que no demostraremos con gran detalle, pero que es razonable y permite simplificar los razonamientos que seguirán. Para calcular integrales dobles, dado un dominio

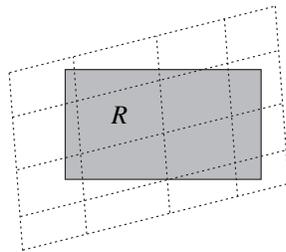
$D \subset \mathbb{R}^2$ con una frontera buena, lo que hacemos es subdividir el dominio en rectángulos, de lados paralelos a los ejes. Mientras más pequeños sean los rectángulos, mejor es la aproximación del dominio, y una integral la pensamos como límite de las sumas

$$\lim \sum f(P_i)\mu(R_i)$$

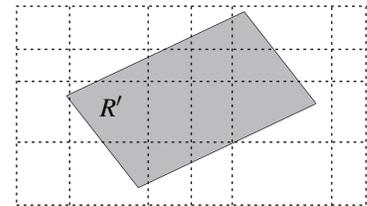
donde $P_i \in R_i$. En particular, el área de D se calcula sumando las áreas de los rectángulos. Sin embargo, la elección de estas figuras (rectángulos de lados paralelos a los ejes) es en cierta medida arbitraria. Es consecuencia de que el área de un rectángulo es fácil de calcular. Pero esto último también es cierto para muchas otras figuras, y nada impide calcular integrales usando estas figuras, es decir, enmarcando al dominio D con estas figuras, particionándolo y calculando las sumas como arriba, y luego tomando el límite al refinar la partición. Un caso concreto es el siguiente: podemos tomar paralelogramos, es decir, figuras de lados paralelos dos a dos, y hacer una partición de D como indica la figura:



Puede probarse que existe la integral de f en D usando rectángulos comunes si y sólo si existe la integral usando estos paralelogramos. La idea de por qué funciona esto es la siguiente: dado un rectángulo R de lados paralelos a los ejes, puede tomarse una partición del mismo usando paralelogramos, que aproxime tanto como uno quiera al rectángulo R .



Recíprocamente, dado un paralelogramo R' , se lo puede aproximar tanto como uno quiera usando rectángulos de lados paralelos a los ejes.



Vamos a usar este hecho: si f es integrable en D , entonces su integral se puede calcular como

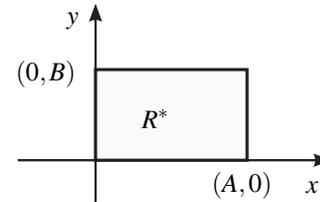
$$\int \int_D f = \lim \sum f(P_i)\mu(R'_i)$$

donde R'_i es una familia de paralelogramos que aproxima la región D , y $P_i \in R'_i$.

8.3. Transformaciones lineales

Observemos primero que ocurre en el plano, con una transformación lineal $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ inversible. Dados dos vectores linealmente independientes, estos van a parar por T a otros dos vectores linealmente independientes. Y entonces todo rectángulo (o mejor dicho, todo paralelogramo) va a parar a otro paralelogramo. Veamos la relación entre las áreas de R^* y $R = T(R^*)$, razonando primero sobre un caso particular.

Dado un rectángulo de lados paralelos a los ejes, su área se calcula como base por altura, $\mu(R^*) = AB$. Podemos suponer que el rectángulo tiene uno de sus vértices en el origen, como indica la figura de abajo, pues una traslación no modifica su área. Es decir, los lados miden exactamente A y B .



Si la transformación lineal T está dada por la matriz $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ en la base canónica, entonces se tiene

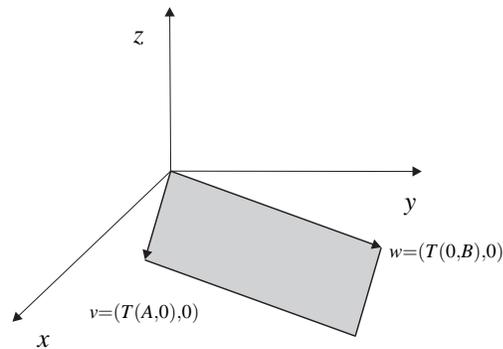
$$T(A,0) = T(A(1,0)) = AT(1,0) = A(a,c) = (Aa,Ac)$$

mientras que

$$T(0,B) = BT(0,1) = B(b,d) = (Bb,Bd).$$

Entonces nuestro rectángulo R^* va a parar al paralelogramo $R = T(R^*)$ que está generado por $T(A,0)$ y $T(0,B)$.

Ahora recordemos que dados dos vectores $v, w \in \mathbb{R}^3$, el área del paralelogramo determinado por v, w se puede calcular como la norma del producto vectorial entre v y w , es decir $area = \|v \times w\|$. Tomemos entonces el paralelogramo imagen, pero pensado en el plano xy en \mathbb{R}^3 como indica la figura de la derecha. Los vectores son simplemente $v = (Aa,Ac,0)$, $w = (Bb,Bd,0)$.



El área de $R = T(R^*)$ es entonces

$$\|v \times w\| = \left\| \det \begin{vmatrix} i & j & k \\ Aa & Ac & 0 \\ Bb & Bd & 0 \end{vmatrix} \right\| = \|(0,0,AaBd - AcBb)\| = |AB| |ad - bc| = \mu(R^*) |\det T|.$$

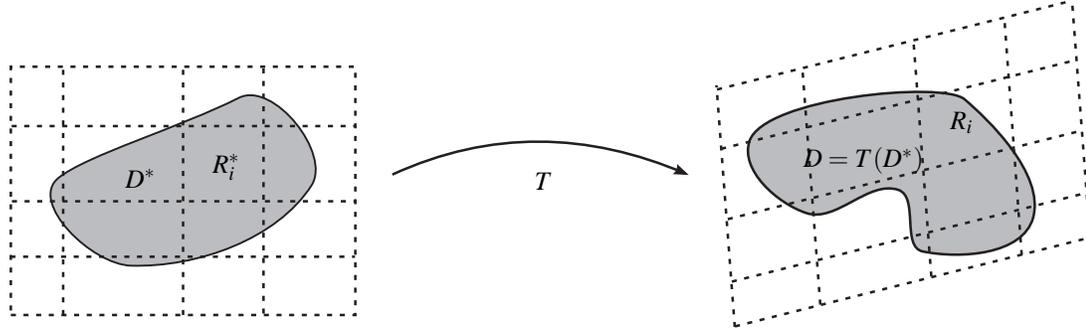
Hemos descubierto el siguiente hecho fundamental:

Proposición 8.3.1. Si $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ es una transformación lineal inversible, y R^* es un rectángulo de lados paralelos a los ejes, entonces el área del paralelogramo $R = T(R^*)$ que se obtiene como la imagen de R^* por

T es

$$\mu(R) = |\det T| \mu(R^*).$$

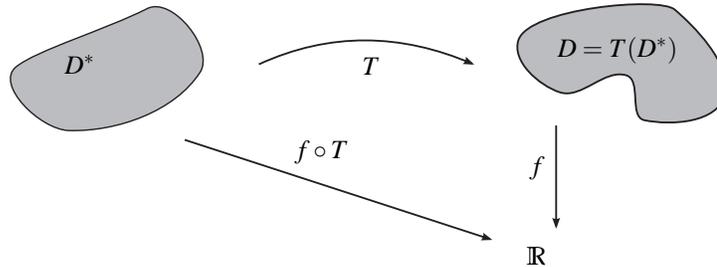
Pero entonces, dada cualquier figura $D^* \subset \mathbb{R}^2$ que sea medible, si $D = T(D^*)$ indica la imagen de D^* por una transformación lineal T inversible, podemos cubrir D^* con rectángulos R_i^* que aproximen la figura, y así estaremos llenando $D = T(D^*)$ con paralelogramos $R_i = T(R_i^*)$ que aproximan esa figura, como en el dibujo:



En consecuencia, $\mu(D) = \mu(T(D^*)) = |\det T| \mu(D^*)$. En efecto,

$$\mu(T(D^*)) = \lim \sum \mu(R_i) = \lim \sum \mu(T(R_i^*)) = \lim \sum |\det T| \mu(R_i^*) = |\det T| \mu(D^*).$$

Esta misma idea se puede aplicar para calcular integrales. Observemos en la figura, que si $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ está definida en $D = T(D^*)$, entonces $f \circ T$ está definida en D^* :



Teorema 8.3.2 (Cambio de variable, versión lineal). Si $D^* \subset \mathbb{R}^2$, $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ es una transformación lineal inversible y $f : D = T(D^*) \rightarrow \mathbb{R}$ es integrable, entonces

$$\int_D f = \int_{D^*} (f \circ T) |\det T|.$$

Demostración. Tomamos R_i^* rectángulos que aproximen D^* , de manera que $R_i = T(R_i^*)$ aproximen $D =$

$T(D^*)$. Entonces, si $P_i \in R_i$, se puede escribir $P_i = T(Q_i)$ con $Q_i \in R_i^*$, luego

$$\begin{aligned} \int_{T(D^*)} f &= \lim \sum f(P_i) \mu(R_i) = \lim \sum (f \circ T)(Q_i) \mu(T(R_i^*)) \\ &= \lim \sum |\det T| (f \circ T)(Q_i) \mu(R_i^*) = \int_{D^*} (f \circ T) |\det T|. \end{aligned}$$

□

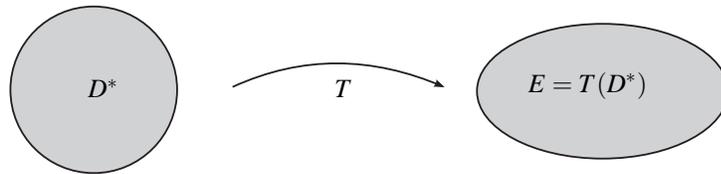
Ejemplo 8.3.3. Supongamos sabido que el área de un disco de radio R en el plano es πR^2 . Se quiere calcular el área de una elipse E de radios $a, b > 0$ dada por la parte interna de la curva de ecuación

$$\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1.$$

Si consideramos $T(x, y) = (ax, by)$, es una transformación lineal cuya matriz es simplemente

$$T = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}.$$

Como $a, b \neq 0$, T es inversible y en particular $|\det T| = |ab| = ab$. Observemos que la circunferencia unitaria $D^* = \{x^2 + y^2 \leq 1\}$ va a parar a E si le aplicamos T :

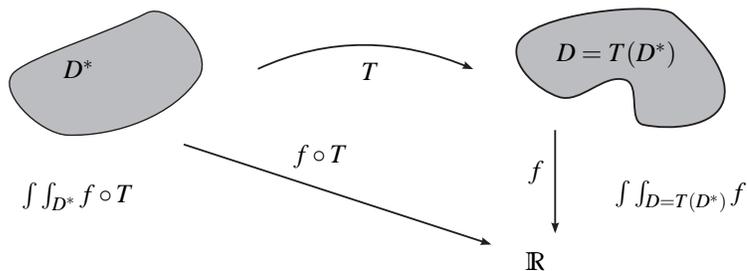


Entonces

$$\mu(E) = \int_E 1 = \int_{T(D^*)} 1 = \int_{D^*} 1 |\det T| = \mu(D^*) |\det T| = \pi ab.$$

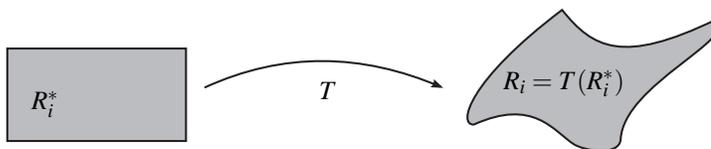
8.4. Cambio de variable

Observemos la analogía en la fórmula del Teorema 8.3.2 con la fórmula (8.1), que representa el caso de una variable: $|\det T|$ juega el papel de g' . Por ahora este teorema tiene utilidad limitada. Lo que queremos es extender el resultado a funciones inyectivas cuya diferencial sea inversible salvo un conjunto muy pequeño (de medida nula). Es decir, dado un dominio D^* y una función $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ inyectiva, queremos hallar primero la relación entre $\mu(D^*)$ y $\mu(T(D^*)) = \mu(D)$,



y luego ver cuál es la relación entre la integral de f en un dominio y la de $f \circ T$ en el otro.

Si T es inyectiva, el dominio $D = T(D^*)$ se puede pensar cubierto con las imágenes de los rectángulos que cubren D^* , es decir si $\cup R_i^*$ aproximan D^* entonces $\cup T(R_i^*)$ aproximan $D = T(D^*)$. ¿Cómo calcular la medida de cada $T(R_i^*)$? En principio es un problema difícil, ya que T no es lineal y cada $R_i = T(R_i^*)$ está deformado como indica la figura:



La idea ahora es la siguiente. Podemos escribir, dado $P_i \in \mathbb{R}^2$, una expresión aproximada

$$T(X) = T(P_i) + DT(P_i)X + O(X),$$

donde $O(X)$ es simplemente "lo que falta después del término lineal". Esta afirmación es imprecisa y luego haremos una demostración formal, pero por ahora pensemos que T se puede escribir así. Tenemos como hipótesis que DT es inversible, salvo tal vez un conjunto de medida nula. Entonces podemos imaginar que, como el primer término es sólo un desplazamiento por un vector fijo $T(P_i) \in \mathbb{R}^2$, que no modifica el área, mientras que $O(X)$ es suficientemente pequeño, y no aporta nada sustancial al área, entonces para un rectángulo pequeño R_i^* con vértice en P_i ,

$$\mu(T(R_i^*)) \approx \mu(DT(P_i)R_i^*) = |\det DT(P_i)|\mu(R_i^*).$$

Siguiendo con esta línea de pensamiento, se tendría para un dominio D^* y un conjunto de rectángulos R_i^* que aproximen D^* ,

$$\begin{aligned} \mu(T(D^*)) &= \lim \sum \mu(T(R_i^*)) \approx \lim \sum \mu(DT(P_i)R_i^*) \\ &= \lim \sum |\det DT(P_i)|\mu(R_i^*) = \int_{D^*} |\det DT|, \end{aligned}$$

es decir

$$\mu(D) = \mu(T(D^*)) = \int_{D^*} |\det DT|.$$

Mientras que si $f : D = T(D^*) \rightarrow \mathbb{R}$ es integrable, con un razonamiento análogo también se tendría

$$\int_D f = \int_{T(D^*)} f = \int_{D^*} (f \circ T) |\det DT|.$$

Vamos a llamar a $JT = |\det DT|$ el **determinante Jacobiano** de la función T .

Antes de pasar a una demostración veamos unos ejemplos para ver como se usa este resultado, conocido como **teorema de cambio de variable**.

Ejemplo 8.4.1. Una transformación que se usa frecuentemente es la dada por el cambio de coordenadas polares, donde

$$T(r, \theta) = (r \cos(\theta), r \sin(\theta)).$$

En este caso se tiene

$$DT = \begin{pmatrix} \cos(\theta) & -r \sin(\theta) \\ \sin(\theta) & r \cos(\theta) \end{pmatrix},$$

con lo cual $JT = r \cos^2(\theta) + r \sin^2(\theta) = r$.

1. Calculamos el área del disco B_R de radio $R > 0$. Se tiene que, si $0 \leq \theta < 2\pi$ y $0 \leq r < R$, entonces $T(r, \theta)$ recorre la circunferencia. Esto es $D^* = [0, R] \times [0, 2\pi]$ mientras que $T(D^*) = B_R$. En consecuencia,

$$\mu(B_R) = \int_0^{2\pi} \int_0^R r dr d\theta = 2\pi \frac{1}{2} r^2 \Big|_0^R = \pi R^2.$$

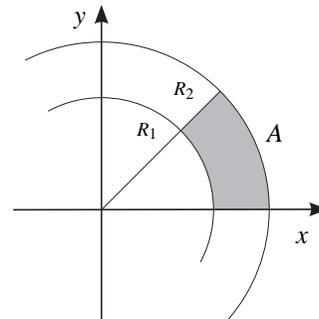
2. Se quiere integrar la función

$$f(x, y) = \frac{xy}{(x^2 + y^2)^{\frac{5}{2}}}$$

en la región anular A de arco $\pi/4$ entre R_1 y R_2 según indica la figura de la derecha. Luego

$$D^* = [R_1, R_2] \times [0, \frac{\pi}{4}],$$

mientras que $T(D^*) = A$, con lo cual



$$\begin{aligned} \iint_A f &= \iint_{D^*} (f \circ T) JT = \int_0^{\frac{\pi}{4}} \int_{R_1}^{R_2} \frac{r \cos(\theta) r \sin(\theta)}{r^5} r dr d\theta \\ &= \int_0^{\frac{\pi}{4}} \cos(\theta) \sin(\theta) d\theta \int_{R_1}^{R_2} \frac{1}{r^2} dr = \frac{1}{2} \sin^2(\theta) \Big|_0^{\frac{\pi}{4}} \left(-\frac{1}{r}\right) \Big|_{R_1}^{R_2} \\ &= -\frac{1}{2} \frac{1}{2} \left(\frac{1}{R_2} - \frac{1}{R_1}\right) = \frac{R_2 - R_1}{4R_1R_2}. \end{aligned}$$

Pasemos ahora a la demostración del teorema. La idea central está extraída del libro “Cálculo en variedades”, de M. Spivak [9].

Teorema 8.4.2. Sea $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ inyectiva y C^1 . Sea $D^* \subset \mathbb{R}^2$ acotado y $f : D = T(D^*) \rightarrow \mathbb{R}$ integrable. Si DT es invertible en D^* y $JT = |\det DT|$, entonces

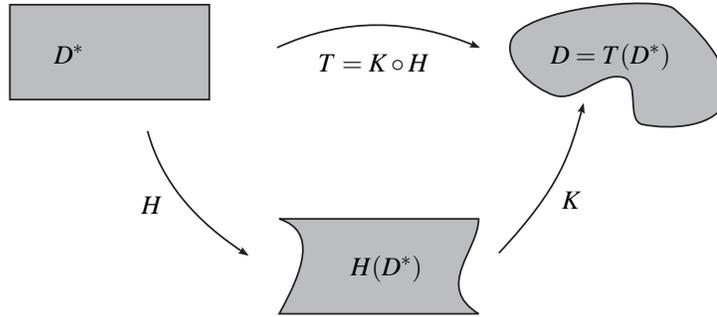
$$\int_D f = \int_{D^*} (f \circ T)JT.$$

Demostración. Basta probar, dado $P \in D^*$, que existe un entorno de R^* de P donde vale el teorema es decir

$$\int_{T(R^*)} f = \int_{R^*} (f \circ T)JT.$$

En efecto, si esto es cierto entonces dividiendo D^* en un conjunto finito de entornos R_i^* donde vale el teorema, y sumando las integrales, se tiene el resultado en el conjunto total D^* .

La idea de la demostración es descomponer a la función T como la composición de dos funciones, cada una moviendo una sola variable, y usar sustitución en una variable en cada una de las funciones.



Sea $P = (a, b) \in D^*$, y $T(x, y) = (t_1(x, y), t_2(x, y))$ y supongamos primero que $DT(P) = I_2$. En particular $\nabla t_1(P) = (1, 0)$. Si ponemos $H(x, y) = (t_1(x, y), y)$, H es una función C^1 y también se tiene $DH(P) = I_2$, puesto que

$$DH = \begin{pmatrix} \frac{\partial t_1}{\partial x} & \frac{\partial t_1}{\partial y} \\ 0 & 1 \end{pmatrix}.$$

Observemos que $JH = |\det DH| = \left| \frac{\partial t_1}{\partial x} \right| = \frac{\partial t_1}{\partial x}$ pues como esta derivada parcial vale uno en P , y es una función continua pues T es C^1 , podemos suponer que estamos en un entorno donde es positiva. Es decir

$$JH = \frac{\partial t_1}{\partial x}.$$

Como $DH(P) = I_2$, por el teorema de la función inversa existe un entorno U (que vamos a suponer que es un rectángulo abierto) de P donde H es invertible. Ahora ponemos

$$K(x, y) = (x, t_2(H^{-1}(x, y)))$$

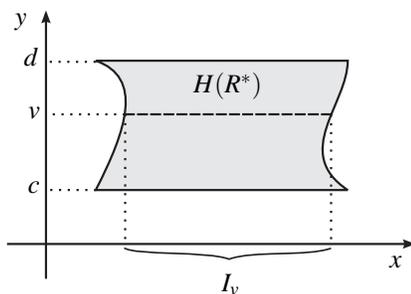
definida en un entorno de $H(P)$. Reemplazando se ve que $K \circ H = (t_1(x, y), t_2(x, y)) = T$. Observemos que $DT = DK_H \circ DH$, donde por DK_H queremos decir la diferencial de K evaluada en $H(X)$. Luego $\det DT = \det DK_H \det DH$, con lo cual

$$JT = JK_H \cdot JH.$$

Tomemos un conjunto $W^* \subset U$ alrededor de P , y una función g integrable allí. Podemos suponer que W^* es suficientemente pequeño para que se pueda tomar un rectángulo $R^* = [a, b] \times [c, d] \subset U$ que contenga W^* , y extendemos g como cero en $R^* - W^*$ como es habitual. Entonces por el Teorema de Fubini,

$$\int_{H(W^*)} g = \int_{[c, d]} \int_{I_v} g(u, v) du dv,$$

donde I_v es el intervalo dado por la imagen de $[a, b]$ via $g_1(u, v)$ con $v \in [c, d]$ fijo, según indica la figura, pues H fija $[c, d]$:



Ahora hacemos la sustitución $u = u_v(x) = t_1(x, v)$. Derivando respecto de x se tiene

$$du = \frac{\partial t_1}{\partial x}(x, v) dx,$$

y cuando x recorre $[a, b]$, u recorre I_v . Entonces

$$\int_{I_v} g(u, v) du = \int_{[a, b]} g(t_1(x, v), v) \frac{\partial t_1}{\partial x}(x, v) dx,$$

con lo cual

$$\int_{H(W^*)} g = \int_{[c, d]} \int_{[a, b]} g(t_1(x, v), v) \frac{\partial t_1}{\partial x}(x, v) dx dv = \int_{W^*} (g \circ H) JH.$$

Con una cuenta análoga se tiene, para cualquier entorno abierto V^* de $H(P)$ suficientemente pequeño,

$$\int_{K(V^*)} f = \int_{V^*} (f \circ K) JK.$$

Luego

$$\int_{T(R^*)} f = \int_{K \circ H(R^*)} f = \int_{H(R^*)} (f \circ K) JK.$$

LLamando $g = (f \circ K) JK$ y usando que $\int_{H(W^*)} g = \int_{W^*} (g \circ H) JH$ se deduce que

$$\int_{H(R^*)} (f \circ K) JK = \int_{W^*} (f \circ K \circ H) JK_H JH.$$

Como $DT(X) = DK_{H(X)}DH(X)$, tomando determinante se tiene $JK_H JH = JT$ (donde con JK_H queremos indicar al Jacobiano de K , evaluado en H), y juntando las últimas dos ecuaciones se tiene

$$\int_{T(R^*)} f = \int_{R^*} (f \circ T) JT.$$

Por último, si $DT(P) \neq I_2$, basta llamar $A = DT(P)$ que es una transformación lineal inversible, y llamando $\bar{T} = A^{-1} \circ T$, se tiene

$$\int_{T(R^*)} f = \int_{A \circ \bar{T}(R^*)} f = \int_{\bar{T}(R^*)} (f \circ A) |\det A|,$$

usando el resultado para transformaciones lineales A . Si aplicamos lo que probamos en el párrafo de arriba a la función \bar{T} (que ahora verifica $D\bar{T}(P) = I_2$), se tiene entonces

$$\int_{\bar{T}(R^*)} (f \circ A) |\det A| = \int_{R^*} (f \circ A \circ A^{-1} \circ T) J\bar{T} |\det A| = \int_{R^*} (f \circ T) JT$$

pues $J\bar{T} = |\det(A^{-1})| |\det T| = |\det A|^{-1} JT$. □

Terminemos esta sección con un ejemplo sutil.

Ejemplo 8.4.3. Se quiere calcular el volumen encerrado bajo el cono $z^2 = x^2 + y^2$ y sobre el piso $z = 0$, pero en el rectángulo $R = [0, 1] \times [0, 1]$ del primer cuadrante. En resumen, se quiere calcular

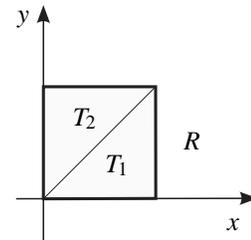
$$\int \int_R \sqrt{x^2 + y^2} dx dy.$$

Presenta ciertas dificultades la integral en coordenadas cartesianas, luego hacemos el cambio de variable a coordenadas polares, es decir $T(r, \theta) = (r \cos \theta, r \sin \theta)$. El integrando queda como

$$(f \circ T) JT = rr = r^2,$$

muy sencillo de integrar.

El problema es el dominio R . Tenemos que identificar el dominio D^* dado en coordenadas polares tal que $T(D^*) = R$. Para ello es conveniente dibujar R y partirlo al medio por la diagonal, como en la figura de la derecha:



Se tienen dos triángulos T_1, T_2 de igual área, y por la forma del cono sabemos que el volumen lo podemos calcular como

$$vol = \int \int_{T_1} f + \int \int_{T_2} f.$$

El triángulo T_1 está determinado por las condiciones $0 \leq y \leq x$, $0 \leq x \leq 1$. En coordenadas polares, se observa que $0 \leq \theta \leq \frac{\pi}{4}$, mientras que para cada θ fijo r se mueve entre cero y la recta vertical $x = 1$. Pero esta recta en coordenadas polares es $r \cos(\theta) = 1$, es decir

$$r = \frac{1}{\cos(\theta)}.$$

Luego debe ser D_1^* el dominio en (r, θ) dado por las condiciones

$$0 \leq r \leq \frac{1}{\cos(\theta)}, \text{ mientras que } 0 \leq \theta \leq \frac{\pi}{4}.$$

Se tiene entonces

$$\frac{1}{2} \text{vol} = \iint_{T_1} f = \iint_{T(D_1^*)} f = \iint_{D_1^*} (f \circ T) JT = \int_0^{\frac{\pi}{4}} \int_0^{1/\cos(\theta)} r^2 dr d\theta.$$

Luego

$$\begin{aligned} \frac{1}{2} \text{vol} &= \frac{1}{3} \int_0^{\frac{\pi}{4}} \frac{1}{\cos^3(\theta)} d\theta = \frac{1}{3} \frac{1}{2} \left(\frac{\sec(r) + \ln(\sec(r) + \tan(r)) \cos^2(r)}{\cos^2(r)} \right) \Big|_0^{\frac{\pi}{4}} \\ &= \frac{1}{3} \frac{1}{2} (\sqrt{2} + \ln(1 + \sqrt{2})), \end{aligned}$$

donde usamos una tabla para calcular la primitiva.

8.4.1. Cambio de variable en \mathbb{R}^3

Para el espacio, se tiene un resultado análogo. En primer lugar, no es difícil probar que si R^* es un paralelogramo sólido en \mathbb{R}^3 , y T es una transformación lineal inversible, entonces $R = T(R^*)$ es otro paralelogramo sólido y además

$$\mu(R) = |\det T| \mu(R^*).$$

Con esto, se tiene un enunciado equivalente al del Teorema 8.3.2, con una demostración análoga:

Teorema 8.4.4. Si $D^* \subset \mathbb{R}^3$ y $T : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ es una transformación lineal inversible, entonces si D^* es medible, $D = T(D^*)$ es medible y además

$$\mu(D) = |\det T| \mu(D^*).$$

Finalmente, usando las mismas ideas que antes, toda función inyectiva $T : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ con diferencial inversible nos permite hacer un cambio de variable que transforma una integral en D^* en una integral en $D = T(D^*)$, de la siguiente manera.

Teorema 8.4.5. Sea $T : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ inyectiva y C^1 . Sea $D^* \subset \mathbb{R}^3$ acotado y $f : D = T(D^*) \rightarrow \mathbb{R}$ integrable. Si DT es inversible en D^* y $JT = |\det DT|$, entonces

$$\int_D f = \int_{D^*} (f \circ T) JT.$$

La demostración usa las mismas ideas que el caso $n = 2$ y la omitimos. El único comentario a tener en cuenta es que ahora hay que descomponer a T como una aplicación que primero mueve dos variables y después una, y usar el resultado que ya tenemos para dos variables. Es decir, es una demostración por inducción.

Para terminar este libro, calculamos integrales en algunos ejemplos que ilustran el poder del teorema combinado con los cambios de variables cilíndricos y esféricos ¿Puede haber mejor manera de cerrar un apunte? Nos gusta creer que, como dijo el gran matemático ruso V. I. Arnold (1937-2010), la matemática es una ciencia experimental, con un laboratorio muy fácil de mantener.

Ejemplo 8.4.6. 1. *Coordenadas cilíndricas.* La transformación T está dada por

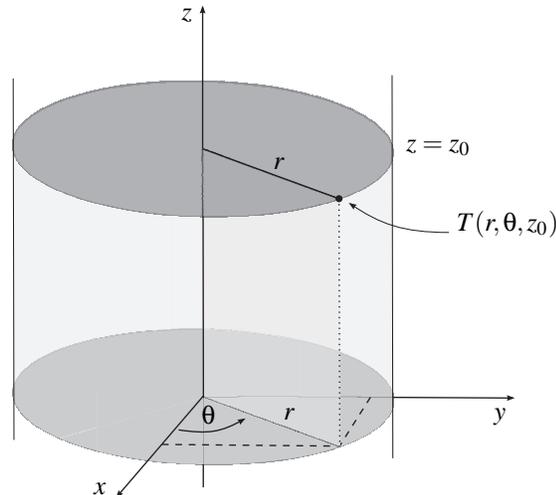
$$T(r, \theta, z) = (r \cos(\theta), r \operatorname{sen}(\theta), z).$$

Esto es $x = r \cos(\theta)$, $y = r \operatorname{sen}(\theta)$, $z = z$.

Es muy conveniente cuando la región D a integrar presenta simetría alrededor de un eje, el “eje de rotación”. En este caso alrededor del eje z , pero con pequeñas modificaciones pueden considerarse otras rectas como eje de rotación. Aquí

$$DT = \begin{pmatrix} \cos(\theta) & -r \operatorname{sen}(\theta) & 0 \\ \operatorname{sen}(\theta) & r \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

luego $JT = |\det T| = r$. Esta transformación manda un ladrillo vertical, en el espacio (r, θ, z) (de lados $r, 2\pi, z_0$) en un cilindro vertical en el espacio (x, y, z) (de radio r y altura z_0) como indica la figura:



Si fijamos la altura $z = z_0$, la transformación $T(r, \theta, z_0)$ recorre una circunferencia horizontal de radio r en la altura del plano $z = z_0$, donde x e y se pueden calcular observando que en la figura, el vector $T(r, \theta, z_0)$ mirado en el piso tiene componentes

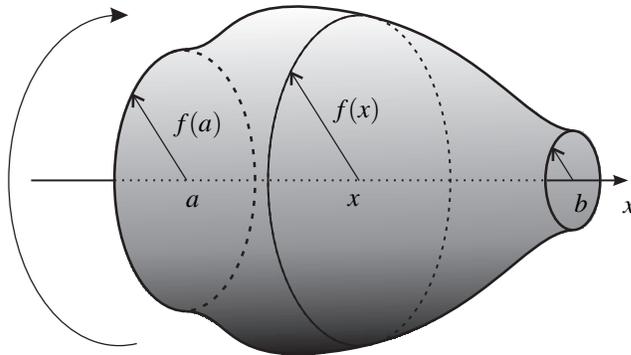
$$x = r \cos(\theta), \quad y = r \operatorname{sen}(\theta).$$

El ángulo θ se mide desde el eje x , en contra del reloj como si mirásemos el plano xy desde arriba. El número $r = \sqrt{x^2 + y^2}$ nos da la distancia del punto (x, y, z) al punto $(0, 0, z)$, es decir, la distancia al eje z .

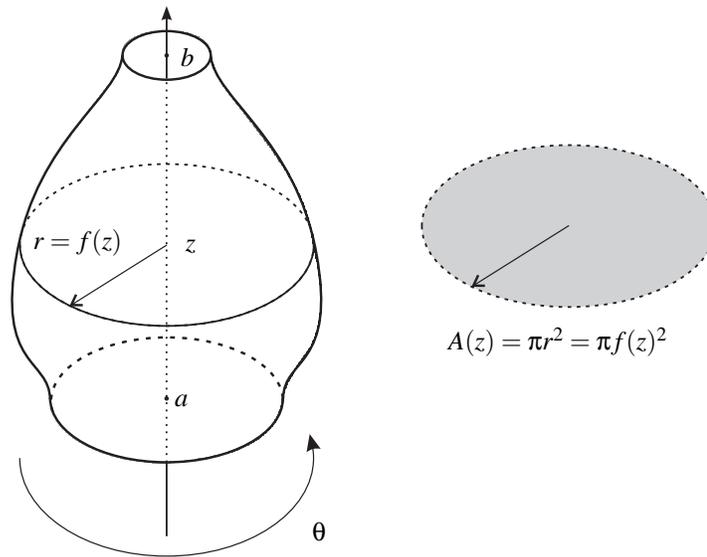
Aquí z está libre, pero θ debe ser un número entre $[0, 2\pi]$ para que sea una función inyectiva (sino estamos pasando dos veces por el mismo ángulo), y r un número positivo pues representa una distancia.

Es decir, si (r, θ, z) viven en $[0, +\infty) \times [0, 2\pi) \times \mathbb{R}$, entonces $T(r, \theta, z)$ recorre todo el espacio en forma inyectiva, con una salvedad: si $r = 0$, y $z = z_0$ está dado, cualquiera sea el ángulo θ no nos moveremos del punto $(0, 0, z_0)$. Esto no es un problema pues $r = 0$ representa la recta del eje z , que es una región que no tienen ningún volumen que pueda afectar el cálculo de una integral.

- a) Se quiere calcular el volumen del **sólido de revolución** que se obtiene al girar el área bajo el gráfico de una función f en el intervalo $[a, b]$. Se supone que f es positiva allí, de manera que se obtiene una figura sólida:



El volumen es independiente de la posición, así que para aprovechar las coordenadas cilíndricas pensamos a f como función de z , en el plano yz , es decir $y = f(z) > 0$ como indica la figura:



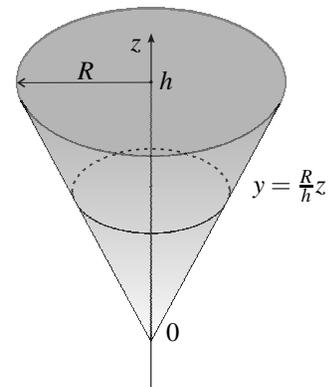
Al hacerla girar alrededor del eje z obtenemos el sólido de revolución. Fijado un z concreto entre a y b , el corte con el sólido está dado por la circunferencia $0 \leq r \leq f(z)$. Luego integrando en coordenadas polares se tiene

$$\begin{aligned} \text{vol} &= \int_0^{2\pi} \int_a^b \int_0^{f(z)} r dr dz d\theta \\ &= 2\pi \int_a^b \frac{1}{2} f^2(z) dz = \int_a^b \pi f^2(z) dz, \end{aligned}$$

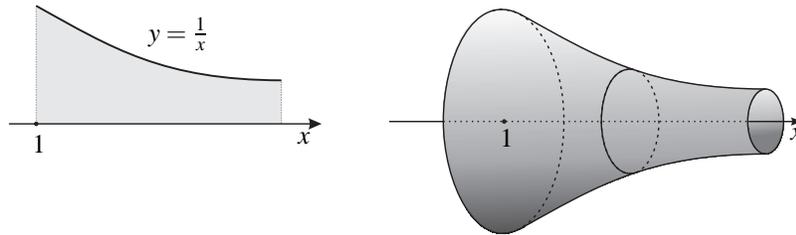
fórmula general que también puede interpretarse como que integramos entre a y b las superficies de los discos de radio $f(z)$, que valen exactamente $\pi f(z)^2$.

- 1) Calculemos el volumen de un cono sólido de base R y altura h . Dibujado con eje en el eje z , el cono se obtiene girando el área bajo la recta $y = \frac{R}{h}z$, con z entre 0 y h . Luego

$$\begin{aligned} \text{vol}(\text{cono}) &= \int_0^h \pi \frac{R^2}{h^2} z^2 dz \\ &= \frac{1}{3} \pi R^2 h = \frac{1}{3} \text{vol}(\text{cilindro}). \end{aligned}$$



- 2) Calculamos el volumen del sólido no acotado que se obtiene al girar el área bajo la gráfica de $y = \frac{1}{x}$, para $x \geq 1$:



Se tiene

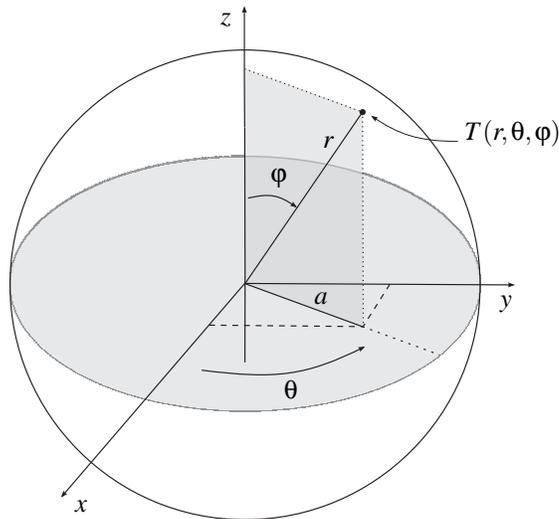
$$\text{vol} = \lim_{r \rightarrow +\infty} \int_1^r \pi \frac{1}{z^2} dz = \pi \lim_{r \rightarrow +\infty} \left. -\frac{1}{z} \right|_1^r = \pi \lim_{r \rightarrow +\infty} \left(-\frac{1}{r} + 1 \right) = \pi.$$

Sorprendentemente, aunque el sólido es no acotado, ¡su volumen es finito! Resulta más sorprendente aún si recordamos que el área bajo la gráfica de $f(x) = \frac{1}{x}$ no es finita pues la integral impropia $\int_1^{+\infty} \frac{1}{x} dx$ **no es convergente**.

2. **Coordenadas esféricas.** En este caso se trata de una transformación que es útil para regiones que presentan simetrías alrededor de un punto del espacio, el caso más sencillo es una esfera. La transformación T está dada por

$$T(r, \theta, \varphi) = (r \cos(\theta) \sin(\varphi), r \sin(\theta) \sin(\varphi), r \cos(\varphi)).$$

Para interpretarla, hay que considerar que un cubo de lado $r, 2\pi, \pi$ se transforma en una esfera de radio r en el espacio (x, y, z) . Para leer estas coordenadas en una esfera, es conveniente entender su representación gráfica: r representa el radio, es decir, la distancia al origen. Debe ser entonces un número positivo. El ángulo θ una vez más se mide desde el eje x en sentido antihorario, es un número entre 0 y 2π . Por último, el ángulo φ se mide desde el eje z en forma vertical, **pero basta tomar un número entre 0 y π** para recorrer todo el espacio, ya que los ángulos mayores se alcanzan tomando θ entre π y 2π .



Se observa que $a = r \sin(\varphi)$ representa la distancia de $(x, y, 0)$ al origen, mientras que $r \cos(\varphi)$ es la altura o coordenada z del punto.

La diferencial de T es la matriz

$$DT = \begin{pmatrix} \cos(\theta) \operatorname{sen}(\varphi) & -r \operatorname{sen}(\theta) \operatorname{sen}(\varphi) & r \cos(\theta) \cos(\varphi) \\ \operatorname{sen}(\theta) \operatorname{sen}(\varphi) & r \cos(\theta) \operatorname{sen}(\varphi) & r \operatorname{sen}(\theta) \cos(\varphi) \\ \cos(\varphi) & 0 & -r \operatorname{sen}(\varphi) \end{pmatrix},$$

luego desarrollando por la última fila se tiene

$$\begin{aligned} JT = |\det T| &= |\cos(\varphi) [-r^2 \operatorname{sen}^2(\theta) \operatorname{sen}(\varphi) \cos(\varphi) - r^2 \cos^2(\theta) \cos(\varphi) \operatorname{sen}(\varphi)] \\ &\quad - r \operatorname{sen}(\varphi) [r \cos^2(\theta) \operatorname{sen}^2(\varphi) + r \operatorname{sen}^2(\theta) \operatorname{sen}^2(\varphi)]| \\ &= |r^2 \cos^2(\varphi) \operatorname{sen}(\varphi) + r^2 \operatorname{sen}^2(\varphi) \operatorname{sen}(\varphi)| = |r^2 \operatorname{sen}(\varphi)|. \end{aligned}$$

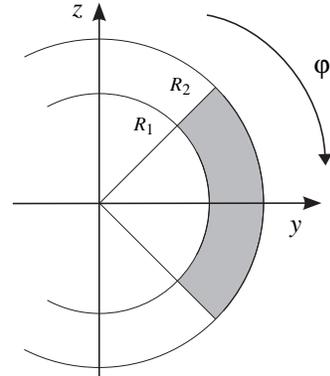
Como en $[0, \pi]$ el seno es positivo, se tiene

$$JT = r^2 \operatorname{sen}(\varphi).$$

a) Para empezar, calculamos el volumen de una esfera de radio R .

$$\begin{aligned} \operatorname{vol}(B_R) &= \int_0^{2\pi} \int_0^\pi \int_0^R r^2 \operatorname{sen}(\varphi) dr d\varphi d\theta \\ &= 2\pi \frac{1}{3} R^3 \int_0^\pi \operatorname{sen}(\varphi) d\varphi = \frac{2}{3} \pi R^3 (-\cos(\varphi)|_0^\pi) = \frac{4}{3} \pi R^3. \end{aligned}$$

b) De acuerdo a las leyes del electromagnetismo, la carga total eléctrica Q_T de un objeto es simplemente la suma de las cargas puntuales. Al igual que en el caso de la masa, en el caso de un objeto sólido D provisto de una densidad de carga ρ , la carga total se consigue integrando ρ en el sólido D . Queremos calcular la carga total de una cáscara esférica según las restricciones $0 < R_1 < r < R_2$, $|z| \leq \sqrt{x^2 + y^2}$ cuya densidad de carga es $\rho(x, y, z) = \sqrt{x^2 + y^2} / (x^2 + y^2 + z^2)$. La segunda ecuación es la de un cono, luego en el corte con el plano yz veremos la región anular encerrada por las rectas $z = \pm y$.



Entonces, como $\rho(r, \theta, \varphi) = \frac{r \operatorname{sen} \varphi}{r^2} = \frac{\operatorname{sen} \varphi}{r}$,

$$\begin{aligned} Q_T &= \int_0^{2\pi} \int_{\pi/4}^{3\pi/4} \int_{R_1}^{R_2} \frac{\operatorname{sen} \varphi}{r} r^2 \operatorname{sen} \varphi dr d\varphi d\theta \\ &= 2\pi \int_{\pi/4}^{3\pi/4} \operatorname{sen}^2 \varphi d\varphi \frac{1}{2} (R_2^2 - R_1^2) = \pi (R_2^2 - R_1^2) \left(\frac{\pi}{4} + \frac{1}{2} \right). \end{aligned}$$

Ocurre aquí un fenómeno curioso. Observemos que la densidad de carga tiende a infinito cuando $r \rightarrow 0$. Sin embargo las esferas de radio pequeño aportan poca carga total, con lo cual, haciendo

tender $R_1 \rightarrow 0$, se tiene

$$Q_T = \pi R_2^2 \left(\frac{\pi}{4} + \frac{1}{2} \right)$$

que nos dice que la carga total es finita aún en el caso de que la región toque el origen.

En el juicio final, cuando me pidan justificar mi vida, me levantaré orgulloso y diré “Fui uno de los senescales de la matemática, y no sufrió ningún daño bajo mi cuidado”.

U. DUDLEY

- [1] G. Birkhoff, S. MacLane, *Algebra Moderna*. Ed. Vicens Vives, Barcelona, 1963.
- [2] R. Courant, J. Fritz, *Introducción al cálculo y el análisis matemático*. Vol. 1 y 2, Ed. Limusa-Wiley, Méjico, 1998.
- [3] S. Lang, *Introducción al álgebra lineal*. Ed. Addison-Wesley Iberoamericana, Argentina, 1990.
- [4] J.E. Marsden, A.J. Tromba, *Cálculo vectorial*. Ed. Addison-Wesley Iberoamericana, Argentina, 1991.
- [5] R. J. Noriega, *Cálculo diferencial e integral*. Ed. Docencia, Buenos Aires, 1987.
- [6] J. Rey Pastor, P. Pi Calleja, C. Trejo, *Análisis Matemático*. Vol. I. Ed. Kapelusz, 1973.
- [7] J. Rey Pastor, P. Pi Calleja, C. Trejo, *Análisis Matemático*. Vol. II. Ed. Kapelusz, 1973.
- [8] M. Spivak, *Cálculo infinitesimal*. Ed. Reverté, Barcelona, 1991.
- [9] M. Spivak, *Cálculo en variedades*. Ed. Reverté, Barcelona, 1970.