

Apuntes de Cálculo y métodos numéricos, Parte 2

Edoardo Provenzi y Gloria Haro

Prólogo

Estos apuntes quieren ser una guía al curso de cálculo del segundo trimestre y un soporte a las clases magistrales.

En esta segunda parte del curso de Cálculo se desarrolla el cálculo diferencial e integral para funciones $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$, con $n, m \geq 1$ además de los métodos numéricos del descenso por gradiente y la generalización del método de Newton a funciones de más de una variable, que puede ser utilizado para aproximar la solución de sistemas no lineales.

La presencia de varias dimensiones espaciales permite desarrollar de forma más completa la teoría de la optimización, considerando problemas interesantes y donde pueden aparecer vínculos.

Como para los apuntes del primer trimestre, también en estas notas intentaremos conectar los conceptos que iremos introduciendo con aplicaciones prácticas, señalando en particular los conceptos físicos que el estudiante podrá examinar con más detalle en el curso de ondas y electromagnetismo.

Los autores.

Índice general

1. Funciones de varias variables reales	4
1.1. Motivaciones para estudiar funciones de varias variables reales	4
1.2. Notación abstracta y funciones componentes	6
1.2.1. Dualidad interpretativa de \mathbb{R}^n	7
1.3. Dominios de funciones de varias variables	8
2. Curvas, superficies y campos vectoriales	10
2.1. Curvas en forma paramétrica en \mathbb{R}^m	10
2.2. Curvas en forma cartesiana o implícita. El ejemplo de las cónicas	14
2.3. Curvas suaves, vector y recta tangente	16
2.4. Longitud de una curva suave	19
2.5. Integrales de línea (o integrales curvilíneas)	21
2.6. Superficies y curvas de nivel	22
2.6.1. Superficies en forma implícita. Las cuádricas	25
2.7. Superficies en forma paramétrica	26
2.8. Campos vectoriales	27
3. Cálculo diferencial para funciones de varias variables	30
3.1. Límites y continuidad	30
3.2. Derivadas direccionales y parciales	31
3.3. El gradiente y su significado geométrico	35
3.3.1. Consecuencias computacionales de la fórmula (3.1) . .	35
3.3.2. Consecuencias teóricas de la fórmula (3.1)	36
3.4. La matriz Jacobiana	38
3.4.1. Derivación de una función compuesta en varias variables	39
3.5. Diferenciabilidad de una función de varias variables reales y plano tangente a una superficie	42
3.5.1. Vector normal a una superficie en un punto	46
3.6. Derivadas de orden superior y matriz Hessiana	50

3.7. La fórmula de Taylor para funciones de varias variables . . .	52
4. Optimización libre y vinculada en \mathbb{R}^n	54
4.1. Extremos libres	54
4.2. Extremos vinculados	59
5. Métodos numéricos iterativos para funciones de varias variables	66
5.1. Método de Newton para funciones de n variables reales . . .	66
5.2. Método del descenso/ascenso por gradiente	73
6. Cálculo integral para funciones de 2 y 3 variables: integrales dobles y triples	78
6.1. Integrales dobles	78
6.1.1. Integrales dobles en rectángulos	78
6.1.2. Integrales dobles en dominios simples o regulares . . .	80
6.1.3. Cambio de variable en las integrales dobles	86
6.2. Integrales triples	89
7. Integrales de superficie y de campos vectoriales, teorema de la divergencia y del rotacional	91
7.1. Integrales de superficie	91
7.2. Los operadores del cálculo vectorial: gradiente, divergencia, rotacional, laplaciano	97
7.3. El flujo de un campo vectorial a través de una superficie . . .	100
7.4. El teorema de la divergencia	101
7.4.1. Significado del operador divergencia e interpretación física del teorema de la divergencia	102
7.5. Trabajo de un campo vectorial y teorema del rotacional . . .	103

Capítulo 1

Funciones de varias variables reales

1.1. Motivaciones para estudiar funciones de varias variables reales

En el capítulo 1 de las notas del curso de Cálculo del primer trimestre hemos introducido la definición general de función, que recordamos aquí debajo.

Def. general de función (Dirichlet-Lobachevsky): sean X e Y dos conjuntos cualquiera. Una función f entre X e Y es una *ley que asocia a cada elemento $x \in X$ un único elemento $y = f(x) \in Y$* . Se suele escribir

$$\begin{aligned} f : X &\longrightarrow Y \\ x &\longmapsto y = f(x), \end{aligned}$$

o bien $f : X \longrightarrow Y, x \in X \longmapsto y = f(x) \in Y$.

$y = f(x)$ se llama **imagen de x** a través de f . x se llama **pre-imagen** o **anti-imagen** de y vía f . X se llama **dominio** de f y el conjunto de todas las imágenes de elementos de X a través de f se llama **rango** o **imagen de f** y se escribe $f(X)$ (o $\text{Im}f$ en algunos libros). El **gráfico** de f es el subconjunto de $X \times Y$ definido por

$$\text{gráfico}(f) = \{(x, y) : x \in X, y = f(x)\},$$

o sea, el conjunto de las parejas ordenadas donde el primer elemento es un

punto del dominio de f y el segundo es la imagen de ese punto a través de f .

Si f es una función real de variable real, o sea si $X, Y \subseteq \mathbb{R}$, su gráfico es una **curva en el plano real** \mathbb{R}^2 .

Ya hemos visto que el concepto de función traduce matemáticamente la idea de transición de un dato de entrada x a un dato de salida $y = f(x)$ que, en general, pueden pertenecer a dos conjuntos diferentes X e Y , respectivamente.

En la primera parte del curso hemos examinado las funciones reales de variable real, por las cuales X e Y son dos subconjuntos, propios o impropios, de \mathbb{R} . Sin embargo, tanto desde un punto de vista teórico, como para las aplicaciones prácticas, resulta interesante considerar funciones cuyo dominio y/o codominio pertenecen a espacios con dimensiones superiores a 1.

Ejemplos:

1. Dado un triángulo cualquiera T en el plano real \mathbb{R}^2 , sabemos que su área depende de la base b y de la altura h según la relación: $\text{Área}(T) = b \cdot h/2$. La función que, a partir de b y h , nos devuelve el área de T se escribe como sigue:

$$\begin{aligned} \text{Área} : \mathbb{R}^+ \times \mathbb{R}^+ &\longrightarrow \mathbb{R}^+ \\ (b, h) &\longmapsto \text{Área}(b, h) = b \cdot h/2. \end{aligned}$$

b y h se llaman **grados de libertad** de la función Área.

2. La función que asocia la posición de un cuerpo en el espacio tridimensional con su temperatura es una función de \mathbb{R}^3 a \mathbb{R} porque hacen falta 3 coordenadas reales (x, y, z) para determinar la posición del cuerpo y un solo número real para describir su temperatura T . La función correspondiente será:

$$\begin{aligned} T : \mathbb{R}^3 &\longrightarrow \mathbb{R} \\ (x, y, z) &\longmapsto T(x, y, z). \end{aligned}$$

Los grados de libertad de T son 3: x , y y z .

3. Si la temperatura del cuerpo varía en el tiempo, hará falta considerar una función de \mathbb{R}^4 a \mathbb{R} , porque en este caso hace falta especificar las 3 coordenadas reales de antes (x, y, z) más la coordenada temporal t para describir su temperatura:

$$\begin{aligned} T : \mathbb{R}^4 &\longrightarrow \mathbb{R} \\ (x, y, z, t) &\longmapsto T(x, y, z, t). \end{aligned}$$

En este caso los grados de libertad de T son 4: x, y, z y t .

4. La función que asocia un punto del espacio-tiempo cuatridimensional donde está una partícula con su velocidad de desplazamiento es una función de \mathbb{R}^4 a \mathbb{R}^3 . En este caso la función velocidad se escribirá como sigue:

$$\begin{aligned} \vec{v}: \quad \mathbb{R}^4 &\longrightarrow \mathbb{R}^3 \\ (x, y, z, t) &\longmapsto \vec{v}(x, y, z). \end{aligned}$$

También aquí los grados de libertad de T son 4: x, y, z y t .

5. La función que asocia cada vector $\vec{x} \in \mathbb{R}^n$ con su norma (que, como el estudiante sabe del curso de álgebra, es la generalización del concepto de ‘módulo’ de un vector en el plano o en el espacio) es una función de \mathbb{R}^n a \mathbb{R}_0^+ :

$$\begin{aligned} \|\cdot\|: \quad \mathbb{R}^n &\longrightarrow \mathbb{R}_0^+ \\ \vec{x} = (x_1, \dots, x_n) &\longmapsto \|\vec{x}\| = \sqrt{x_1^2 + \dots + x_n^2}. \end{aligned}$$

Los grados de libertad de $\|\cdot\|$ son las n componentes de \vec{x} : x_1, \dots, x_n .

1.2. Notación abstracta y funciones componentes

En general, en los problemas prácticos, aparecen funciones que dependen de $n \geq 1$ *grados de libertad* representados por n variables que pertenecen a determinados conjuntos. En este curso supondremos que estos grados de libertad son n variables reales, que se pueden considerar por separado: x_1, \dots, x_n , o bien como componentes de un vector de \mathbb{R}^n : $\vec{x} = (x_1, \dots, x_n)$.

Estaremos interesados en las funciones de estas n variables a valores reales escalares o vectoriales, es decir, respectivamente:

$$\begin{aligned} f: \quad \mathbb{R}^n &\longrightarrow \mathbb{R} \\ (x_1, \dots, x_n) &\longmapsto y = f(\vec{x}). \end{aligned}$$

o bien,

$$\begin{aligned} \vec{f}: \quad \mathbb{R}^n &\longrightarrow \mathbb{R}^m \\ \vec{x} = (x_1, \dots, x_n) &\longmapsto \vec{y} = \vec{f}(\vec{x}). \end{aligned}$$

Observamos que, en este último caso, $\vec{y} = \vec{f}(\vec{x})$ es un vector de \mathbb{R}^m , entonces se puede escribir como $\vec{y} = (y_1, \dots, y_m)$. Definimos ahora m funciones

a valores reales escalares como sigue:

$$\begin{cases} f_1(\vec{x}) = y_1 \\ f_2(\vec{x}) = y_2 \\ \vdots \\ f_m(\vec{x}) = y_m, \end{cases}$$

es evidente que las escrituras $\vec{f}(\vec{x}) = (f_1(\vec{x}), \dots, f_m(\vec{x}))$ son equivalentes, por eso se suele escribir:

$$\boxed{\vec{f} \equiv (f_1, \dots, f_m)},$$

notación que hay que interpretar según la igualdad antecedente.

Las funciones

$$\begin{aligned} f_j : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ \vec{x} &\longmapsto f_j(\vec{x}) = y_j \end{aligned}$$

se llaman **funciones componentes** de la función a valores vectoriales \vec{f} . Obsérvese que *las funciones componentes son funciones a valores reales escalares*, por eso hace falta que sean m para determinar unívocamente \vec{f} .

Ejemplo. Las funciones componentes de la función de dos variables reales a valores vectoriales definida por $\vec{f} : \mathbb{R}^2 \rightarrow \mathbb{R}^3$:

$$\vec{f}(x, y) = (2x, y^2, \log(\sqrt{xy}))$$

son $f_i : \mathbb{R}^2 \rightarrow \mathbb{R}$, $f_1(x, y) = 2x$, $f_2(x, y) = y^2$, $f_3(x, y) = \log(\sqrt{xy})$, entonces $\vec{f} = (f_1, f_2, f_3)$.

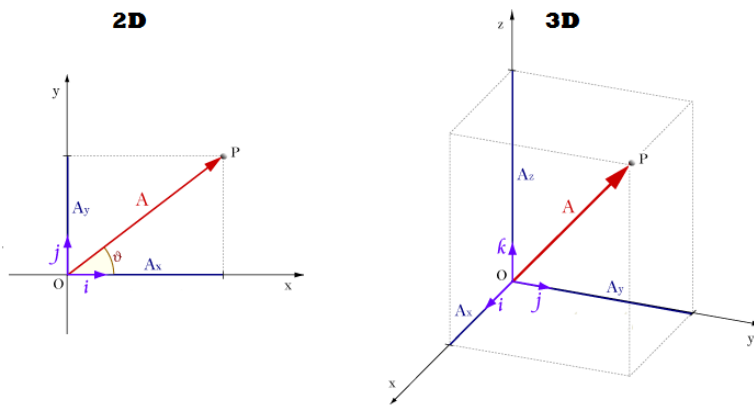
En el capítulo 1 de los apuntes de Cálculo del primer trimestre hemos definido la composición y las operaciones algebraicas entre funciones abstractas $f : X \rightarrow Y$, con lo cual no hace falta dar una nueva definición para definir la composición y las operaciones algebraicas entre funciones de varias variables reales vectoriales o escalares, sino simplemente elegir $X = \mathbb{R}^n$ e $Y = \mathbb{R}$ o bien $Y = \mathbb{R}^m$, $n, m \geq 1$. Proporcionaremos unos ejemplos en las clases de prácticas y seminarios.

1.2.1. Dualidad interpretativa de \mathbb{R}^n

El espacio \mathbb{R}^n se puede interpretar de dos formas diferentes, que es importante remarcar desde el comienzo del curso:

1. los elementos de \mathbb{R}^n se pueden pensar como **puntos** que viven en un espacio n -dimensional;
2. los elementos de \mathbb{R}^n se pueden pensar como **vectores** con n componentes.

Si $n = 2, 3$ podemos visualizar esta dualidad interpretativa en la figura aquí debajo:



Se observa que el punto P determina y está determinado por un vector que lo conecta con el origen de los ejes cartesianos ortogonales.

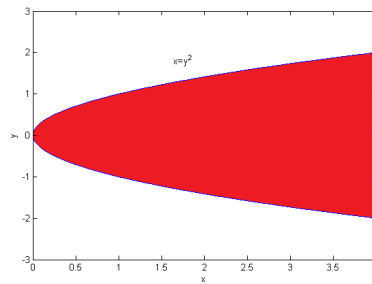
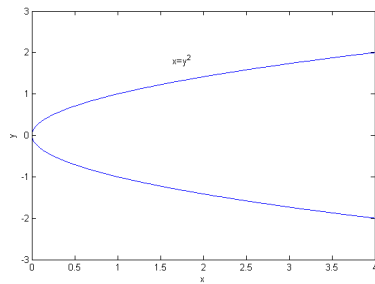
1.3. Dominios de funciones de varias variables

Destacamos, por su importancia, una observación sobre el dominio de funciones de n variables reales: esta tarea puede resultar complicada si n es grande. Sin embargo, en el caso de dos variables, a veces es posible determinar tanto analíticamente como geoméricamente el dominio de una función, como muestra el siguiente ejemplo.

Determinar el dominio D de $f : D \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$, $f(x, y) = \frac{y}{\sqrt{x-y^2}}$.

Dado que en la expresión analítica de f aparece una fracción con una raíz cuadrada en el denominador, para determinar el dominio de f hace falta simplemente imponer que el radicando sea estrictamente mayor que cero: $x - y^2 > 0$, es decir, $x > y^2$. La región del plano real correspondiente a esta condición está dada por todos los puntos que tienen la coordenada x

estrictamente mayor que la coordenada y elevada al cuadrado. La ecuación $x = y^2$ define una parábola relativa al eje x y con vértice en el origen (figura debajo a la izquierda). Resulta claro que los puntos (x, y) con $x > y^2$ están contenidos en la región (estrictamente) interna a la parábola, como muestra la figura debajo a la derecha (el dominio es la región en rojo). Como comprobación adicional, tenemos que $y^2 \geq 0 \forall y \in \mathbb{R}$, entonces el dominio de f será una porción del semiplano dado por la unión del primer y cuarto cuadrante como remarca la figura.



Capítulo 2

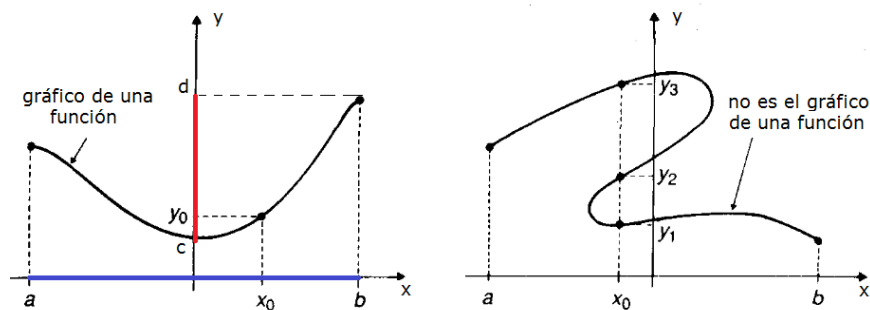
Curvas, superficies y campos vectoriales

En este capítulo examinaremos el significado geométrico y físico de unas funciones particulares que involucran varias variables en el dominio y/o en el codominio.

2.1. Curvas en forma paramétrica en \mathbb{R}^m

Comenzamos el análisis de las curvas considerando curvas planas, es decir, curvas contenidas en \mathbb{R}^2 , después generalizaremos la discusión al caso de curvas contenidas en \mathbb{R}^m .

Como hemos visto en el primer capítulo de los apuntes de teoría de la primera parte de curso, el gráfico de una función $f : D \subset \mathbb{R} \rightarrow \mathbb{R}$, o sea el conjunto $\{(x, f(x)), x \in D\}$ es una curva (plana) en \mathbb{R}^2 . Sin embargo, existen curvas planas que no pueden ser el gráfico de una función, como muestra la figura aquí debajo.



Por supuesto, una partícula puede moverse tanto a lo largo de la curva de izquierda (que es el gráfico de una función), como a lo largo de la curva de derecha (que no lo es), por eso, si queremos calcular la velocidad de desplazamiento de la partícula o la longitud de su recorrido, surge el problema de encontrar una caracterización más general del concepto de curva en el plano. Esta generalización no debe entrar en conflicto con la interpretación de curva como gráfico de una función cuando se de este caso.

Comencemos el análisis de este problema observando que una curva plana es un objeto que vive en \mathbb{R}^2 , por eso no podemos identificarla ni con una función real de variable real, ni con su dominio o codominio: las dimensiones no serían coherentes! Sin embargo, el gráfico de una función real de variable real es un subconjunto de \mathbb{R}^2 , por eso podemos identificarlo con una curva plana.

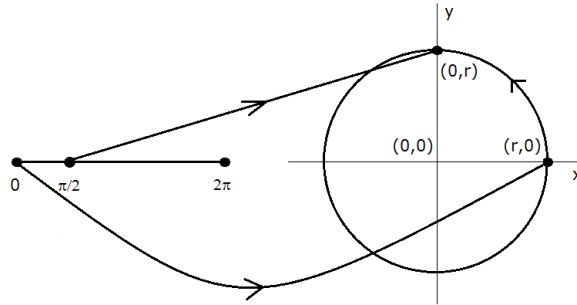
Ahora, si en vez de atarnos a la situación muy limitante de las funciones reales de variable real, consideramos funciones más generales, no es difícil intuir que podemos obtener un objeto bidimensional considerando una función genérica de variable real que toma valores en \mathbb{R}^2 : si es cierto que el gráfico de esta función está contenido en \mathbb{R}^3 (y entonces no puede representar una curva plana), el codominio de esta función vive en \mathbb{R}^2 y podría, por lo menos a nivel de coherencia dimensional, representar una curva plana.

Para saber si esta intuición es correcta, examinamos la posibilidad de incorporar la interpretación de curva plana como gráfico de una función real de variable real en esta nueva visión. No es difícil reconocer que, dada la función $f : D \subseteq \mathbb{R} \rightarrow \mathbb{R}$, $x \mapsto f(x)$, su gráfico coincide con el codominio de la función $F : D \rightarrow D \times \mathbb{R}$, definida así: $x \mapsto F(x) = (x, f(x))$. Tanto el gráfico de f como el codominio de F coinciden con el subconjunto de \mathbb{R}^2 dado por $\{(x, f(x)), x \in D\}$.

El último ‘test’ que tenemos que hacer para saber si estamos yendo hacia la generalización correcta del concepto de curva plana es averiguar si, con esta nueva interpretación, es posible describir curvas que no son gráficos de funciones reales de variable real. Para ello es suficiente proporcionar uno de los ejemplos más sencillos: la *circunferencia de radio r y centro en $(0, 0)$ recorrida en sentido antihorario* puede ser interpretada como la imagen de la función

$$\begin{aligned} \vec{c}_r : [0, 2\pi] &\longrightarrow \mathbb{R}^2 \\ \alpha &\longmapsto \vec{c}_r(\alpha) = (r \cos \alpha, r \sin \alpha), \end{aligned}$$

como se puede averiguar en la figura aquí debajo.



Todas estas consideraciones motivan la siguiente definición.

Def. Una **curva plana \mathcal{C} en forma paramétrica** es la imagen de una función continua de una variable real que toma valores en \mathbb{R}^2 :

$$\begin{aligned} \vec{c}: [a, b] \subset \mathbb{R} &\longrightarrow \mathbb{R}^2 \\ t &\longmapsto \vec{c}(t) = (c_1(t), c_2(t)), \end{aligned}$$

- $\mathcal{C} = \{\vec{c}(t) \in \mathbb{R}^2, t \in [a, b]\}$ (conjunto de todos los puntos imagen de \vec{c});
- la función \vec{c} se llama *parametrización* de la curva \mathcal{C} ;
- t se llama *parámetro* de la \mathcal{C} ;
- $[a, b]$ se llama *intervalo de parametrización* de \mathcal{C} .
- $\vec{c}(a)$: **punto inicial** de \mathcal{C} , $\vec{c}(b)$: **punto final** de \mathcal{C} .

La definición se puede extender de forma inmediata a una **curva en forma paramétrica en \mathbb{R}^m** , sustituyendo simplemente \mathbb{R}^2 por \mathbb{R}^m , en este caso la función \vec{c} se escribe así

$$\begin{aligned} \vec{c}: [a, b] \subset \mathbb{R} &\longrightarrow \mathbb{R}^m \\ t &\longmapsto \vec{c}(t) = (c_1(t), \dots, c_m(t)). \end{aligned}$$

Se suelen denotar las funciones componentes de \vec{c} como sigue:

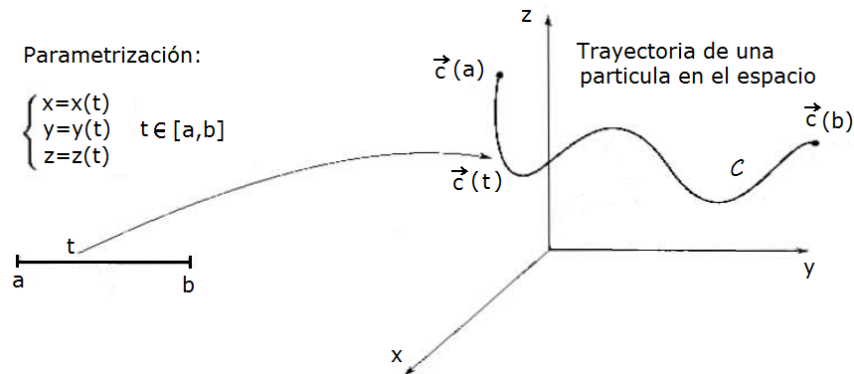
$$\begin{aligned} c_i: [a, b] \subset \mathbb{R} &\longrightarrow \mathbb{R} \\ t &\longmapsto c_i(t) = x_i(t), \end{aligned}$$

para cada $i = 1, \dots, m$. Como consecuencia, en los libros una curva en \mathbb{R}^m se encuentra escrita muchas veces con la notación siguiente:

$$\vec{c}(t) = (x_1(t), \dots, x_m(t)), \quad t \in [a, b].$$

El hecho de que aparezca un único parámetro t implica que una **curva es un objeto matemático con un solo grado de libertad** (se dice que su dimensión ‘topológica’ es 1), independientemente del espacio en el cual vive.

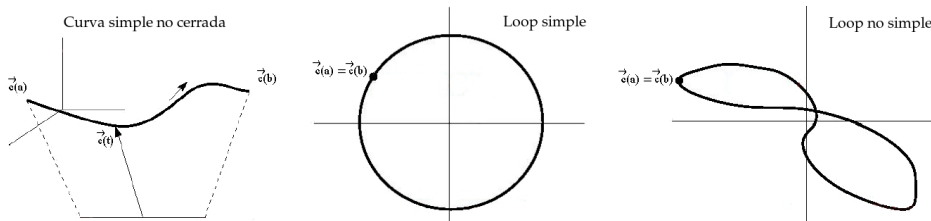
Interpretación física de una curva en forma paramétrica: como se puede ver en la figura aquí debajo, una curva en forma paramétrica en \mathbb{R}^3 coincide con la **trayectoria de una partícula que se mueve en el espacio** mientras el tiempo avanza.



Una *analogía* que puede ayudar es la siguiente: si pensamos en la curva como la trayectoria espacial de un tren, $b-a$ es el tiempo que el tren necesita para ir del punto inicial $\vec{c}(a)$ al punto final $\vec{c}(b)$ y \vec{c} es la tabla que dice dónde estará el tren en cada instante $t \in [a, b]$.

Nomenclatura: una curva se dice

- **simple** si \vec{c} es inyectiva, es decir, si ningún punto se recorre dos veces;
- **cerrada** o **loop** si $\vec{c}(a) = \vec{c}(b)$, es decir, el punto inicial y final coinciden. Un loop es simple si, excepto en los puntos extremos, es una curva simple.



2.2. Curvas en forma cartesiana o implícita. El ejemplo de las cónicas

Es importante remarcar que el adjetivo ‘paramétrica’ que hemos usado en la sección anterior sirve para distinguir esa forma de describir una curva de la representación **implícita** o **cartesiana**, en la cual una curva plana se considera como el lugar geométrico dado por todos los puntos del plano (x, y) que cumplen con la ecuación $f(x, y) = 0$ que viene dada por una cierta función $f : \mathbb{R}^2 \rightarrow \mathbb{R}$. Si f está formada únicamente por las operaciones de suma, diferencia, producto, división, potencia y raíz, la curva se llama **algebraica**, si no se llama **trascendente**.

Ejemplos:

- Sea $f(x, y) = x^2 + y^2 - r^2$, la ecuación $f(x, y) = 0$ corresponde a $x^2 + y^2 = r^2$, que es la ecuación cartesiana de una circunferencia con centro en $(0, 0)$ y radio $r > 0$;
- Sea $f(x, y) = ax + by + c$, la ecuación $f(x, y) = 0$ corresponde a $ax + by + c = 0$, que es la ecuación cartesiana de una recta con pendiente $-a/b$, $b \neq 0$.

Generalizando, una curva en forma cartesiana o implícita en \mathbb{R}^m está dada por los puntos de \mathbb{R}^m que satisfacen la ecuación $f(x_1, \dots, x_m) = 0$ dada por una cierta función $f : \mathbb{R}^m \rightarrow \mathbb{R}$.

Queremos remarcar, por su importancia, el caso en el que $m = 2$ y f es un polinomio de grado 2 en las variables x, y . En este caso tenemos 4 curvas notables que se llaman **cónicas** ya que, geoméricamente, representan todas las posibles curvas que se pueden obtener cortando un cono doble de base circular con un plano.

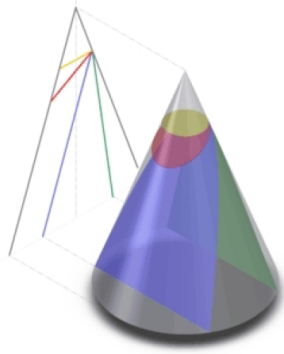
Las ecuaciones que definen las cónicas son las siguientes:

$$\left\{ \begin{array}{ll} x^2 + y^2 = r^2 & \text{Circunferencia con centro en } (0, 0) \text{ y radio } r \\ \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 & \text{Elipse con centro en } (0, 0), \text{ semieje horizontal } a \text{ y semieje vertical } b \\ y^2 = 2px & \text{Parábola} \\ \frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 & \text{Hipérbola.} \end{array} \right.$$

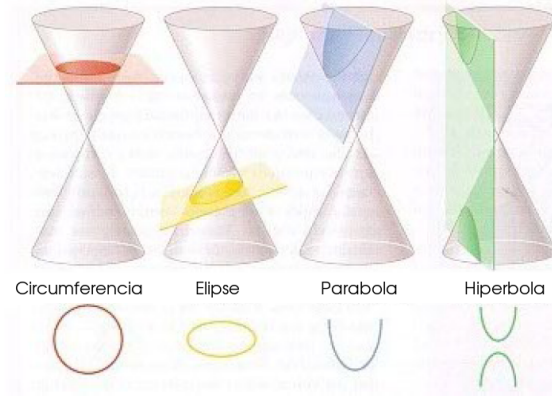
En la figura siguiente podemos visualizar las cónicas.

α : ángulo entre la vertical y la directriz del cono

β : ángulo entre el eje vertical y el plano que corta el cono



a,b,p: parametros reales



Circunferencia

Elipse

Parabola

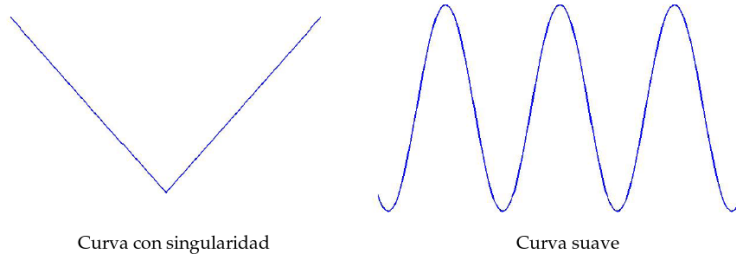
Hiperbola



$\beta = 90^\circ$	$\beta > \alpha$	$\beta = \alpha$	$\beta < \alpha$
$x^2 + y^2 = r^2$	$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$	$y^2 = 2px$	$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$

2.3. Curvas suaves, vector y recta tangente

Una curva puede recorrer una trayectoria ‘suave’ (smooth en inglés) o bien presentar ‘singularidades’ en su trayectoria como en la figura aquí debajo.



No debería resultar sorprendente que la caracterización matemática de la suavidad de una curva se hace a través del concepto de derivabilidad, dado que sabemos que la derivada proporciona una medida de la rapidez de variación de una función.

Def. Decimos que la curva con parametrización $\vec{c} \equiv (x_1, \dots, x_m)$ es **suave** (o **smooth**) si todas sus funciones componentes (que son funciones reales de variable real) son derivables en el intervalo de parametrización de la curva misma.

Ejemplos:

1. La circunferencia definida por $\vec{c}_r(t) = (r \cos t, r \sin t)$, $t \in [0, 2\pi]$, $r > 0$, es una curva suave porque sus funciones componentes $x(t) = r \cos t$ y $y(t) = r \sin t$ son derivables para cada $t \in [0, 2\pi]$;
2. La recta en \mathbb{R}^2 que pasa por el punto $\vec{x}_0 = (x_0, y_0)$ con dirección dada por el vector unitario $\vec{v} = (v_x, v_y)$, $\|\vec{v}\| = 1$, se parametriza con $\vec{c}_{\vec{x}, \vec{v}}(t) = (x_0 + tv_x, y_0 + tv_y)$, $t \in \mathbb{R}$, es una curva suave porque sus funciones componentes son lineales en t , entonces derivables;
3. La curva que proporciona el gráfico del valor absoluto en el intervalo $[-1, 1]$, es decir $\vec{c}(t) = (t, |t|)$, no es suave porque la segunda función componente, $y(t) = |t|$, no es derivable en $t = 0$, dado que el valor absoluto tiene un punto anguloso en el origen.

A partir de ahora consideraremos solo curvas suaves y definiremos los conceptos de vector y recta tangente, velocidad y longitud de una curva suave.

Def. Dada la curva suave \mathcal{C} con parametrización

$$\begin{aligned} \vec{c}: [a, b] &\longrightarrow \mathbb{R}^m \\ t &\longmapsto \vec{c}(t) = (x_1(t), \dots, x_m(t)), \end{aligned}$$

definimos

- El **vector tangente** a \mathcal{C} o **velocidad** en el punto $\vec{c}(t_0)$, $t_0 \in [a, b]$ como el vector de \mathbb{R}^m cuyas componentes son las derivadas de las funciones componentes de \vec{c} calculadas en $t = t_0$:

$$\vec{v}(t_0) = \vec{c}'(t_0) = (x_1'(t_0), \dots, x_m'(t_0));$$

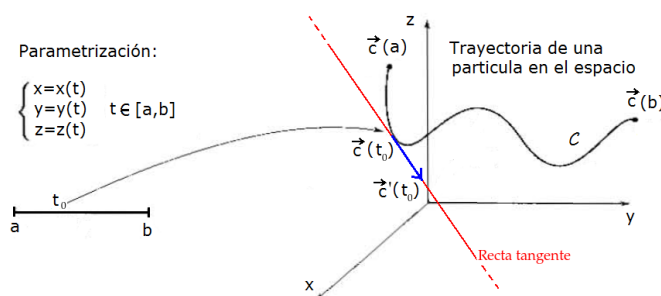
La norma del vector velocidad $\vec{c}'(t_0)$

$$\|\vec{c}'(t_0)\| = \sqrt{x_1'(t_0)^2 + \dots + x_m'(t_0)^2},$$

representa, físicamente, la rapidez de desplazamiento en el instante t_0 de una partícula hipotética que se mueve a lo largo de la curva;

- Si se sustituye la derivada primera por la derivada segunda se obtiene la **aceleración** (dejamos como ejercicio la escritura de las ecuaciones correspondientes);
- La **recta tangente** a la curva \mathcal{C} en el punto $\vec{c}(t_0)$ es el lugar geométrico en \mathbb{R}^m definido como sigue:

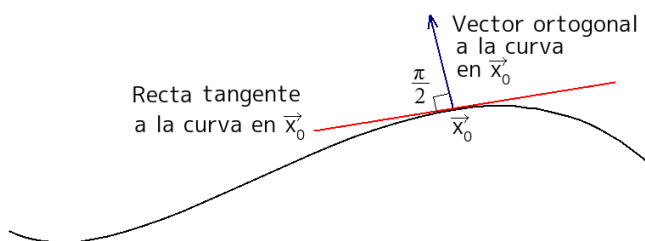
$$\vec{r}(s) = \{\vec{c}(t_0) + s\vec{c}'(t_0), s \in \mathbb{R}\}.$$



Hay que observar que $\vec{r}(s)$ es un lugar geométrico ilimitado: si $s = 0$, $\vec{r}(0)$ se reduce al punto en \mathbb{R}^m dado por $\vec{c}(t_0)$, si $s \neq 0$, nos desplazamos de $\vec{c}(t_0)$ a lo largo de la dirección dada por el vector tangente $\vec{v}(t_0)$. Podemos definir también el concepto de vector ortogonal a una curva en un punto.

Def.: Sea \mathcal{C} una curva suave en \mathbb{R}^m con parametrización $\vec{c}: [a, b] \rightarrow \mathbb{R}^m$ tal que $\vec{c}(t_0) = \vec{x}_0$, $t_0 \in [a, b]$ y $\vec{x}_0 \in \mathcal{C}$. Decimos que un **vector** \vec{v} en \mathbb{R}^m es **ortogonal a la curva** \mathcal{C} en \vec{x}_0 si es ortogonal al vector tangente $\vec{c}'(t_0)$.

La figura siguiente muestra la ortogonalidad entre un vector y una curva.



Utilizaremos este concepto en el capítulo 3. Ahora vamos a ver un ejemplo de cálculo de vector tangente.

1. Dada la circunferencia unitaria en el plano real $\vec{c}(t) = (\cos t, \sin t)$, $t \in [0, 2\pi]$, calcular el vector tangente y la intensidad de la velocidad de recorrido en los instantes $t_0 = 0$, $t_1 = \pi/4$, $t_2 = \pi/2$. Antes de todo tenemos que calcular el vector tangente $\vec{v}(t)$ en un instante genérico $t \in [0, 2\pi]$: $\vec{v}(t) = (-\sin t, \cos t)$.

- $\vec{v}(0) = (-\sin 0, \cos 0) = (0, 1)$;
- $\vec{v}(\pi/4) = (-\sin \pi/4, \cos \pi/4) = (-\sqrt{2}/2, \sqrt{2}/2)$;
- $\vec{v}(\pi/2) = (-\sin \pi/2, \cos \pi/2) = (-1, 0)$.

Observemos que la norma de todos estos vectores vale 1, lo cual significa que si una partícula recorre la circunferencia parametrizada como antes, su velocidad se mantiene constante;

2. Dada la curva parametrizada como sigue $\vec{c}(t) = (t, \cos t)$, $t \in [0, 2\pi]$, calcular la recta tangente en $t_0 = \pi$.

Necesitamos dos informaciones: cuál es el punto de la curva que corresponde al instante $t_0 = \pi$ y el vector tangente en $t_0 = \pi$: $\vec{c}(\pi) =$

$(\pi, \cos \pi) = (\pi, -1)$, $\vec{v}(t) = (1, -\sin t)$, $\vec{v}(\pi) = (1, 0)$. Entonces la recta tangente se escribe:

$$\vec{r}(t) = \{(\pi, -1) + t(1, 0) = (\pi, -1) + (t, 0) = (\pi + t, -1), t \in \mathbb{R}\}.$$

La interpretación geométrica es la siguiente: independientemente de t , la ordenada de la recta vale -1 , entonces la recta tangente a la curva en $t_0 = \pi$ es la recta paralela al eje horizontal y que intersecta el eje vertical en -1 (dejamos como ejercicio dibujar tanto la curva como la recta tangente que hemos calculado).

2.4. Longitud de una curva suave

En esta sección queremos determinar cómo medir la longitud de una curva. Para motivar la técnica que introduciremos pensamos en la siguiente analogía: imaginamos que es necesario medir un hilo curvilíneo y solamente disponemos de un metro lineal rígido. Lo que podemos hacer es desplegar el hilo, rectificándolo, y ponerlo al lado del metro lineal para comparar la longitud del hilo con el metro para determinar la longitud.

En matemática el metro lineal rígido está dado por la norma Euclidiana, mientras que la ‘rectificación’, como hemos visto en la primera parte del curso, es posible solo localmente y está dada por la diferenciación. La diferenciación presupone el cálculo de derivadas, este es el enlace entre la longitud de una curva y su suavidad.

En la primera parte del curso de cálculo vimos que la variación infinitésima df de una función real de variable real $f(x)$ a lo largo de su recta tangente vale $df(x) = f'(x)dx$. No podemos escribir la longitud infinitésima de una curva como $\vec{c}'(t)dt$ porque $\vec{c}'(t)$ es un vector, mientras que la longitud es una cantidad escalar. Por eso tenemos que medir la longitud del vector $\vec{c}'(t)$ con el ‘metro lineal’ representado por la norma Euclidiana y definiremos la **longitud infinitésima de la curva \mathcal{C}** como sigue:

$$d\ell = \|\vec{c}'(t)\|dt = \sqrt{x_1'(t)^2 + \dots + x_m'(t)^2}dt.$$

Naturalmente, la longitud total estará dada ‘sumando’ todas las contribuciones de las longitudes infinitésimas, que, como sabemos, corresponde al proceso de integración. Por eso definimos la longitud de una curva suave de la forma siguiente:

$$\ell(\mathcal{C}) = \int_a^b \|\vec{c}'(t)\|dt = \int_a^b \sqrt{x_1'(t)^2 + \dots + x_m'(t)^2}dt.$$

Observamos que $\|\vec{c}'(t)\|$ representa la intensidad de la velocidad de una hipotética partícula que recorre la curva en el instante t , por eso la fórmula de la longitud de una curva se puede interpretar como una generalización de la fórmula: espacio = velocidad \times tiempo.

Por supuesto, esta construcción tiene sentido si la fórmula para calcular la longitud de una curva es coherente con las longitudes conocidas de la geometría elemental y con la longitud del gráfico de una función que ya vimos en la sección 5.6 de los apuntes de Cálculo 1. Como ejemplo de cálculo de curvas de longitudes conocidas, vamos a comprobar la coherencia con la longitud de la circunferencia de radio $r > 0$, que sabemos que es $\ell = 2\pi r$.

Ya sabemos que la parametrización de la circunferencia de radio r y centro en la origen es $(r \cos t, r \sin t)$, $t \in [0, 2\pi]$. El vector tangente a la circunferencia es $\vec{c}'(t) = (-r \sin t, r \cos t)$ y su norma vale

$$\|\vec{c}'(t)\| = \sqrt{r^2 \sin^2 t + r^2 \cos^2 t} = \sqrt{r^2(\sin^2 t + \cos^2 t)} = \sqrt{r^2 \cdot 1} = r.$$

Entonces la longitud infinitésima de la circunferencia vale $r dt$ y la longitud total $\ell = \int_0^{2\pi} r dt = r \int_0^{2\pi} dt = 2\pi r$, coherente con la longitud de la circunferencia.

Consideramos ahora la relación entre la longitud de una curva definida en esta sección y la longitud del gráfico de una función real de variable real $f : [a, b] \rightarrow \mathbb{R}$. Como ya hemos visto, la parametrización de la curva \mathcal{C} dada por el gráfico de f es $\vec{c}(t) = (t, f(t))$, $t \in [a, b]$, por lo tanto $\vec{c}'(t) = (1, f'(t))$ y $\|\vec{c}'(t)\| = \sqrt{1 + (f'(t))^2}$ entonces $L(\mathcal{C}) = \int_a^b \sqrt{1 + (f'(t))^2} dt$, que coincide con la definición de longitud del gráfico de f dada en la sección 5.6 de los apuntes de Cálculo 1.

2.5. Integrales de línea (o integrales curvilíneas)

Supongamos que tenemos un hilo de material no homogéneo distribuido en el espacio tridimensional y representado por una curva parametrizada $\vec{c} : [a, b] \rightarrow \mathbb{R}^3$. Supongamos que su densidad lineal, o sea la distribución de masa por unidad de línea, se representa por la función $\rho : \vec{c}([a, b]) \rightarrow \mathbb{R}$. Sabemos que la longitud infinitésima del hilo vale $\|\vec{c}'(t)\|dt$, entonces su masa infinitésima será $\rho(\vec{c}(t))\|\vec{c}'(t)\|dt$, por eso la masa total valdrá $\int_a^b \rho(\vec{c}(t))\|\vec{c}'(t)\|dt$.

Más en general, cuando una función f está definida sobre la trayectoria de una curva \mathcal{C} en \mathbb{R}^n (o un dominio que la contiene), es posible definir **la integral de f a lo largo de la curva \mathcal{C}** con parametrización $\vec{c}(t) = (x_1(t), \dots, x_n(t))$ como sigue:

$$\int_{\mathcal{C}} f dl = \int_a^b f(\vec{c}(t))\|\vec{c}'(t)\|dt,$$

siendo $f(\vec{c}(t))$ la evaluación de f a lo largo de la curva \mathcal{C} , es decir:

$$f(\vec{c}(t)) = f(x_1(t), \dots, x_n(t)).$$

Es posible demostrar que, bajo unas condiciones técnicas que no explicamos por falta de tiempo, esta definición es independiente de la parametrización de \mathcal{C} que se escoge.

Vamos a ver un ejemplo de cálculo de integral de línea a través de un ejercicio interesante. Supongamos que la distribución espacial de masa en cada punto de coordenadas (x, y, z) de un alambre se pueda modelar con la función $f(x, y, z) = x^2 + y^2 + z^2$. Supongamos también que el alambre tenga la forma de una **hélice circular** que se puede parametrizar como sigue:

$$\vec{c}(t) = (\cos t, \sin t, t), \quad t \in [0, 2\pi].$$

Queremos calcular cuanto vale la masa global del alambre. Para hacer eso tenemos que integrar la función que nos proporciona la masa en cada punto del alambre, a lo largo de la curva que describe en el espacio tridimensional el alambre mismo.

Para hacer eso, como primera cosa tenemos que evaluar la función f a lo largo de la curva. Eso se hace observando que $(\cos t, \sin t, t) = (x(t), y(t), z(t))$ y sustituyendo estas expresiones de x, y, z en la función: $f(x(t), y(t), z(t)) = \cos^2 t + \sin^2 t + t^2 = 1 + t^2$ en virtud de la relación fundamental de la trigonometría. Después tenemos que calcular el elemento de longitud infinitésima

de línea de la hélice circular $\|\vec{c}'(t)\|dt$: $\vec{c}'(t) = (-\sin t, \cos t, 1)$, entonces $\|\vec{c}'(t)\| = \sqrt{(-\sin t)^2 + (\cos t)^2 + 1} = \sqrt{2}$. Por lo tanto:

$$\int_0^{2\pi} (1+t^2)\sqrt{2} dt = \sqrt{2} \left[t + \frac{t^3}{3} \right]_0^{2\pi} = 2\sqrt{2}\pi \left(1 + \frac{4}{3}\pi^2 \right)$$

es el valor de la masa global del alambre.

2.6. Superficies y curvas de nivel

Consideremos de nuevo el gráfico de una función $f : X \rightarrow Y$, $\text{gráfico}(f) = \{(x, f(x)), x \in X\} \subset X \times Y$.

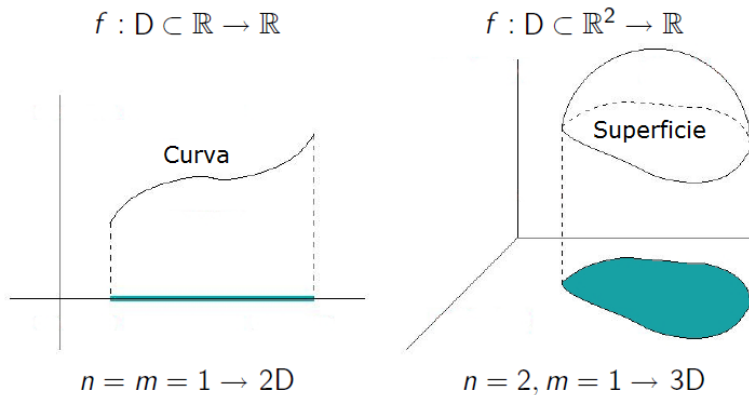
Si $f : D \subseteq \mathbb{R} \rightarrow \mathbb{R}$, sabemos que el gráfico de f es un subconjunto particular del plano real $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ dado por una curva.

Vamos a examinar qué pasa si consideramos una función real de dos variables:

$$\begin{aligned} f : D \subseteq \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ (x, y) &\longmapsto z = f(x, y), \end{aligned}$$

$\text{gráfico}(f) = \{(x, y, f(x, y)), (x, y) \in D\} \subset (\mathbb{R}^2 \times \mathbb{R}) = \mathbb{R}^3$. Podemos ver que el gráfico de $f(x, y)$ es un subconjunto del espacio tridimensional al cual damos el nombre de **superficie**.

Resumimos en esta figura la diferencia entre el gráfico de una función real de una variable y de una de dos variables.



Las proyecciones de los gráficos (en azul) de las dos funciones representan los espacios donde viven los parámetros (o grados de libertad) de sus gráficos: para una curva este espacio es un subconjunto de \mathbb{R} , para una superficie es un subconjunto de \mathbb{R}^2 . Por eso se suele decir que una curva es un objeto

geométrico con 1 grado de libertad, mientras que **una superficie tiene 2 grados de libertad** (o que tiene dimensión topológica 2).

Superficies y curvas están relacionadas a través del concepto de curva de nivel.

Def. Dada la superficie definida por el gráfico de la función $f : D \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$ y un valor k que pertenece al codominio de f , es decir $k \in f(D)$, definimos la curva de nivel k , C_k , de la superficie como el conjunto de todos los puntos de D que f transforma en el valor k :

$$C_k = \{(x, y) \in D : f(x, y) = k\}.$$

Dicho de otra forma, una curva de nivel k de una superficie es la curva que conecta todos los puntos sobre la superficie que tienen altura k .

Ejemplo gráfico:

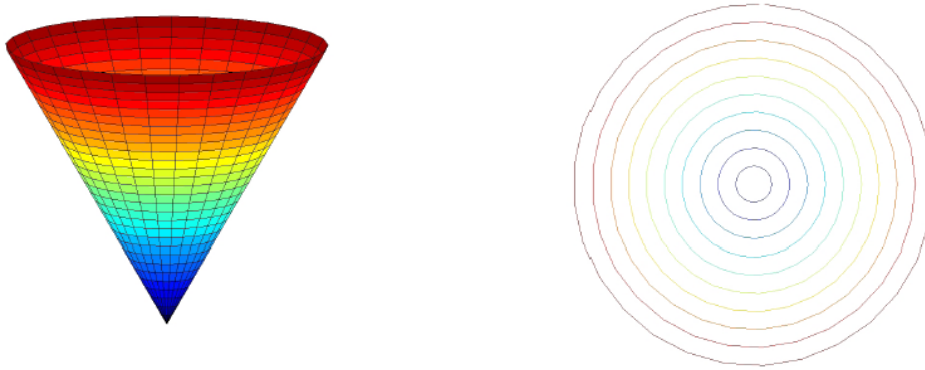


Figura 2.1: Izquierda: gráfico de la función $f(x, y) = \sqrt{x^2 + y^2}$ donde a cada punto $(x, y, f(x, y))$ se le asocia un color relacionado con la tercera coordenada. Derecha: algunas curvas de nivel con los diferentes colores asociados a cada nivel (o valor de la función en esos puntos). Se deja como ejercicio comprobar que la curva de nivel k de f es una circunferencia de radio k .

Para determinar analíticamente las curvas de nivel hay que resolver las ecuaciones $f(x, y) = k, \forall k \in f(D)$ o, análogamente, encontrar los ceros de la función auxiliar $\tilde{f}(x, y) = f(x, y) - k$. Esta tarea se puede llevar a cabo analíticamente solo en algunos casos, generalmente hay que utilizar métodos numéricos para aproximar el resultado, como el método de Newton para funciones de varias variables, que veremos más tarde en el curso.

Una propiedad fundamental de **las curvas de nivel** es que **nunca se intersectan**: $\forall k, k' \in f(D), k \neq k', C_k \cap C_{k'} = \emptyset$. Esto viene de la definición de función: si un punto (x_0, y_0) del dominio de f perteneciera tanto a C_k como a $C_{k'}$, f tomaría dos valores distintos en (x_0, y_0) , con lo que se violaría la unicidad de la relación entre dato de ingreso y de salida de una función.

La longitud y la cercanía de las curvas de nivel nos dan informaciones importantes sobre la geometría de una superficie: analizando la figura 2.2 se puede entender sin esfuerzo que curvas de nivel cortas corresponden a los ‘picos’ de las superficies, curvas de nivel más cercanas entre si corresponden a zonas de la superficie con mayor pendiente (mayor crecimiento o decrecimiento) mientras que curvas más alejadas corresponden a zonas de variación más suave de la superficie. Estas informaciones son muy importantes porque permiten determinar, por lo menos cualitativamente, las áreas donde están situados los extremos locales de las superficies.

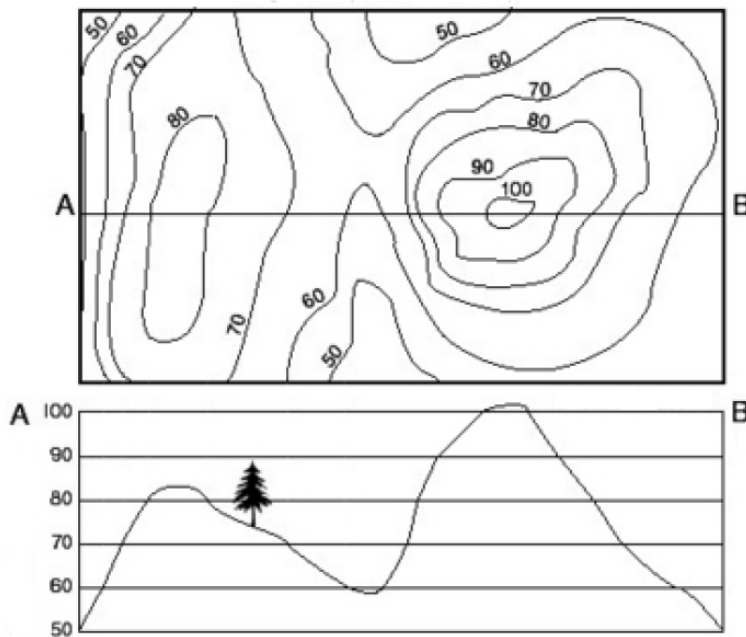


Figura 2.2: Ejemplo de curvas de nivel (arriba) de la montaña cuyo perfil se muestra en la imagen inferior.

2.6.1. Superficies en forma implícita. Las cuádricas

Como para las curvas, existe una forma implícita, o Cartesiana, de representar una superficie a través de una ecuación del tipo $f(x, y, z) = 0$, siendo f una función real. Si f es un polinomio cuadrático en las variables x, y, z , se pueden generar 16 superficies notables llamadas **cuádricas**. Mostramos las siete más comunes en la figura 2.3.

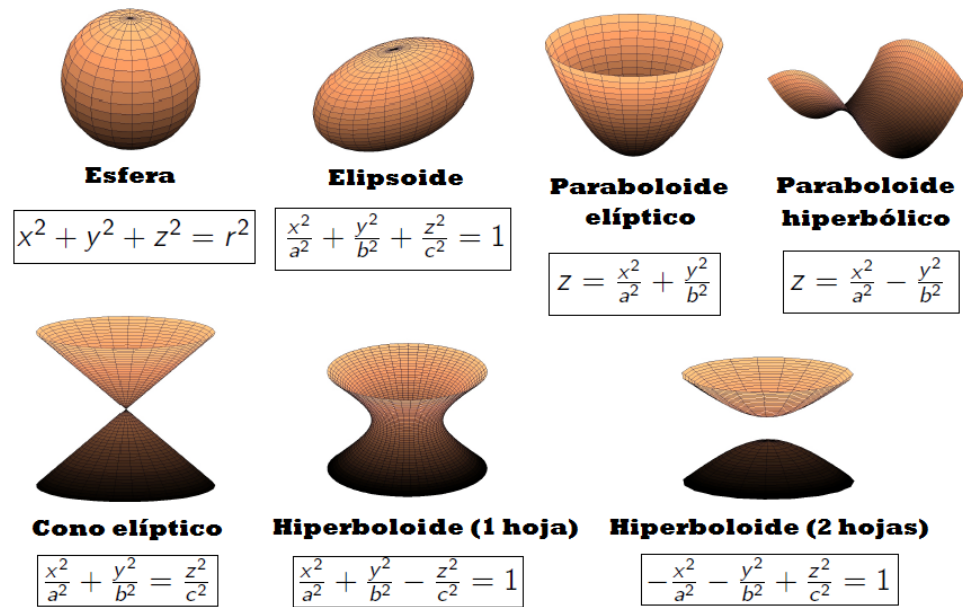


Figura 2.3: Cuádricas.

2.7. Superficies en forma paramétrica

Hemos visto que existen curvas en el plano que no son el gráfico de ninguna función real de variable real. De la misma forma, existen superficies en el espacio tridimensional que no son el gráfico de ninguna función de dos variables. Las cuádricas que hemos visto antes nos proporcionan unos ejemplos de semejantes superficies. Consideremos la más simple, la esfera: es evidente que esta superficie no puede ser el gráfico de una función $f(x, y)$ porque, por ejemplo, el polo norte y el polo sur, con alturas z_N y z_S respectivamente, tales que $z_N > z_S$, se proyectan sobre el mismo punto (x, y) en el plano. Como ya hemos remarcado, una función no puede asociar a un elemento de ingreso dos elementos distintos de salida!!

Al igual que para las curvas, la solución a este problema se encuentra definiendo las superficies en forma paramétrica como imágenes de funciones adecuadas:

Def. Llamamos **superficie en forma paramétrica** \mathcal{S} a la imagen de una función continua definida en un dominio de \mathbb{R}^2 a valores en \mathbb{R}^3 :

$$\begin{aligned} \vec{r}: D \subset \mathbb{R}^2 &\longrightarrow \mathbb{R}^3 \\ (s, t) &\longmapsto \vec{r}(s, t) = (x(s, t), y(s, t), z(s, t)), \end{aligned}$$

la función vectorial \vec{r} se llama parametrización de la superficie \mathcal{S} y $D \subseteq \mathbb{R}^2$ se llama dominio de parametrización.

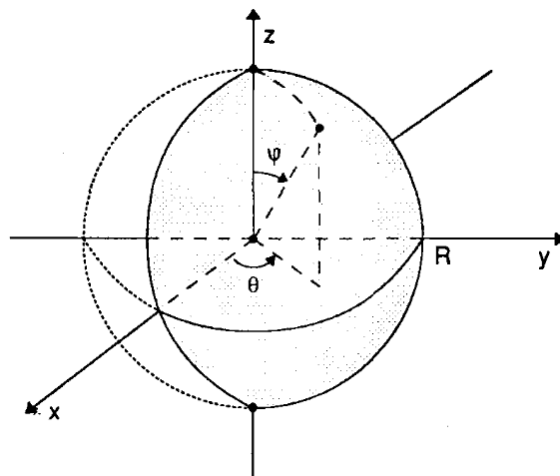
Ejemplo: la esfera de radio R y centro en el origen de coordenadas dibujada a continuación tiene expresión paramétrica dada por

$$\begin{aligned} \vec{r}: [0, \pi] \times [0, 2\pi] &\longrightarrow \mathbb{R}^3 \\ (\varphi, \theta) &\longmapsto \vec{r}(\varphi, \theta) = (x(\varphi, \theta), y(\varphi, \theta), z(\varphi, \theta)), \end{aligned}$$

con:

$$\begin{cases} x(\varphi, \theta) = R \sin \varphi \cos \theta \\ y(\varphi, \theta) = R \sin \varphi \sin \theta \\ z(\varphi, \theta) = R \cos \varphi \end{cases}$$

φ se llama ángulo polar y θ se llama ángulo acimutal, como se ve en la figura debajo. En este caso D es un rectángulo en el plano real: $D = [0, \pi] \times [0, 2\pi]$, siendo R fijo, los parámetros, o grados de libertad, son φ y θ , coherentemente con el hecho de que una superficie tiene 2 grados de libertad.



Observamos que, así como una curva en forma paramétrica se reduce al gráfico de una función escalar de una variable $f : [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}$ cuando se escribe la parametrización como

$$t \mapsto (t, f(t)), \quad t \in [a, b]$$

también una **superficie en forma paramétrica se reduce al gráfico de una función escalar de dos variables**, $f : D \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ cuando se escribe la parametrización como

$$(t, s) \mapsto (t, s, f(t, s)), \quad t, s \in D.$$

Naturalmente, también las superficies en forma paramétrica tienen curvas de nivel. En este caso, C_k , una curva de nivel k de \mathcal{S} , se define como la intersección entre \mathcal{S} y un plano Π_k paralelo al plano xy con altura k :

$$C_k = \mathcal{S} \cap \Pi_k.$$

En las clases de prácticas y seminarios veremos ejemplos de intersecciones entre superficies y planos.

2.8. Campos vectoriales

Terminamos la presentación de los diferentes objetos geométricos y físicos asociados con funciones de varias variables con los campos vectoriales,

cuya importancia surgió primeramente en el estudio de los fenómenos electromagnéticos y de la dinámica de los fluidos y que después se ha extendido a prácticamente todos los campos de investigación científica.

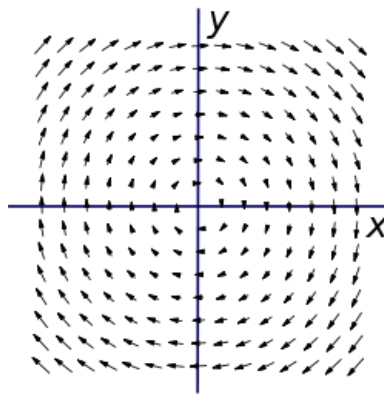
El movimiento de un fluido, por ejemplo del agua en un canal, se puede describir asignando, en cada instante y en cada punto, la velocidad de cada ‘partícula de agua’. La velocidad es un vector que vive en \mathbb{R}^3 , por tanto la asignación de esa velocidad será una función de este tipo:

$$\begin{aligned} \vec{v}: \quad \mathbb{R}^4 &\longrightarrow \mathbb{R}^3 \\ (t, x, y, z) &\longmapsto \vec{v}(t, x, y, z). \end{aligned}$$

Se dice que el movimiento de un fluido es **estacionario** si \vec{v} no depende de t y se dice **plano** si \vec{v} es independiente de una de las tres coordenadas espaciales. Un movimiento estacionario plano (por ejemplo en el plano xy) entonces se describe a través de una función como ésta

$$\begin{aligned} \vec{v}: \quad \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 \\ (x, y) &\longmapsto \vec{v}(x, y), \end{aligned}$$

que se puede visualizar dibujando un cierto número de vectores a partir de unos puntos de una región del plano, como en la figura aquí debajo, que muestra un movimiento rotacional dado por $\vec{v}(x, y) = (-y, x) = -y\hat{i} + x\hat{j}$.



Es importante observar que, aunque el conjunto de partida y de llegada sean lo mismo, o sea \mathbb{R}^2 , la interpretación de sus elementos es diferente: *los elementos del dominio de \vec{v} se consideran como puntos del plano, los elementos del codominio se consideran como vectores.*

Generalizando, dada una función $\vec{f}: D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$, si interpretamos los elementos de D como puntos y los elementos de \mathbb{R}^m como vectores m -dimensionales, tenemos que interpretar f como la asignación de un vector

m -dimensional a cada punto del dominio n -dimensional D . A esta asignación se le llama **campo vectorial**. Si $m = 1$ se llama **campo escalar**.

Ejemplos comunes de campos vectoriales son los campos de fuerza y de velocidad. Un ejemplo físico de importancia fundamental es el **campo electromagnético**:

$$(\vec{E}, \vec{B}) : \mathbb{R}^4 \longrightarrow \mathbb{R}^6 \\ (t, x, y, z) \longmapsto (\vec{E}, \vec{B})(t, x, y, z) = (\vec{E}(t, x, y, z), \vec{B}(t, x, y, z)),$$

donde $\vec{E}(t, x, y, z), \vec{B}(t, x, y, z) \in \mathbb{R}^3$ son los campos eléctrico y magnético, respectivamente, en el instante t y en el punto de coordenadas (x, y, z) .

Observamos que la diferencia entre un campo vectorial $\vec{f} : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ con $n = 1$ y una curva $\vec{c} : [a, b] \rightarrow \mathbb{R}^m$ está en la interpretación del codominio: para \vec{f} el codominio es un conjunto de vectores, para \vec{c} es un conjunto de puntos.

Es posible relacionar también los campos vectoriales con las curvas: llamamos **curva (o línea) integral de un campo vectorial** la curva que, en cada punto, es tangente a los vectores definidos por el campo mismo. Si el campo tiene significado físico de campo de fuerzas, las curvas integrales se llaman **líneas de fuerza**; si tiene significado físico de campo de velocidad, se llaman **líneas de flujo**.

El estudiante podrá ver ejemplos notables de curvas integrales en el curso de física y ondas en relación al campo gravitacional y electrostático. Aquí simplemente decimos que las líneas de flujo del campo de velocidad presentado en la figura anterior son circunferencias concéntricas.

Capítulo 3

Cálculo diferencial para funciones de varias variables

En este capítulo extenderemos los conceptos de límite, continuidad, derivabilidad y diferenciabilidad a funciones de varias variables reales.

3.1. Límites y continuidad

Como para las funciones reales de variable real, también para las funciones de varias variables el fundamento de todo el cálculo diferencial es el concepto de límite.

Recordamos, antes de todo, que en \mathbb{R} el entorno abierto de radio $r > 0$ de x_0 está definido como $U_r(x_0) = \{x \in \mathbb{R} : |x - x_0| < r\}$. Dada la función $f : D \subseteq \mathbb{R} \rightarrow \mathbb{R}$, con D subconjunto *abierto* de \mathbb{R} , y $x_0 \in D$ o bien x_0 punto de la frontera de D , vale que $\lim_{x \rightarrow x_0} f(x) = \ell$ si para cada entorno $U_\varepsilon(\ell) \subseteq f(D)$, $\varepsilon > 0$, existe un entorno $U_{\delta_\varepsilon}(x_0) \subseteq D$ tal que para $x \in U_{\delta_\varepsilon}(x_0) \setminus \{x_0\}$, se tiene que $f(x) \in U_\varepsilon(\ell)$.

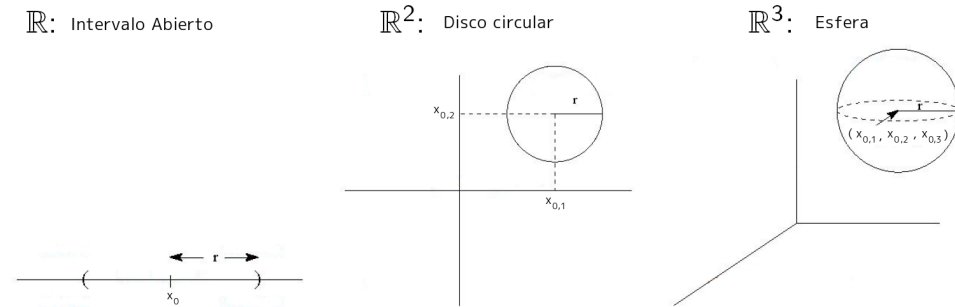
La definición de límite para una función de varias variables $\vec{f} : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ es análoga, la única diferencia consiste en la definición de los entornos: ya no tiene sentido medir la distancia entre puntos con el valor absoluto, hay que sustituirlo por la norma.

Def.: dado $\vec{x}_0 \in \mathbb{R}^N$, llamamos entorno de radio $r > 0$ de x_0 el conjunto de puntos así definido:

$$U_r(\vec{x}_0) = \{\vec{x} \in \mathbb{R}^N : \|\vec{x} - \vec{x}_0\| < r\},$$

donde $\|\vec{x} - \vec{x}_0\| = \sqrt{(x_1 - x_{0,1})^2 + \dots + (x_N - x_{0,N})^2}$, siendo $x_{0,i}$ la coordenada i -ésima de $\vec{x}_0 \in \mathbb{R}^N$.

Si $N = 1, 2, 3$ podemos representar un entorno como un intervalo real abierto y simétrico, un disco circular y una esfera, respectivamente, con centro en \vec{x}_0 :



La definición de límite y continuidad para una función $\vec{f} : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$, es la siguiente.

Def.: se dice que $\lim_{\vec{x} \rightarrow \vec{x}_0} \vec{f}(x) = \vec{\ell}$ si para cada entorno $U_\varepsilon(\vec{\ell}) \subseteq \vec{f}(D)$, $\varepsilon > 0$, existe un entorno $U_{\delta_\varepsilon}(\vec{x}_0) \subseteq D$ tal que para $\vec{x} \in U_{\delta_\varepsilon}(\vec{x}_0) \setminus \{\vec{x}_0\}$, se tiene que $\vec{f}(\vec{x}) \in U_\varepsilon(\vec{\ell})$. Decimos que \vec{f} es continua en \vec{x}_0 si $\lim_{\vec{x} \rightarrow \vec{x}_0} \vec{f}(x) = \vec{f}(\vec{x}_0)$.

La interpretación del concepto de continuidad es lo mismo que vimos para funciones de una variable real: una función $\vec{f} : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ es continua en \vec{x}_0 si f mapea puntos ‘cercaños’ a \vec{x}_0 a puntos ‘cercaños’ a $\vec{f}(\vec{x}_0)$, donde, en este caso, la cercanía se mide a través de la norma Euclidiana y no con el valor absoluto.

El cálculo de límites para funciones de más variables no es una tarea trivial y queda fuera de los objetivos de este curso. Nos limitamos a decir que los límites de funciones de varias variables siguen teniendo todas las propiedades de los límites de funciones de una variable, en particular la **linealidad**.

3.2. Derivadas direccionales y parciales

Hemos visto que los conceptos de derivabilidad y diferenciabilidad en un punto son equivalentes para una función real de variable real $f : D \subseteq \mathbb{R} \rightarrow \mathbb{R}$.

La derivabilidad de f en $x_0 \in D$ corresponde a la existencia de la recta tangente al gráfico de la función en x_0 , mientras que la diferenciabilidad de f en x_0 corresponde a la existencia de un entorno de x_0 en el cual podemos aproximar f con su recta tangente por menos de un error desechable (más precisamente, un infinitésimo de orden superior con respecto al desplazamiento $x - x_0$).

Para funciones de más variables reales esta equivalencia ya no vale y la diferenciabilidad resulta una condición más fuerte que la derivabilidad.

Sabemos que la construcción de la derivada pasa por el límite de la fracción incremental, o sea de la diferencia entre el valor que f toma cuando se evalúa en dos puntos cercanos dividida por la distancia entre estos dos puntos. En \mathbb{R} hay solo una forma de desplazarse de un punto x_0 : sumar un valor adecuado $h \geq 0$, pero en más de una dimensión esto ya no vale porque podemos desplazarnos de \vec{x}_0 siguiendo infinitas direcciones. Resulta por lo tanto imprescindible especificar en qué dirección nos estamos desplazando para poder definir la derivada de una función de varias variables, entonces el concepto de derivabilidad en más variables es un concepto ‘direccional’. Para ello resulta útil la definición siguiente.

Def.: dado el vector unitario $\vec{u} \in \mathbb{R}^n$, $\|\vec{u}\| = 1$, llamamos recta en \mathbb{R}^n que pasa por el punto $\vec{x}_0 \in \mathbb{R}^n$ con dirección \vec{u} al conjunto definido de esta forma:

$$r_{\vec{x}_0, \vec{u}} = \{\vec{x} \in \mathbb{R}^n : \vec{x} = \vec{x}_0 + t\vec{u}\}.$$

La idea que está en la base de la definición de derivabilidad de $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ en $\vec{x}_0 \in D$ a lo largo de la dirección definida por un vector unitario \vec{u} , $\|\vec{u}\| = 1$, consiste en medir la rapidez de variación de f cuando nos alejamos de \vec{x}_0 haciendo un ‘paso de longitud infinitésima’ a lo largo de la recta $r_{\vec{x}_0, \vec{u}}$, es decir:

Def.: llamamos **derivada de $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ en $\vec{x}_0 \in D$ a lo largo de la dirección $\vec{u} \in \mathbb{R}^n$, $\|\vec{u}\| = 1$** el valor del siguiente límite, si existe y es finito,

$$D_{\vec{u}}f(\vec{x}_0) = \lim_{h \rightarrow 0} \frac{f(\vec{x}_0 + h\vec{u}) - f(\vec{x}_0)}{h}.$$

La derivada direccional expresa la rapidez de variación de una función f cuando nos alejamos del punto en el que se calcula la derivada a lo largo de una cierta dirección.

Siendo definida a través de un límite, la derivada direccional resulta automáticamente una operación **lineal**.

Si $n = 2$, podemos explicitar la definición de derivada direccional como sigue: dado $\vec{x}_0 = (x_0, y_0)$ y $\vec{u} = (u_1, u_2)$, $\sqrt{u_1^2 + u_2^2} = 1$,

$$D_{\vec{u}}f(\vec{x}_0) = \lim_{h \rightarrow 0} \frac{f(x_0 + hu_1, y_0 + hu_2) - f(x_0, y_0)}{h}.$$

En la sección siguiente veremos una técnica muy sencilla para calcular las derivadas direccionales.

Como los estudiantes saben del curso de Álgebra, en \mathbb{R}^n existen n direcciones ‘privilegiadas’, dadas por las direcciones de la base canónica $(\vec{e}_1, \dots, \vec{e}_n)$, es decir, las direcciones de los ejes Cartesianos ortogonales. \vec{e}_i es el vector unitario que tiene todas las componentes nulas, excepto la componente i -ésima, que vale 1.

Las derivadas direccionales calculadas a lo largo de las direcciones de la base canónica de \mathbb{R}^n tienen un nombre y un significado geométrico particular:

Def.: llamamos **derivada parcial** i -ésima, $i = 1, \dots, n$, de $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ en $\vec{x}_0 \in D$ el valor del siguiente límite, si existe y es finito

$$D_{\vec{e}_i}f(\vec{x}_0) = \frac{\partial f}{\partial x_i}(\vec{x}_0) = \lim_{h \rightarrow 0} \frac{f(\vec{x}_0 + h\vec{e}_i) - f(\vec{x}_0)}{h},$$

de forma más explícita, dado que $\vec{x}_0 + h\vec{e}_i = (x_0, \dots, x_i + h, \dots, x_n)$, podemos escribir

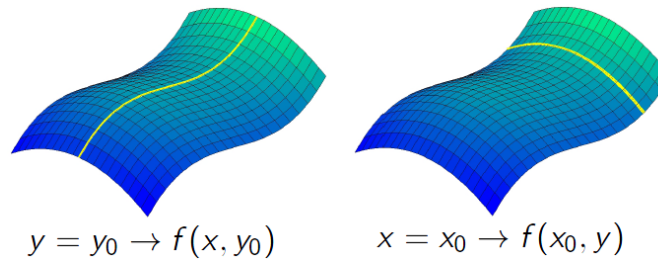
$$\boxed{\frac{\partial f}{\partial x_i}(\vec{x}_0) = \lim_{h \rightarrow 0} \frac{f(x_0, \dots, x_i + h, \dots, x_n) - f(x_0, \dots, x_i, \dots, x_n)}{h}}.$$

Otras notaciones que se encuentran a menudo en los libros para las derivadas parciales son: $\partial_{x_i}f(\vec{x}_0)$, $f_{x_i}(\vec{x}_0)$. En el cálculo de la derivada parcial i -ésima solo la componente i -ésima varía, las otras se quedan fijas. Este comportamiento se puede explicitar de forma clara considerando el caso $n = 2$: dada $f : D \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$, y $\vec{x}_0 = (x_0, y_0)$, las derivadas parciales de f en \vec{x}_0 son

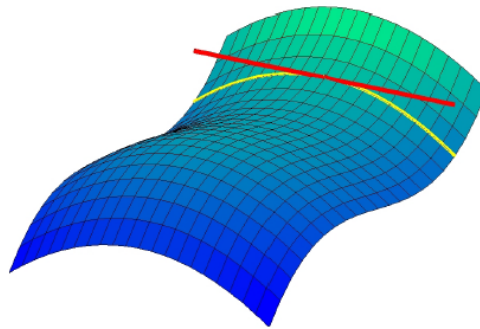
$$\boxed{\frac{\partial f}{\partial x}(\vec{x}_0) = \lim_{h \rightarrow 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h}}$$

$$\boxed{\frac{\partial f}{\partial y}(\vec{x}_0) = \lim_{h \rightarrow 0} \frac{f(x_0, y_0 + h) - f(x_0, y_0)}{h}}.$$

La interpretación geométrica de las derivadas parciales viene directamente del significado geométrico de la derivada de una función de una variable real. Antes de todo, si fijamos la variable $x = x_0$ o bien $y = y_0$ y variamos la otra, obtenemos dos curvas sobre la superficie de ecuación $z = f(x, y)$, como se ve en la figura aquí debajo:



Geoméricamente, las derivadas parciales representan las pendientes de las rectas tangentes en cada punto a las curvas mencionadas arriba.



De la definición de derivada parcial con respecto a la variable x_i sigue que esta se calcula considerando f como una función de una sola variable x_i y considerando las otras variables como constantes.

Veamos un ejemplo: calcular las derivadas parciales de $f(x, y) = x^2y^3$ en el punto $(2, -1)$ y establecer si, cuando estamos en un entorno de ese punto, f varía más en la dirección de x o de y . Comenzamos a calcular las derivadas parciales en un punto genérico (x, y) con la regla vista arriba:

- $f_x(x, y) = 2xy^3$;
- $f_y(x, y) = 3x^2y^2$.

Entonces $f_x(2, -1) = -4$, $f_y(2, -1) = 12$, lo cual quiere decir que la función f varía más rápidamente (en valor absoluto) cuando nos desplazamos en la dirección del eje y .

3.3. El gradiente y su significado geométrico

En esta sección comenzamos a ver por primera vez una intersección entre el cálculo diferencial y el álgebra lineal, esta intersección nos permitirá obtener resultados importantes con una sencillez sorprendente.

Observamos que las derivadas parciales de una función $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ en $\vec{x}_0 \in D$ se pueden organizar en un vector que se suele llamar **gradiente de f en \vec{x}_0** :

$$\nabla f(\vec{x}_0) = \text{grad}f(\vec{x}_0) = \left(\frac{\partial f}{\partial x_1}(\vec{x}_0), \dots, \frac{\partial f}{\partial x_n}(\vec{x}_0) \right).$$

Ejemplo: calcular el gradiente de $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, $f(x, y) = \log(1 + x^2 + y^2)$ en un punto genérico de coordenadas (x, y) . Antes de todo calculemos las derivadas parciales en (x, y) : $f_x(x, y) = 2x/(1 + x^2 + y^2)$, $f_y(x, y) = 2y/(1 + x^2 + y^2)$, entonces:

$$\nabla f(x, y) = \left(\frac{2x}{1 + x^2 + y^2}, \frac{2y}{1 + x^2 + y^2} \right).$$

El gradiente no es simplemente una forma de organizar las derivadas parciales, sino que tiene un significado geométrico profundo e importante, que sigue del siguiente teorema.

Teorema: $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, $\vec{x}_0 \in D$, $\vec{u} \in \mathbb{R}^n$, $\|\vec{u}\| = 1$. Vale la siguiente relación entre gradiente y derivada direccional:

$$D_{\vec{u}}f(\vec{x}_0) = \langle \nabla f(\vec{x}_0), \vec{u} \rangle, \quad (3.1)$$

donde $\langle \cdot, \cdot \rangle$ representa el producto escalar en el espacio Euclídeo¹ entre dos vectores de \mathbb{R}^n .

Este teorema tiene consecuencias tanto computacionales como teóricas.

3.3.1. Consecuencias computacionales de la fórmula (3.1)

Gracias a la fórmula (3.1) podemos calcular el gradiente a través de la derivada direccional y viceversa, lo cual resulta extremadamente útil porque

¹Recordamos que $\langle \vec{v}, \vec{w} \rangle = v_1 w_1 + \dots + v_n w_n$ para cualquier pareja de vectores $\vec{v}, \vec{w} \in \mathbb{R}^n$.

en algunas situaciones es más fácil el cálculo de la derivada direccional (típicamente para funciones de matrices, aunque este tipo de funciones quedan afuera del alcance de este curso) y en otras el del gradiente (típicamente para funciones a valores escalares).

Veamos un ejemplo: calcular la derivada direccional de la función $f(x, y) = x^2e^{-y}$ en el punto genérico (x, y) a lo largo de la dirección definida por el vector unitario $\vec{u} = (1/\sqrt{2}, 1/\sqrt{2})$. La derivada direccional es

$$D_{\vec{u}}f(x, y) = \lim_{h \rightarrow 0} \frac{f(x + hu_x, y + hu_y) - f(x, y)}{h}$$

es decir

$$\lim_{h \rightarrow 0} \frac{\left(x + \frac{h}{\sqrt{2}}\right)^2 e^{-(y + \frac{h}{\sqrt{2}})} - x^2 e^{-y}}{h},$$

este límite es complicado de calcular, por eso resulta más fácil calcular la derivada direccional pasando por el cálculo del gradiente y usando la fórmula (3.1). Para calcular el gradiente hacen falta las derivadas parciales: $f_x(x, y) = 2xe^{-y}$, $f_y(x, y) = -x^2e^{-y}$, entonces $\nabla f(x, y) = (2xe^{-y}, -x^2e^{-y})$ y

$$D_{\vec{u}}f(x, y) = \langle \nabla f(x, y), \vec{u} \rangle = \frac{2xe^{-y}}{\sqrt{2}} + \frac{(-x^2e^{-y})}{\sqrt{2}} = \frac{xe^{-y}}{\sqrt{2}}(2 - x).$$

3.3.2. Consecuencias teóricas de la fórmula (3.1)

Recordamos que, dados dos vectores cualquiera $\vec{v}, \vec{w} \in \mathbb{R}^n$, el producto escalar es aquel número real (positivo, negativo o nulo), dado por el producto de las normas de los dos vectores multiplicada por el coseno del ángulo α entre ellos:

$$\langle \vec{v}, \vec{w} \rangle = \|\vec{v}\| \|\vec{w}\| \cos \alpha.$$

Por eso, podemos escribir la fórmula (3.1) como

$$D_{\vec{u}}f(\vec{x}_0) = \|\nabla f(\vec{x}_0)\| \|\vec{u}\| \cos \alpha, \quad \alpha = \text{ángulo entre } \nabla f(\vec{x}_0) \text{ y } \vec{u},$$

pero $\|\vec{u}\| = 1$, entonces:

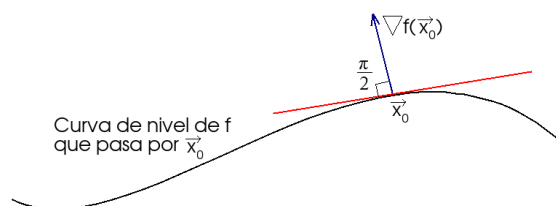
$$\boxed{D_{\vec{u}}f(\vec{x}_0) = \|\nabla f(\vec{x}_0)\| \cos \alpha}.$$

Sabemos que el coseno es una función acotada entre -1 y +1, por lo tanto:

$$\boxed{-\|\nabla f(\vec{x}_0)\| \leq D_{\vec{u}}f(\vec{x}_0) \leq +\|\nabla f(\vec{x}_0)\|}, \quad \forall \vec{u} \in \mathbb{R}^n.$$

En las tres situaciones notables en las que el coseno toma su valor inferior -1, su valor superior +1 y el valor 0, la derivada direccional resulta, respectivamente: mínima (máximo valor en negativo), máxima (máximo valor en positivo) y nula, vamos a examinar en detalle lo que significan estas condiciones.

- Si ha habido un desplazamiento a lo largo de una dirección \vec{u} , el gradiente $\|\nabla f(\vec{x}_0)\|$ no puede ser nulo y dado que $\|u\| = 1$, la única forma de que la derivada direccional $D_{\vec{u}}f(\vec{x}_0)$ sea nula es que $\cos \alpha = 0$. Esto pasa si y solo si $\alpha = \frac{\pi}{2}, \frac{3}{2}\pi$, es decir, si y solo si $\nabla f(\vec{x}_0)$ y \vec{u} son ortogonales ($\nabla f(\vec{x}_0) \perp \vec{u}$). Esta consideración nos permite dar una definición alternativa de las curvas de nivel de una función: dado que las curvas de nivel de f se definen como los conjuntos de puntos $\vec{x} \in \mathbb{R}^n$ tales que $f(\vec{x}) = k$, donde k es constante, es evidente que a lo largo de una curva de nivel la variación de f es nula. Sigue que, si calculamos la derivada direccional de f en un punto \vec{x}_0 en la dirección del vector tangente a la (única) curva de nivel de f que pasa por \vec{x}_0 , esta derivada es nula. Pero, por lo que acabamos de ver, la derivada direccional es nula solo en las direcciones ortogonales al gradiente de f en \vec{x}_0 . Podemos por lo tanto definir **las curvas de nivel de f como aquellas curvas cuyo vector tangente es ortogonal al gradiente de f en cada punto**. Con un lenguaje impropio se suele resumir este concepto diciendo que *‘el gradiente es ortogonal a las curvas de nivel’*. La figura siguiente visualiza el concepto que acabamos de describir.



- $\cos \alpha = +1$ si y solo si $\nabla f(\vec{x}_0)$ y \vec{u} son paralelos ($\nabla f(\vec{x}_0) \parallel \vec{u}$), de forma que $\alpha = 0$. Esto implica que la **dirección de crecimiento más rápido** de la función f , cuando nos desplazamos del punto \vec{x}_0 , es la dirección del gradiente de f en \vec{x}_0 . El vector unitario que representa esta dirección es:

$$\vec{u}_{\text{max. crec.}} = \frac{\nabla f(\vec{x}_0)}{\|\nabla f(\vec{x}_0)\|}.$$

- Análogamente, $\cos \alpha = -1$ si y solo si $\nabla f(\vec{x}_0)$ y \vec{u} son antiparalelos, de forma que $\alpha = \pi$. Esto implica que la **dirección de decrecimiento más rápido** de la función f , cuando nos desplazamos del punto \vec{x}_0 , es la dirección opuesta al gradiente de f en \vec{x}_0 . El vector unitario que la representa es:

$$\vec{u}_{\text{max. decrec.}} = -\frac{\nabla f(\vec{x}_0)}{\|\nabla f(\vec{x}_0)\|}.$$

3.4. La matriz Jacobiana

En esta sección examinamos la extensión del concepto de gradiente a funciones que toman valores vectoriales. Consideramos por lo tanto una función $\vec{f} : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$, $\vec{f} = (f_1, \dots, f_m)$, siendo $f_i : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, $i = 1, \dots, m$, sus funciones componentes.

Como ya sabemos, las funciones componentes f_i de \vec{f} son funciones a valores escalares, para las cuales podemos definir las derivadas parciales y el gradiente como hicimos antes, es decir:

$$\frac{\partial f_i}{\partial x_j}(\vec{x}_0) = \lim_{h \rightarrow 0} \frac{f_i(\vec{x}_0 + h\vec{e}_j) - f_i(\vec{x}_0)}{h}, \quad i = 1, \dots, m, \quad j = 1, \dots, n,$$

$$\nabla f_i(\vec{x}_0) = \left(\frac{\partial f_i}{\partial x_1}(\vec{x}_0), \dots, \frac{\partial f_i}{\partial x_n}(\vec{x}_0) \right), \quad i = 1, \dots, m.$$

Dejando variar los índices i y j tenemos $m \cdot n$ derivadas parciales posibles, que se pueden organizar en m vectores gradientes cada uno con n componentes.

Ejemplo: calcular derivadas parciales y gradientes de las funciones componentes de la siguiente función

$$\begin{aligned} \vec{f}: \quad \mathbb{R}^3 &\longrightarrow \mathbb{R}^2 \\ (x, y, z) &\longmapsto \vec{f}(x, y, z) = (x + y + z, xyz^3). \end{aligned}$$

Dado que $n = 3$ y $m = 2$, tendremos 6 derivadas parciales y 2 vectores gradientes con 3 componentes. Antes de todo identificamos las funciones componentes: $f_1(x, y, z) = x + y + z$, $f_2(x, y, z) = xyz^3$, entonces

$$\frac{\partial f_1}{\partial x}(x, y, z) = \frac{\partial f_1}{\partial y}(x, y, z) = \frac{\partial f_1}{\partial z}(x, y, z) = 1,$$

$$\frac{\partial f_2}{\partial x}(x, y, z) = yz^3, \quad \frac{\partial f_2}{\partial y}(x, y, z) = xz^3, \quad \frac{\partial f_2}{\partial z}(x, y, z) = 3xyz^2.$$

Los gradientes de las funciones componentes son:

$$\nabla f_1(x, y, z) = (1, 1, 1), \quad \nabla f_2(x, y, z) = (yz^3, xz^3, 3xyz^2).$$

Las $m \cdot n$ derivadas parciales de una función vectorial $\vec{f} : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ se pueden organizar en una matriz llamada² **Jacobiana**, que es una matriz $m \times n$ con filas dadas por los gradientes de las funciones componente calculados en \vec{x} :

$$J_{\vec{f}}(\vec{x}) = \begin{bmatrix} \nabla f_1(\vec{x}) \\ \vdots \\ \nabla f_m(\vec{x}) \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\vec{x}) & \cdots & \frac{\partial f_1}{\partial x_n}(\vec{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\vec{x}) & \cdots & \frac{\partial f_m}{\partial x_n}(\vec{x}) \end{bmatrix}.$$

Recordar:

- El número de columnas de $J_{\vec{f}}(\vec{x})$ es la dimensión del dominio de \vec{f} ;
- El número de filas de $J_{\vec{f}}(\vec{x})$ es la dimensión del codominio de \vec{f} .

Como observamos en relación al gradiente, también la matriz Jacobiana no es simplemente una forma de organizar las derivadas parciales, sino que tiene una utilidad práctica que ya comenzaremos a ver a partir de la subsección siguiente.

Para tener un ejemplo práctico de matriz Jacobiana, podemos considerar la función $\vec{f}(x, y, z) = (x + y + z, xyz^3)$ examinada en el ejemplo anterior:

$$J_{\vec{f}}(x, y, z) = \begin{bmatrix} \nabla f_1(x, y, z) \\ \nabla f_2(x, y, z) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ yz^3 & xz^3 & 3xyz^2 \end{bmatrix}.$$

3.4.1. Derivación de una función compuesta en varias variables

Empezamos recordando cómo se deriva una función compuesta en una dimensión: sea $g \circ f$ la función compuesta de dos funciones f y g . Si f es derivable en x_0 y g es derivable en $y = f(x_0)$ entonces $g \circ f$ es derivable en x_0 y se cumple la regla de la cadena:

$$(g \circ f)'(x_0) = g'(f(x_0)) \cdot f'(x_0)$$

²En honor al gran matemático alemán Carl Gustav Jacobi (1804-1851) que fue el primero en entender la utilidad de la matriz que ahora lleva su nombre.

que dice que la rapidez de variación de la función compuesta $g \circ f$ es el producto de la rapidez de variación de f y de g .

La regla de la cadena sigue siendo válida para funciones de más variables, bajo la condición de sustituir las derivadas con las matrices Jacobianas.

Consideramos $\vec{f} : D \subseteq \mathbb{R}^n \rightarrow \vec{f}(D) \subseteq \mathbb{R}^m$ y $\vec{g} : \vec{f}(D) \subseteq \mathbb{R}^m \rightarrow \mathbb{R}^p$, podemos definir la función compuesta $\vec{g} \circ \vec{f} : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^p$. Supongamos que tanto \vec{f} como \vec{g} tienen derivadas parciales en cada punto de sus dominios.

La matriz Jacobiana asociada a \vec{f} es una matriz $m \times n$:

$$J_{\vec{f}}(\vec{x}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\vec{x}) & \cdots & \frac{\partial f_1}{\partial x_n}(\vec{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\vec{x}) & \cdots & \frac{\partial f_m}{\partial x_n}(\vec{x}) \end{bmatrix}, \quad \vec{x} \in D.$$

Análogamente, la matriz Jacobiana asociada a \vec{g} es una matriz $p \times m$:

$$J_{\vec{g}}(\vec{y}) = \begin{bmatrix} \frac{\partial g_1}{\partial y_1}(\vec{y}) & \cdots & \frac{\partial g_1}{\partial y_m}(\vec{y}) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_p}{\partial y_1}(\vec{y}) & \cdots & \frac{\partial g_p}{\partial y_m}(\vec{y}) \end{bmatrix}, \quad \vec{y} = \vec{f}(\vec{x}).$$

La única multiplicación que tiene sentido entre estas dos matrices es $J_{\vec{g}}(\vec{f}(\vec{x}))J_{\vec{f}}(\vec{x})$ y el resultado es una matriz $p \times n = (p \times m) \cdot (m \times n)$ (en el curso de álgebra se ha visto que dos matrices se pueden multiplicar si y solo si el número de columnas de la primera coincide con el número de filas de la segunda).

Se puede demostrar que la matriz producto que acabamos de escribir es la matriz Jacobiana asociada a la función compuesta:

Teorema de la cadena en varias variables: sean $\vec{f} : D \subseteq \mathbb{R}^n \rightarrow \vec{f}(D) \subseteq \mathbb{R}^m$ y $\vec{g} : \vec{f}(D) \subseteq \mathbb{R}^m \rightarrow \mathbb{R}^p$, dos funciones que admiten derivadas parciales en cada punto de sus dominios, entonces

$$\boxed{J_{\vec{g} \circ \vec{f}}(\vec{x}) = J_{\vec{g}}(\vec{f}(\vec{x}))J_{\vec{f}}(\vec{x})}.$$

Este resultado muestra la utilidad de organizar las derivadas parciales en la matriz Jacobiana. Podemos explicitar la fórmula de la cadena poniendo $\vec{z} = \vec{g} \circ \vec{f}$, $\vec{y} = \vec{f}$:

$$\frac{\partial z_j}{\partial x_i}(\vec{x}) = \sum_{k=1}^m \frac{\partial z_j}{\partial y_k}(\vec{f}(\vec{x})) \frac{\partial y_k}{\partial x_i}(\vec{x}),$$

donde $i = 1, \dots, n$, $j = 1, \dots, k$.

Ejemplo: dadas las funciones

$$\begin{aligned} \vec{f}: \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 \\ (x, y) &\longmapsto \vec{f}(x, y) = (x^2 + 1, y^2), \end{aligned}$$

$$\begin{aligned} \vec{g}: \mathbb{R}^2 &\longrightarrow \mathbb{R}^3 \\ (u, v) &\longmapsto \vec{g}(u, v) = (u + v, u, v^2), \end{aligned}$$

calcular la matriz Jacobiana de $\vec{g} \circ \vec{f}$ como producto de las matrices Jacobianas de \vec{g} y \vec{f} .

Antes de todo explicitamos $(\vec{g} \circ \vec{f})(x, y)$ sustituyendo $u = x^2 + 1$ e $v = y^2$:

$$\begin{aligned} \vec{g} \circ \vec{f}: \mathbb{R}^2 &\longrightarrow \mathbb{R}^3 \\ (x, y) &\longmapsto \vec{g} \circ \vec{f}(x, y) = (1 + x^2 + y^2, x^2 + 1, y^4). \end{aligned}$$

Las funciones no tienen problemas de diferenciabilidad porque involucran funciones polinomiales, entonces podemos construir las matrices Jacobianas. Sabemos que $J_{\vec{f}}(x, y) \in \mathbb{R}^{2,2}$, es decir, es una matriz real 2×2 y que $J_{\vec{g}}(u, v) \in \mathbb{R}^{3,2}$, entonces $J_{\vec{g} \circ \vec{f}}(x, y) \in \mathbb{R}^{3,2}$.

Tenemos que

$$J_{\vec{f}}(x, y) = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{bmatrix} (x, y) = \begin{bmatrix} 2x & 0 \\ 0 & 2y \end{bmatrix},$$

$$J_{\vec{g}}(u, v) = \begin{bmatrix} \frac{\partial g_1}{\partial u} & \frac{\partial g_1}{\partial v} \\ \frac{\partial g_2}{\partial u} & \frac{\partial g_2}{\partial v} \\ \frac{\partial g_3}{\partial u} & \frac{\partial g_3}{\partial v} \end{bmatrix} (u, v) = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 2v \end{bmatrix},$$

entonces

$$J_{\vec{g}}(\vec{f}(x, y)) = J_{\vec{g}}(u = x^2 + 1, v = y^2) = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 2y^2 \end{bmatrix}.$$

Usando la regla de la cadena obtenemos:

$$J_{\vec{g} \circ \vec{f}}(x, y) = J_{\vec{g}}(\vec{f}(x, y)) J_{\vec{f}}(x, y) = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 2y^2 \end{bmatrix} \begin{bmatrix} 2x & 0 \\ 0 & 2y \end{bmatrix} = \begin{bmatrix} 2x & 2y \\ 2x & 0 \\ 0 & 4y^3 \end{bmatrix}.$$

Como ejercicio, invitamos a comprobar la regla de la cadena construyendo directamente la matriz Jacobiana de la función compuesta $\vec{g} \circ \vec{f}$ y comparándola con la matriz que acabamos de determinar.

3.5. Diferenciabilidad de una función de varias variables reales y plano tangente a una superficie

En la primera parte del curso de cálculo hemos visto que una función es derivable en un punto de su dominio si su gráfico admite una recta tangente con pendiente finita en ese punto. Hemos visto también que una función es diferenciable en un punto si existe un entorno del punto en el cual la función se puede aproximar con su recta tangente con menos de un error que tiende a cero más rápidamente que el error que hacemos desplazándonos del punto inicial. Hemos demostrado que estas dos condiciones son equivalentes y que son condiciones suficientes para la continuidad de una función en un punto.

Es totalmente natural preguntarse si vale lo mismo para funciones de más variables. Gracias a la fórmula (3.1) sabemos que para calcular cualquier derivada direccional es suficiente conocer el gradiente, pero el gradiente está compuesto por las derivadas parciales, entonces una función que admite todas las derivadas parciales se puede derivar a lo largo de cualquier dirección. Es esta una condición suficiente para que una función sea continua? La respuesta es: en general no, como prueba el caso de la función

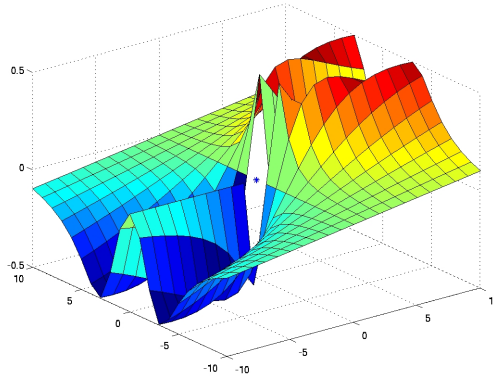
$$f(x, y) = \begin{cases} \frac{xy^2}{x^2+y^4} & \text{si } (x, y) \neq (0, 0) \\ 0 & \text{si } (x, y) = (0, 0) \end{cases}$$

se puede demostrar que esta función es derivable a lo largo de todas las direcciones en $(0, 0)$ pero no es continua en ese punto! Podemos comprobarlo de forma gráfica en la siguiente figura donde se muestra el gráfico de la función.

Esta diferencia importante entre derivabilidad en una variable y en más variables motiva la búsqueda de una definición de derivabilidad ‘global’ y no relacionada con la direccionalidad, que resulte, como en el caso monodimensional, una condición suficiente de continuidad.

Resulta que la definición de derivabilidad ‘global’ para funciones de varias variables coincide con la generalización de la definición de diferenciable, que por eso es conveniente recordar:

Def.: Dada la función $f : (a, b) \subseteq \mathbb{R} \rightarrow \mathbb{R}$, se dice que f es **diferenciable** en $x_0 \in (a, b)$ si la función $h \mapsto o(h) = f(x_0 + h) - f(x_0) - f'(x_0)h$ es un infinitésimo de orden superior con respecto a h para $h \rightarrow 0$, es decir si vale



que:

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

Si consideramos un punto genérico x en un entorno suficientemente pequeño de x_0 y llamamos el incremento $h = x - x_0$, podemos explicitar el significado geométrico del concepto de diferenciabilidad como sigue: una función real de una variable real f es diferenciable en un punto x_0 si $f(x)$ se puede aproximar mediante la función lineal dada por la recta tangente a la gráfica de f en x_0 , $y(x) = f(x_0) + f'(x_0)(x - x_0)$, con un error $o(x - x_0)$ que tiende a 0 más rápidamente que la distancia lineal entre x y x_0 . Dado que la aproximación de f se hace con una función lineal y dado que esta aproximación produce un error ‘despreciable’ solo en puntos cercanos a x_0 , se suele decir que la diferenciación es un proceso de linealización local de una función.

Si queremos generalizar la definición de diferenciabilidad a funciones de varias variables tenemos que hacer unos cambios debido a que ya no tiene sentido considerar una recta tangente, sino que hay que introducir el concepto de *plano* o *hiperplano tangente*; además, hay que tener en consideración el hecho de que las informaciones sobre la rapidez de variación local de una función de varias variables están contenidas en el vector gradiente o en la matriz Jacobiana. Esto implica que la fórmula de la diferenciabilidad tendrá que ser modificada para que las dimensiones resulten coherentes.

Iremos introduciendo la definición que extiende el concepto de diferenciabilidad a funciones de varias variables por pasos: antes consideraremos funciones a valores reales de dos variables, después pasaremos al caso de

n variables y finalmente daremos la definición más general que involucra funciones de \mathbb{R}^n a \mathbb{R}^m .

Def. de diferenciabilidad para funciones reales de 2 variables: la función $f : D \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$ es **diferenciable** en $\vec{x}_0 = (x_0, y_0) \in D$ si la función

$$\|\vec{x} - \vec{x}_0\| \mapsto o(\|\vec{x} - \vec{x}_0\|) = f(\vec{x}) - f(\vec{x}_0) - \langle \nabla f(\vec{x}_0), (\vec{x} - \vec{x}_0) \rangle \quad (3.2)$$

es un infinitésimo de orden superior con respecto a $\|\vec{x} - \vec{x}_0\|$ cuando $\vec{x} \rightarrow \vec{x}_0$, es decir:

$$\lim_{\vec{x} \rightarrow \vec{x}_0} \frac{o(\|\vec{x} - \vec{x}_0\|)}{\|\vec{x} - \vec{x}_0\|} = 0.$$

La ecuación (3.2) significa que, en un entorno suficientemente pequeño de \vec{x}_0 , f se puede escribir como

$$\boxed{f(\vec{x}) = f(\vec{x}_0) + \langle \nabla f(\vec{x}_0), (\vec{x} - \vec{x}_0) \rangle + o(\|\vec{x} - \vec{x}_0\|)},$$

de forma que si desarrollamos el producto escalar y la norma Euclídea del vector desplazamiento tenemos:

$$\boxed{f(\vec{x}) = f(\vec{x}_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0) + o(\sqrt{(x - x_0)^2 + (y - y_0)^2})}.$$

A partir de la igualdad anterior podemos decir que f es diferenciable en \vec{x}_0 si y solo si se puede aproximar con una función lineal de dos variables con menos de un error despreciable descrito por $o(\|\vec{x} - \vec{x}_0\|)$, que tiende a 0 más rápidamente que la distancia Euclídea entre \vec{x} y \vec{x}_0 . Esto traduce el concepto de ‘linealización local’ a funciones de 2 variables.

Es interesante comparar la definición de linealización local en una y dos dimensiones:

$$1D : \quad f(x) = f(x_0) + f'(x_0)(x - x_0) + o(x - x_0)$$

$$2D : \quad f(\vec{x}) = f(\vec{x}_0) + \langle \nabla f(\vec{x}_0), (\vec{x} - \vec{x}_0) \rangle + o(\|\vec{x} - \vec{x}_0\|)$$

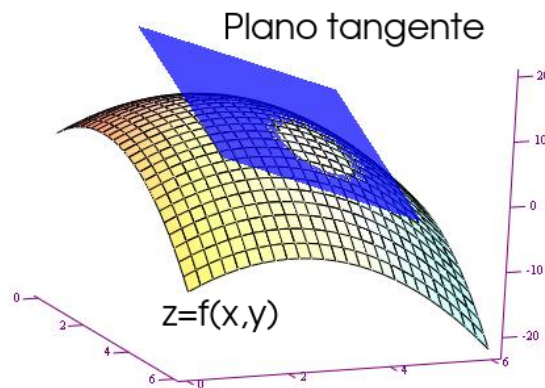
estas fórmulas son conceptualmente idénticas, las únicas diferencias consisten en la sustitución de la derivada por el gradiente, del producto puntual por el producto escalar y de la distancia lineal con la norma, todas operaciones necesarias para que las dimensiones sean coherentes.

Así como la recta de ecuación $y = f(x_0) + f'(x_0)(x - x_0)$ define la recta tangente al gráfico de f en x_0 , podemos definir el **plano tangente** a la superficie dada por el gráfico de f en \vec{x}_0 a través de la ecuación:

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

lo cual nos permite definir de forma equivalente la diferenciabilidad de una función f de 2 variables reales en un punto de su dominio como la condición de que f sea **aproximable por su plano tangente en ese punto con menos de un error despreciable** representado por la función $o(\|\vec{x} - \vec{x}_0\|)$.

Sabemos que una función de una variable no diferenciable en $x = 0$ es el valor absoluto $f(x) = |x| = \sqrt{x^2}$, la correspondiente función de 2 variables no diferenciable en $(0, 0)$ es el cono con vértice en el origen: $f(x, y) = \sqrt{x^2 + y^2}$.



Observación importante: el hecho de poder representar localmente f a través de una función lineal nos permite utilizar las potentes herramientas del álgebra lineal para obtener informaciones sobre la acción de f . El precio que hay que pagar es que eso se puede hacer con precisión solo en un entorno pequeño del punto donde se está linealizando la función. Por eso, si queremos estudiar la acción de una función a través de su linealización, hay que iterar el proceso de linealización a medida que nos desplazamos del punto inicial. Esta consideración es importante y será la base para comprender los métodos numéricos que estudiaremos en el capítulo 5.

Veamos ahora cómo generalizar la definición de diferenciabilidad al caso de funciones de \mathbb{R}^n a \mathbb{R} y de \mathbb{R}^n a \mathbb{R}^m . La generalización a una función

$f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ es inmediata, la fórmula (3.2) sigue siendo válida, con la única diferencia que, ahora, el gradiente es un vector de n componentes, entonces la fórmula explícita se escribe así:

$$f(\vec{x}) = f(\vec{x}_0) + \sum_{i=1}^n \frac{\partial f}{\partial x_i}(\vec{x}_0)(x_i - x_{0,i}) + o\left(\sqrt{\sum_{i=1}^n (x_i - x_{0,i})^2}\right).$$

En cambio, si queremos generalizar la definición a funciones $\vec{f} : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$, tenemos que utilizar la matriz Jacobiana:

$$\vec{f}(\vec{x}) = \vec{f}(\vec{x}_0) + J_{\vec{f}}(\vec{x}_0)(\vec{x} - \vec{x}_0) + \vec{o}(\|\vec{x} - \vec{x}_0\|),$$

la fórmula es coherente a nivel dimensional: si representamos $\vec{f}(\vec{x})$ como un vector columna $m \times 1$, el producto matricial $J_{\vec{f}}(\vec{x}_0)(\vec{x} - \vec{x}_0)$ tiene dimensiones $(m \times n) \times (n \times 1) = m \times 1$ y $\vec{o}(\|\vec{x} - \vec{x}_0\|)$ también es un vector columna $m \times 1$.

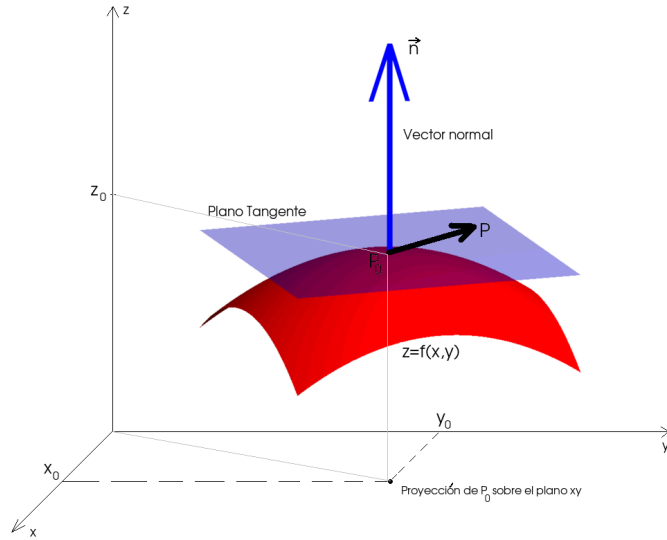
Se puede demostrar que **la diferenciabilidad es condición suficiente para la continuidad**. Determinar la diferenciabilidad de una función en un punto a través de la definición de arriba puede ser una tarea difícil, por eso resulta extremadamente útil el siguiente resultado, que proporciona una **condición suficiente de diferenciabilidad a través del estudio de la continuidad de las derivadas parciales**.

Teorema: si $\vec{f} : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ admite todas las derivadas parciales en $\vec{x}_0 \in D$ y son continuas, entonces \vec{f} es diferenciable en \vec{x}_0 .

3.5.1. Vector normal a una superficie en un punto

En esta subsección consideraremos solo funciones de 2 variables reales $f : D \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$ diferenciables. Queremos llegar a definir el concepto análogo al de vector ortogonal a una curva en un punto para una superficie: hemos visto que el vector ortogonal a una curva en un punto es aquel vector perpendicular a la dirección de la recta tangente a la curva en el mismo punto. Esto sugiere definir el vector perpendicular a una superficie en un punto como aquel vector ortogonal al plano tangente a la superficie en ese punto, como en la figura de abajo. El vector definido de esta forma se suele llamar **vector normal a una superficie en un punto**. La determinación del vector normal es importante en muchos campos de la ciencia (electromagnetismo, procesamiento de imágenes digitales, dinámica de los fluidos, etc.).

Para definir el vector normal consideramos la figura siguiente:



Llamamos $P_0 = (x_0, y_0, f(x_0, y_0))$ al punto de interés y $P = (x, y, z)$ a un punto cualquiera del plano tangente, y de forma que $P \neq P_0$. Para obtener la expresión del vector normal resultará útil comparar la definición del plano tangente que hemos visto antes con la que se obtiene imponiendo que el vector normal, que denotamos con $\vec{n}((x_0, y_0, f(x_0, y_0))) = (a, b, c)$, tenga un producto escalar nulo con un vector cualquiera del plano (por ser vectores ortogonales). Elegimos como vector del plano $P - P_0 = (x - x_0, y - y_0, z - z_0)$, es decir el vector que va de P_0 a P . La ecuación que determina el plano tangente es entonces $\langle \vec{n}, P - P_0 \rangle = 0$, que podemos explicitar así $a(x - x_0) + b(y - y_0) + c(z - f(x_0, y_0)) = 0$, es decir,

$$-cz = -cf(x_0, y_0) + a(x - x_0) + b(y - y_0).$$

Comparando esta última expresión con la del plano tangente a la superficie en $(x_0, y_0, f(x_0, y_0))$,

$$z = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0),$$

podemos ver que las dos expresiones coinciden si $a = f_x(x_0, y_0)$, $b = f_y(x_0, y_0)$, y $c = -1$ o bien si $a = -f_x(x_0, y_0)$, $b = -f_y(x_0, y_0)$, y $c = 1$. Estas condiciones determinan de forma unívoca las componentes del vector normal exceptuando el signo, o dicho de otra forma, el sentido del vector a lo largo de su dirección. Se suele normalizar el vector normal para obtener el vector

unitario normal, para ello tenemos simplemente que dividirlo por su norma:

$$\begin{aligned} \|(-f_x(x_0, y_0), -f_y(x_0, y_0), 1)\| &= \|(f_x(x_0, y_0), f_y(x_0, y_0), -1)\| \\ &= \sqrt{f_x(x_0, y_0)^2 + f_y(x_0, y_0)^2 + 1} \\ &= \sqrt{1 + \|\nabla f(x_0, y_0)\|^2}. \end{aligned}$$

Se definen los vectores:

$$\boxed{\vec{n}(x_0, y_0, f(x_0, y_0)) = \frac{(-f_x(x_0, y_0), -f_y(x_0, y_0), 1)}{\sqrt{1 + \|\nabla f(x_0, y_0)\|^2}}} \quad \text{Normal externa,}$$

$$\boxed{\vec{n}(x_0, y_0, f(x_0, y_0)) = \frac{(f_x(x_0, y_0), f_y(x_0, y_0), -1)}{\sqrt{1 + \|\nabla f(x_0, y_0)\|^2}}} \quad \text{Normal interna.}$$

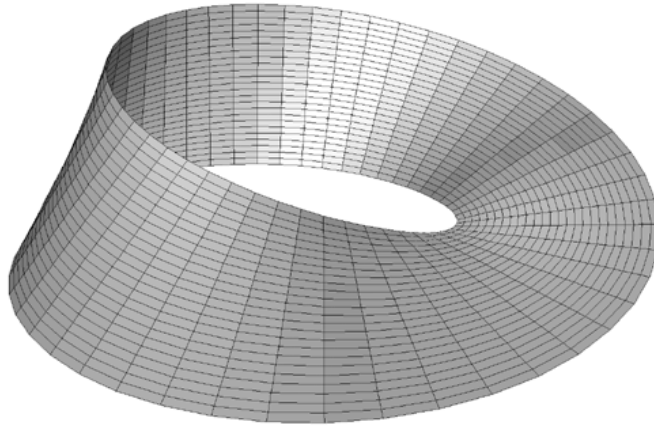
Mediante *externa* e *interna* se definen dos sentidos opuestos sobre la dirección del vector normal.

La elección del nombre externa o interna puede parecer arbitraria, en realidad la definición dada arriba es la más natural si se considera que el vector normal a una superficie esférica es el vector radial que conecta el centro de la esfera con cada punto sobre su superficie. Dicho vector tiene el mismo sentido que la normal externa, mientras que el vector que conecta cada punto de la superficie esférica con el centro tiene el mismo sentido que la normal interna.

Si es posible definir de forma coherente el sentido de la normal en cada punto de una superficie, de forma tal que se pueda desplazar con continuidad a lo largo de la superficie manteniendo el sentido, se dice que la **superficie es orientable** y que la elección del sentido de la normal es una **orientación de la superficie**.

Un ejemplo muy famoso de superficie no orientable es la banda de Möbius (August Ferdinand Möbius (1790-1868), matemático alemán), representada en la figura siguiente. Se puede ver que si empezamos con una normal en un punto y recorremos la banda a lo largo de su perfil, volvemos al punto inicial con la normal que tiene un sentido opuesto al inicial.

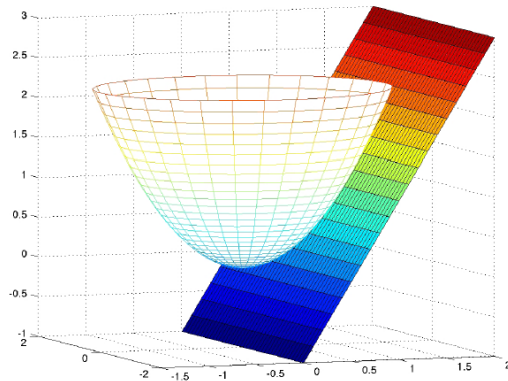
Veamos un ejemplo de cálculo de plano tangente y normal externa: dada la función $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, $f(x, y) = x^2 + y^2$, ($z = x^2 + y^2$ es un paraboloides con vértice en el origen), determinar la ecuación del plano tangente al paraboloides en el punto $(1, 0, f(1, 0))$ y la normal externa al paraboloides en ese punto. Antes de todo calculamos $f(1, 0) = 1$, entonces el punto sobre la superficie del paraboloides donde queremos determinar el plano tangente y



vector normal es $(1, 0, 1)$. Necesitamos las derivadas parciales de f en $(1, 0)$: $f_x(x, y) = 2x$, $f_y(x, y) = 2y$, entonces $f_x(1, 0) = 2$, $f_y(1, 0) = 0$. El plano tangente tiene por ecuación:

$$z = f(1, 0) + f_x(1, 0)(x-1) + f_y(1, 0)(y-0) = 1 + 2(x-1) + 0 = 1 + 2x - 2 = 2x - 1,$$

es decir $z = 2x - 1$. La normal externa es $\vec{n}(1, 0, 1) = \frac{(-2, 0, 1)}{\|(-2, 0, 1)\|} = \frac{1}{\sqrt{5}}(-2, 0, 1)$. La siguiente figura muestra el paraboloides y su plano tangente en $(1, 0, 1)$.



3.6. Derivadas de orden superior y matriz Hessiana

Como para las funciones de una variable real, podemos definir derivadas parciales de orden superior derivando las funciones derivadas parciales. Para entenderlo, consideremos las derivadas parciales de una función de dos variables reales $f(x, y)$ a valores escalares:

$$\begin{aligned} \frac{\partial f}{\partial x} : \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ (x, y) &\longmapsto \frac{\partial f}{\partial x}(x, y) \end{aligned}$$

$$\begin{aligned} \frac{\partial f}{\partial y} : \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ (x, y) &\longmapsto \frac{\partial f}{\partial y}(x, y) \end{aligned}$$

como se ve, en este caso las *funciones derivadas parciales* son funciones de 2 variables y por lo tanto podemos derivarlas parcialmente otra vez:

$$\frac{\partial}{\partial x} \frac{\partial f}{\partial x}(x, y) = \frac{\partial^2 f}{\partial x^2}(x, y) = f_{xx}(x, y)$$

$$\frac{\partial}{\partial y} \frac{\partial f}{\partial x}(x, y) = \frac{\partial^2 f}{\partial y \partial x}(x, y) = f_{yx}(x, y)$$

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

$$\frac{\partial}{\partial y} \frac{\partial f}{\partial y}(x, y) = \frac{\partial^2 f}{\partial y^2}(x, y) = f_{yy}(x, y)$$

Se denotan como

- $f_{xx}(x, y), f_{yy}(x, y)$: **derivadas parciales de orden 2 puras**;
- $f_{yx}(x, y), f_{xy}(x, y)$: **derivadas parciales de orden 2 mixtas o cruzadas**.

Obviamente podemos seguir definiendo derivadas de orden superior hasta cualquier orden y también para funciones de n variables.

Las derivadas cruzadas tienen una propiedad muy bonita que facilita su computación.

*Teorema de Schwarz*³: si $f : D \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$ tiene derivadas primeras y segundas cruzadas en un entorno de $\vec{x}_0 \in D$ y además f_{xy} es continua en

³Karl Hermann Schwarz (1843-1921), matemático alemán.

\vec{x}_0 , entonces también f_{yx} es continua en \vec{x}_0 y vale la igualdad

$$f_{xy}(\vec{x}_0) = f_{yx}(\vec{x}_0).$$

Vale un enunciado análogo intercambiando los papeles de las derivadas cruzadas.

Gracias al teorema de Schwarz podemos calcular solo una derivada segunda mixta, la otra es idéntica, bajo la condición de continuidad.

También las derivadas segundas se pueden organizar en una matriz, cuya importancia veremos en el capítulo siguiente, la **matriz Hessiana**⁴:

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

Si valen las hipótesis del teorema de Schwarz, **la matriz Hessiana es simétrica**.

Ejemplo: calcular la matriz Hessiana de la función $f(x, y) = \sin(x^2y)$. Hay que calcular las derivadas primeras y después las derivadas segundas:

$$f_x(x, y) = 2xy \cos(x^2y), \text{ continua}$$

$$f_y(x, y) = x^2 \cos(x^2y), \text{ continua}$$

entonces vale el teorema de Schwarz:

$$f_{xx}(x, y) = 2y \cos(x^2y) - 4x^2y^2 \sin(x^2y)$$

$$f_{xy}(x, y) = f_{yx}(x, y) = 2x \cos(x^2y) - 2x^3y \sin(x^2y)$$

$$f_{yy}(x, y) = -x^4 \sin(x^2y)$$

entonces la matriz Hessiana es:

$$H_f(x, y) = \begin{bmatrix} 2y \cos(x^2y) - 4x^2y^2 \sin(x^2y) & 2x \cos(x^2y) - 2x^3y \sin(x^2y) \\ 2x \cos(x^2y) - 2x^3y \sin(x^2y) & -x^4 \sin(x^2y) \end{bmatrix}.$$

⁴Ludwig Otto Hesse (1811-1874), matemático alemán.

3.7. La fórmula de Taylor para funciones de varias variables

Recordamos que la fórmula de Taylor para funciones de una variable real permite aproximar una función complicada con la función no lineal más sencilla, un polinomio, con menos de infinitésimos de orden superior, alrededor de cada punto de su dominio donde la función sea derivable hasta un cierto orden n :

$$\begin{aligned} f(x) \underset{x \rightarrow x_0}{=} & f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2!} f''(x_0)(x - x_0)^2 + \dots \\ & \dots + \frac{1}{n!} f^{(n)}(x_0)(x - x_0)^n + o((x - x_0)^n), \end{aligned}$$

donde $\underset{x \rightarrow x_0}{=}$ es una notación que significa que la igualdad vale $\forall x \in U(x_0)$ y donde $o((x - x_0)^n)$ es un infinitésimo superior con respecto a $(x - x_0)^n$:

$$\frac{o((x - x_0)^n)}{(x - x_0)^n} \underset{x \rightarrow x_0}{\rightarrow} 0.$$

$o((x - x_0)^n)$ se llama resto de Peano. Los términos de orden superior de la fórmula de Taylor añaden detalles cada vez más finos a la aproximación de f .

Existe una forma alternativa del resto, llamado de Lagrange, que se puede escribir si f es derivable $n + 1$ veces en todo su dominio con derivadas continuas

$$\begin{aligned} f(x) = & f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2!} f''(x_0)(x - x_0)^2 + \dots \\ & \dots + \frac{1}{n!} f^{(n)}(x_0)(x - x_0)^n + \frac{1}{(n + 1)!} f^{(n+1)}(\xi)(x - x_0)^{n+1}. \end{aligned}$$

Ambas fórmulas se pueden generalizar a funciones de varias variables, pero en este caso la fórmula del resto se vuelve extremadamente complicada y su expresión va más allá del objetivo de este curso. Por eso nos concentramos únicamente sobre el polinomio aproximante de Taylor sin dar la fórmula de los restos.

Teorema: si $f : D \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$ tiene derivadas parciales continuas en D hasta el orden n , entonces, para cada $(x_0, y_0) \in D$ existe un entorno en el cual $f(x_0, y_0)$ se puede aproximar por el polinomio de Taylor de orden n con

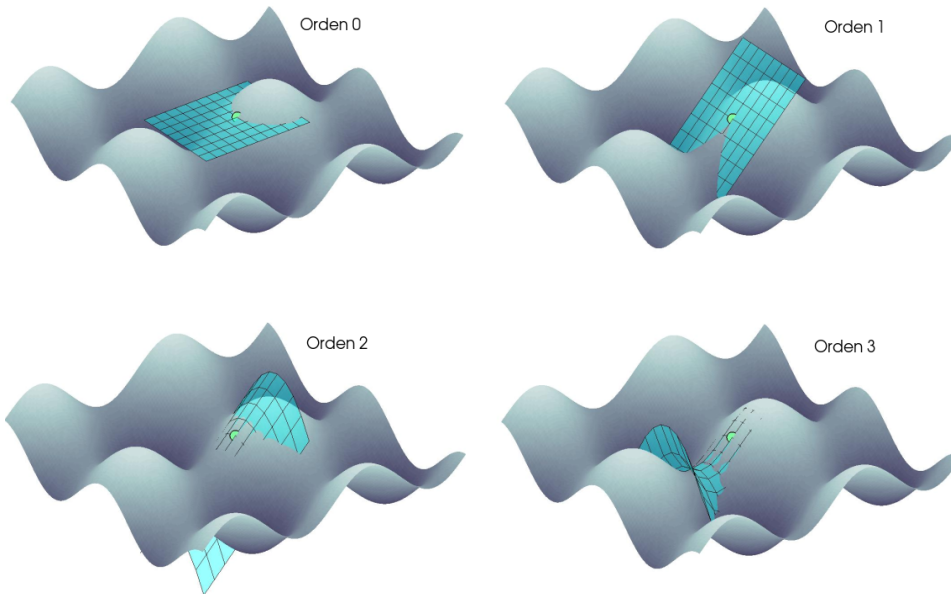
menos de un error despreciable:

$$\begin{aligned}
 f(x, y) &\underset{(x,y) \rightarrow (x_0,y_0)}{\simeq} f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) + \\
 &\frac{1}{2!} (f_{xx}(x_0, y_0)(x - x_0)^2 + 2f_{xy}(x_0, y_0)(x - x_0)(y - y_0) + f_{yy}(x_0, y_0)(y - y_0)^2) + \dots \\
 &\dots + \frac{1}{n!} f^{(n)} \sum_{k=0}^n \binom{n}{k} \frac{\partial f}{\partial x^{n-k} \partial y^k}(x_0, y_0)(x - x_0)^{n-k}(y - y_0)^k.
 \end{aligned}$$

El polinomio de Taylor de orden 2 se puede expresar a través de una fórmula matricial muy compacta:

$$\begin{aligned}
 T_{(x_0, y_0)}^{(2)}(x, y) &= f(x_0, y_0) + \langle J_f(x_0, y_0), (x - x_0, y - y_0) \rangle + \\
 &+ \frac{1}{2} \langle H_f(x_0, y_0)(x - x_0, y - y_0), (x - x_0, y - y_0) \rangle.
 \end{aligned}$$

En las figuras siguientes podemos ver como el polinomio de Taylor aproxima, localmente, cada vez mejor la superficie dada por el gráfico de una función de 2 variables a medida que se sube el grado del polinomio de 0 hasta 3.



Capítulo 4

Optimización libre y vinculada en \mathbb{R}^n

En este capítulo extendaremos las técnicas de análisis de extremos (máximos o mínimos) a funciones de varias variables reales. Veremos que la diferencia más importante con el caso 1D es que en más variables se pueden añadir vínculos no triviales que modifican la búsqueda de los extremos de una función de forma sustancial.

Por sencillez desarrollaremos la teoría solo para funciones de 2 variables, pero los resultados se pueden extender sin problemas a n variables.

Comencemos con los extremos ‘libres’, es decir, sin restricciones o vínculos.

4.1. Extremos libres

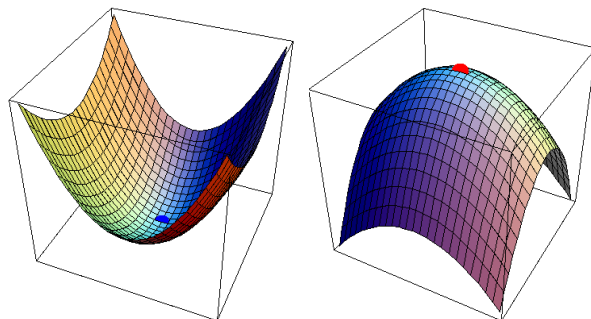
Las definiciones de extremos libres (sin vínculos) para una función de 2 o más variables reales es análoga a la definición dada en el caso de una variable:

Def.: Dada $f : D \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$ y $(x_0, y_0) \in D$ decimos que

- (x_0, y_0) es un punto de **mínimo global** de f si $f(x_0, y_0) \leq f(x, y) \forall (x, y) \in D$;
- (x_0, y_0) es un punto de **máximo global** de f si $f(x_0, y_0) \geq f(x, y) \forall (x, y) \in D$;
- (x_0, y_0) es un punto de **mínimo local** de f si existe un entorno abierto $U((x_0, y_0))$ tal que $f(x_0, y_0) \leq f(x, y) \forall (x, y) \in U((x_0, y_0))$;

- (x_0, y_0) es un punto de **máximo local** de f si existe un entorno abierto $U((x_0, y_0))$ tal que $f(x_0, y_0) \geq f(x, y) \forall (x, y) \in U((x_0, y_0))$.

A continuación podemos ver dos ejemplos gráficos de mínimo y máximo:



El resultado análogo al teorema de Fermat en más variables es el siguiente.

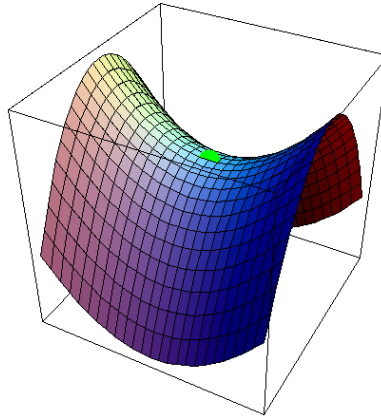
Teorema: $f : D \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$ es parcialmente derivable con respecto a x e y en $(x_0, y_0) \in D$ y (x_0, y_0) es un punto extremo de f , entonces f tiene ambas derivadas parciales nulas en (x_0, y_0) , es decir: $\nabla f(x_0, y_0) = (0, 0)$.

La consecuencia de este teorema es que, exactamente como para funciones de una sola variable real, los extremos de una función de más variables reales se pueden hallar solo:

1. En los puntos del **borde** de D ;
2. En los puntos donde f **no es derivable o diferenciable**, es decir, respectivamente, donde no existe una de las derivadas parciales de f o ambas, y donde el gráfico de f no se puede aproximar con su plano tangente con menos de errores infinitésimos de orden ≥ 2 . Estos puntos se llaman **puntos críticos** de f ;
3. En los puntos donde ambas derivadas parciales de f se anulan, es decir, **en los puntos que anulan el vector gradiente**, estos puntos se llaman **puntos críticos o estacionarios** de f :

$$\nabla f(x_0, y_0) = (0, 0) \iff f_x(x_0, y_0) = f_y(x_0, y_0) = 0.$$

Remarcamos que, como en el caso 1D, el hecho de que f tenga derivadas parciales nulas en un punto es una **condición solo necesaria y no suficiente** para que el punto sea un extremo. En particular, se pueden encontrar casos como el que se muestra en la figura siguiente



que se llama **punto silla**: (x_0, y_0) es un punto silla si es un *máximo local a lo largo de una dirección* y un *mínimo local a lo largo de otra dirección*. Un punto silla es un ejemplo de punto estacionario que no es ni mínimo ni máximo (porque debería serlo en un entorno y no solo a lo largo de una dirección particular!).

Como para las funciones de una variable real, si combinamos la estacionariedad de un punto con una condición sobre la derivada segunda podemos encontrar condiciones suficientes para hallar extremos de una función. Por supuesto, en el caso de 2 variables, no habrá solo una derivada segunda sino 4, que, como sabemos, se pueden organizar en la matriz Hessiana, por eso no debería resultar sorprendente el hecho de que aparezca esta matriz en el siguiente teorema.

Teorema del Hessiano: sea $f : D \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$ una función parcialmente derivable dos veces con respecto a x e y en $(x_0, y_0) \in D$ y sea $H_f(x_0, y_0)$ la matriz Hessiana de f en (x_0, y_0) . Supongamos además que:

1. (x_0, y_0) sea un punto estacionario para f : $\nabla f(x_0, y_0) = (0, 0)$;
2. las derivadas parciales segundas sean continuas en un entorno abierto de (x_0, y_0) ;

entonces:

- si $\det H_f(x_0, y_0) > 0$ y $f_{xx}(x_0, y_0) > 0$, entonces (x_0, y_0) es un mínimo local de f ;
- si $\det H_f(x_0, y_0) > 0$ y $f_{xx}(x_0, y_0) < 0$, entonces (x_0, y_0) es un máximo local de f ;
- si $\det H_f(x_0, y_0) < 0$ entonces (x_0, y_0) es un punto silla de f ;
- si $\det H_f(x_0, y_0) = 0$ no se puede decir nada a priori sobre la naturaleza de (x_0, y_0) y hay que usar otras técnicas.

Ejemplo: determinar los puntos estacionarios de la función $f(x, y) = -x^3 + 4xy - 2y^2 + 1$ y establecer su naturaleza usando el teorema del Hessiano. f es una función polinomial en las variables x, y entonces su dominio es todo \mathbb{R}^2 y f es derivable infinitas veces, entonces los únicos puntos donde f puede tener extremos son los puntos estacionarios, que calculamos anulando su gradiente: $f_x(x, y) = -3x^2 + 4y$, $f_y(x, y) = 4x - 4y$. La ecuación $\nabla f(x, y) = (-3x^2 + 4y, 4x - 4y) = (0, 0)$ corresponde al sistema no lineal:

$$\begin{cases} -3x^2 + 4y = 0; \\ 4x - 4y = 0. \end{cases}$$

No todos los sistemas no lineales resolubles tienen soluciones determinables analíticamente, eso motiva el interés de los métodos numéricos iterativos que presentaremos en el próximo capítulo. Afortunadamente, el sistema de arriba es resoluble analíticamente gracias a una simple observación: en la primera ecuación aparece $4y$ y en la segunda $-4y$, entonces, si sumamos las dos ecuaciones podemos eliminar la dependencia de y : $-3x^2 + 4x = 0$, cuyas soluciones son $x_1 = 0$, $x_2 = 4/3$. Introducimos x_1 e x_2 en la ecuación más simple (la segunda) y encontramos los valores de y correspondientes:

- $4x_1 - 4y = 0 \iff -4y = 0$, o sea $y = y_1 = 0$;
- $4x_2 - 4y = 0 \iff 16/3 - 4y = 0$, o sea $y = y_2 = 4/3$.

Los dos únicos puntos estacionarios son $(x_1, y_1) = (0, 0)$ y $(x_2, y_2) = (4/3, 4/3)$. Determinamos la naturaleza de estos puntos con el método del Hessiano. Tenemos que calcular las derivadas segundas, recordando que en este caso valen las hipótesis del teorema de Schwarz, entonces las derivadas segundas mixtas coinciden. $f_{xx}(x, y) = -6x$, $f_{yy} = -4$, $f_{xy}(x, y) = f_{yx}(x, y) = 4$, entonces la matriz Hessiana es:

$$H_f(x, y) = \begin{bmatrix} -6x & 4 \\ 4 & -4 \end{bmatrix}.$$

Tenemos que analizar el primer elemento de matriz de $H_f(x, y)$ y su determinante cuando (x, y) valen $(0, 0)$ y $(4/3, 4/3)$:

$$(0, 0) : H_f(0, 0) = \begin{bmatrix} 0 & 4 \\ 4 & -4 \end{bmatrix}$$

aprovechamos la ocasión para observar que **cada vez que la matriz Hessiana tiene un elemento de la diagonal nulo, su determinante es negativo**, en particular vale $-f_{xy}^2$, como en este caso que vale -16. Entonces si la matriz Hessiana calculada en un punto estacionario no es nula pero tiene un elemento nulo en la diagonal, el punto estacionario en cuestión es un punto silla, como en este caso es $(0, 0)$.

Pasamos ahora a analizar la matriz Hessiana calculada en el segundo punto estacionario:

$$(4/3, 4/3) : H_f(4/3, 4/3) = \begin{bmatrix} -8 & 4 \\ 4 & -4 \end{bmatrix}, \det H_f(4/3, 4/3) = 32 - 16 = 16 > 0$$

el determinante de la matriz Hessiana es positivo y el primer elemento de matriz es negativo, por lo tanto, en virtud del teorema del Hessiano, podemos decir que $(4/3, 4/3)$ es un punto de máximo local de f . No es un punto de máximo absoluto porque f diverge cuando sus variables tienden al ∞ .

Antes de pasar a examinar los extremos vinculados, citamos, por su importancia, el teorema de Weierstrass para funciones de más variables reales.

Teorema de Weierstrass: si $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ es una función continua y D es compacto, es decir limitado (está enteramente contenido en un entorno esférico en \mathbb{R}^n de radio finito) y cerrado (todas las sucesiones convergentes que se pueden definir en D tienen límite en D mismo), entonces f tiene mínimo y máximo absolutos en D mismo.

4.2. Extremos vinculados

El objetivo del cálculo de los extremos vinculados, o condicionados, es encontrar los extremos de una función bajo alguna restricción. Un ejemplo de la vida real que motiva este análisis está dado por la situación de una empresa que quiere maximizar la producción de un producto bajo una restricción de presupuesto o de material.

Los vínculos que trataremos en este curso son los que se pueden expresar a través de ecuaciones¹ que involucran funciones de las mismas variables de las que dependen las funciones que hay que optimizar. La idea, por lo tanto, es resolver el problema siguiente:

$$\begin{cases} \min_{\vec{x} \in \mathbb{R}^n} \vec{f}(\vec{x}) & \text{o} & \max_{\vec{x} \in \mathbb{R}^n} \vec{f}(\vec{x}) \\ \vec{g}(\vec{x}) = \vec{k}, \end{cases}$$

donde:

- $\vec{f}: D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ es la función que se quiere optimizar;
- $\vec{g}: D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ es la **función vinculante**;
- $\vec{g}(\vec{x}) = \vec{k}$ es la *ecuación vinculante* o simplemente **vínculo**.

A veces, en lugar del valor mínimo o máximo tomado por la función \vec{f} , nos puede interesar conocer el punto \vec{x} en el cual se halla el mínimo o el máximo, a este valor se le llama ‘**argmin**’ o ‘**argmax**’, respectivamente (que proviene de argumento del mínimo o argumento del máximo).

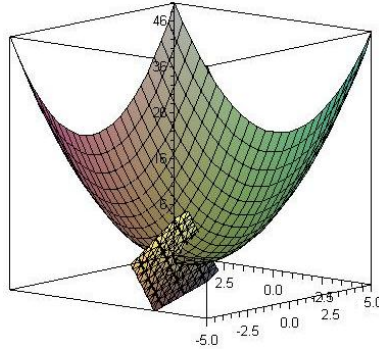
Veremos que la condición que permite encontrar los posibles extremos vinculados ya no es la estacionariedad, es decir, el hecho de que el gradiente de la función se anule, sino que los gradientes de la función que se quiere optimizar y de la función vinculante sean *paralelos*.

Para entender por qué pasa esto presentamos un ejemplo que involucra unas funciones muy sencillas. Consideramos el problema de minimización vinculada siguiente:

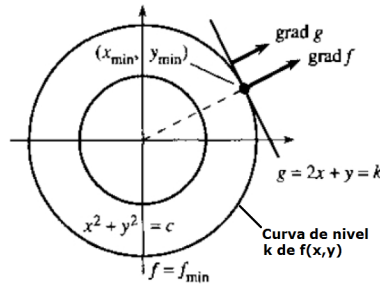
$$\begin{cases} \min_{(x,y) \in \mathbb{R}^2} f(x,y) = x^2 + y^2 \\ 2x + y = 5. \end{cases}$$

La función vinculante es la función lineal $g(x,y) = 2x + y$, el gráfico de f es un paraboloide con vértice (mínimo) en el origen y el vínculo $g(x,y) = 5$ define un plano que intersecta el paraboloide.

¹Los vínculos también podrían estar representados por inecuaciones (desigualdades), pero quedan fuera del alcance de este curso.



Podemos entender el problema con más sencillez si lo proyectamos sobre el plano xy y consideramos las curvas de nivel del paraboloides, que son obviamente círculos concéntricos, con centro $(0, 0)$ y radio variable dado que están dadas por la ecuación: $x^2 + y^2 = c^2$ (siendo c el nivel de la curva). La proyección del plano sobre el eje xy es la recta de ecuación $y = -2x + 5$. La figura siguiente visualiza la geometría del problema.



Existen curvas de nivel que no intersectan la recta y que corresponden a valores de f menores que el valor expresado por el vínculo. Después hay curvas de nivel que la intersectan dos veces porque corresponden a valores de f mayores que el valor expresado por el vínculo. Sin embargo, existe solo una curva de nivel que intersecta la recta vinculante en un solo punto, satisfaciendo el vínculo. Como se ve en la figura 4.1.

La característica que determina esta última situación es que los gradientes de f y de g son paralelos, dado que los dos son ortogonales a dos curvas de nivel que tienen la misma tangente. En este caso la tangente coincide con la recta vinculante, pero en general, si hay un vínculo no lineal, el núcleo del razonamiento no cambiaría, porque la curva de nivel de la función vinculante g podría no ser una recta, pero el gradiente de g sería perpendicular a la recta tangente a la curva de nivel y paralelo al gradiente de f .

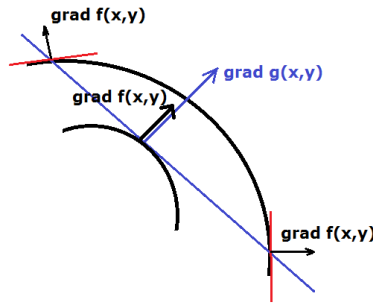


Figura 4.1: Relación entre los gradientes de f y de g en el punto solución al problema del mínimo vinculado. Curvas de nivel de f en negro y curva de nivel 5 de g en azul.

Resumiendo, el sistema que corresponde al mínimo vinculado de f es:

$$\begin{cases} \nabla f(x, y) = \lambda \nabla g(x, y) & \lambda \in \mathbb{R} \setminus \{0\} \\ 2x + y = 5, \end{cases}$$

donde la primera condición expresa el paralelismo entre los dos vectores tangentes.

Si calculamos los gradientes de f y g obtenemos: $\nabla f(x, y) = (2x, 2y)$, $\nabla g(x, y) = (2, 1)$, entonces el sistema se escribe

$$\begin{cases} 2x = 2\lambda \\ 2y = \lambda \\ 2x + y = 5, \end{cases}$$

simplificando y sustituyendo en la tercera ecuación queda

$$\begin{cases} x = \lambda \\ y = \frac{\lambda}{2} \\ 2\lambda + \frac{\lambda}{2} = 5, \end{cases}$$

la última ecuación se cumple para un valor de $\lambda = 2$ que permite explicitar el punto de mínimo vinculado: $(x, y) = (2, 1)$. Éste resulta ser el único mínimo vinculado porque f es una función cuadrática, entonces convexa (posee solo un mínimo). Por supuesto $(2, 1)$ difiere del mínimo absoluto y no vinculado que es $(0, 0)$.

Podemos resumir lo que este ejemplo nos ha enseñado diciendo que, **para encontrar los posibles extremos vinculados de una función**

f , sujeta a la función vinculante g , tenemos que imponer que los gradientes de f y g sean paralelos, además de imponer el vínculo, lo que resulta en un sistema de tres ecuaciones (dos provenientes de las componentes de los gradientes y una del vínculo).

Lagrange encontró una forma muy elegante y sencilla de plantear este sistema introduciendo una función auxiliar que hoy en día llamamos **Lagrangiana** en su honor y que, para funciones de 2 variables reales con un solo vínculo, tiene esta expresión:

$$\boxed{\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda[g(x, y) - k]} \quad k \in \mathbb{R}, \lambda \in \mathbb{R} \setminus \{0\},$$

observamos que \mathcal{L} es una función de 3 variables: x, y, λ , entonces su gradiente es un vector de 3 componentes: $\nabla \mathcal{L}(x, y, \lambda) = (\mathcal{L}_x, \mathcal{L}_y, \mathcal{L}_\lambda)(x, y, \lambda)$.

Teorema: dada $f : D \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$, diferenciable, entonces, una condición necesaria para que valga

$$\begin{cases} \min_{(x,y) \in D} f(x, y) & \text{o} & \max_{(x,y) \in D} f(x, y) \\ g(x, y) = k, \end{cases}$$

es que $\nabla \mathcal{L} = (0, 0, 0)$.

Prueba: calculamos el gradiente de \mathcal{L} :

$$\begin{cases} \mathcal{L}_x(x, y, \lambda) = f_x(x, y) - \lambda g_x(x, y) \\ \mathcal{L}_y(x, y, \lambda) = f_y(x, y) - \lambda g_y(x, y) \\ \mathcal{L}_\lambda(x, y, \lambda) = -g(x, y) + k \end{cases}$$

entonces $\nabla \mathcal{L}(x, y, \lambda) = (\mathcal{L}_x, \mathcal{L}_y, \mathcal{L}_\lambda)(x, y, \lambda) = (0, 0, 0)$ si y solo si

$$\begin{cases} f_x(x, y) = \lambda g_x(x, y) \\ f_y(x, y) = \lambda g_y(x, y) \\ g(x, y) = k \end{cases}$$

las primeras dos condiciones equivalen a pedir que ∇f sea paralelo a ∇g , la tercera es el vínculo, lo cual prueba la tesis del teorema. \square

El sistema determinado por $\nabla \mathcal{L}(x, y, \lambda) = (\mathcal{L}_x, \mathcal{L}_y, \mathcal{L}_\lambda)(x, y, \lambda) = (0, 0, 0)$ es un sistema de 3 ecuaciones con 3 incógnitas que puede ser *lineal* o *no*

lineal, esta es una motivación más para estudiar las técnicas de resolución aproximada de sistemas no lineales que veremos en el capítulo siguiente.

Una vez encontrados los posibles extremos gracias al teorema que acabamos de demostrar, tenemos que determinar si estos puntos son máximos o mínimos. Para hacer esto tenemos una condición suficiente que usa la matriz Hessiana de la Lagrangiana, es decir:

$$H_{\mathcal{L}}(x, y, \lambda) = \begin{bmatrix} \mathcal{L}_{xx} & \mathcal{L}_{xy} & \mathcal{L}_{x\lambda} \\ \mathcal{L}_{yx} & \mathcal{L}_{yy} & \mathcal{L}_{y\lambda} \\ \mathcal{L}_{\lambda x} & \mathcal{L}_{\lambda y} & \mathcal{L}_{\lambda\lambda} \end{bmatrix} (x, y, \lambda)$$

Teorema: si $\nabla \mathcal{L}(x^*, y^*, \lambda^*) = (0, 0, 0)$ y si

$$\begin{cases} \det H_{\mathcal{L}}(x^*, y^*, \lambda^*) > 0 & \implies (x^*, y^*) \text{ es un punto de máximo vinculado} \\ \det H_{\mathcal{L}}(x^*, y^*, \lambda^*) < 0 & \implies (x^*, y^*) \text{ es un punto de mínimo vinculado} \\ \det H_{\mathcal{L}}(x^*, y^*, \lambda^*) = 0 & \implies \text{no se puede decir nada.} \end{cases}$$

Observamos que estas condiciones son diferentes de las contenidas en el teorema del Hessiano, este hecho no tiene que sorprender dado que se trata de dos problemas diferentes que involucran matrices Hessianas diferentes.

La condición necesaria para encontrar extremos vinculados se puede extender al caso de funciones de n variables y con más de un vínculo, vamos a ver cómo. Sea $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ una función diferenciable y sean $g_1, \dots, g_p : D \rightarrow \mathbb{R}$ funciones continuas. Consideramos el problema de optimización vinculada definido por:

$$\begin{cases} \min_{\vec{x} \in D} f(\vec{x}) \quad \text{o} \quad \max_{\vec{x} \in D} f(\vec{x}) \\ g_1(\vec{x}) = k_1, \dots, g_p(\vec{x}) = k_p \quad k_i \in \mathbb{R} \quad \forall i = 1, \dots, p. \end{cases}$$

Definimos el vector p -dimensional $\vec{\lambda} = (\lambda_1, \dots, \lambda_p)$ y la Lagrangiana (función de $n + p$ variables) como sigue:

$$\mathcal{L}(\vec{x}, \vec{\lambda}) = f(\vec{x}) - \sum_{i=1}^p \lambda_i [g_i(\vec{x}) - k_i],$$

como antes, la condición necesaria para que \vec{x}^* sea un extremo vinculado de f es que $\nabla \mathcal{L} = \vec{0}$. La condición suficiente determinada por la matriz Hessiana en cambio no se puede extender a este caso.

Veamos un ejemplo. Determinar el área máxima que puede tener un rectángulo bajo la restricción de que su perímetro sea de 10 metros.

Llamamos x e y a los dos lados del rectángulo. Su perímetro es $p = 2x + 2y$ y su área es $A = xy$. La función f en este caso es el área: $f(x, y) = xy$, la función vinculante es el perímetro: $g(x, y) = 2x + 2y$ y el valor de k es 10. Entonces tenemos que resolver el problema de maximización vinculada siguiente:

$$\begin{cases} \text{máx}_{(x,y) \in \mathbb{R}^2} xy \\ 2x + 2y = 10. \end{cases}$$

Construimos la función Lagrangiana:

$$\mathcal{L}(x, y, \lambda) = xy - \lambda(2x + 2y - 10),$$

e igualamos a cero su gradiente:

$$\nabla \mathcal{L}(x, y, \lambda) = (y - 2\lambda, x - 2\lambda, -2x - 2y + 10) = (0, 0, 0),$$

que se cumple si y solo si

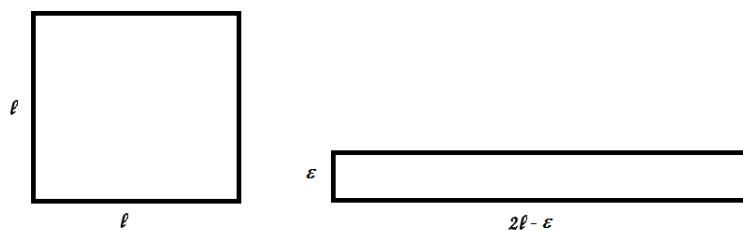
$$\begin{cases} x = 2\lambda, \\ y = 2\lambda, \\ 2x + 2y = 10. \end{cases}$$

Sustituyendo $x = 2\lambda$ e $y = 2\lambda$ en el vínculo encontramos $\lambda = 5/4$, así que el único punto que puede ser un extremo vinculado de la función área es $(x^*, y^*) = (5/2, 5/2)$. Comprobemos que sea un máximo con la condición suficiente dada por la matriz Hessiana, que se puede calcular muy fácilmente en este caso:

$$H_{\mathcal{L}}(x, y, \lambda) = \begin{bmatrix} 0 & 1 & -2 \\ 1 & 0 & -2 \\ -2 & -2 & 0 \end{bmatrix}$$

es una matriz constante con respecto a x, y, λ y su determinante vale 8 para cada punto, así que también para $(x^*, y^*) = (5/2, 5/2)$, que resulta ser un punto de máximo vinculado para la función área con la restricción de tener un perímetro de 10 metros.

Por lo tanto el área máxima compatible con un perímetro de 10 metros es $25/4 \text{ m}^2$, que es el área de un cuadrado de 2.5 metros de lado. La información que hemos obtenido no es trivial: **la geometría que maximiza un área rectangular con un perímetro de longitud fija es la geometría del cuadrado!** En las clases de prácticas y seminarios extenderemos este resultado a la geometría tridimensional.



La figura de arriba muestra por qué la geometría cuadrática maximiza el área: el perímetro del cuadrado de lado l es $4l$, igual que el perímetro del rectángulo de lados $2l - \epsilon$ y ϵ , pero el área del cuadrado es l^2 , mientras que el área del rectángulo es $(2l - \epsilon)\epsilon = 2l\epsilon - \epsilon^2$. Si hacemos tender ϵ a cero, podemos descartar el término ϵ^2 y nos quedamos con un área de $2l\epsilon < l^2$.

Capítulo 5

Métodos numéricos iterativos para funciones de varias variables

En este capítulo examinaremos la extensión del método de Newton para encontrar ceros de funciones al caso n -dimensional y lo aplicaremos a la resolución aproximada de sistemas no lineales. Veremos además el método del descenso/ascenso por gradiente para acercarse a extremos locales de una función. Ya hemos visto más de una motivación para estudiar estas técnicas en los capítulos anteriores.

Observación: excepcionalmente, en este capítulo, no escribiremos el símbolo de vector para evitar de complicar aún más la notación. Resultará claro desde el contexto si un objeto es un vector o un escalar.

5.1. Método de Newton para funciones de n variables reales

En esta sección veremos cómo se puede extender el método de Newton 1-D a funciones de $n > 1$ variables a valores reales y vectoriales y después veremos cómo se puede utilizar este algoritmo para aproximar las soluciones de sistemas no lineales y para el cálculo de extremos con restricciones.

Método de Newton para funciones reales de varias variables reales

Consideremos una función real y diferenciable $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ de n variables reales que representamos como un vector columna:

$$x = (x_1, \dots, x_n)^t = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

donde 't' define la transposición matricial. Queremos aproximar el valor de un cero de f . El método de Newton 1-D, o sea $x^k = x^{k-1} - \frac{f(x^{k-1})}{f'(x^{k-1})}$, no se puede extender de forma inmediata sustituyendo $f'(x^{k-1})$ por el gradiente de f porque, dimensionalmente, la fórmula no tendría sentido, siendo x^k , x^{k-1} y $\nabla f(x^{k-1})$ vectores de \mathbb{R}^n y $f(x^{k-1})$ un número real. Sin embargo, observemos que si reescribimos el método de Newton multiplicando ambos lados de la fórmula 1-D por $f'(x^{k-1})$ obtenemos:

$$f'(x^{k-1})x^k = f'(x^{k-1})x^{k-1} - f(x^{k-1}), \quad (5.1)$$

veremos en breve que esta versión del método de Newton se puede extender al caso de funciones reales de varias variables sustituyendo $f'(x^{k-1})$ por $\nabla f(x^{k-1})$ y considerando el producto escalar en lugar del producto puntual.

Para llegar a esa fórmula consideramos, como en el caso de funciones de una sola variable, la linealización de f . En este caso la linealización local de f en x^0 se realiza aproximando la gráfica de f con su hiperplano tangente en x^0 , que, como sabemos, se halla a través del polinomio de Taylor de orden 1 con centro en $x^0 = (x_1^0, \dots, x_n^0)^t$:

$$T_{x^0}(x) = f(x^0) + \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x^0)(x_i - x_i^0) = f(x^0) + \langle \nabla f(x^0), x - x^0 \rangle,$$

donde $\nabla f(x^0)$ es el gradiente de f en x^0 escrito como vector columna: $\nabla f(x^0) = (f_{x_1}(x^0), \dots, f_{x_n}(x^0))^t$ y $\langle \nabla f(x^0), x - x^0 \rangle$ es el *producto escalar* entre el gradiente de f en x^0 y el vector $x - x^0$.

Sea $x^1 \in \mathbb{R}^n$ un cero de $T_{x^0}(x)$:

$$0 = T_{x^0}(x^1) = f(x^0) + \langle \nabla f(x^0), x^1 - x^0 \rangle = f(x^0) + \langle \nabla f(x^0), x^1 \rangle - \langle \nabla f(x^0), x^0 \rangle,$$

donde en la última igualdad se ha usado la linealidad del producto escalar. Resolviendo con respecto a x^1 obtenemos:

$$\langle \nabla f(x^0), x^1 \rangle = \langle \nabla f(x^0), x^0 \rangle - f(x^0).$$

que es un sistema *lineal* que se puede resolver exactamente con las técnicas vistas en el curso de álgebra, o de forma aproximada con los métodos de Jacobi y Gauss-Seidel. Este es el primer paso del algoritmo de Newton para funciones de más variables, el paso k -ésimo es:

$$\boxed{\langle \nabla f(x^{k-1}), x^k \rangle = \langle \nabla f(x^{k-1}), x^{k-1} \rangle - f(x^{k-1})},$$

que es la extensión natural de la fórmula (5.1) a funciones de varias variables.

Método de Newton para funciones de varias variables reales a valores vectoriales

El paso siguiente es la extensión de la fórmula iterativa de Newton al caso de funciones $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$,

$$f(x) = (f_1(x), \dots, f_m(x))^t = \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix} \quad x \in D,$$

con $n, m > 1$ y $f_1, \dots, f_m : D \rightarrow \mathbb{R}$ son las funciones componentes de f . Recordemos que, en este caso, la derivada de f en un punto $x \in D$ es la matriz Jacobiana $J_f(x)$, matrix $m \times n$ con filas dadas por los gradientes de las funciones componentes calculados en x :

$$J_f(x) = \begin{bmatrix} \nabla f_1(x) \\ \vdots \\ \nabla f_m(x) \end{bmatrix} = \begin{bmatrix} \partial_{x_1} f_1(x) & \cdots & \partial_{x_n} f_1(x) \\ \vdots & \ddots & \vdots \\ \partial_{x_1} f_m(x) & \cdots & \partial_{x_n} f_m(x) \end{bmatrix}.$$

En este caso veremos que tanto la fórmula $x^k = x^{k-1} - \frac{f(x^{k-1})}{f'(x^{k-1})}$ como la fórmula $f'(x^{k-1})x^k = f'(x^{k-1})x^{k-1} - f(x^{k-1})$ se pueden generalizar sustituyendo $f(x^{k-1})$ por $f(x^{k-1})^t$, $f'(x^{k-1})$ por $J_f(x^{k-1})$, $\frac{1}{f'(x^{k-1})}$ por $J_f(x^{k-1})^{-1}$ (cuando $(n = m)$) y los productos puntuales por los productos matriciales.

Como en el caso 1-D y en el caso de funciones de varias variables a valores reales, también en este caso consideramos la linealización local de f en un

punto inicial x^0 a través del polinomio de Taylor de orden 1 con centro en x^0 :

$$T_{x^0} f(x) = f(x^0)^t + J_f(x^0)(x - x^0), \quad (5.2)$$

observemos que las dimensiones son correctas, dado que $f(x^0)$ es un vector columna ($m \times 1$) y el resultado de $J_f(x^0)(x - x^0)$, $(m \times n) \times (n \times 1)$, es un vector columna de dimensión $(m \times 1)$.

Si ahora consideramos un cero x^1 de $T_{x^0} f$, es decir, $T_{x^0}(x^1) = (0, \dots, 0)^t$, cero en \mathbb{R}^m , y utilizamos la linealidad del producto matricial, podemos escribir la fórmula (5.2) como sigue:

$$0 = f(x^0) + J_f(x^0)(x^1 - x^0) = f(x^0) + J_f(x^0)x^1 - J_f(x^0)x^0$$

o también

$$J_f(x^0)x^1 = J_f(x^0)x^0 - f(x^0),$$

que es el primer paso del método de Newton. La iteración k -ésima se escribe:

$$\boxed{J_f(x^{k-1})x^k = J_f(x^{k-1})x^{k-1} - f(x^{k-1})}, \quad (5.3)$$

que es un sistema lineal de m ecuaciones con n incógnitas. Si $n = m$ y $\det(J_f(x^{k-1})) \neq 0$, entonces la matriz Jacobiana se puede invertir y la fórmula antecedente se puede escribir como:

$$\boxed{x^k = x^{k-1} - \left(J_f(x^{k-1})\right)^{-1} f(x^{k-1})} \quad n = m, \det(J_f(x^{k-1})) \neq 0 \quad (5.4)$$

Si $n \gg 1$, la expresión (5.3) resulta computacionalmente más viable que la (5.4) porque la inversión de una matriz de grandes dimensiones es una operación costosa. En ambos casos, es oportuno observar que *la técnica de Newton para encontrar ceros de funciones de varias variables reales a valores vectoriales consiste en resolver iterativamente un cierto número \bar{k} de sistemas lineales.*

El valor de \bar{k} de parada se puede determinar con las condiciones vistas para el método de Newton 1-D: definida una tolerancia $T > 0$,

1. \bar{k} es el mínimo k tal que

$$\frac{\|x^k - x^{k-1}\|}{\|x^k\|} \leq T,$$

(la distancia relativa entre las soluciones está limitada por T);

2. \bar{k} es el mínimo k tal que

$$\|f(x^k)\| \leq T,$$

(la norma de f en x^k , es decir $\|f(x^k)\| = \sqrt{f_1(x^k)^2 + \dots + f_m(x^k)^2}$, está entre 0 y T);

3. definimos un número máximo de iteraciones $\bar{k} = k_{\text{máx}}$ (se usa cuando se prefiere un algoritmo rápido a uno preciso o como condición de parada cuando el método no converge).

Método de Newton para aproximar las soluciones de cualquier sistema de ecuaciones (lineales o no)

Lo que acabamos de ver se puede aplicar directamente para aproximar las soluciones de sistemas lineales o no lineales de n ecuaciones con n incógnitas.

Consideremos

- $x = (x_1, \dots, x_n)^t = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix};$

- n funciones a valores reales $f_j(x) = f_j \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, j = 1, \dots, n;$

- el sistema

$$\begin{cases} f_1(x) = 0 \\ \vdots \\ f_n(x) = 0; \end{cases}$$

- *interpretamos las funciones f_1, \dots, f_m como las funciones componentes de la función vectorial*

$$F(x) = (f_1(x), \dots, f_n(x))^t = \begin{pmatrix} f_1(x) \\ \vdots \\ f_n(x) \end{pmatrix}.$$

Esta interpretación nos permite aplicar las fórmulas de Newton vistas antes para aproximar las soluciones del sistema, dado que aproximar un cero de F equivale a aproximar los ceros de todas sus funciones componentes, es decir resolver el sistema.

La solución iterativa en el paso k del sistema anterior se puede determinar a través de la resolución (exacta o aproximada) del sistema *lineal*:

$$\boxed{J_F(x^{k-1})x^k = J_F(x^{k-1})x^{k-1} - F(x^{k-1})}, \quad (5.5)$$

en la variable vectorial x^k , o bien, si $\det(J_F(x^{k-1})) \neq 0$, con la fórmula

$$\boxed{x^k = x^{k-1} - \left(J_F(x^{k-1})\right)^{-1} F(x^{k-1})}, \quad (5.6)$$

siendo $J_F(x^{k-1})$ la matriz Jacobiana calculada en x^{k-1} :

$$J_F(x^{k-1}) = \begin{bmatrix} \partial_{x_1} f_1(x^{k-1}) & \cdots & \partial_{x_n} f_1(x^{k-1}) \\ \vdots & \ddots & \vdots \\ \partial_{x_1} f_n(x^{k-1}) & \cdots & \partial_{x_n} f_n(x^{k-1}) \end{bmatrix}.$$

Terminemos observando que la resolución de un problema no lineal a través de la iteración de métodos lineales es una técnica que se encuentra en muchos campos diferentes de la matemática y de la ciencia. Muchas veces *se comienza con un modelo lineal (impreciso) de un problema y se refina ese modelo iterando correcciones lineales.*

Aplicación del método de Newton para aproximar la solución de sistemas a la optimización con restricciones

Ya hemos visto que para un problema de extremos condicionados del tipo:

$$\min_{x \in \mathbb{R}^n} f(x) \text{ o } \max_{x \in \mathbb{R}^n} f(x)$$

con las restricciones de igualdad $g_1(x) = k_1, \dots, g_m(x) = k_m$, los posibles extremos de f vinculados a esas restricciones son los puntos que anulan el gradiente de la función *Lagrangiana*. Se trata de una función de las n variables $x = (x_1, \dots, x_n)$ originales del problema y de los m multiplicadores de Lagrange $\lambda = (\lambda_1, \dots, \lambda_m)$, $\mathcal{L} : \mathbb{R}^{n+m} \rightarrow \mathbb{R}$, $\mathcal{L}(x, \lambda) =$

$f(x) - \sum_{j=1}^m \lambda_j (g_j(x) - k_j)$. Para calcular los puntos críticos de la Lagrangiana hay que resolver el sistema (en general *no lineal*):

$$\begin{cases} \mathcal{L}_{x_1}(x, \lambda) = 0 \\ \vdots \\ \mathcal{L}_{x_n}(x, \lambda) = 0 \\ \mathcal{L}_{\lambda_1}(x, \lambda) = 0 \\ \vdots \\ \mathcal{L}_{\lambda_m}(x, \lambda) = 0. \end{cases}$$

Para resolverlo podemos usar el método de Newton que acabamos de ver. Como antes, definimos la función vectorial

$$F(x, \lambda) = \begin{pmatrix} \mathcal{L}_{x_1}(x, \lambda) \\ \vdots \\ \mathcal{L}_{x_n}(x, \lambda) \\ \mathcal{L}_{\lambda_1}(x, \lambda) \\ \vdots \\ \mathcal{L}_{\lambda_m}(x, \lambda) \end{pmatrix}.$$

Su matriz Jacobiana coincide con la matriz Hessiana de la función Lagrangiana:

$$H_{\mathcal{L}}(x, \lambda) = \begin{pmatrix} \mathcal{L}_{x_1x_1} & \mathcal{L}_{x_1x_2} & \cdots & \mathcal{L}_{x_1x_n} & \mathcal{L}_{x_1\lambda_1} & \cdots & \mathcal{L}_{x_1\lambda_m} \\ \mathcal{L}_{x_2x_1} & \mathcal{L}_{x_2x_2} & \cdots & \mathcal{L}_{x_2x_n} & \mathcal{L}_{x_2\lambda_1} & \cdots & \mathcal{L}_{x_2\lambda_m} \\ \vdots & \vdots & \cdots & \cdots & \cdots & \cdots & \vdots \\ \mathcal{L}_{x_nx_1} & \mathcal{L}_{x_nx_2} & \cdots & \mathcal{L}_{x_nx_n} & \mathcal{L}_{x_n\lambda_1} & \cdots & \mathcal{L}_{x_n\lambda_m} \\ \mathcal{L}_{\lambda_1x_1} & \mathcal{L}_{\lambda_1x_2} & \cdots & \mathcal{L}_{\lambda_1x_n} & \mathcal{L}_{\lambda_1\lambda_1} & \cdots & \mathcal{L}_{\lambda_1\lambda_m} \\ \mathcal{L}_{\lambda_2x_1} & \mathcal{L}_{\lambda_2x_2} & \cdots & \mathcal{L}_{\lambda_2x_n} & \mathcal{L}_{\lambda_2\lambda_1} & \cdots & \mathcal{L}_{\lambda_2\lambda_m} \\ \vdots & \vdots & \cdots & \cdots & \cdots & \cdots & \vdots \\ \mathcal{L}_{\lambda_mx_1} & \mathcal{L}_{\lambda_mx_2} & \cdots & \mathcal{L}_{\lambda_mx_n} & \mathcal{L}_{\lambda_m\lambda_1} & \cdots & \mathcal{L}_{\lambda_m\lambda_m} \end{pmatrix} (x, \lambda).$$

Si denotamos $z = (x, \lambda)^t = (x_1, \dots, x_n, \lambda_1, \dots, \lambda_m)^t$, gracias al método de Newton podemos calcular el valor aproximado en el paso k del punto crítico de la Lagrangiana \mathcal{L} aproximando la solución del sistema *lineal*

$$H_{\mathcal{L}}(z^{k-1})z^k = H_{\mathcal{L}}(z^{k-1})z^{k-1} - F(z^{k-1}).$$

o bien, si $\det(H_{\mathcal{L}}(z^{k-1})) \neq 0$, a través de la fórmula

$$z^k = z^{k-1} - \left(H_{\mathcal{L}}(z^{k-1}) \right)^{-1} F(z^{k-1}).$$

5.2. Método del descenso/ascenso por gradiente

Sea $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ una función real de n variables reales y $x^0 \in D$. Recordemos que recta en \mathbb{R}^n que pasa por x^0 con dirección $u \in \mathbb{R}^n$, $\|u\| = 1$, es el lugar geométrico $r_{x^0, u}$ definido por

$$r_{x^0, u} = \{x \in \mathbb{R}^n \mid x = x^0 + tu, t \in \mathbb{R}\},$$

cuando $t = 0$ estamos en el punto x^0 , cuando $|t| > 0$ nos alejamos de x^0 sobre la recta exactamente de la cantidad $|t|$: $d(x^0 + tu, x^0) = \|x^0 + tu - x^0\| = \|tu\| = |t|\|u\| = |t|$. Llamamos $|t|$ el *paso* del alejamiento de x^0 sobre la recta $r_{x^0, u}$.

Durante el curso hemos visto que si nos alejamos de x^0 siguiendo rectas con diferentes direcciones, f varía (en general) de forma diferente. En particular, localmente, es decir, en un entorno abierto $U_r(x^0)$ de radio $r > 0$, se cumple que¹:

- la *dirección de máximo decrecimiento* de f en $x^0 \in D$ es la dirección $-\frac{\nabla f(x^0)}{\|\nabla f(x^0)\|}$ del vector opuesto al gradiente de f en x^0 ;
- la *dirección de máximo crecimiento* de f en $x^0 \in D$ es la dirección $+\frac{\nabla f(x^0)}{\|\nabla f(x^0)\|}$ del gradiente de f en x^0 .

Estas consideraciones permiten formular el método numérico de minimización (o maximización) más fácil: el descenso (o ascenso) por gradiente.

Algoritmo de descenso por gradiente:

- Sea x^0 un punto cualquiera en el dominio de f y x^* un punto donde f alcanza un mínimo.
- Nos alejamos de x^0 siguiendo la recta que pasa por x^0 y que tiene dirección opuesta al gradiente de f en x^0 con un paso de longitud \tilde{t}^0 :

$$x^1 = x^0 - \tilde{t}^0 \frac{\nabla f(x^0)}{\|\nabla f(x^0)\|} \equiv x^0 - t^0 \nabla f(x^0),$$

hemos incorporado la normalización en la longitud del incremento definiendo $t^0 = \frac{\tilde{t}^0}{\|\nabla f(x^0)\|}$, eso significa simplemente que el paso de alejamiento con respecto al punto inicial x^0 escrito con respecto a t^0 es:

¹Hay que observar que fuera del entorno esférico $U_r(x^0)$ podría no ser verdadero que $\pm \frac{\nabla f(x^0)}{\|\nabla f(x^0)\|}$ es la dirección que maximiza el crecimiento/decrecimiento.

$t^0 = t^0 \|\nabla f(x^0)\|$. t^0 tiene que ser tal que $f(x^1) < f(x^0)$ para que la función decrezca con la iteración. Semejante t^0 siempre existe en un entorno suficientemente pequeño de x^0 por lo que hemos dicho arriba, pero si se escoge un t^0 demasiado grande podríamos salir de ese entorno y la función podría hasta crecer. Si la función no decrece hay que repetir el cálculo de x^1 con un t^0 más pequeño (por ejemplo la mitad).

- Ahora nos alejamos de x^1 siguiendo la recta que pasa por x^0 y que tiene dirección opuesta al gradiente de f en x^1 con un paso de longitud t^1 :

$$x^2 = x^1 - t^1 \nabla f(x^1),$$

de nuevo, el valor de t^1 tiene que ser tal que $f(x^2) < f(x^1)$, y si no se cumple hay que reducir el valor de t^1 .

- Iterando este razonamiento el paso genérico k será:

$$\boxed{x^k = x^{k-1} - t^{k-1} \nabla f(x^{k-1})}, \quad \text{Descenso por gradiente, paso } k \quad (5.7)$$

con t^{k-1} tal que $f(x^k) < f(x^{k-1})$, $\forall k \geq 1$.

La fórmula del ascenso por gradiente es análoga, con las únicas diferencias que aparece el signo + delante del gradiente y que en cada paso tenemos que escoger el paso t^{k-1} de forma que el valor de la función en cada iteración crezca:

$$\boxed{x^k = x^{k-1} + t^{k-1} \nabla f(x^{k-1})}, \quad \text{Ascenso por gradiente, paso } k \quad (5.8)$$

con t^{k-1} tal que $f(x^k) > f(x^{k-1}) \forall k \geq 1$.

Los valores de t pueden también ser constantes, en ese caso se dice que el método del descenso por gradiente tiene *paso único*, en este caso $t^{k-1} = t$ en las fórmulas (5.7) y (5.8). Si se permite a t variar en cada paso k , el método se denomina *multipaso*.

Un esquema numérico del tipo multipaso para el descenso por gradiente puede ser el siguiente:

1. Escogemos un valor inicial x^0 (en el dominio de f) y un valor de t .
2. Iteramos hasta la convergencia (o hasta un número máximo de iteraciones):

$$\bar{x}^k = x^{k-1} - t \nabla f(x^{k-1})$$

Mientras $f(\bar{x}^k) \geq f(x^{k-1})$

$t = \alpha t$ (donde $\alpha < 1$, por ejemplo $\alpha = 0,8$)

$\bar{x}^k = x^{k-1} - t\nabla f(x^{k-1})$

$x^k = \bar{x}^k$

Existen otras variantes donde el paso t^k se busca de forma distinta. Para el método de ascenso por gradiente bastaría con usar el signo $+$ en lugar del $-$ delante de t y comprobar la desigualdad \leq (en lugar de \geq).

Gráficamente, sabiendo que el gradiente de f es ortogonal a las curvas de nivel de f , el descenso (o ascenso) por gradiente presenta las estructuras que muestran los dos ejemplos de la Figura 1.

Las *condiciones de parada* del algoritmo de descenso o ascenso por gradiente son las siguientes:

- El error relativo entre dos iteraciones sucesivas está por debajo de una cierta tolerancia $T > 0$:

$$\frac{\|x^k - x^{k-1}\|}{\|x^k\|} < T;$$

- k ha llegado a un número máximo de iteraciones $k_{\text{máx}}$. Con esta condición podríamos no haber llegado al extremo de f , entonces este criterio de parada se usa cuando se prefiere rapidez a precisión. Normalmente se fija $k_{\text{máx}}$ a un valor suficientemente alto para asegurarnos de llegar a una solución con buena precisión en caso de convergencia del algoritmo y que evite que el algoritmo itere de forma infinita en caso de no convergencia de éste.

Observemos que si la función tiene más extremos de la misma naturaleza, por ejemplo dos o más mínimos, dependiendo de la elección de x^0 podríamos encontrar un *mínimo local* y no el mínimo global.

El *descenso por gradiente converge linealmente* (véase el archivo sobre la convergencia de los algoritmos iterativos), siempre que la función tenga al menos un extremo, entonces es *más lento que el método de Newton*; sin embargo el algoritmo de Newton para converger necesita una elección del punto inicial x^0 más precisa.

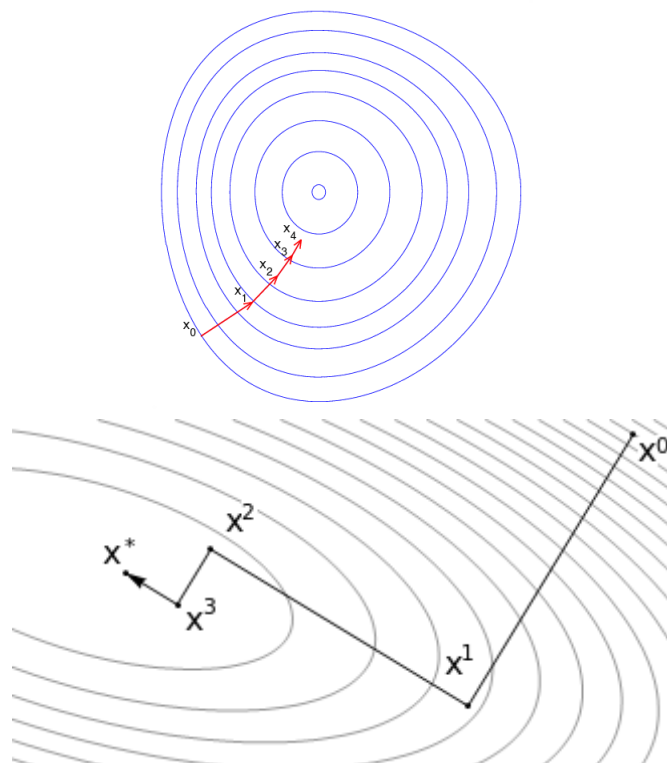


Figura 5.1: Dos ejemplos de la evolución de la aproximación de la solución x^k en el método del descenso de gradiente y su relación con las curvas de nivel. En el ejemplo de la derecha el paso t se busca de tal forma que se llegue al mínimo de f en la dirección de búsqueda de cada iteración (es decir, $\pm \nabla f(x^{k-1}) / \|\nabla f(x^{k-1})\|$). Con esta variante del método, llamada búsqueda lineal, el método presenta la típica estructura de ‘zig-zag’ que muestra la imagen de la derecha y el algoritmo correspondiente se llama, en inglés, ‘steepest descent’.

Aplicación del método del descenso por gradiente a la aproximación de soluciones de sistemas no lineales

El método del descenso por gradiente se puede utilizar para aproximar las soluciones de sistemas no lineales en sustitución del método de Newton (por ejemplo si es difícil determinar el punto inicial del método de Newton).

Consideremos el sistema

$$\begin{cases} f_1(x) = 0 \\ \vdots \\ f_n(x) = 0; \end{cases}$$

con $x \in \mathbb{R}^n$ y f_i diferenciable $\forall i = 1, \dots, n$. Definimos la función $F : \mathbb{R}^n \rightarrow \mathbb{R}$,

$$F(x) = \sum_{i=1}^n f_i(x)^2,$$

F es una función cuadrática, entonces tiene un solo mínimo x^* , en el cual toma el valor 0: $F(x^*) = 0$, pero esto es posible si y solo si $f_i(x^*) = 0 \forall i = 1, \dots, n$, entonces x^* es la solución del sistema anterior. Aplicando el método del descenso por gradiente a F tenemos que el paso de actualización en cada iteración es:

$$x^k = x^{k-1} - t^{k-1} \nabla F(x^{k-1}), \quad (5.9)$$

donde $x^k \in \mathbb{R}^n$ (al igual que x) y es por lo tanto un vector de n componentes (nótese que $\nabla F(x^{k-1})$ es también un vector de n componentes). Iterando el cálculo dado por la fórmula (5.9) podremos aproximar el valor de x^* , es decir, la solución al sistema de ecuaciones.

Observamos que el hecho de que F se defina como la suma de los cuadrados de las funciones f_i es crucial para poder asegurar la convergencia a x^* y no a otro eventual mínimo local.

Terminemos con una observación técnica: dado que el método de Newton necesita que se cumplan unas condiciones sobre el vector de inicialización x^0 para converger, puede ser útil usar como x^0 el vector proporcionado por unas iteraciones del método del descenso por gradiente. Encadenar descenso (o ascenso) por gradiente y método de Newton es una estrategia genérica que no se limita solo a la resolución de sistemas de ecuaciones.

Capítulo 6

Cálculo integral para funciones de 2 y 3 variables: integrales dobles y triples

En este capítulo extenderemos la definición de integral a funciones de 2 y 3 variables. Comenzamos con 2 variables.

6.1. Integrales dobles

En la primera parte del curso vimos la definición de Riemann de las integrales definidas de funciones continuas $f : [a, b] \rightarrow \mathbb{R}$: $\int_a^b f(x)dx$ es el límite de las sumas de Cauchy-Riemann:

$$\int_a^b f(x)dx = \lim_{n \rightarrow +\infty} \sum_{k=1}^n \frac{b-a}{n} f(t_k)$$

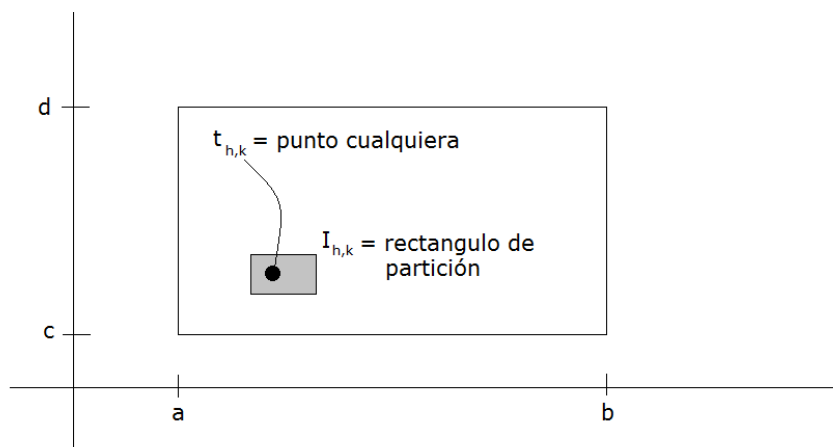
donde $\frac{b-a}{n}$ es la amplitud de los intervalos de la partición de $[a, b]$ y t_k son n puntos cualesquiera elegidos en cada uno de los intervalos de la partición.

La generalización a funciones de 2 variables pasa por diferentes niveles de dificultad, comenzaremos con el nivel más bajo y extenderemos el resultado a casos cada vez más complicados.

6.1.1. Integrales dobles en rectángulos

Consideramos una función de 2 variables definida sobre un rectángulo en el plano, es decir: $f : [a, b] \times [c, d] \rightarrow \mathbb{R}$, donde \times es el producto cartesiano.

Si consideramos una partición de $[a, b]$ en n subintervalos de longitud $\frac{b-a}{n}$ y una de $[c, d]$ en n subintervalos de longitud $\frac{d-c}{n}$, obtenemos una partición bidimensional del rectángulo $[a, b] \times [c, d]$ en n^2 subrectángulos $I_{j,k}$ con área dada por el producto de las longitudes horizontal y vertical de los intervalos de partición $|I_{j,k}| = \frac{(b-a)(d-c)}{n^2}$. Consideramos unos puntos cualesquiera $t_{j,k} \in I_{j,k}$ como en la figura siguiente.



Iterando el teorema de Riemann es posible demostrar que vale el siguiente resultado:

Teorema: si $f : [a, b] \times [c, d] \rightarrow \mathbb{R}$ es continua, el límite

$$\lim_{n \rightarrow +\infty} \sum_{j=1}^n \sum_{k=1}^n |I_{j,k}| f(t_{j,k}) = \lim_{n \rightarrow +\infty} \sum_{k=1}^n \sum_{j=1}^n |I_{j,k}| f(t_{j,k})$$

existe, es finito y es independiente de la elección de los puntos $t_{j,k}$.

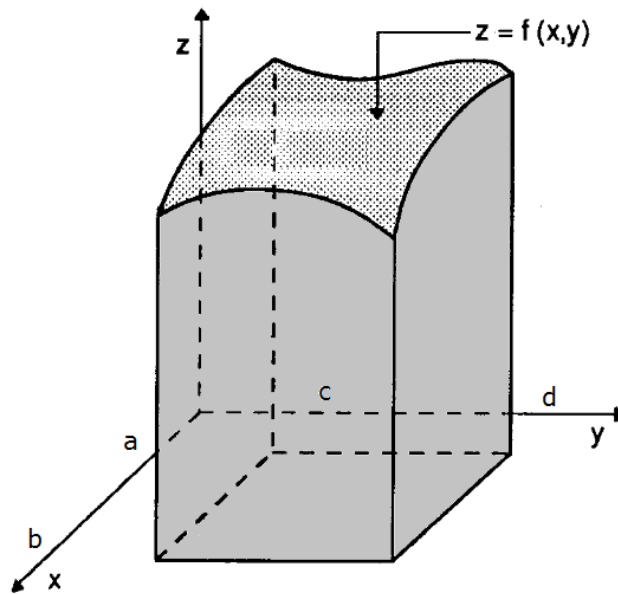
Llamaremos a este límite la integral doble de f en el rectángulo $[a, b] \times [c, d]$ y escribiremos:

$$\iint_{[a,b] \times [c,d]} f(x, y) dx dy = \lim_{n \rightarrow +\infty} \sum_{j=1}^n \sum_{k=1}^n |I_{j,k}| f(t_{j,k})$$

Como se ha visto, la definición de la integral doble de una función continua en un rectángulo es una generalización inmediata de la definición de Riemann,

sustituyendo los intervalos de la partición y su longitud por los rectángulos de la partición y su área.

Podemos generalizar también la interpretación geométrica de la integral de Riemann, pero en este caso solo para funciones no negativas: $f(x, y) \geq 0 \forall (x, y) \in [a, b] \times [c, d]$. En este caso la cantidad $|I_{j,k}|f(t_{j,k})$ representa el volumen del paralelepípedo de base $|I_{j,k}|$ y altura $f(t_{j,k})$. A medida que afinamos la partición con la operación de límite nos acercamos cada vez mejor al valor del **volumen de la región del espacio tridimensional contenida entre el gráfico de f y el rectángulo $[a, b] \times [c, d]$** , como muestra la figura siguiente:



6.1.2. Integrales dobles en dominios simples o regulares

Por supuesto definir la integral (doble) de una función de dos variables en un rectángulo es una limitación bastante fuerte porque, como hemos visto, el dominio de una función de dos variables, en general, tiene una forma más complicada. Por eso necesitamos extender esta definición a dominios con un perímetro curvilíneo.

La respuesta definitiva a esta exigencia fue dada por el matemático francés Henri Léon **Lebesgue** (1875-1941), pero la teoría de la integración que él desarrolló es demasiado complicada para ser discutida en este curso.

Por eso nos limitamos a extender la definición que hemos dado a regiones del plano que pueden tener perfiles curvilíneos, pero estos perfiles tienen que ser el gráfico de funciones continuas. Veremos que si consideramos estas regiones del plano podremos generalizar fácilmente la definición que acabamos de ver y, además, presentar una fórmula para el cálculo de integrales dobles a través de la iteración de dos integrales de una variable.

Introducimos las definiciones siguientes:

Def.: un subconjunto del plano $E \subset \mathbb{R}^2$ se dice **y-simple** si existen dos funciones continuas $g_1, g_2 : [a, b] \rightarrow \mathbb{R}$ tales que E se puede escribir como sigue

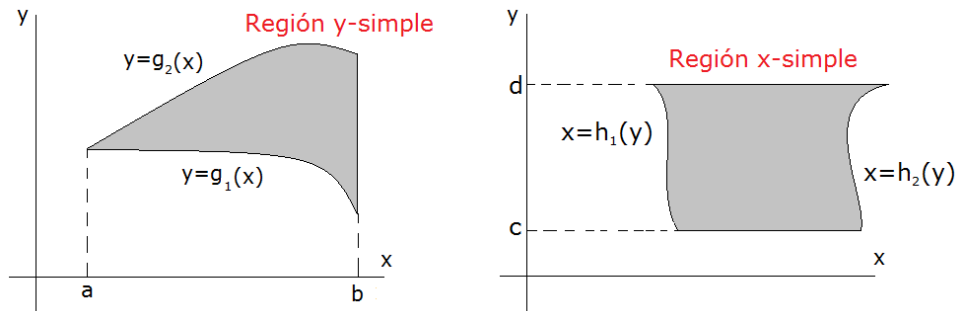
$$E = \{(x, y) \in \mathbb{R}^2 : x \in [a, b], g_1(x) \leq y \leq g_2(x)\}$$

es decir, mientras x varía de a a b , y varía entre los gráficos de la función g_1 (que proporciona el perfil inferior de E) y g_2 (que proporciona el perfil superior de E), como se ve en la figura de abajo a la izquierda.

Análogamente, un subconjunto del plano $E \subset \mathbb{R}^2$ se dice **x-simple** si existen dos funciones continuas $h_1, h_2 : [c, d] \rightarrow \mathbb{R}$ tales que E se puede escribir como sigue

$$E = \{(x, y) \in \mathbb{R}^2 : y \in [c, d], h_1(y) \leq x \leq h_2(y)\}$$

es decir, mientras y varía de c a d , x varía entre los gráficos de la función h_1 (que proporciona el perfil izquierdo de E) y h_2 (que proporciona el perfil derecho de E), como se ve en la figura de abajo a la derecha.

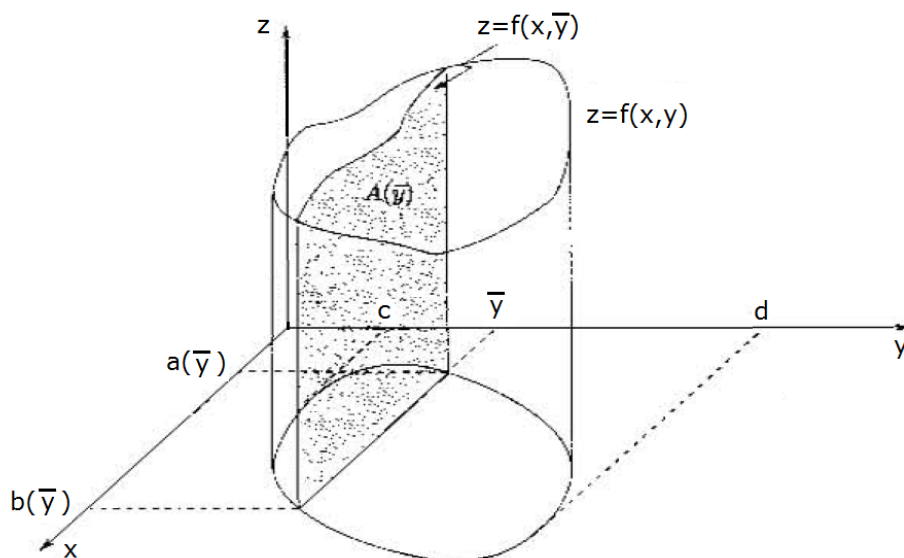


Finalmente, definimos un subconjunto del plano $E \subset \mathbb{R}^2$ como **regular**, si está dado por la **unión de un número finito de subconjuntos x y/o y simples**.

Observaciones:

1. E puede ser x y y -simple a la vez (por ejemplo un triángulo);
2. Un conjunto x y/o y simple es necesariamente limitado, es decir, no puede extenderse hasta el infinito. Gracias al teorema de Weierstrass sabemos que una función continua tiene máximo y mínimo absolutos en un dominio limitado y cerrado, como es un conjunto simple o regular, entonces la integral doble definida sobre un dominio simple o regular tiene que ser finita.

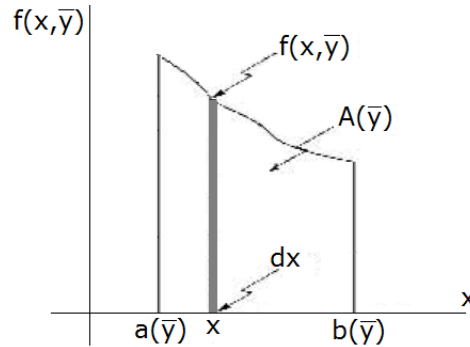
Ahora queremos encontrar una fórmula para calcular la integral doble de una función definida sobre un dominio simple o regular: resulta útil dejarse guiar por la interpretación geométrica de integral doble como volumen por debajo del gráfico de una función de dos variables. Para fijar las ideas consideramos una función f definida en un dominio x -simple y cortamos el volumen por debajo del gráfico de f con un plano paralelo al plano xz como en la figura siguiente.



El plano corta el volumen a lo largo de una región con un área que denotamos con $A(y)$, para cada $y = \bar{y}$, fijado entre c y d . La ‘loncha’ del volumen sólido que tiene como base el segmento infinitésimo $(y, y + dy)$ tiene un volumen infinitésimo igual a $A(y)dy$, entonces el volumen total valdrá

$$V = \int_c^d A(y)dy.$$

Para saber cómo calcular el área $A(y)$ consideramos la figura siguiente. Dado



que hemos fijado $y = \bar{y}$, la función f , mientras nos desplazamos en el plano de área $A(y)$, es una función de la sola variable x , por lo tanto el área $A(y)$ estará dada por la integral

$$A(y) = \int_{a(y)}^{b(y)} f(x, y) dx.$$

Sigue que el volumen total, es decir la integral doble, vale

$$V = \int_c^d \left(\int_{a(y)}^{b(y)} f(x, y) dx \right) dy.$$

Este razonamiento geométrico intuitivo puede ser formalizado en el teorema siguiente, que es un caso particular de un teorema de validez más general debido a los dos matemáticos italianos Guido Fubini (1879-1943) y Leonida Tonelli (1885-1946).

Teorema de Fubini-Tonelli: sea $\Omega \subset \mathbb{R}^2$ un dominio x -simple y $f : \Omega \rightarrow \mathbb{R}$ una función continua, entonces la integral doble de f en Ω existe y es finita y se puede calcular como integral iterada de la siguiente forma

$$\iint_{\Omega} f(x, y) dx dy = \int_c^d \left(\int_{h_1(y)}^{h_2(y)} f(x, y) dx \right) dy,$$

por otro lado si Ω es y -simple entonces vale la fórmula

$$\iint_{\Omega} f(x, y) dx dy = \int_a^b \left(\int_{g_1(x)}^{g_2(x)} f(x, y) dy \right) dx.$$

Si Ω es x -simple e y -simple, entonces valen ambas fórmulas (que, obviamente, dan el mismo resultado). Si Ω es un dominio regular dado por la unión de n dominios simples Ω_k , que coinciden solo en los bordes y cubren todo Ω , entonces vale la fórmula

$$\iint_{\Omega} f(x, y) dx dy = \sum_{k=1}^n \iint_{\Omega_k} f(x, y) dx dy.$$

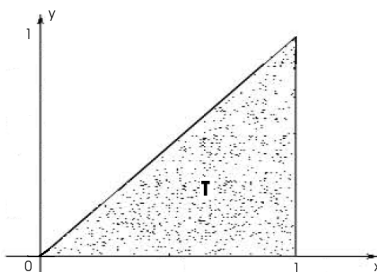
Observaciones:

1. El teorema de Fubini-Tonelli permite reconducir el cálculo de una integral doble a la iteración de dos integrales monodimensionales. Observamos que, *mientras se calcula la integral más ‘interna’, la variable diferente con respecto a la cual se está integrando tiene que ser pensada como una constante*;
2. Integrando la función constante $\mathbf{1}(x) \equiv 1$, podemos definir el **área de un dominio simple o regular D en \mathbb{R}^2** como sigue:

$$\text{Área}(D) = \iint_D dx dy.$$

Observamos que, como ya se anticipó en el capítulo 2 de la primera parte del curso, el cálculo diferencial permite resolver el problema de calcular el área de una figura plana con perímetro curvilíneo.

Ejemplo de cálculo de integral doble. Calcular la integral doble $\iint_T xy dx dy$, donde T es el triángulo representado en la figura de abajo.



Comenzamos definiendo el triángulo a través del dominio y -simple:

$$T = \{(x, y) \in \mathbb{R}^2 : 0 < x < 1, 0 < y < x\},$$

identificamos: $f(x, y) = xy$, $a = 0$, $b = 1$, $g_1 : [0, 1] \rightarrow \mathbb{R}$, $g_1(x) \equiv 0$, $g_2 : [0, 1] \rightarrow \mathbb{R}$, $g_2(x) = x$.

El cálculo de la integral doble se hace fijando x entre 0 y 1 (y considerándola como una constante) e integrando con respecto a la variable y entre 0 y x ; después se integra el resultado obtenido con respecto a la variable x entre 0 y 1:

$$\iint_T xy \, dx dy = \int_0^1 \left(\int_0^x xy dy \right) dx = \int_0^1 \left[\frac{1}{2} xy^2 \right]_0^x dx = \int_0^1 \frac{1}{2} x^3 dx = \left[\frac{x^4}{8} \right]_0^1 = \frac{1}{8}.$$

Ahora definimos el triángulo a través del dominio x -simple:

$$T = \{(x, y) \in \mathbb{R}^2 : 0 < y < 1, y < x < 1\},$$

identificamos: $f(x, y) = xy$, $c = 0$, $d = 1$, $h_1 : [0, 1] \rightarrow \mathbb{R}$, $h_1(y) = y$, $h_2 : [0, 1] \rightarrow \mathbb{R}$, $h_2(y) \equiv 1$.

El cálculo de la integral doble en este caso se hace fijando y entre 0 y 1 (y considerándola como una constante) e integrando con respecto a la variable x entre y y 1; después se integra el resultado obtenido con respecto a la variable y entre 0 y 1:

$$\begin{aligned} \iint_T xy \, dx dy &= \int_0^1 \left(\int_y^1 xy dx \right) dy = \int_0^1 \left[\frac{1}{2} x^2 y \right]_y^1 dy = \int_0^1 \frac{1}{2} (1 - y^2) y dy \\ &= \left[\frac{1}{4} y^2 - \frac{1}{8} y^4 \right]_0^1 = \frac{1}{8}. \end{aligned}$$

Coherentemente con lo que esperábamos, las dos integrales dobles devuelven el mismo valor.

En las clases de prácticas y seminarios veremos cómo aplicar las integrales dobles para calcular baricentros y momentos de inercia.

Terminamos esta sección con un teorema que describe unas propiedades importantes de las integrales dobles.

Teorema: sea D un conjunto regular, f, g funciones continuas en D y sea $c \in \mathbb{R}$.

1. La integral doble es **lineal**:

$$\begin{aligned} \iint_D [f(x, y) + g(x, y)] \, dx dy &= \iint_D f(x, y) \, dx dy + \iint_D g(x, y) \, dx dy, \\ \iint_D cf(x, y) \, dx dy &= c \iint_D f(x, y) \, dx dy. \end{aligned}$$

2. La integral doble es **monótona**: si $f \geq g$ en D , entonces:

$$\iint_D f(x, y) \, dx dy \geq \iint_D g(x, y) \, dx dy,$$

en particular, si $f \geq 0$ en D :

$$\iint_D f(x, y) \, dx dy \geq 0.$$

3. La integral doble es **aditiva con respecto al dominio de integración**, es decir, si D_1 y D_2 son dominios simples y disjuntos, entonces

$$\iint_{D_1 \cup D_2} f(x, y) \, dx dy = \iint_{D_1} f(x, y) \, dx dy + \iint_{D_2} f(x, y) \, dx dy.$$

4. Si D es *conexo* vale el **teorema de la media** (para integrales): existe un punto $(x_0, y_0) \in D$ tal que

$$\frac{1}{\text{Área}(D)} \iint_D f(x, y) \, dx dy = f(x_0, y_0).$$

6.1.3. Cambio de variable en las integrales dobles

Resulta útil en muchas aplicaciones sustituir las variables originales x, y de una función con variables auxiliares que permiten resolver las integrales dobles de forma más simple. Un ejemplo notable está dado por las **coordenadas polares**. Estas coordenadas resultan particularmente útiles cuando se examinan problemas con una geometría circular. La transformación a coordenadas polares es:

$$\begin{cases} x = \rho \cos \vartheta \\ y = \rho \sin \vartheta \end{cases}$$

con $\rho \in [0, +\infty)$ y $\vartheta \in [0, 2\pi)$. Esta transformación se puede interpretar como una función

$$\begin{aligned} T : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 \\ (\rho, \vartheta) &\longmapsto T(\rho, \vartheta) = (\rho \cos \vartheta, \rho \sin \vartheta) \equiv (x(\rho, \vartheta), y(\rho, \vartheta)), \end{aligned}$$

las funciones componentes de T son estas:

$$\begin{aligned} T_1 : \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ (\rho, \vartheta) &\longmapsto T_1(\rho, \vartheta) = \rho \cos \vartheta \end{aligned}$$

$$\begin{aligned} T_2 : \quad \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ (\rho, \vartheta) &\longmapsto T_1(\rho, \vartheta) = \rho \sin \vartheta. \end{aligned}$$

La matriz Jacobiana de T es:

$$J_T(\rho, \vartheta) = \begin{bmatrix} \partial_\rho T_1 & \partial_\vartheta T_1 \\ \partial_\rho T_2 & \partial_\vartheta T_2 \end{bmatrix}(\rho, \vartheta) = \begin{bmatrix} \cos \vartheta & -\rho \sin \vartheta \\ \sin \vartheta & \rho \cos \vartheta \end{bmatrix},$$

y su determinante es $\det(J_T(\rho, \vartheta)) = \rho \cos^2 \vartheta + \rho \sin^2 \vartheta = \rho$. Se puede demostrar que, bajo el cambio de coordenadas $(x, y) \mapsto (\rho, \vartheta)$, el área infinitésima $dxdy$ se transforma en $\rho d\rho d\vartheta$, de forma que la integral doble de una función continua f en un dominio regular o simple D se puede escribir:

$$\boxed{\iint_D f(x, y) \, dxdy = \iint_{\tilde{D}} f(T(\rho, \vartheta)) \, \rho \, d\rho d\vartheta},$$

donde $\tilde{D} = T^{-1}(D)$ es el dominio D expresado en términos de las variables polares ρ, ϑ . Más en general, dada la transformación de coordenadas $(u, v) = T(x, y)$, T función derivable, se puede demostrar que el área infinitésima $dxdy$ se transforma en $|\det J_T(u, v)|dudv$ así que vale la fórmula del cambio de variable para las integrales dobles:

$$\boxed{\iint_D f(x, y) \, dxdy = \iint_{\tilde{D}} f(T(u, v)) |\det J_T(u, v)| \, dudv},$$

donde $\tilde{D} = T^{-1}(D)$ es el dominio D expresado en términos de las nuevas variables u, v .

Ejemplo: calcular la integral doble

$$I_R = \iint_D e^{-(x^2+y^2)} \, dxdy,$$

$D = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < R^2\}$. El dominio D es el interior de un círculo de radio R y centro en el origen de los ejes. En coordenadas polares D se representa como $\tilde{D} = \{(\rho, \vartheta) : 0 \leq \rho < R, 0 \leq \vartheta < 2\pi\}$, entonces podemos escribir:

$$I_R = \int_0^{2\pi} \left(\int_0^R e^{-\rho^2} \rho \, d\rho \right) d\vartheta$$

$\int_0^R e^{-\rho^2} \rho \, d\rho$ es una integral semi-inmediata y vale $\left[-\frac{1}{2}e^{-\rho^2}\right]_0^R = \frac{1-e^{-R^2}}{2}$, entonces

$$I_R = \int_0^{2\pi} \frac{1-e^{-R^2}}{2} d\vartheta = \frac{1-e^{-R^2}}{2} \int_0^{2\pi} d\vartheta = \pi(1-e^{-R^2}).$$

Aprovechamos la oportunidad para decir que se pueden definir también las **integrales dobles generalizadas**, como en el caso monodimensional, tomando límites oportunos. No queremos entrar en detalle, solo queremos observar que, si tomamos el límite para $R \rightarrow +\infty$ de la integral doble que acabamos de calcular obtenemos:

$$\lim_{R \rightarrow +\infty} \iint_{\{(x,y) \in \mathbb{R}^2 : x^2+y^2 < R^2\}} e^{-(x^2+y^2)} dx dy = \iint_{\mathbb{R}^2} e^{-(x^2+y^2)} dx dy,$$

pero $\lim_{R \rightarrow +\infty} I_R = \lim_{R \rightarrow +\infty} (\pi(1 - e^{-R^2})) = \pi$, es decir

$$\boxed{\iint_{\mathbb{R}^2} e^{-(x^2+y^2)} dx dy = \pi}.$$

Esta integral doble permite calcular una integral monodimensional notable: $\int_{\mathbb{R}} e^{-x^2} dx$, notable porque la función e^{-x^2} no tiene primitiva!

Tenemos que

$$\iint_{\mathbb{R}^2} e^{-(x^2+y^2)} dx dy = \int_{\mathbb{R}} e^{-x^2} dx \int_{\mathbb{R}} e^{-y^2} dy$$

pero la variable de integración es muda, por lo tanto

$$\iint_{\mathbb{R}^2} e^{-(x^2+y^2)} dx dy = \left(\int_{\mathbb{R}} e^{-x^2} dx \right)^2 = \pi,$$

entonces

$$\boxed{\int_{\mathbb{R}} e^{-x^2} dx = \sqrt{\pi}}.$$

Esta es la integral de la densidad Gaussiana y juega un papel de enorme importancia en probabilidad, estadística y varias otras disciplinas científicas.

Otras coordenadas importantes en las aplicaciones son las **coordenadas elípticas**:

$$\begin{cases} x = a\rho \cos \vartheta \\ y = b\rho \sin \vartheta \end{cases}$$

con $a, b > 0$, $\rho \in [0, +\infty)$ y $\vartheta \in [0, 2\pi)$, en este caso el elemento infinitésimo de área se escribe como

$$\boxed{dx dy = ab\rho d\rho d\vartheta} \quad \text{Coordenadas elípticas.}$$

6.2. Integrales triples

Todo lo que acabamos de ver sobre las integrales dobles se puede generalizar a dimensiones superiores. En particular consideramos muy brevemente, por su importancia práctica en las aplicaciones, el caso de las integrales triples, o sea de funciones de tres variables reales. Lo único que cambia en este caso es que los rectángulos $[a, b] \times [c, d]$ se sustituyen con paralelepípedos $[a_1, b_1] \times [a_2, b_2] \times [a_3, b_3]$.

Un dominio z -simple en \mathbb{R}^3 es un dominio del tipo

$$\Omega = \{(x, y, z) \in \mathbb{R}^3 : g_1(x, y) \leq z \leq g_2(x, y), (x, y) \in D\},$$

donde D es un dominio simple en \mathbb{R}^2 y g_1, g_2 son funciones continuas en D . Los dominios x e y -simples en \mathbb{R}^3 se definen análogamente.

Si $f : \Omega \subset \mathbb{R}^3 \rightarrow \mathbb{R}$ es continua en Ω , dominio z -simple en \mathbb{R}^3 , entonces es integrable y la integral triple de f se puede calcular con la fórmula de iteración dada por:

$$\boxed{\iiint_{\Omega} f(x, y, z) \, dx \, dy \, dz = \iint_D \left(\int_{g_1(x, y)}^{g_2(x, y)} f(x, y, z) \, dz \right) \, dx \, dy.}$$

Esta fórmula reduce el cálculo de la integral triple a la iteración de una integral monodimensional con respecto a z , en la cual x e y se tienen que considerar como constantes, y de una integral doble en el dominio D que se puede calcular con las fórmulas vistas en la sección anterior.

Para calcular las integrales triples resulta útil en muchas situaciones el uso de coordenadas diferentes de las cartesianas, destacamos, en particular, las

- **Coordenadas esféricas:**

$$\begin{cases} x = \rho \cos \varphi \sin \vartheta \\ y = \rho \sin \varphi \sin \vartheta \\ z = \rho \cos \vartheta \end{cases}$$

con $\rho \in [0, +\infty)$, $\varphi \in [0, 2\pi]$, $\vartheta \in [0, \pi]$;

- **Coordenadas cilíndricas:**

$$\begin{cases} x = \rho \cos \vartheta \\ y = \rho \sin \vartheta \\ z = z \end{cases}$$

$\rho \in [0, +\infty)$, $\vartheta \in [0, 2\pi]$, $z \in \mathbb{R}$.

Los elementos de volumen en coordenadas esféricas y cilíndricas se pueden calcular directamente y resultan:

$$\boxed{dx dy dz = \rho^2 \sin \varphi d\rho d\varphi d\vartheta} \quad \text{Coordenadas esféricas}$$

$$\boxed{dx dy dz = \rho d\rho d\vartheta dz} \quad \text{Coordenadas cilíndricas}$$

No tenemos tiempo suficiente para proporcionar ejemplos o hacer ejercicios sobre integrales triples porque consideramos más importante introducir las integrales de superficie y de campos vectoriales, que es el tema del próximo capítulo.

Capítulo 7

Integrales de superficie y de campos vectoriales, teorema de la divergencia y del rotacional

En este último capítulo definiremos las integrales de superficies y de campos vectoriales, además discutiremos unos teoremas de importancia fundamental tanto en matemáticas como en las aplicaciones a la física y a la ingeniería: el teorema de la divergencia y del rotacional.

7.1. Integrales de superficie

Como vimos en el capítulo anterior, una vez formulada la teoría de la integral doble, es bastante natural generalizarlo a dimensiones superiores. En esta sección veremos que es también bastante natural generalizar la teoría de la integral doble para definir la integral de una función definida sobre una superficie, que sigue siendo un objeto geométrico bidimensional, como en la integral doble, pero en este caso resulta curvo y no plano.

Una motivación para recalcar la necesidad de saber integrar una función sobre una superficie viene dada, por ejemplo, por la necesidad de calcular la carga eléctrica de un conductor metálico: en estos conductores la carga eléctrica se mantiene sobre la superficie, por lo tanto, si se conoce la función densidad de carga eléctrica superficial, su integración proporciona la carga eléctrica total del conductor. En el curso de física y ondas se verán aún más

ejemplos prácticos.

Para definir la integral de superficie necesitamos recordar brevemente el concepto de **producto vectorial** (o **wedge**): sean $\vec{v}, \vec{w} \in \mathbb{R}^3$ dos vectores tridimensionales cualesquiera, el producto vectorial entre ellos es el vector tridimensional $\vec{v} \wedge \vec{w}$ tal que:

- $\|\vec{v} \wedge \vec{w}\| = \|\vec{v}\|\|\vec{w}\| \sin \alpha$, siendo α el ángulo más pequeño formado por \vec{v} y \vec{w} ;
- la dirección de $\vec{v} \wedge \vec{w}$ es ortogonal al plano π que contiene \vec{v} y \vec{w} ;
- el sentido de $\vec{v} \wedge \vec{w}$ es el sentido tal que, si miramos el plano π desde la punta del vector $\vec{v} \wedge \vec{w}$, vemos \vec{v} que se acerca a \vec{w} cumpliendo una rotación antihoraria (regla ‘de la mano derecha’ o del ‘sacacorchos’).

Resulta útil una fórmula para poder calcular el producto vectorial una vez dadas las componentes de \vec{v} y \vec{w} . Se puede demostrar que, si $\vec{v} = (v_x, v_y, v_z)$ y $\vec{w} = (w_x, w_y, w_z)$, entonces

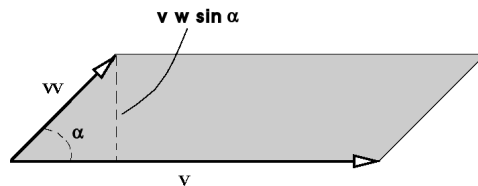
$$\vec{v} \wedge \vec{w} = \det \begin{bmatrix} \hat{i} & \hat{j} & \hat{k} \\ v_x & v_y & v_z \\ w_x & w_y & w_z \end{bmatrix},$$

donde $\hat{i}, \hat{j}, \hat{k}$ son los vectores unitarios de los ejes x, y, z , respectivamente. Explícitamente tenemos:

$$\vec{v} \wedge \vec{w} = (v_y w_z - v_z w_y, v_z w_x - v_x w_z, v_x w_y - v_y w_x).$$

Se puede averiguar directamente que el producto vectorial es antisimétrico, es decir: $\vec{v} \wedge \vec{w} = -\vec{w} \wedge \vec{v}$.

La norma del producto vectorial tiene un significado geométrico de importancia fundamental para nuestros fines: como se ve en la figura aquí debajo, **la norma del producto vectorial entre dos vectores representa el área del paralelogramo que tiene los dos vectores por lados.**

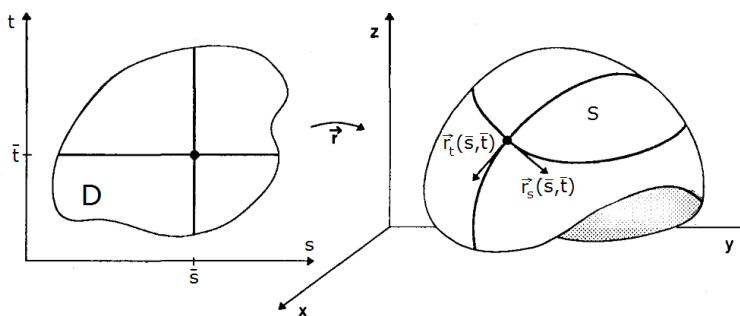


Recordamos ahora que la definición más general de superficie que hemos dado es la de superficie parametrizada \mathcal{S} (que incluye como caso particular

el gráfico de una función escalar de dos variables reales), dada por la imagen de la función vectorial de dos variables reales $\vec{r} : D \subset \mathbb{R}^2 \rightarrow \mathbb{R}^3$, $(t, s) \mapsto \vec{r}(s, t) = (x(s, t), y(s, t), z(s, t))$.

Exactamente como para las integrales curvilíneas, necesitamos definir el concepto de elemento de área infinitésima sobre la superficie \mathcal{S} a través de un proceso de linealización local. Una vez hecho esto, extenderemos el área infinitésima a un área finita a través de una integral doble en un dominio de \mathbb{R}^2 .

Comenzamos observando que, si fijamos una de las dos coordenadas $s = \bar{s}$ o bien $t = \bar{t}$ en D obtenemos unos segmentos rectilíneos que se cruzan en un punto. La parametrización \vec{r} transformará estos segmentos rectilíneos en curvas sobre la superficie que se cruzan en $\vec{r}(\bar{s}, \bar{t})$, como se puede ver en la figura aquí debajo:



Ya sabemos que los vectores tangentes al punto de cruce están dados por $\vec{r}_s(\bar{s}, \bar{t}) = (x_s(\bar{s}, \bar{t}), y_s(\bar{s}, \bar{t}), z_s(\bar{s}, \bar{t}))$ y $\vec{r}_t(\bar{s}, \bar{t}) = (x_t(\bar{s}, \bar{t}), y_t(\bar{s}, \bar{t}), z_t(\bar{s}, \bar{t}))$ donde los subíndices s y t se refieren a las derivadas parciales con respecto a s y t , respectivamente.

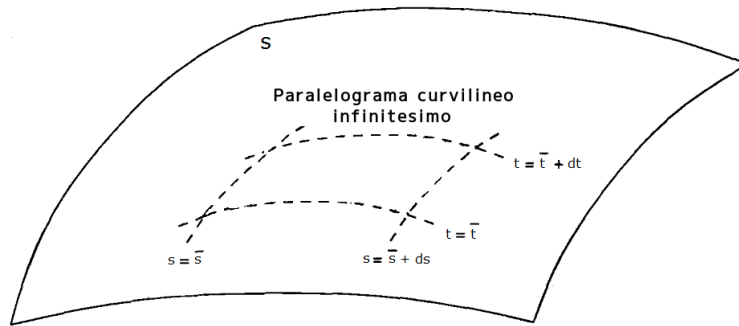
Sabemos también que los dos vectores tangentes viven en el plano tangente a \mathcal{S} en el punto $\vec{r}(\bar{s}, \bar{t})$ por lo tanto, dado que son vectores planares, podemos calcular el área del paralelogramo que los tiene como lados y el vector unitario normal a la superficie en el punto $\vec{r}(\bar{s}, \bar{t})$ a través del producto vectorial:

Área del paralelogramo de lados $\vec{r}_s(\bar{s}, \bar{t})$ y $\vec{r}_t(\bar{s}, \bar{t}) = \|\vec{r}_s(\bar{s}, \bar{t}) \wedge \vec{r}_t(\bar{s}, \bar{t})\|$

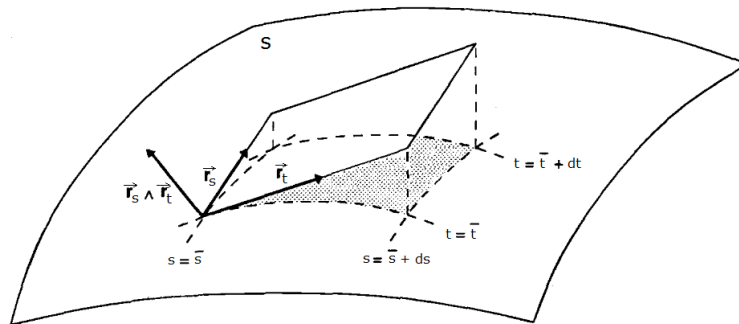
$$\text{Normal externa a } \mathcal{S} \text{ en } \vec{r}(\bar{s}, \bar{t}) = \frac{\vec{r}_s(\bar{s}, \bar{t}) \wedge \vec{r}_t(\bar{s}, \bar{t})}{\|\vec{r}_s(\bar{s}, \bar{t}) \wedge \vec{r}_t(\bar{s}, \bar{t})\|}$$

Si ahora nos desplazamos del punto $\vec{r}(\bar{s}, \bar{t})$ con dos pasos infinitésimos ds y dt a lo largo de las curvas $\vec{r}(s, \bar{t})$ y $\vec{r}(\bar{s}, t)$, respectivamente, podemos

definir el elemento de área superficial infinitésimo como el área del paralelogramo curvilíneo que tiene por lados (curvilíneos) los dos desplazamientos infinitésimos, como en la figura debajo:



Nuestro problema es que no sabemos cómo calcular el área de un paralelogramo curvilíneo. Sin embargo, si \vec{r} es una función diferenciable, podemos aproximar los lados del paralelogramo curvilíneo, a menos de infinitésimos de orden superior a ds y dt , con los lados del paralelogramo rectilíneo dados por $\vec{r}_s ds$ y $\vec{r}_t dt$.



Como ya vimos muchas veces durante el curso, la ventaja de la aproximación lineal es que nos permite usar técnicas de cálculos conocidas: en este caso, sabemos que el área del paralelogramo rectilíneo infinitésimo es la norma del producto vectorial entre $\vec{r}_s ds$ y $\vec{r}_t dt$, a este área le daremos el nombre de

Elemento de área infinitésimo sobre la superficie parametrizada S :

$$dS = \|\vec{r}_s(s, t) \wedge \vec{r}_t(s, t)\| ds dt.$$

Tiene particular importancia el caso de una superficie dada por el gráfico de una función $f : D \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ que, como sabemos tiene representación paramétrica dada por $\vec{r}(s, t) = (s, t, f(s, t))$, $(s, t) \in D$. En este caso $\vec{r}_s(s, t) = (1, 0, f_s(s, t))$, $\vec{r}_t(s, t) = (0, 1, f_t(s, t))$, entonces

$$\vec{r}_s(s, t) \wedge \vec{r}_t(s, t) = \det \begin{bmatrix} \hat{i} & \hat{j} & \hat{k} \\ 1 & 0 & f_s(s, t) \\ 0 & 1 & f_t(s, t) \end{bmatrix} = (-f_s(s, t), -f_t(s, t), 1)$$

lo cual confirma el resultado sobre el vector normal que hemos obtenido en el capítulo 2, y

$$\|\vec{r}_s(s, t) \wedge \vec{r}_t(s, t)\| = \sqrt{1 + \|\nabla f(s, t)\|^2},$$

entonces llamamos

Elemento de área infinitésimo sobre la superficie dada por el gráfico de f :

$$dS = \sqrt{1 + \|\nabla f(s, t)\|^2} ds dt.$$

Ahora tenemos todos los ingredientes para definir el área de una superficie y la integral de superficie de una función: dada la superficie en forma paramétrica $\vec{r} : D \subset \mathbb{R}^2 \rightarrow \mathbb{R}^3$, $(s, t) \in D$ y una función $g : \mathbb{R}^3 \rightarrow \mathbb{R}$,

Área de una superficie:

$$\iint_S dS = \iint_D \|\vec{r}_s(s, t) \wedge \vec{r}_t(s, t)\| ds dt$$

Integral de g sobre la superficie S :

$$\iint_S g dS = \iint_D g(\vec{r}(s, t)) \|\vec{r}_s(s, t) \wedge \vec{r}_t(s, t)\| ds dt,$$

donde $g(\vec{r}(s, t))$ es la evaluación de la función g en los puntos de la superficie.

Si \mathcal{S} está dada por el gráfico de una función $f : D \subset \mathbb{R}^2 \rightarrow \mathbb{R}$, entonces

Área de una superficie:

$$\iint_{\mathcal{S}} dS = \iint_D \sqrt{1 + \|\nabla f(s, t)\|^2} ds dt$$

Integral de g sobre la superficie \mathcal{S} :

$$\iint_{\mathcal{S}} g dS = \iint_D g(s, t, f(s, t)) \sqrt{1 + \|\nabla f(s, t)\|^2} ds dt.$$

Observamos la similitud entre la integral curvilínea y la integral de superficie de una función:

$$\int_C g dl = \int_a^b g(\vec{c}(t)) \|\vec{c}'(t)\| dt$$

$$\int_{\mathcal{S}} g dS = \int_D g(\vec{r}(s, t)) \|\vec{r}_s(s, t) \wedge \vec{r}_t(s, t)\| ds dt,$$

en ambos casos hay que

1. evaluar g en los puntos de la curva o superficie;
2. multiplicar por el elemento de longitud o área infinitésima;
3. integrar con respecto al parámetro (curva) o a los parámetros (superficie).

Ejemplo: calculamos el área de un paraboloides elíptico de altura 2 metros, que podemos definir a través de la ecuación $z = f(x, y) = x^2 + y^2$, $z \leq 2$, y que representa la superficie de una antena parabólica. Dado que $\nabla f(x, y) = (2x, 2y)$, el elemento de área infinitésimo es $dS = \sqrt{1 + 4(x^2 + y^2)} dx dy$, por lo tanto

$$A = \text{Área paraboloides elíptico de 2 metros} = \iint_D \sqrt{1 + 4(x^2 + y^2)} dx dy$$

donde $D = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 2\}$. La computación de la integral doble resulta más sencilla si pasamos a las coordenadas polares (ρ, ϑ) , con respecto a las cuales:

- $dx dy = \rho d\rho d\vartheta$;
- $x^2 + y^2 = \rho^2$;
- D se transforma en $\tilde{D} = \{(\rho, \vartheta) \in \mathbb{R}^2 : 0 \leq \rho \leq \sqrt{2}, 0 \leq \vartheta \leq 2\pi\}$, donde la limitación sobre ρ deriva del hecho de que $0 \leq x^2 + y^2 = \rho^2 \leq 2$ si y solo si $\rho \leq \sqrt{2}$.

Entonces:

$$A = \int_0^{2\pi} d\vartheta \int_0^{\sqrt{2}} \sqrt{1 + 4\rho^2} \rho d\rho d\vartheta = 2\pi \int_0^{\sqrt{2}} \sqrt{1 + 4\rho^2} \rho d\rho.$$

Para resolver la integral en $d\rho$ usamos este cambio de variable $u = 1 + 4\rho^2$, $du = 8\rho d\rho$, entonces $\rho d\rho = \frac{1}{8} du$ y

$$\int \sqrt{1 + 4\rho^2} \rho d\rho = \frac{1}{8} \int \sqrt{u} du = \frac{1}{8} \frac{u^{\frac{3}{2}}}{\frac{3}{2}} = \frac{u^{\frac{3}{2}}}{12} = \frac{(1 + 4\rho^2)^{\frac{3}{2}}}{12},$$

entonces

$$\int_0^{\sqrt{2}} \sqrt{1 + 4\rho^2} \rho d\rho = \frac{1}{12} \left[(1 + 4\rho^2)^{\frac{3}{2}} \right]_0^{\sqrt{2}} = \frac{26}{12} = \frac{13}{6}$$

y, finalmente:

$$A = \text{Área paraboloides elíptico de 2 metros} = 2\pi \frac{13}{6} = \frac{13\pi}{3} \text{ metros}^2.$$

7.2. Los operadores del cálculo vectorial: gradiente, divergencia, rotacional, laplaciano

Un elemento matemático de gran importancia en esta segunda parte del curso de cálculo ha sido el gradiente de una función. Dado un campo escalar $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, podemos ver el **gradiente como un operador que mapea el campo escalar f a un campo vectorial**:

$$f \mapsto_{\nabla} \nabla f = (f_{x_1}, \dots, f_{x_n})$$

donde

$$\begin{aligned} \nabla f : \mathbb{R}^n &\longrightarrow \mathbb{R}^n \\ \vec{x} &\longmapsto \nabla f(\vec{x}) = (f_{x_1}(\vec{x}), \dots, f_{x_n}(\vec{x})). \end{aligned}$$

Se pueden definir otros dos operadores fundamentales que actúan sobre campos escalares o vectoriales.

- **Operador divergencia:** div transforma un campo vectorial en un campo escalar. Sea $\vec{v} : \mathbb{R}^n \rightarrow \mathbb{R}^n$, $\vec{v}(\vec{x}) = (v_1, \dots, v_n)$, un campo vectorial, la divergencia de \vec{v} es el campo escalar $\text{div} \vec{v}$ definido de la forma siguiente

$$\begin{aligned} \text{div } \vec{v} : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ \vec{x} &\longmapsto \text{div } \vec{v}(\vec{x}) = \frac{\partial v_1}{\partial x_1}(\vec{x}) + \dots + \frac{\partial v_n}{\partial x_n}(\vec{x}), \end{aligned}$$

si $n = 3$, $\text{div } \vec{v}(x, y, z) = \frac{\partial v_x}{\partial x}(x, y, z) + \frac{\partial v_y}{\partial y}(x, y, z) + \frac{\partial v_z}{\partial z}(x, y, z)$. Observamos que, a nivel algebraico, el operador divergencia se puede definir como el producto escalar entre el operador gradiente y el campo vectorial \vec{v} :

$$\boxed{\text{div } v = \langle \nabla, \vec{v} \rangle}$$

dato que, por definición de producto escalar Euclidiano, tenemos que $\langle \nabla, \vec{v} \rangle = \frac{\partial}{\partial x_1} v_1 + \dots + \frac{\partial}{\partial x_n} v_n$. Remarcamos que esta es solo una escritura algebraica ‘formal’, que tiene que ser interpretada con el significado que hemos comentado anteriormente.

- **Operador rotacional:** rot (‘curl’ en los libros ingleses) es un operador que transforma un campo vectorial tridimensional $\vec{v} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$, $\vec{v} = (v_x, v_y, v_z)$ en otro campo vectorial tridimensional $\text{rot } \vec{v}$ definido de esta forma

$$\begin{aligned} \text{rot } \vec{v} : \mathbb{R}^3 &\longrightarrow \mathbb{R}^3 \\ \vec{x} &\longmapsto \text{rot } \vec{v}(x, y, z), \end{aligned}$$

donde $\text{rot } \vec{v}(x, y, z) = (\partial_y v_z - \partial_z v_y, \partial_z v_x - \partial_x v_z, \partial_x v_y - \partial_y v_x)(x, y, z)$. Podemos reconocer fácilmente que

$$\text{rot } \vec{v}(x, y, z) = \det \begin{bmatrix} \hat{i} & \hat{j} & \hat{k} \\ \partial_x & \partial_y & \partial_z \\ v_x & v_y & v_z \end{bmatrix} (x, y, z)$$

por lo tanto

$$\boxed{\text{rot } \vec{v} = \nabla \wedge \vec{v}}.$$

Vimos que el gradiente de un campo escalar tiene un significado geométrico muy importante: es el campo vectorial que, en cada punto, proporciona la dirección de variación más rápida del campo escalar mismo. También la divergencia y el rotacional tienen significados geométricos importantes. Para entender el significado geométrico de la divergencia tenemos que presentar el teorema homónimo; pero ya podemos introducir el significado geométrico

del rotacional y justificar su nombre: consideramos el campo de velocidad del *movimiento rotacional uniforme alrededor de un eje*: si \hat{k} es el vector unitario del eje de rotación, $\vec{\omega} = \omega\hat{k}$ es el vector de velocidad angular y $\vec{r} = x\hat{i} + y\hat{j}$ es el vector que da la posición de un punto con respecto al eje \hat{k} , el campo vectorial de velocidad en el punto (x, y) es $\vec{v} = \vec{\omega} \wedge \vec{r} = -y\omega\hat{i} + x\omega\hat{j}$ (véase el capítulo 2 Sección 8). El rotacional de este campo de velocidad es $\text{rot}\vec{v} = \hat{k}(\partial_x(x\omega) - \partial_y(-y\omega)) = 2\omega\hat{k}$, entonces

$$\boxed{\vec{\omega} = \frac{1}{2}\text{rot}\vec{v}},$$

lo cual dice que el vector velocidad angular es proporcional al rotacional del vector velocidad.

Este resultado pierde su validez global para un campo de velocidad \vec{v} genérico, pero mantiene su validez local: si \vec{v} es un campo genérico de velocidad de un medio (por ejemplo de agua), el vector $\frac{1}{2}\text{rot}\vec{v}$ representa la **velocidad angular local** del movimiento, es decir, es un **indicador de la vorticidad local** del movimiento. Dicho de otra forma, el rotacional muestra la tendencia de un campo vectorial a inducir rotación alrededor de un punto.

Si un campo \vec{v} tiene un rotacional (o rotor) nulo en todos puntos, significa que ese campo no tiene vórtices en ningún punto, se dice en este caso que \vec{v} es **irrotacional**: $\text{rot}\vec{v}=0$.

Ahora observamos que podemos componer los operadores de gradiente, divergencia y rotacional: dado el campo escalar f y el campo vectorial \vec{v} , resulta bien definido considerar

$$\text{div rot } \vec{v}, \quad \text{rot } \nabla f, \quad \text{div } \nabla f,$$

pero de estas combinaciones, solo la última no es nula, porque las otras se anulan debido al teorema de Schwarz, es decir, al hecho de que las derivadas segundas mixtas son iguales y al restarlas se anulan:

$$\boxed{\text{div rot } \vec{v} = 0} \quad \forall \vec{v}$$

$$\boxed{\text{rot } \nabla f = \vec{0}} \quad \forall f.$$

En cambio,

$$\text{div } \nabla f = \sum_{i=1}^n \frac{\partial}{\partial x_i} \left(\frac{\partial f}{\partial x_i} \right) = \sum_{i=1}^n \frac{\partial^2 f}{\partial x_i^2}$$

a partir de esta identidad podemos definir un operador que tiene gran importancia en las aplicaciones físicas e ingenieriles, es el **operador Laplaciano** (Pierre Simon Laplace, francés, (1749-1827) uno de los matemáticos más importantes de la historia) que transforma un campo escalar f en otro campo escalar:

$$f \xrightarrow{\nabla^2} \nabla^2 f$$

$$\nabla^2 f : \mathbb{R}^n \longrightarrow \mathbb{R}^3$$

$$\vec{x} \longmapsto \nabla^2 f(\vec{x}) = \sum_{i=1}^n \frac{\partial^2 f}{\partial x_i^2}(\vec{x}),$$

si $n = 3$,

$$\boxed{\nabla^2 f(x, y, z) = \frac{\partial^2 f}{\partial x^2}(x, y, z) + \frac{\partial^2 f}{\partial y^2}(x, y, z) + \frac{\partial^2 f}{\partial z^2}(x, y, z)},$$

en muchos libros, sobre todo escritos por matemáticos, el Laplaciano se encuentra escrito como Δ , mientras que físicos e ingenieros suelen preferir la notación ∇^2 . Resumiendo, los posibles símbolos usados para denotar el Laplaciano son:

$$\boxed{\operatorname{div} \nabla = \nabla^2 = \Delta}.$$

7.3. El flujo de un campo vectorial a través de una superficie

El concepto de flujo de un campo vectorial \vec{v} nace del problema práctico de describir el flujo de un fluido, por ejemplo agua, en un canal. El flujo es la rapidez de variación del volumen de agua que pasa a través de una superficie. Para medir el flujo de agua a través de una porción de superficie infinitésima dS necesitamos calcular la componente del campo \vec{v} a lo largo de la dirección normal a dS , por eso calculamos el producto escalar $\langle \vec{v}, \vec{n} \rangle$, siendo \vec{n} el vector unitario normal a dS . El flujo infinitésimo será:

$$d\Phi = \langle \vec{v}, \vec{n} \rangle dS$$

que podemos extender a una superficie finita \mathcal{S} con la integral de superficie que definimos en este capítulo:

Flujo del campo vectorial \vec{v} a través de la superficie \mathcal{S}

$$\Phi = \iint_{\mathcal{S}} \langle \vec{v}, \vec{n} \rangle dS.$$

El flujo de un campo vectorial \vec{v} tiene varios significados físicos: según las aplicaciones puede representar

- el **caudal de un fluido** si \vec{v} es la velocidad de movimiento de las partículas de fluido;
- la **corriente eléctrica** si \vec{v} es la densidad de corriente en un conductor;
- el **calor** que pasa a través de una membrana si \vec{v} es la densidad térmica de un sistema termodinámico;
- etc.

7.4. El teorema de la divergencia

El teorema de la divergencia es uno de los teoremas más importantes de las matemáticas y reúne muchos de los conceptos que hemos desarrollado hasta ahora: las derivadas parciales, las superficies y sus integrales, los campos vectoriales y sus flujos. Este teorema permite deducir muchas ecuaciones fundamentales de la física y de la ingeniería y da la posibilidad de transformar relaciones integrales entre objetos matemáticos a relaciones diferenciales o viceversa. Lamentablemente no tenemos el tiempo de discutir en este curso las aplicaciones del teorema de la divergencia, pero el estudiante podrá ya apreciar su importancia en el curso de ondas y electromagnetismo.

No sorprende que un teorema tan importante y útil haya sido desarrollado y demostrado por el ‘príncipe de los matemáticos’: Gauss.

Teorema de la divergencia (Gauss):

- Sea Ω un dominio cerrado y limitado de \mathbb{R}^3 y sea $\partial\Omega$ su superficie, que suponemos ser parametrizada, orientable y con normal externa \vec{n} .
- Sea \vec{v} un campo vectorial diferenciable con derivadas parciales continuas en Ω , entonces

$$\boxed{\iint_{\partial\Omega} \langle \vec{v}, \vec{n} \rangle dS = \iiint_{\Omega} \operatorname{div} \vec{v} dx dy dz,}$$

es decir: **el flujo de un campo vectorial que sale de una superficie cerrada es igual a la integral de la divergencia del campo en el volumen que la superficie encierra.**

7.4.1. Significado del operador divergencia e interpretación física del teorema de la divergencia

El teorema de la divergencia dice que la integral de volumen de la divergencia de un campo devuelve el flujo del campo a través de la superficie que contiene el volumen. Sigue que **la divergencia representa la densidad de flujo del campo vectorial por unidad de volumen.**

El flujo de un campo representa, como dice su mismo nombre, como fluye el campo a lo largo del espacio. Por supuesto, para que exista un campo, tiene que existir por lo menos una **fuentes** del campo y para que el campo deje de fluir tiene que existir por lo menos un **sumidero**.

Vamos a considerar una región del espacio con volumen Ω y superficie $\partial\Omega$ (frontera de Ω). Supongamos que en Ω haya fuentes y sumideros del campo, las fuentes generarán un flujo del campo que *sale* a través de $\partial\Omega$ y los sumideros un flujo del campo que *entra* en Ω . El teorema de la divergencia dice que el flujo neto a través de la superficie $\partial\Omega$, que es la resta entre el flujo que sale y el que entra, se mide a través de la integral de la divergencia en el volumen Ω . En este sentido, **el teorema de la divergencia es una abstracción del concepto de continuidad física, representando un balance entre flujo saliente y entrante.** El flujo saliente de $\partial\Omega$ se considera positivo porque tiene el mismo sentido que la normal externa \vec{n} a $\partial\Omega$, entonces el producto escalar que define el flujo es positivo. Recíprocamente, el flujo entrante en $\partial\Omega$ se considera negativo porque tiene sentido opuesto a la normal externa \vec{n} a $\partial\Omega$, entonces el producto escalar que define el flujo en este caso es negativo.

Como hemos visto, el operador divergencia es el instrumento matemático que permite ‘medir’ la intensidad de las fuentes y de los sumideros del campo, según su signo. Por ejemplo, la divergencia no nula del campo gravitacional señala la presencia de materia en un punto, la divergencia no nula del campo electrostático señala la presencia de una carga en un punto, etc.

Un campo vectorial que tiene **divergencia nula en cada punto**, es decir, que no tiene ni fuentes ni sumideros, se llama **solenoidal**.

No sorprende que una de las aplicaciones del teorema de la divergencia

sea la formulación de la ecuación de continuidad, que expresa el principio de conservación de la masa y de la carga eléctrica. El teorema de la divergencia juega también un rol fundamental en el desarrollo de la teoría de Maxwell del campo electromagnético (James Clerk Maxwell (1831-1879) físico y matemático británico, considerado uno de los tres físicos más importantes de la historia junto con Isaac Newton y Albert Einstein).

El teorema de la divergencia se puede formular en dimensión n cualquiera, pero si $n = 1$ el teorema de la divergencia se reduce a un resultado que ya conocemos, vamos a ver cuál es. Sea \vec{e} el vector unitario orientado como la recta real positiva y sea (a, b) un intervalo abierto en \mathbb{R} , su frontera es el conjunto dado por los puntos $\{a, b\}$ y la normal externa en los dos puntos de frontera es \vec{e} en $x = b$ y $-\vec{e}$ en $x = a$. Si $f(x)\vec{e}$ es un campo vectorial definido en un abierto de \mathbb{R} que contiene (a, b) , la divergencia de $f(x)\vec{e}$ es simplemente $f'(x)$, la derivada de f . Para calcular el flujo a través de la frontera, recordemos que una integral no es nada más que un límite de sumas, en este caso la frontera está dada por los puntos a y b , por eso calculamos el flujo a través de la frontera de (a, b) como la siguiente suma: $f(a)\langle \vec{e}, -\vec{e} \rangle + f(b)\langle \vec{e}, \vec{e} \rangle = f(b) - f(a)$. Si introducimos estos datos en el teorema de la divergencia monodimensional encontramos el resultado: $\int_a^b f'(x)dx = f(b) - f(a)$, es decir: **la versión monodimensional del teorema de la divergencia es el teorema fundamental del cálculo integral.**

7.5. Trabajo de un campo vectorial y teorema del rotacional

Para introducir el teorema del rotacional tenemos que discutir antes el concepto de trabajo de un campo vectorial. Podríamos considerar un campo en \mathbb{R}^n pero preferimos quedarnos en \mathbb{R}^3 para ser más concretos. Consideramos un campo (vectorial) de fuerzas $\vec{F} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ (típicamente $n = 3$) que actúa sobre unas partículas desplazándolas de su sitio. El **trabajo elemental** del campo de fuerza para desplazar una partícula una cantidad infinitésima $d\vec{s} = (dx, dy, dz)$ está dado por el producto escalar entre \vec{F} y $d\vec{s}$:

$$dW = \langle \vec{F}, d\vec{s} \rangle = F_x dx + F_y dy + F_z dz,$$

W como ‘work’, trabajo en ingles. Se considera el producto escalar por su significado geométrico: únicamente la componente de la fuerza en la dirección de $d\vec{s}$ produce el desplazamiento.

Si queremos conocer el trabajo global que el campo de fuerza hace para desplazar una partícula a lo largo de una curva \mathcal{C} con parametrización $\vec{c} : [a, b] \rightarrow \mathbb{R}^3$, $\vec{c}(t) = (x(t), y(t), z(t))$, es $W = \int_{\mathcal{C}} dW$, es decir:

$$\begin{aligned} W &= \int_{\mathcal{C}} \langle \vec{F}(\vec{c}(t)), d\vec{c}(t) \rangle = \int_{\mathcal{C}} \langle \vec{F}(\vec{c}(t)), \vec{c}'(t) \rangle dt \\ &= \int_a^b [F_x(\vec{c}(t))x'(t) + F_y(\vec{c}(t))y'(t) + F_z(\vec{c}(t))z'(t)] dt. \end{aligned}$$

Si \mathcal{C} es una curva cerrada se escribe

$$W = \oint_{\mathcal{C}} \langle \vec{F}(\vec{c}(t)), \vec{c}'(t) \rangle dt$$

y la integral se llama **circulación**.

Particularmente importante es el caso en el cual el campo \vec{F} proviene del gradiente de un campo escalar V : si $\vec{F} = \nabla V$, se dice que el campo \vec{F} es **conservativo**. Ejemplos notables de campos conservativos son el campo gravitacional, el electrostático, etc.

Vamos a calcular el trabajo de un campo conservativo:

$$W = \int_a^b \langle \vec{F}(\vec{c}(t)), \vec{c}'(t) \rangle dt = \int_a^b \langle \nabla V(\vec{c}(t)), \vec{c}'(t) \rangle dt$$

pero

$$\langle \nabla V(\vec{c}(t)), \vec{c}'(t) \rangle = \frac{\partial V}{\partial x}(\vec{c}(t))x'(t) + \frac{\partial V}{\partial y}(\vec{c}(t))y'(t) + \frac{\partial V}{\partial z}(\vec{c}(t))z'(t) = \frac{d}{dt}(V(\vec{c}(t)))$$

habiendo usado la regla de la cadena. Entonces tenemos

$$W = \int_a^b d(V(\vec{c}(t))) = V(\vec{c}(b)) - V(\vec{c}(a)),$$

lo cual indica que el **trabajo de un campo de fuerzas conservativo** que actúa sobre una partícula, desplazándola de un punto P a un punto Q a lo largo de una curva en el espacio, depende solo de los extremos de la curva y está dado por la **diferencia de potencial entre los puntos extremos** de la curva. En particular, la circulación de un campo conservativo, es decir el trabajo a lo largo de una curva cerrada, es nulo.

Tenemos todas las informaciones para formular el teorema del rotacional.

*Teorema del rotacional o de Stokes*¹: sea \mathcal{S} una superficie orientable con borde $\partial\mathcal{S}$ parametrizado por \vec{c} y sea \mathcal{F} un campo vectorial definido en una región del espacio que contiene \mathcal{S} , entonces

$$\boxed{\iint_{\mathcal{S}} \langle \text{rot}\vec{F}, \vec{n} \rangle dS = \oint_{\partial\mathcal{S}} \langle \vec{F}(\vec{c}(t)), \vec{c}'(t) \rangle dt .}$$

El teorema del rotacional dice que el flujo del rotacional de un campo vectorial a través de una superficie es igual al trabajo del mismo campo a lo largo del borde de la superficie.

En física el teorema del rotacional tiene una importancia fundamental para la formulación de la teoría del electromagnetismo. En matemática el teorema es importante porque relaciona integrales dobles con integrales curvilíneas.

THE END.

¹George Gabriel Stokes (1819-1903), físico y matemático irlandés.