

Cálculo en varias variables

Alejandro Zubiri

ADVE

AVE

Índice

1. Bibliografía	4
Chapter 1. El espacio \mathbb{R}^n	5
1. El conjunto \mathbb{R}^n	5
2. El símbolo de Levi-Civita o símbolo totalmente antisimétrico	8
3. Topología en \mathbb{R}^n	8
4. Funciones en \mathbb{R}^n	9
Chapter 2. Continuidad	11
1. Continuidad	11
Chapter 3. Diferenciabilidad	13
1. Derivadas de campos escalares	13
2. Derivadas parciales	13
3. La diferencial	14
4. Cálculo de planos tangentes a funciones en puntos	15
5. Diferencial para campos vectoriales	15
6. Derivadas parciales de orden superior	16
7. Polinomio de Taylor	16
8. Regla de la cadena	16
Chapter 4. Funciones Implícitas	17
1. Funciones Inversas	17
2. Extremos de funciones escalares	18
3. Cálculo de extremos en conjuntos compactos	21
4. Extremos condicionados y multiplicadores de Lagrange	21
Chapter 5. La integral de Lebesgue	23
1. Sobre la integral de Riemann	23
2. Una nueva integral	23
3. Medida de Lebesgue	24
4. Integración de Lebesgue	25
5. Función integrable Lebesgue	25

1. Bibliografía

- Stewart, J. Cálculo en varias variables.
- Apuntes de Pepe Aranda.
- Tom M. Apostol, "Calculus".
- Tom M. Apostol, Análisis Matemático.

ANEXO

El espacio \mathbb{R}^n

1. El conjunto \mathbb{R}^n

\mathbb{R}^n es el conjunto

$$\mathbb{R}^n = \{(x_1, \dots, x_n) / x_i \in \mathbb{R}, 1 \leq i \leq n\}$$

De momento, este conjunto no tiene ninguna estructura. Para ello, se introduce la noción de espacio vectorial (EV a partir de ahora).

- La suma vectorial:

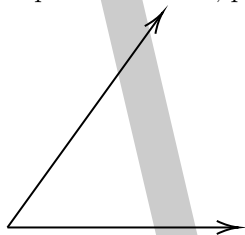
$$\vec{a} + \vec{b} = (a_1 + b_1, \dots, a_n + b_n)$$

- Producto por escalar:

$$k\vec{a} = (ka_1, \dots, ka_n)$$

Con estas dos operaciones, \mathbb{R}^n es EV, y a sus elementos, los vamos a llamar vectores, y los denotaremos por $\vec{x} = (x_1, \dots, x_n)$.

Con esta estructura, los elementos de \mathbb{R}^n se pueden ordenar, por ejemplo, en una cuadrícula.



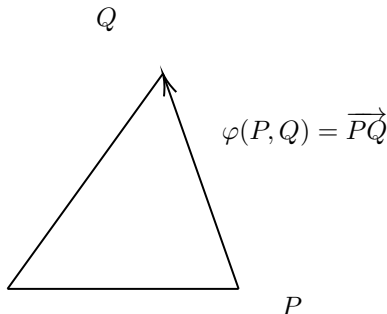
1.1. \mathbb{R}^n como espacio afín. Nos será útil para definir direcciones desde cualquier punto de \mathbb{R}^n . La estructura afín en \mathbb{R}^n se define por la aplicación

$$\begin{aligned} \varphi : \mathbb{R}^n \times \mathbb{R}^n &\mapsto \mathbb{R}^n \\ (\vec{u}, \vec{v}) &\mapsto \varphi(\vec{u}, \vec{v}) \end{aligned}$$

donde $\varphi(\vec{u}, \vec{v})$ se representará como un vector cuyo punto de aplicación está en el extremo de \vec{u} y el extremo, el de $\varphi(\vec{u}, \vec{v})$ en el extremo de \vec{v} . Es fácil comprobar que

$$\varphi(\vec{u}, \vec{v}) = \vec{v} - \vec{u}$$

En este contexto es conveniente llamar puntos a los vectores con punto de aplicación en el $\vec{0}$ y vectores a los vectores cuyo punto de aplicación es arbitrario. Los puntos también los denotaremos mediante letras mayúsculas, y los vectores con letras minúsculas.



A recordar que punto - punto define un vector, y que punto + vector = punto. Con esto ya podemos definir direcciones.

1.2. \mathbb{R}^n como espacio métrico. Para medir longitudes, ángulos y distancias introduciremos en \mathbb{R}^n el **producto escalar**.

DEFINICIÓN 1. El producto escalar entre dos vectores \vec{u} y \vec{v} se define como

$$\begin{aligned} \cdot : \mathbb{R}^n \times \mathbb{R}^n &\mapsto \mathbb{R} \\ (1) \quad \vec{u} \cdot \vec{v} &= \sum_{i=1}^n u_i v_i \end{aligned}$$

El producto escalar tiene las propiedades siguientes:

- $\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$
- $\vec{u} \cdot (\vec{v} + \lambda \vec{w}) = \vec{u} \cdot \vec{v} + \lambda \vec{u} \cdot \vec{w}$
- $\vec{u} \cdot \vec{u} \geq 0 \implies \vec{u} \cdot \vec{u} = 0 \iff \vec{u} = \vec{0}$

Debido a la propiedad 3, podemos definir la longitud (o norma) de un vector como

DEFINICIÓN 2. La longitud o norma de un vector se define como:

$$(2) \quad |\vec{u}| = \sqrt{\vec{u} \cdot \vec{u}} = \sqrt{\sum_{i=1}^n u_i^2}$$

Las propiedades son que:

- $|\vec{v}| = 0 \iff \vec{v} = \vec{0}$
- $|\lambda \vec{u}| = |\lambda| |\vec{u}|$
- Desigualdad de Cauchy-Schwarz:

$$|\vec{u} \cdot \vec{v}| \leq |\vec{u}| |\vec{v}|$$

DEMOSTRACIÓN. Observemos que si uno de los vectores es el vector nulo, entonces la desigualdad se satisface por igualdad, o también si ambos vectores son proporcionales entre sí. Supongamos entonces que u y v son LI. Eso significa que la ecuación $u = \lambda v$ no tiene solución.

$$u - \lambda v = \vec{0}$$

$$(u - \lambda v)(u - \lambda v) = 0$$

$$u \cdot u - 2\lambda u \cdot v + \lambda^2 v \cdot v = 0$$

Recordando la definición de norma queda

$$x^2 |v|^2 - 2xu \cdot v + |u|^2 = 0$$

Ahora tenemos una ecuación de segundo grado en x . Como no puede tener solución, $b^2 - 4ac < 0$.

$$(2u \cdot v)^2 - 4|v|^2 |u|^2 < 0$$

$$2|u \cdot v| < 2|v| |u|$$

$$|u \cdot v| < |v| |u|$$

Como deben cumplirse ambas

$$(3) \quad |u \cdot v| \leq |v| |u|$$

□

- Desigualdad triangular

$$|\vec{u} + \vec{v}| \leq |\vec{u}| + |\vec{v}|$$

DEMOSTRACIÓN. Partimos de $u + v$

$$(u + v) \cdot (u + v) = |u|^2 + 2u \cdot v + |v|^2 = |u + v|^2 \geq 0$$

Por Cauchy-Schwarz, $u \cdot v \leq |u \cdot v| \leq |u| \cdot |v|$

$$|u + v|^2 \leq |u|^2 + |v|^2 + 2|u||v|$$

$$|u + v|^2 \leq (|u| + |v|)^2$$

$$|u + v| \leq |u| + |v|$$

□

De especial importancia son los vectores con norma 1, denominados como **vectores unitarios**.

DEFINICIÓN 3. Definimos como vectores unitarios a esos vectores \vec{v} que cumplen que

$$(4) \quad |\vec{v}| = 1$$

Mediante la desigualdad de Cauchy-Schwarz, podemos obtener un método para medir ángulos. Observemos que de C-S se deduce

$$(5) \quad ||u||v|| \leq u \cdot v \leq |u||v|$$

Si ninguno de los vectores es el nulo, podemos dividir entre las normas

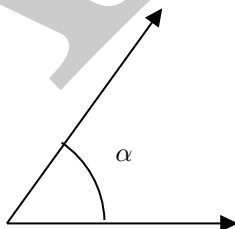
$$(6) \quad -1 \leq \frac{u \cdot v}{|u||v|} \leq 1$$

Entonces, como está acotada en $[-1, 1]$, podemos definir el ángulo α entre u y v como

DEFINICIÓN 4. El ángulo α entre u y v como el ángulo que satisface

$$(7) \quad \cos \alpha = \frac{u \cdot v}{|u||v|}$$

donde $\alpha \in [0, \pi]$. Es de relevancia que α siempre se mide como el ángulo "interior" o el más pequeño.



Si $\alpha = \frac{\pi}{2} \implies u \cdot v = 0 \implies u$ y v son **ortogonales**.

Aunque el ángulo entre $\vec{0}$ y otro vector cualquiera no está definido, sin embargo, se suele decir que $\vec{0}$ es ortogonal a todos los vectores de \mathbb{R}^n

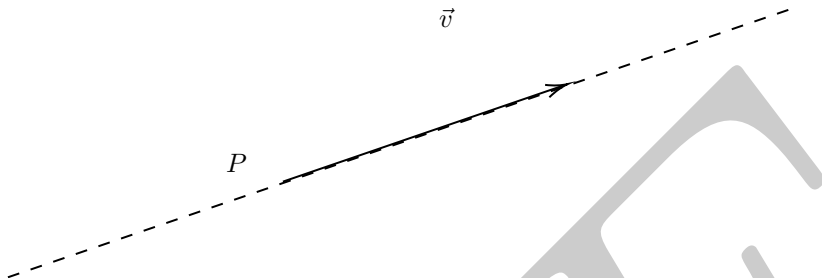
Ahora en \mathbb{R}^2 ya podemos dibujarlos "correctamente". Ahora falta definir distancias entre puntos de \mathbb{R}^n .

DEFINICIÓN 5. Definimos la distancia entre dos puntos P y Q como

$$(8) \quad d(P, Q) = |\overrightarrow{PQ}| = |Q - P|$$

1.3. Rectas e hiperplanos en \mathbb{R}^n . Una recta en \mathbb{R}^n que pasa por un punto P y tiene la dirección $\vec{v} \in \mathbb{R}^n$ se define como los puntos X que satisfacen

$$(9) \quad X = P + t\vec{v}$$



La recta está descrita por un solo parámetro libre, por lo que es un objeto de dimensión 1. Si $v_j \neq 0$ con $j \in \{1, \dots, n\}$ podemos eliminar t

$$(10) \quad t = \frac{x_j - p_j}{v_j}$$

Y entonces

$$(11) \quad x_i = p_i + \frac{x_j - p_j}{v_j} v_i \quad / \quad i \neq j$$

Por tanto, la recta está definida por $n - 1$ ecuaciones. Por tanto, la recta es un objeto de codimensión $n - 1$.

2. El símbolo de Levi-Civita o símbolo totalmente antisimétrico

En \mathbb{R}^n se define por:

$$\varepsilon_{i_1, \dots, i_n} = \begin{cases} 1 & \text{si } i_1, \dots, i_n \text{ es una permutación par} \\ -1 & \text{si es una permutación impar} \\ 0 & \text{en el resto de casos, cuando un índice se repite} \end{cases}$$

i_1, \dots, i_n es una permutación par $123, \dots, n$ si para transformar una en otra necesitamos un número par de trasposiciones. Una trasposición es una permutación en la que solo se intercambian dos y solo dos elementos (no tienen por qué ser adyacentes).

Si queremos abusar todavía más de la notación, podemos usar el criterio de suma de Einstein. Este consiste en suponer que hay una suma siempre que un índice se repita, y la suma se hace en todo el rango de definición del índice:

$$\mathbf{v} \cdot \mathbf{w} = v_i w_i$$

2.1. La delta de Kronecker.

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Los vectores de la base canónica cumplen que:

$$\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}$$

Las bases cuyos vectores cumplan esto se dice que son bases **ortonormales**.

3. Topología en \mathbb{R}^n

DEFINICIÓN 6. Definimos un conjunto abierto de \mathbb{R}^n a partir de la definición de bolas abiertas. Dado $\mathbf{x} \in \mathbb{R}^n$ y dado $r \in \mathbb{R}^+$ se define la bola abierta centrada en \mathbf{x} y de radio r como el conjunto

$$B_r(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^n : d(\mathbf{x}, \mathbf{y}) < r\}$$

A veces convendrá usar bolas abiertas sin el punto central (perforadas), y se denotan por $B_r^*(\mathbf{x})$:

$$B_r^*(\mathbf{x}) = B_r(\mathbf{x}) - \{\mathbf{x}\}$$

Un conjunto abierto $A \subset \mathbb{R}^n$ es abierto si y solo si para todo $\mathbf{x} \in A \exists \delta > 0 : B_\delta(\mathbf{x}) \subset A$. Además, se verifica que $\text{ext}(S) = \mathbb{R}^n - (S \cup \partial S)$

DEFINICIÓN 7. Un conjunto $B \subset \mathbb{R}^n$ es cerrado si y solo si $\mathbb{R}^n - B$ es abierto. Además se verifica que $\text{ext}(B) = \mathbb{R}^n - B$

Dado cualquier conjunto $S \subset \mathbb{R}^n$, podemos clasificar los puntos de \mathbb{R}^n en:

- Puntos interiores de S : si existe una bola centrada en dicho punto contenida en S .
- Puntos exteriores de S : si existe una bola centrada en el punto tal que $B(\mathbf{x}, r) \cap S = \emptyset$.
- Puntos frontera de S : si $\forall r \in \mathbb{R}^+ B(\mathbf{x}, r) \cap S \neq \emptyset$ y $B(\mathbf{x}, r) \cap (\mathbb{R}^n - S) \neq \emptyset$.

Todos estos conjuntos son disjuntos entre sí. Al conjunto de todos los puntos interiores de S se les llama interior de S y lo denotaremos por $\text{int}S$. Al conjunto de todos los puntos exteriores a S se les llama exterior de S y se denota por $\text{ext}S$. A los puntos frontera de S se les llama frontera de S y se denota por ∂S .

DEFINICIÓN 8. A $\text{int}(S) \cup \partial S$ se la denomina adherencia de S y se denota por \bar{S} .

DEFINICIÓN 9. A los puntos \mathbf{x} que cumplen $B^*(\mathbf{x}, r) \cap S \neq \emptyset$ se les llama puntos límite de S .

4. Funciones en \mathbb{R}^n

Sea

$$f: D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$$

Si $n = m = 1$ entonces es una función de variable real. Si $n > 1$ y $m = 1$, entonces f es una función real de variable vectorial o campo escalar. Para hacernos una idea del comportamiento de estas funciones es conveniente dibujar estas funciones (cuando sea posible). Para ello nos serán útiles los **conjuntos de nivel**:

DEFINICIÓN 10. Sea $f: \mathbb{R}^n \rightarrow \mathbb{R}$

$$L_f(c) = \{\mathbf{x} \in D : f(\mathbf{x}) = c\} = f^{-1}(c)$$

Que es precisamente la preimagen de c .

4.1. Campos vectoriales. Estas funciones son de la forma

$$\mathbf{f}: D \subset \mathbb{R}^n \rightarrow \mathbb{R}^{m>1}$$

Estas funciones toman un vector \mathbf{x} como entrada y se mapean a un vector $\mathbf{f}(\mathbf{x})$. f_i son componentes (campos escalares). Si $n = 1$, \mathbf{f} es una curva en \mathbb{R}^m . Si $n = 2$, \mathbf{f} es una superficie en \mathbb{R}^m . En general, si $n < m$, \mathbf{f} es una variedad n -dimensional dentro de \mathbb{R}^m . Si $n = m$, entonces se puede representar a \mathbf{f} sobre su dominio. Si $n > m > 1$, ya no podemos dibujar a \mathbf{f} .

AVE

CHAPTER 2

Continuidad

DEFINICIÓN 11. Dada $\mathbf{f} : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$, si \mathbf{a} es un punto límite de D , decimos que $\mathbf{f}(\mathbf{x})$ tiende a \mathbf{b} cuando \mathbf{x} tiende a \mathbf{a} , y se denota por

$$\lim_{\mathbf{x} \rightarrow \mathbf{a}} \mathbf{f}(\mathbf{x}) = \mathbf{b}$$

Si se cumple que

$$\forall \varepsilon > 0 \exists \delta > 0 : |\mathbf{x} - \mathbf{a}| < \delta \implies |\mathbf{f}(\mathbf{x}) - \mathbf{b}| < \varepsilon$$

Esta definición es equivalente al límite

$$\lim_{|\mathbf{x} - \mathbf{a}| \rightarrow 0} |\mathbf{f}(\mathbf{x}) - \mathbf{b}| = 0$$

En estas definiciones no decimos cómo el punto \mathbf{x} se acerca a \mathbf{a} . Por tanto, si el límite depende de por donde nos acerquemos a \mathbf{a} , entonces este no existe.

Para evitar utilizar la definición a la hora de calcular límites, utilizaremos los siguientes teoremas.

TEOREMA 1. Sean $\mathbf{f} : D \rightarrow \mathbb{R}^m$ y $\mathbf{g} : C \rightarrow \mathbb{R}^m$ con $D \cap C \neq \emptyset$, y sea \mathbf{a} un punto límite de $D \cap C$, y además suponemos que

$$\begin{aligned} \lim_{\mathbf{x} \rightarrow \mathbf{a}} \mathbf{f}(\mathbf{x}) &= \mathbf{b} \\ \lim_{\mathbf{x} \rightarrow \mathbf{a}} \mathbf{g}(\mathbf{x}) &= \mathbf{c} \end{aligned}$$

Entonces se cumple que - $\lim_{\mathbf{x} \rightarrow \mathbf{a}} \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x}) = \mathbf{b} + \mathbf{c}$

- $\lim_{\mathbf{x} \rightarrow \mathbf{a}} \lambda \mathbf{f}(\mathbf{x}) = \lambda \mathbf{b}$

- $\lim_{\mathbf{x} \rightarrow \mathbf{a}} \mathbf{f}(\mathbf{x}) \cdot \mathbf{g}(\mathbf{x}) = \mathbf{b} \cdot \mathbf{c}$

- $\lim_{\mathbf{x} \rightarrow \mathbf{a}} |\mathbf{f}(\mathbf{x})| = |\mathbf{b}|$

1. Continuidad

DEFINICIÓN 12. Una función \mathbf{f} es continua en un punto \mathbf{a} si

$$\lim_{\mathbf{x} \rightarrow \mathbf{a}} \mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{a})$$

AVE

Diferenciabilidad

1. Derivadas de campos escalares

La derivada de un campo escalar, f en un punto \mathbf{x} sería entonces

$$Df(\mathbf{x}) = \lim_{\mathbf{h} \rightarrow 0} \frac{f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x})}{\mathbf{h}}$$

Pero la división por vectores no está definida, por lo que vamos a necesitar otra forma. Por tanto, vamos a definir la **derivada respecto de un vector**.

DEFINICIÓN 13. La derivada de una función f respecto de un vector \mathbf{v} se define como

$$f'(\mathbf{x})_{\mathbf{v}} = \lim_{h \rightarrow 0} \frac{f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x})}{h}$$

Como estas derivadas son un verdadero dolor de calcular, vamos a definir otra forma de calcularlas

TEOREMA 2. Sea $g(t) = f(\mathbf{x} + t\mathbf{v})$. Si $g'(t)$ o $f'(\mathbf{x} + t\mathbf{v}, \mathbf{v})$ existe, entonces también existe la otra y son iguales. En particular, cuando $t = 0$,

$$f'(\mathbf{x}, \mathbf{v}) = g'(0)$$

TEOREMA 3. Sean f y \mathbf{v} tales que $f(\mathbf{x} + t\mathbf{v})$ está definida para todo $t \in [0, 1]$ y supongamos que $\exists f'(\mathbf{x} + t\mathbf{v}, \mathbf{v}) \forall t \in [0, 1]$, entonces

$$\exists \tau \in (0, 1) : f(\mathbf{x} + \mathbf{v}) - f(\mathbf{x}) = f'(\mathbf{x} + \tau\mathbf{v}, \mathbf{v})$$

El problema de estas derivadas respecto de un vector es que no se tiene en cuenta la distancia que separa los puntos donde evaluamos f , pues esta es

$$|\mathbf{x} + h\mathbf{v} - \mathbf{x}| = |h\mathbf{v}|$$

Para solucionar este problema, utilizaremos vectores de norma 1. Entonces la derivada se llama **derivada direccional**.

2. Derivadas parciales

Las derivadas parciales son aquellas que se hacen respecto a un eje. Si como vector unitario usamos uno de la base canónica de \mathbb{R}^n

$$\{\mathbf{e}_i\}$$

donde

$$(\mathbf{e}_i)_j = \delta_{ij}$$

entonces la correspondiente derivada direccional en un punto \mathbf{x} es

$$f'(\mathbf{x}; \mathbf{e}_i)$$

es la **derivada parcial** respecto a \mathbf{e}_i , y se denota por $D_i f = \partial_i f$. Observemos que

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

Como todos los componentes que no estén alineados con \mathbf{e}_i no cambian, podemos tomarlos como si **fuesen constantes**, es decir, como una derivada ordinaria respecto a x_i .

3. La diferencial

Las derivadas direccionales nos dan una información muy pobre del comportamiento de una función. Queremos buscar algo análogo a

$$Df(\mathbf{x}) = \lim_{\mathbf{h} \rightarrow 0} \frac{f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x})}{\mathbf{h}}$$

El problema es que tenemos un vector dividiendo. Vamos a tomar que \mathbf{h} es un vector finito

$$(12) \quad \begin{aligned} Df(\mathbf{x}) + \varepsilon(\mathbf{x}, \mathbf{h}) &= \frac{f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x})}{\mathbf{h}} \\ \mathbf{h}(Df(\mathbf{x}) + \varepsilon(\mathbf{x}, \mathbf{h})) &= f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) \end{aligned}$$

Se tiene que cumplir que

$$\lim_{\mathbf{h} \rightarrow 0} \varepsilon(\mathbf{x}, \mathbf{h}) = 0$$

ahora sigue que

$$\frac{d}{dt}(f \circ \mathbf{c}) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(\mathbf{x}) \frac{dx_i}{dt}$$

Para que el lado izquierdo sea escalar, \mathbf{h} debe ser una matriz $n \times 1$, y Df otra de $1 \times n$. Podemos reescribir esto como

$$f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) = T_{\mathbf{x}}(\mathbf{h}) + |\mathbf{h}|E(\mathbf{x}, \mathbf{h})$$

Como $\varepsilon \rightarrow 0$ si $h \rightarrow 0$, entonces también lo hará $E(\mathbf{x}, \mathbf{h})$. Entonces obtenemos

$$f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) = T_{\mathbf{x}}(\mathbf{h}) + o(|\mathbf{h}|)$$

DEFINICIÓN 14. Sean $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$ y $\mathbf{x} \in \text{int}(S)$, $B_r(\mathbf{x}) \subset S$ y $\mathbf{v} \in \mathbb{R}^n$ tal que $\mathbf{x} + \mathbf{v} \in B_r(\mathbf{x})$. ¿Decimos que f es diferenciable en \mathbf{x} si existe una transformación lineal $T_{\mathbf{x}} : \mathbb{R}^n \rightarrow \mathbb{R}$ tal que

$$f(\mathbf{x} + \mathbf{v}) - f(\mathbf{x}) = T_{\mathbf{x}}(\mathbf{v}) + o(|\mathbf{v}|)$$

Vamos a ver si la diferencial es capaz de detectar discontinuidades. Supongamos que $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ es diferenciable en $\mathbf{x} \in \text{int}(D)$. Veamos si esto implica continuidad en \mathbf{x} : Si f es continua en \mathbf{x} se cumplirá

$$\lim_{\mathbf{h} \rightarrow 0} f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x})$$

Cálculemos entonces este límite:

$$(13) \quad \begin{aligned} \lim_{\mathbf{h} \rightarrow 0} f(\mathbf{x} + \mathbf{h}) &= \lim_{\mathbf{h} \rightarrow 0} f(\mathbf{x}) + T_{\mathbf{x}}(\mathbf{h}) + o(|\mathbf{h}|) \\ &= f(\mathbf{x}) + T_{\mathbf{x}}(\mathbf{0}) + \lim_{\mathbf{h} \rightarrow 0} |\mathbf{h}|o(1) \\ &= f(\mathbf{x}) \end{aligned}$$

Por tanto, si f es diferenciable en \mathbf{x} , también es continua en ese punto. Ahora, ¿cómo son las derivadas de una función diferenciable? Calculemos la derivada de f respecto de un \mathbf{v} en \mathbf{x} :

$$(14) \quad \begin{aligned} f'(\mathbf{x}, \mathbf{v}) &= \lim_{h \rightarrow 0} \frac{f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x})}{h} \\ &= \lim_{h \rightarrow 0} \frac{T_{\mathbf{x}}(h\mathbf{v}) + o(h)}{h} \\ &= \lim_{h \rightarrow 0} \frac{hT_{\mathbf{x}}(\mathbf{v}) + h \cdot o(1)}{h} \\ &= \lim_{h \rightarrow 0} T_{\mathbf{x}}(\mathbf{v}) + o(1) \\ &= T_{\mathbf{x}}(\mathbf{v}) \\ f'(\mathbf{x}, \mathbf{v}) &= T_{\mathbf{x}}(\mathbf{v}) \end{aligned}$$

Si $\mathbf{v} = (v_1, \dots, v_n) \in \mathbb{R}^n$, entonces se puede escribir como combinación lineal de vectores de la base canónica de \mathbb{R}^n .

$$\mathbf{v} = \sum v_i \mathbf{e}_i$$

entonces

$$T_{\mathbf{x}}(\mathbf{v}) = T_{\mathbf{x}}\left(\sum v_i \mathbf{e}_i\right) = \sum v_i T_{\mathbf{x}}(\mathbf{e}_i)$$

Sabemos que una aplicación se define por cómo actúa en los vectores de la base. Como sabemos que $f'(\mathbf{x}, \mathbf{v}) = T_{\mathbf{x}}(\mathbf{v})$, entonces

$$= \sum v_i f'(\mathbf{x}, \mathbf{e}_i) = \sum v_i \frac{\partial f}{\partial x_i}$$

La conclusión de esto es que

$$f'(\mathbf{x}, \mathbf{v}) = \sum v_i \frac{\partial f}{\partial x_i}$$

DEFINICIÓN 15. El operador Nabla es un operador lineal que se define como

$$\nabla = \left(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n} \right)$$

Luego ∇f es el **gradiente** de f .

Por tanto, si f es diferenciable entonces

$$f'(\mathbf{x}, \mathbf{v}) = \nabla f \cdot \mathbf{v}$$

Con el gradiente es fácil determinar en qué direcciones f crece más deprisa. Sea \mathbf{u} un vector unitario, entonces:

$$f'(\mathbf{x}, \mathbf{u}) = \mathbf{u} \cdot \nabla f = |\nabla f| \cos \theta$$

TEOREMA 4. Si $\mathbf{v} = \lambda \mathbf{w} \neq \mathbf{0}$ y f es un campo escalar para el que existe $f'(\mathbf{x}, \mathbf{v})$ entonces

$$f'(\mathbf{x}, \mathbf{v}) = \lambda f'(\mathbf{x}, \mathbf{w})$$

Para reconocer si una función es diferenciable usaremos el siguiente teorema:

TEOREMA 5. Si existen las derivadas parciales $\frac{\partial f}{\partial x_i}$ en una cierta bola $B_r(\mathbf{x})$ y son continuas en \mathbf{x} entonces f es diferenciable en \mathbf{x} .

4. Cálculo de planos tangentes a funciones en puntos

Supongamos que tenemos una función $f(x, y)$. Vamos a definir una función de la forma

$$g(x, y, z) = f(x, y) - z$$

Ahora, evaluaremos la función $f(x_1, x_2)$, y calcularemos

$$\nabla g(x_1, x_2, f(x_1, x_2)) = (\alpha, \beta, \gamma)$$

Ahora definiremos un plano como

$$\pi \equiv \alpha x + \beta y + \gamma z + c = 0$$

Para calcular c , sustituimos en el plano los valores de $x_1, x_2, f(x_1, x_2)$, y despejamos.

5. Diferencial para campos vectoriales

Cada una de las componentes de un campo vectorial $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_m(\mathbf{x}))$ es un campo escalar, y si cada una de ellas es diferenciable en \mathbf{x} , podemos introducir la matriz **jacobiana** como

$$D\mathbf{f}(\mathbf{x}) = \begin{pmatrix} \nabla f_1(\mathbf{x}) \\ \vdots \\ \nabla f_m(\mathbf{x}) \end{pmatrix}$$

Geométricamente, si tenemos una curva $\mathbf{c}(t)$, que se mappea a \mathbb{R}^m como $\mathbf{f}(\mathbf{c}(t))$, se tiene que su vector tangente \mathbf{v} en \mathbf{x} es $D[\mathbf{f}(\mathbf{x})]\mathbf{v}$.

6. Derivadas parciales de orden superior

DEFINICIÓN 16. Una función $f : A \rightarrow \mathbb{R}$ es de clase \mathcal{C}^r en A , con $r \in \mathbb{N}_0$ si todas sus derivadas parciales de orden r existen y son continuas.

TEOREMA 6. Si $f \in \mathcal{C}^2$, entonces

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$$

7. Polinomio de Taylor

Sea $A \subset \mathbb{R}^n$ un conjunto abierto y convexo (si $\mathbf{a}, \mathbf{b} \in A$ está contenido en A , entonces \mathbf{ab} está contenido en A). Si $f \in \mathcal{C}^k$, entonces, definimos el polinomio de Taylor de grado k centrado en \mathbf{a} como

$$\begin{aligned} P_{\mathbf{a},k}f(\mathbf{x}) &= f(\mathbf{a}) + \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x_i - a_i) \\ &+ \frac{1}{2!} \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j^2}(x_i - a_i)(x_j - a_j) + \dots \\ &+ \frac{1}{k!} \sum_{i_1, \dots, i_k=1}^n \frac{\partial^k f}{\partial x_{i_1} \dots \partial x_{i_k}^k}(x_{i_1} - a_{i_1}) \dots (x_{i_k} - a_{i_k}) \end{aligned} \quad (15)$$

8. Regla de la cadena

$$D(\mathbf{f} \circ \mathbf{g})(\mathbf{x}) = D(\mathbf{f}(\mathbf{g}(\mathbf{x})))D\mathbf{g}(\mathbf{x})$$

Cada uno de los componentes se define como

$$[D(\mathbf{f} \circ \mathbf{g})(\mathbf{x})]_{ij} = \frac{\partial f_i(\mathbf{g}(\mathbf{x}))}{\partial x_j} = \sum_{k=1}^p \frac{\partial f_i}{\partial g_k} \frac{\partial g_k}{\partial x_j}$$

8.1. Derivada de un campo escalar a lo largo de una curva.

$$\frac{d}{dt}(f \circ \mathbf{c}) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(\mathbf{x}) \frac{dx_i}{dt}$$

Funciones Implícitas

1. Funciones Inversas

Dada $f : A \rightarrow B$, se dice que f tiene inversa si existe una función $g : B \rightarrow A$ tal que

$$f \circ g = Id_b$$

y similarmente

$$g \circ f = Id_a$$

Si esta función existe, se denota por

$$f^{-1} \neq \frac{1}{f}$$

Se puede demostrar que f^{-1} existe si y solo si f es inyectiva y sobreyectiva (biyectiva). Sin embargo, dadas ciertas funciones que no sean ni inyectivas ni sobreyectivas, se pueden restringir de forma que sean biyectivas.

Ejemplos:

- $f(x) = x^2$. Esta función es sobreyectiva, pero no es inyectiva, y por tanto no tiene inversa. Sin embargo, podemos restringir el dominio tal que $x \geq 0$, y entonces sí que es inyectiva, por tanto biyectiva, y por tanto tiene inversa.
- $g(x) = \cos x$. Esta función es periódica, pero si restringimos la función a un período $[0, \pi]$, entonces es biyectiva.

Sea $\bar{f}(\mathbf{x}) = \mathbf{y}$. Queremos ver si podemos invertir esta relación (\bar{f}^{-1}), y obtener

$$\mathbf{x} = \bar{f}^{-1}(\mathbf{y})$$

Para ello, introducimos el siguiente teorema.

TEOREMA 7 (Teorema de la función inversa). *Dada una función $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$, esta tiene inversa si y solo si el determinante jacobiano (determinante de su diferencial) es distinto a 0.*

Dado un sistema de ecuaciones, no necesariamente lineales, siempre lo podemos escribir de la forma:

$$(16) \quad \begin{aligned} f_1(x_1, \dots, x_{n+k}) &= 0 \\ &\dots \\ f_n(x_1, \dots, x_{n+k}) &= 0 \end{aligned}$$

Vamos a analizar en qué condiciones podemos despejar n incógnitas en función de las k restantes, que serían parámetros libres. En este caso, estas ecuaciones describen un objeto de k dimensiones en \mathbb{R}^{n+k} . Para distinguir entre parámetros libres e incógnitas, introduciremos la siguiente notación:

- Las incógnitas son x_1, \dots, x_n .
- Los parámetros libres serán $t_1 = x_{n+1}, \dots, t_k = x_{n+k}$.

Además, introduciremos la función vectorial $\bar{f} = (f_1, \dots, f_n)$, entonces el sistema a resolver acaba siendo

$$\bar{f}(\mathbf{x}, \mathbf{t}) = \mathbf{0}$$

Ahora, supongamos que $(\mathbf{x}_0, \mathbf{t}_0)$ es una solución del sistema que cumple la ecuación anteriormente establecida, y que además $\bar{f} \in \mathcal{C}^1$. Entonces, podemos hacer la aproximación

$$\bar{f}(\mathbf{x}, \mathbf{t}) \approx \bar{f}(\mathbf{x}_0, \mathbf{t}_0) + D\bar{f}(\mathbf{x}_0, \mathbf{t}_0)(\mathbf{x} - \mathbf{x}_0, \mathbf{t} - \mathbf{t}_0)$$

Como hemos asumido que la anterior tupla era solución del sistema, y queremos una solución a la ecuación, queda

$$D\bar{f}(\mathbf{x}_0, \mathbf{t}_0)(\mathbf{x} - \mathbf{x}_0, \mathbf{t} - \mathbf{t}_0) = \mathbf{0}$$

Si escribimos explícitamente el diferencial, obtenemos que

$$\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_n} & \frac{\partial f_1}{\partial t_1} & \dots & \frac{\partial f_1}{\partial t_k} \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \dots & \frac{\partial f_n}{\partial x_n} & \frac{\partial f_n}{\partial t_1} & \dots & \frac{\partial f_n}{\partial t_k} \end{bmatrix} \cdot (\mathbf{x} - \mathbf{x}_0, \mathbf{t} - \mathbf{t}_0) = \mathbf{0}$$

Evaluado en $(\mathbf{x}_0, \mathbf{t}_0)$

Aquí sigue que

$$(D_x \bar{f}(\mathbf{x}_0, \mathbf{t}_0), D_t \bar{f}(\mathbf{x}_0, \mathbf{t}_0))(\mathbf{x} - \mathbf{x}_0, \mathbf{t} - \mathbf{t}_0) = \mathbf{0}$$

Reordenando y simplificando, obtenemos que debe existir la inversa de la matriz $D_x \bar{f}$, y por tanto su determinante debe ser distinto a 0.

DEFINICIÓN 17. Dada una función $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$, y sea $\mathbf{a} \in D$, f tiene un mínimo absoluto en \mathbf{a} si

$$f(\mathbf{a}) \leq f(\mathbf{x}) \forall \mathbf{x} \in D$$

La definición de máximo es análoga.

DEFINICIÓN 18 (Mínimo local). Esta definición sigue de tomar los puntos en una cierta bola. Es decir, un punto \mathbf{a} será mínimo local en B_δ si y solo si

$$f(\mathbf{a}) \leq f(\mathbf{x}) \forall \mathbf{x} \in B_\delta$$

2. Extremos de funciones escalares

Usaremos con frecuencia el siguiente teorema:

TEOREMA 8. Sea $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ una función diferenciable en $\mathbf{a} \in \text{int}(D)$. Si f tiene un extremo local en \mathbf{a} , entonces

$$\nabla f(\mathbf{a}) = \mathbf{0}$$

DEMOSTRACIÓN. Hallemos la derivada de f respecto de un vector $\mathbf{v} \in \mathbb{R}^n$ cualquiera:

$$(17) \quad f'(\mathbf{a}, \mathbf{v}) = \lim_{h \rightarrow 0} \frac{f(\mathbf{a} + h\mathbf{v}) - f(\mathbf{a})}{h}$$

Sabemos que este límite existe ya que f es diferenciable. Supongamos que f tiene un mínimo local en \mathbf{a} . En particular, también existen los límites laterales en $\mathbf{x} = \mathbf{a}$.

Si f tiene un mínimo local en \mathbf{a} , entonces

$$f(\mathbf{a} + h\mathbf{v}) - f(\mathbf{a}) \geq 0$$

para un h lo suficientemente pequeño. Entonces, la derivada lateral en ese punto, $f'(\mathbf{a}, \mathbf{v})^+ \geq 0$. Y la otra derivada lateral, cumple que $f'(\mathbf{a}, \mathbf{v})^- \leq 0$. Para que ambas ecuaciones se cumplan, y teniendo en cuenta que es diferenciable:

$$f'(\mathbf{a}, \mathbf{v}) = \nabla f(\mathbf{a}) \cdot \mathbf{v} = 0$$

Como esto se debe cumplir para todos los vectores, entonces el gradiente es perpendicular a todos los vectores. El único que lo cumple es

$$\nabla f(\mathbf{a}) = 0$$

□

DEFINICIÓN 19. Si $\mathbf{a} \in \text{int}(D)$ y $\nabla f(\mathbf{a}) = \mathbf{0}$, entonces \mathbf{a} se llama punto crítico o punto estacionario de f . Si desarrollamos la función como polinomio de Taylor de primer orden en \mathbf{a} , veremos que obtenemos

$$P_{1,\mathbf{a}} = f(\mathbf{a})$$

Sin embargo, estas herramientas son muy pobres para analizar correctamente el comportamiento de una función. Si desarrollamos f hasta segundo orden mediante el polinomio de Taylor, obtendremos más información acerca de cómo se comporta la función cerca del punto crítico en el que estamos interesados. Supongamos que \mathbf{a} es un punto crítico:

$$f(\mathbf{x}) = f(\mathbf{a}) + \nabla f(\mathbf{a})\mathbf{h} + \frac{1}{2}\mathbf{h}^T H(\mathbf{a})\mathbf{h} + o(|\mathbf{h}|^2)$$

Donde $\mathbf{h} = \mathbf{x} - \mathbf{a}$. Aplicando el hecho de que \mathbf{a} es punto crítico, el término del gradiente se va

$$= f(\mathbf{a}) + \frac{1}{2}\mathbf{h}^T H(\mathbf{a})\mathbf{h} + o(|\mathbf{h}|^2)$$

Aquí podemos tener los siguientes casos:

- $\mathbf{h}^T H(\mathbf{a})\mathbf{h} > 0 \forall \mathbf{h} \neq \mathbf{0}$. Entonces, f crece en puntos cercanos a \mathbf{a} , entonces \mathbf{a} es un mínimo.
- En el caso de ser menor a 0, análogamente, \mathbf{a} es un máximo.
- Si cambia en función de \mathbf{h} , entonces es un **punto de silla**.
- Si $H(\mathbf{a}) = \mathbf{0}$, nos falta información para saber el tipo de punto crítico que tenemos.

DEFINICIÓN 20. Si $A \in \mathcal{M}_{n \times n}$ es simétrica, entonces la función

$$Q : \mathbb{R}^n \rightarrow \mathbb{R}$$

donde

$$\mathbf{v} \rightarrow Q(\mathbf{v}) = \mathbf{v}^t A \mathbf{v}$$

se llama forma cuadrática simétrica.

TEOREMA 9. Sea $A \in \mathcal{M}_{n \times n}$ una matriz simétrica, y sea d_i el determinante de una de sus submatrices:

- Si $d_i > 0 \forall i$, Q es definida positiva.
- Si $d_i > 0$ para i par y $d_i < 0$ para i impar, es definida negativa

TEOREMA 10. Si tenemos una matriz simétrica $n \times n$, entonces

- Si los autovalores de A son positivos entonces Q es definida positiva.
- Si los autovalores de A son negativos entonces Q es definida negativa.
- Si alguno es 0 entonces Q es singular.
- En el resto de casos Q es indefinida.

También podemos usar el hecho de que, siendo λ_i los autovalores de A , entonces

$$\sum \lambda_i = \text{tr}(A)$$

Ahora combinando todo

TEOREMA 11. Tenemos que f tiene en \mathbf{a} en función de $H(\mathbf{a})$:

- Mínimo local si H es definida positiva.
- Máximo local si H es definida negativa.

- *Punto de silla si H es indefinida.*
- *Nos falta información si H es singular.*

3. Cálculo de extremos en conjuntos compactos

DEFINICIÓN 21. Un conjunto $D \subset \mathbb{R}^n$ es acotado si existe $M \in \mathbb{R}$ tal que

$$\|\mathbf{x}\| \leq M \quad \forall \mathbf{x} \in D$$

Ahora ya podemos definir el siguiente tipo de conjunto:

DEFINICIÓN 22. Un conjunto $D \subset \mathbb{R}^n$ es compacto si es cerrado y acotado.

TEOREMA 12. Sea $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}$ continua en U , tal que U es compacto, existen al menos dos puntos \mathbf{a}, \mathbf{b} tales que:

$$f(\mathbf{a}) \leq f(\mathbf{x}) \leq f(\mathbf{b}) \quad \forall \mathbf{x} \in U$$

4. Extremos condicionados y multiplicadores de Lagrange

Estos problemas consisten en hallar extremos de una función

$$f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}$$

sujetos a una serie de condiciones, que se pueden escribir como ecuaciones.

$$\begin{pmatrix} g_1(\mathbf{x}) = 0 \\ \dots \\ g_n(\mathbf{x}) = 0 \end{pmatrix} : k < n$$

Para ello, utilizaremos la generalización del siguiente teorema:

TEOREMA 13 (Teorema de multiplicadores de Lagrange). Sea $g : U \subset \mathbb{R}^n \rightarrow \mathbb{R}$, es \mathcal{C}^1 en U , y sea

$$S = \{\mathbf{x} \in \mathbb{R}^n : g(\mathbf{x}) = 0, \nabla g(\mathbf{x}) \neq \mathbf{0}\}$$

Entonces sea también $f : U \rightarrow \mathbb{R}$, y que también es \mathcal{C}^1 en U . Entonces si f tiene un extremo sobre S en $\mathbf{p} \in S$, entonces

$$\exists \lambda \in \mathbb{R} : \nabla f(\mathbf{p}) = \lambda \nabla g(\mathbf{p})$$

DEMOSTRACIÓN. Vamos a ver que $\nabla f(\mathbf{p})$ y $\nabla g(\mathbf{p})$ tienen la misma dirección. Vamos a empezar analizando la dirección de $\nabla g(\mathbf{p})$. Vemos que S define un conjunto de nivel de g , por tanto $\forall \mathbf{x} \in S \nabla g(\mathbf{p}) \perp \mathbf{x}$. Para ver que ∇f tiene la misma dirección tomemos una curva cualquiera $\mathbf{c}(t) : \mathbb{R} \rightarrow S$, con la única condición de que pase por \mathbf{p} en $t = t_0$. Como $f|_S$ tiene un extremo en \mathbf{p} y $f \in \mathcal{C}^1$ en S , entonces

$$\left. \frac{df(\mathbf{c})}{dt} \right|_{t=t_0} = 0$$

Por la regla de la cadena, podemos reescribir la ecuación anterior como:

$$\left. \frac{d}{dt} f(\mathbf{c}(t)) \right|_{t=t_0} = \nabla f(\mathbf{c}(t_0)) \cdot \frac{d\mathbf{c}}{dt} = \nabla f(\mathbf{p}) \mathbf{c}'(t_0) = 0$$

Como esta curva es arbitraria, esto significa que ∇f es perpendicular a cualquier curva en S que pase por \mathbf{p} , y por tanto, proporcional a $\nabla g(\mathbf{p})$, lo que significa que

$$\nabla f(\mathbf{p}) = \lambda \nabla g(\mathbf{p})$$

Donde λ se denomina como **multiplicador de Lagrange**. □

Generalicemos el teorema para el caso en el que los extremos han de cumplir más de una condición. Queremos hallar los extremos de $f(\mathbf{x})$ restringida a los puntos que cumplen el conjunto de ecuaciones

$$g_{i < n}(\mathbf{x}) = 0$$

que podemos abreviar por

$$\mathbf{g}(\mathbf{x}) = (g_1, \dots, g_k) = \mathbf{0}$$

obteniendo así el siguiente teorema:

TEOREMA 14. \mathbf{g} de clase \mathcal{C}^1 en un abierto $U \subset \mathbb{R}^n$. Sea S el conjunto

$$S = \{\mathbf{x} \in U : \mathbf{g}(\mathbf{x}) = \mathbf{0}, \text{rg}(D\mathbf{g}(\mathbf{x})) = k\}$$

Queremos que sea igual a k , ya que entonces el rango será máximo, ya que $k \leq n$. Debido a esto, $\dim(S) = n - k$, porque las k ecuaciones $\mathbf{g}(\mathbf{x}) = \mathbf{0}$ son independientes en cada punto. Como $S \subset \mathbb{R}^n$ en cada punto de S podemos escoger una base de \mathbb{R}^n , con $n - k$ vectores tangentes, y el resto, k vectores perpendiculares a S . Cada ∇g_i es perpendicular a S , pues $S \subset \{g_i = 0\}$, como estos gradientes son LI, entonces expanden el subespacio ortogonal a S en cada uno de sus puntos.

Suponiendo que f tiene un extremo en \mathbf{p} , condicionado a por $\mathbf{g}(\mathbf{p}) = \mathbf{0}$, dada cualquier curva $\mathbf{x}(t)$ que pase por \mathbf{p} ($\exists t_0 : \mathbf{x}(t_0) = \mathbf{p}$) contenida en S se cumplirá

$$\frac{d}{dt}f(\mathbf{x}(t)) = 0$$

de donde se deduce, como se hizo anteriormente, que

$$\nabla f(\mathbf{p}) \perp S$$

entonces $\nabla f(\mathbf{p})$ está en el subespacio ortogonal a S , y una base de este subespacio era el conjunto $\{\nabla g_i(\mathbf{p})\}$. Esto quiere decir que existen escalares $\{\lambda_i\}$ tal que:

$$\nabla f(\mathbf{p}) = \sum_i^k \lambda_i \nabla g_i(\mathbf{p})$$

Y aquí, el conjunto $\{\lambda_i\}$ son los multiplicadores de Lagrange.

Para hallar los extremos de una función f condicionados por $g_{i \leq k} = 0$, resolveremos el sistema

$$\begin{aligned} \nabla f(\mathbf{x}) &= \sum_i^k \lambda_i \nabla g_i(\mathbf{x}) \\ g_i(\mathbf{x}) &= 0 \\ &\dots \\ g_k(\mathbf{x}) &= 0 \end{aligned} \tag{18}$$

Donde tendremos $n + k$ incógnitas y también $n + k$ incógnitas. A veces para resolver este sistema se introduce la **función de Lagrange**:

$$F(x_1, \dots, x_n, \lambda_1, \dots, \lambda_k) = f(\mathbf{x}) - \lambda_1 g_1(\mathbf{x}) - \dots - \lambda_k g_k(\mathbf{x})$$

Con esta función, el sistema planteado anteriormente se reduce a

$$\nabla F = \mathbf{0}$$

Donde ahora el gradiente es

$$\nabla F = \left(\frac{\partial F}{\partial x_1}, \dots, \frac{\partial F}{\partial x_n}, \frac{\partial F}{\partial \lambda_1}, \dots, \frac{\partial F}{\partial \lambda_k} \right)$$

Una vez halladas las soluciones de este sistema, evaluamos f en cada una de ellas, lo que nos permitirá saber los máximos y los mínimos de la función.

La integral de Lebesgue

1. Sobre la integral de Riemann

Para integrar una función f en un intervalo $[a, b]$, introducíamos el concepto de partición de $[a, b]$:

$$P = \{x_0 = a, x_1, \dots, x_n = b\}$$

Una partición de $[a, b]$ es un conjunto finito de puntos ordenados. También necesitamos el ínfimo y el supremo de una función en un conjunto:

$$\inf f = \inf\{f(x) : x \in P\}$$

$$\sup f = \sup\{f(x) : x \in P\}$$

Con esto vamos a construir las **sumas de Riemann**, superiores e inferiores, denotadas por S y s , respectivamente, y se definen como: dada f definida en $[a, b]$ introducimos una partición P de este intervalo, y construimos

$$s = \sum_i^n (x_i - x_{i-1}) \inf f$$

es la suma inferior, y

$$S = \sum_i^n (x_i - x_{i-1}) \sup f$$

Esto define el área de la curva bajo la función f en el intervalo $[a, b]$. Si tomamos el límite $n \rightarrow \infty$, vemos que $S = s$, y entonces

$$\int_b^a f = \lim_{n \rightarrow \infty} \sum (x_i - x_{i-1}) \sup f$$

TEOREMA 15. *Una función $f(x)$ es integrable de Riemann en $[a, b]$ si es continua en ese intervalo.*

Sin embargo, la integral de Riemann no es suficientemente buena, ya que:

- No funciona si hay muchas discontinuidades.

2. Una nueva integral

Los problemas de la integral de Riemann se deben a querer calcular áreas sobre intervalos, pero las funciones integrables Riemann han de ser lo suficientemente buenas en dichos intervalos.

En vez de considerar solo la medida de intervalos, vamos a fijarnos en los valores que toma f , y entonces medir las preimágenes:

$$f^{1-}(\{a\}) = \{x : f(x) = a\}$$

De manera informal, queremos

$$f : [a, b] \subset \mathbb{R} \rightarrow \mathbb{R} : f \geq 0 \forall x \in [a, b]$$

para ello, introducimos una partición en la **imagen** de f .

$$P : 0 = y_0 < y_1 < \dots < y_n = \sup f$$

Construimos entonces la suma inferior de Lebesgue:

$$\mathcal{L}(f, P, [a, b]) = \sum_{j=1}^n \mu(f^{-1}([y_{j-1}, y_j])) \cdot y_{j-1}$$

Donde μ es la **medida** del intervalo. La medida que usaremos será una generalización de la medida de longitud de intervalos abiertos de la forma $(a, b) \subset \mathbb{R}$:

$$l((a, b)) = b - a$$

además de que $l(\emptyset) = 0$. También

$$l(-\infty, a) = l(a, \infty) = l(-\infty, \infty) = \infty$$

La generalización para medir cualquier $A \subset \mathbb{R}$ es:

$$|A| = \inf \left\{ \sum_{k=1}^{\infty} l(I_k) : I_k \text{ son intervalos abiertos tales que } A \subset \bigcup I_k \right\}$$

Esta medida no será la que finalmente usemos, ya que existen $A, B \subset \mathbb{R}$, disjuntos, y para los cuales:

$$|A \cup B| \neq |A| + |B|$$

De hecho no existe ninguna medida que cumpla las siguientes propiedades que refleje la idea intuitiva de tamaño:

- $\mu : \mathcal{P}(\mathbb{R}) \rightarrow [0, \infty]$
- Si $\{A_k\}$ son disjuntos entonces

$$\mu\left(\bigcup A_k\right) = \sum |A_k|$$

- Si $A \subset \mathbb{R}, t \in \mathbb{R}$ y

$$t + A = \{t + a : a \in A\}$$

entonces esperamos que $\mu(A) = \mu(A + t)$

- Si I es un intervalo abierto entonces

$$\mu(I) = l(I)$$

Por ello, vamos a eliminar la primera propiedad, usando la medida exterior, y nos restringiremos a una familia de conjuntos conocidos como los **conjuntos de Borel**.

DEFINICIÓN 23. *Un conjunto de Borel es de la familia \mathcal{B} mínima (en sentido de inclusión) que contiene a todos los intervalos abiertos de \mathbb{R} y que cumple:*

- (1) *Cerrada bajo uniones numerables.*
- (2) *Cerrada bajo complementos.*

Con esto, observamos que \emptyset es un boreliano.

3. Medida de Lebesgue

DEFINICIÓN 24. *Cuando restringimos la medida exterior al conjunto de los borelianos, la medida pasa a llamarse **medida de Lebesgue**, y la denotaremos por μ .*

$$\mu : \mathcal{B} \rightarrow [0, \infty]$$

Esta medida cumple:

- (1) $\mu(I) = l(I)$ si I es un intervalo.
- (2) $\mu(t + B) = \mu(B)$ para todo $t \in \mathbb{R}$ y todo $B \in \mathcal{B}$.
- (3) $\mu(\bigcup B_k) = \sum \mu(B_k)$ con $\{B_k\} \in \mathcal{B}$ y disjuntos.

Todos los conjuntos numerables tienen medida 0.

DEFINICIÓN 25. *Una propiedad $P(x)$ con $x \in \mathbb{R}$, decimos que es casi doquier, abreviado como c.d., si*

$$\mu(\{x \in \mathbb{R} : P(x) \text{ es falsa}\}) = 0$$

DEFINICIÓN 26 (Conjuntos medibles Lebesgue). *Son aquellos conjuntos a los que se puede extender μ . Si L es medible Lebesgue entonces existe un boreliano B tal que*

$$B \subset L \wedge \mu(L - B) = 0$$

DEFINICIÓN 27 (Funciones medibles Borel). $f : D \subset \mathbb{R} \rightarrow \mathbb{R}$ es medible Borel si $\forall B \in \mathcal{B}$
 $f^{-1}(B) \in \mathcal{B}$

En particular, como $\mathbb{R} \in \mathcal{B}$, entonces $f^{-1}(\mathbb{R}) = D \in \mathcal{B}$.

Es interesante hacer notar el paralelismo entre conjuntos abiertos y funciones continuas, y conjuntos borelianos y funciones medibles de Borel.

A las funciones medibles Borel también se las llama borelianas. Esta definición se extiende también a funciones que toman valores infinitos, donde exigimos que

$$f^{-1}(\{-\infty\}) \in \mathcal{B} \wedge f^{-1}(\{\infty\}) \in \mathcal{B}$$

Para identificar si una función es boreliana, en vez de usar la definición, podemos usar alguno de los siguientes criterios:

TEOREMA 16. $f : D \subset \mathbb{R} \rightarrow \mathbb{R}$ entonces

- f es boreliana si y solo si

$$f^{-1}((a, b)) \in \mathcal{B}$$

- f es boreliana si y solo si

$$f^{-1}((a, \infty)) \in \mathcal{B}$$

- f es boreliana si y solo si

$$\{x \in D : f(x) < b\} \in \mathcal{B}$$

- Si $\{f_n\}$ son borelianos tales que

$$\lim_{n \rightarrow \infty} f_n = f(x) \text{ c.d.}$$

entonces f es boreliana.

- Si f es continua en D y D es boreliano entonces f es boreliana.

4. Integración de Lebesgue

Para integrar según Lebesgue una función boreliana no negativa $f : D \subset \mathbb{R} \rightarrow [0, \infty]$, dividimos la imagen de f mediante una partición $P = \{y_0, \dots, y_n\}$ (ordenada), formando las sumas inferiores de Lebesgue:

$$\mathcal{L}(f, P) = \sum y_{k-1} \mu(f^{-1}([y_{k-1}, y_k]))$$

La integral de Lebesgue de f en D es entonces:

$$(19) \quad \int_D f \, d\mu = \sup \mathcal{L}(f, P)$$

5. Función integrable Lebesgue

Dada $f : D \subset \mathbb{R} \rightarrow \mathbb{R}$, definimos:

$$(20) \quad f_+(x) = \max\{f(x), 0 : x \in D\}$$

y

$$(21) \quad f_-(x) = \max\{-f(x), 0 : x \in D\}$$

De esta forma nos aseguramos de estar calculando el área bajo la función. Decimos que f es integrable Lebesgue en D , denotado por

$$f \in \mathcal{L}_1(D)^1$$

si

$$\int_D |f| \, d\mu = \int_D f_+ \, d\mu + \int_D f_- \, d\mu < \infty$$

y la integral de f se define por

$$(22) \quad \int_D f \, d\mu = \int_D f_+ \, d\mu - \int_D f_- \, d\mu$$

¹El 1 hace referencia a que integramos el módulo elevado a 1.

Análogamente se definen las funciones $f \in \mathcal{L}_P(D)$

$$\int_D |f|^P d\mu < \infty : p \in [1, \infty)$$

Algunas propiedades de esta integral son:

- Una función integrable de Riemann es también integrable de Lebesgue.
- $f = g$ e.c.t $D \subset \mathbb{R} \implies \int_D f d\mu = \int_D g d\mu$

Comprobemos, para finalizar, que la integral de Lebesgue no tiene los problemas que tenía la integral de Riemann.

- La integral de Lebesgue puede integrar funciones con infinitas discontinuidades (como la función de Dirichlet).

TEOREMA 17. Sea $\{f_n\}$ una sucesión de funciones en $\mathcal{L}_1(\mathbb{R})$, tales que existe

$$\lim_{n \rightarrow \infty} f_n(x) = f(x) \text{ c.d.}$$

Si además existe una función $g \in \mathcal{L}_1(\mathbb{R})$ no negativa tal que para todo $n \in \mathbb{N}$

$$|f_n(x)| \leq g(x) \text{ c.d.}$$

entonces

- $f \in \mathcal{L}_1(\mathbb{R})$
-

$$\int_{\mathbb{R}} f = \lim_{n \rightarrow \infty} \int_{\mathbb{R}} f_n$$

o de otra forma

$$\int \lim = \lim \int$$