Matemáticas para la ciencia de la computación

Apuntes de clase

Hernández García Argenis.
Moreno Mendieta Luis Enrique.
Ramírez Escamilla Marco Antonio.
Rosas Espinosa Edgar Hernan.
Salazar Hernandez Abraham Alejandro.
Segura Morales Cristian Camilo.
Torres Gonzáles José Alfredo.
Vera Cortes Christian Axel.

15A7071

MCC / MCIC
Centro de Investigación en Computación
November 29, 2021

Contents

0.1 Introducción

El trabajo presentado a continuación es el conjunto de ideas como trabajo en equipo. El trabajo es una muestra de como se trabaja en la clase, basándonos en un metodología pragmática que no lleva un orden lineal, pero si completo y estructurado de forma que al retomar temas se puede reforzar el aprendizaje. Por ello, el orden de los temas no siguen del todo el temario; esto para retomar y complementar temas. De este modo podemos ver de forma externa e interna cada tema de la asignatura.

El trabajo consta de de los capítulos que se vieron en las notas que presentó el Doctor en el transcurso de todo el curso. El trabajo será colocado en un repositorio donde alumnos futuros de la asignatura podrán encontrar para consulta. De igual forma se encontrará en el mismo repositorio programas que también fueron vistos y probados en clase. Estos contendrán un Header para hacer más fácil la consulta y descarga con información útil del funcionamiento y ejecución de los mismos.

El repositorio es creado por todo el grupo y tiene como fin apoyar en la consulta de información. Es meramente motivadora para que el estudiante tenga una guía y de este modo aportamos un poco para la comprensión más ligera de las clases. Se espera poder ser de utilidad. También se comenta que las clases junto a la practicidad de las mismas, conjuntarán un aprendizaje más profundo, al igual que el de las redes neuronales, profundo y eficaz. Con técnicas más precisas y útiles.

1 Lógica proposicional y conjuntos

1.1 El contexto de la lógica simbólica

Una variable lógica sólo puede tener dos posibles valores: Falso (F) y Verdadero (V) y puede ser representado por el conjunto F, V o F, T por las iniciales en inglés de las palabras False y True. También puede ser representado por el conjunto de valores 0, 1 de manera correspondiente.

Si la variable *a* tiene un valor verdadero, se dice que es verdadero; en caso contrario es falsa. Las variables lógicas son utilizadas para realizar la representación de proposiciones y su valor de verdad, es decir, si son falsas o verdaderas.

1.2 Álgebra proposicional y cálculo de predicados.

Se pueden realizar tres tipos de operaciones básicas con variables lógicas: Conjunción ($\wedge$), Disyunción ($\vee$) y Negación ($\neg$). La manipulación de las variables lógicas con estas tres operaciones da origen a la lógica proposicional. Estas operaciones funcionan de acuerdo con los cuadros 1.1, 1.2 y 1.3 para cada una de las operaciones.

a	b	a $\wedge$ b
F	F	F
F	T	F
T	F	F
T	T	T

Table 1.1: Tabla de verdad de la operación Conjunción

Estas tres tablas son casos particulares de funciones lógicas. Las tablas 1.2 y 1.2 poseen dos variables y la tabla 1.3 de una sola variable.

Nota. A la función constante verdadera se le llama tautología y a la falsa negación.

a	b	a ∨ b
F	F	F
F	T	T
T	F	T
T	T	T

Table 1.2: Tabla de verdad de la operación Disyunción

a	¬ a
F	T
T	F

Table 1.3: Tabla de verdad de la operación Negación

1.3 Equivalencia de expresiones lógicas

1.3.1 Funciones lógicas

En general una función lógica puede depender de n variables lógicas y como variable dependiente puede tener únicamente dos valores; Falso y Verdadero. La función lógica se puede representar mediante una tabla; su tabla de verdad. En la figura 1.4 se representa una tabla de verdad general para cualquier función.

a_n	...	a_2	a_1	$f(a_1,...,a_n)$
F	...	F	F	Columna
F	...	F	T	de
	.			los
	.			valores
	.			de
T	...	T	F	la
T	...	T	T	función

Table 1.4: Tabla de verdad estándar para funciones

Para cada combinación de valores falso y verdadero en la columna de los valores de la función; se tiene una función diferente.

Puesto que se tienen n variables lógicas en la tabla, se tienen 2^n renglones, por lo tanto, la columna de la función puede tener un total de 2^{2^n} diferentes de falso y verdadero. Es decir, si por cada combinación se tiene una función diferente, entonces para n variables lógicas se tienen 2^{2^n} funciones lógicas diferentes. Las funciones tienen como característica interesante que cualquier función lógica puede ser representada únicamente con combinaciones de las funciones $\wedge$, $\vee$ y $\neg$.

Estas representaciones son conocidas como Forma Normal Conjuntiva y Disyuntiva. Estas dos representaciones se pueden demostrar formalmente sin problema, pero es preferible primero usarlas para familiarizarse con ellas.

1.4 Formas normales (disyuntivas y conjuntivas)

1.4.1 Forma Normal Disyuntiva

Reglas:

1. Identificar en la tabla de verdad, tomando la columna de los valores de la función y obtener los valores verdaderos.

2. Teniendo en cuenta cada uno de los renglones del paso anterior, obtener la conjunción adecuada de las variables que permita obtener el valor verdadero de la función correspondiente. A esta combinación se le conoce como mintérmino.

3. Se forma la disyunción de todos los mintérminos obtenidos.

Ejemplo. Considerando la función lógica de dos variables conocida como XOR, encontrar su Forma Normal Disyuntiva. La tabla de verdad se la función se encuentra en el cuadro 1.5.

a	b	a $\oplus$ b
F	F	F
F	T	T
T	F	T
T	T	F

Table 1.5: Tabla de verdad de la operación XOR

Pasos:

1. Identificar los valores verdaderos de la función ($FT|T$ y $TF|T$)
2. Se obtienen los mintérminos correspondientes ($\neg a \wedge b$ y $a \wedge \neg b$)
3. La función en este caso XOR se representa como una disyunción de los mintérminos. a XOR b = $(\neg a \wedge b) \vee (a \wedge \neg b)$

1.4.2 Forma Normal Conjuntiva

La manera de construir una función lógica a partir de su tabla de verdad usando la forma normal conjuntiva y de manera similar al uso de la forma normal disyuntiva.

Reglas:

1. Identificar en la tabla de verdad, tomando la columna de los valores de la función y obtener los valores falsos.

2. Teniendo en cuenta cada uno de los renglones del paso anterior, construir la disyunción que permita obtener dichos valores falsos de la función. A esta disyunción se le conoce como maxtérmino.

+3. Hacer la conjunción con las disyunciones obtenidas.

Ejemplo. Reutilizando la función XOR del recuadro 1.5.

Pasos: 1. Identificar los valores falsos de la función($FF|F$ y $TT|F$)

2. Se obtienen los maxtérminos correspondientes ($a \vee b$ y $\neg a \vee \neg b$)

3. Se representa la función como una conjunción de los maxtérminos. a XOR b = $(a \vee b) \wedge (\neg a \vee \neg b)$

Como se puede evidenciar se han obtenido dos representaciones diferentes para la misma función XOR mediante la misma tabla de verdad.

Definición. Cuando dos funciones lógicas cuentan con la misma tabla de verdad, se dice que son lógicamente equivalentes o equivalentes, y se usa el símbolo $\equiv$ para indicar esta relación.

1.4.3 Tarea de la sección

1. Usar las formas normales conjuntivas y disyuntiva para probar las leyes de De Morgan.
   $\neg(a \wedge b) = \neg a \vee \neg b$ y también $\neg(a \vee b) = \neg a \wedge \neg b$

2. Usar tablas de verdad para probar las siguientes equivalencias lógicas.

3. Usando las equivalencias del punto 2 probar las sigueintes equivalencias teniendo en cuenta que $p \rightarrow p \equiv \neg p \vee q$

   a. $p \rightarrow q \equiv \neg q \rightarrow \neg p$
   b. $p \vee q \equiv \neg p \rightarrow q$
   c. $\neg(p \rightarrow p) \equiv p \wedge \neg q$
   d. $p \wedge q \equiv \neg(p \rightarrow \neg q)$

Equivalencia	Nombre
$p \wedge T \equiv p$ $p \vee F \equiv p$	Leyes de Identidad
$p \wedge F \equiv F$ $p \vee T \equiv T$	Leyes de dominación
$p \wedge p \equiv p$ $p \vee p \equiv p$	Leyes de idempotencia
$\neg(\neg p) \equiv p$	Leyes de doble negación
$(p \vee q) \vee h \equiv p \vee (q \vee h)$ $(p \wedge q) \wedge h \equiv p \wedge (q \wedge h)$	Leyes de asociación
$p \vee (q \wedge h) \equiv (p \vee q) \wedge (p \vee h)$ $p \wedge (q \vee h) \equiv (p \wedge q) \vee (p \wedge h)$	Leyes de distribución
$p \vee (p \wedge q) \equiv p$ $p \wedge (p \vee q) \equiv p$	Leyes de absorción
$p \vee \neg p \equiv T$ $p \wedge \neg p \equiv F$	Leyes de negación

Table 1.6: Tabla de equivalencias lógicas

e. $(p \to q) \wedge (p \to h) \equiv p \to (p \wedge h)$
f. $(p \to q) \vee (p \to h) \equiv p \to (p \vee h)$
g. $(p \to h) \wedge (q \to h) \equiv (p \vee q) \to h)$
h. $(p \to h) \vee (q \to h) \equiv (p \wedge q) \to h)$

4. Teniendo en cuenta que $p \leftrightarrow q \equiv (p \to q) \wedge (q \to p)$

a. $p \leftrightarrow q \equiv \neg p \leftrightarrow \neg q$
b. $p \leftrightarrow q \equiv (p \wedge q) \vee (\neg p \wedge \neg q)$
c. $\neg(p \leftrightarrow q) \equiv p \leftrightarrow \neg q$

1.5 Técnicas de demostración de teoremas

Demostrar un hecho o proposición matemática, consiste en realizar la utilización de hechos conocidos previamente y realizar su manipulación de manera adecuada para así comprobar la veracidad de la proposición a verificar.

Existes tres técnicas básicas de demostración de postulados.

1.5.1 Demostración Directa

De la forma $A \to B$, donde la hipótesis es A y a partir de procedimientos válidos (representados por una inferencia o cadena de inferencias) se requiere probar B

Ejemplo. Si $n \in \mathbb{N}$ es un número impar, entonces n^2 también es impar.

Prueba. Si n es un número impar de la forma $2R+1$, para $R \in \mathbb{N}$. Por lo tanto

$$\begin{aligned}(2R+1)^2 &= 2(2R^2)+2R+1\\ &= 2(2R^2+R)+1\end{aligned}$$

por lo que su cuadrado también es impar.■

1.5.2 Reducción al absurdo

Esta técnica consiste en la propiedad del operador de la implicación $A \to B$ el cual es equivalente a $\neg B \to \neg A$. En el caso que se supongan A y $A \to F$ mediante algún procedimiento válido, es de notar que $(P \wedge \neg P) \equiv F$ y esto ocurre cuando se llega a una contradicción. Entonces

$$A \to F \equiv \neg F \to \neg A \equiv T \to \neg A \equiv \neg A$$

Es decir, si se supone A y se llega a una contradicción lo que definitivamente ocurre (es válida) es A.

1.5.3 Demostración Contrapositiva

Por propiedades del operador lógico $(\rightarrow)$ condicional o implicación:

$$A \rightarrow B \equiv \neg B \rightarrow \neg A$$

Entonces si se quiere probar B dado A es suficiente probar no A $(\neg A)$ dado que se dio no B $(\neg B)$.

1.5.4 Tarea de la sección

1. De un ejemplo de una demostración por reducción al absurdo.
   2. Demostrar por reducción al absurdo que el valor de $\sqrt{2}$ es un número irracional.
   3. Dar dos ejemplos de la demostración de tipo contrapositiva.

1.6 Conjuntos y sus operaciones

Uno de los conceptos más importantes en las matemáticas es el de conjunto. Sobretodo porque permite expresar de una manera breve y clara una gran cantidad de ideas y conceptos relevantes en las matemáticas. En pocas palabras los conjuntos son entidades agrupadas de manera implícita o explícita por una propiedad. Por ejemplo el conjunto de los números 1,2 y 3, denotado de manera explícita como:

$\{1, 2, 3\}$

O de forma implícita como:

{x: los tres primeros números naturales}

Donde la propiedad que satisface el elemento x es ser alguno de los tres primeros números naturales.

Con los conjuntos se puede tener un álgebra a partir de las tres siguientes operaciones. (desde este punto se denotan a los conjuntos con letras mayúsculas y los elementos de los mismos con letras minúsculas).

1. $A \cup B = \{x : \text{está al menos en uno de los dos conjuntos}\}$
2. $A \cap B = \{x : \text{está en los dos conjuntos}\}$
3. $\bar{A} = \{x : \text{el elemento x no está en el conjunto A}\}$

Para indicar que un elemento x se encuentra en el conjunto A, se escribe como $x \in A$, y para indicar que todos los elementos de A están en B, se denota como $A \subseteq B$. si algún elemento de B no está en A cuando todos los elementos de A están en B, se utiliza $A \subset B$. Para obtener una idea gráfica de lo indicado, se demuestra en los siguientes diagramas de Venn.

Si se considera la función de permanencia actuando sobre los elementos, se tiene que $P_a(x) = 1$ si $x \in A$ ó 0 si $x \notin A$

1.6.1 El conjunto potencia

Definición. Dado un conjunto A, el conjunto potencia corresponde al conjunto de todos los subconjuntos de A. A esto se le llamamos conjunto potencia de A y es denotado como $\wp(A)$. Para el caso finito se puede probar que $|\wp(A)| = 2^{|A|}$.

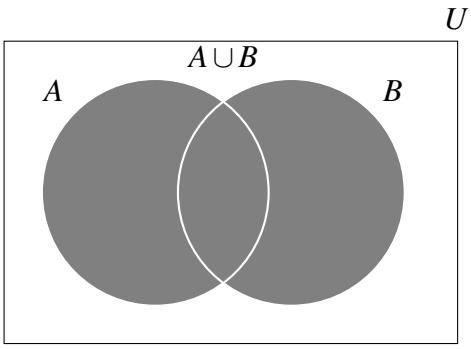


Figure 1.1: Diagrama de Venn para la Unión

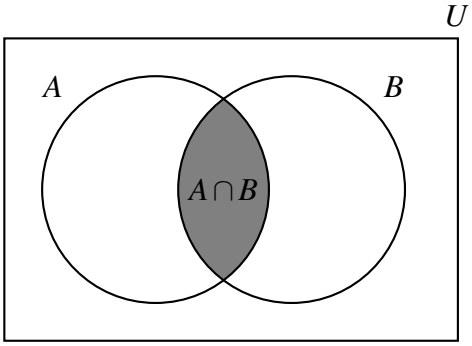


Figure 1.2: Diagrama de Venn para la Intersección

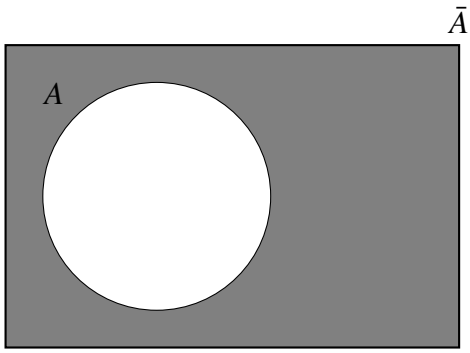


Figure 1.3: Diagrama de Venn para el complemento

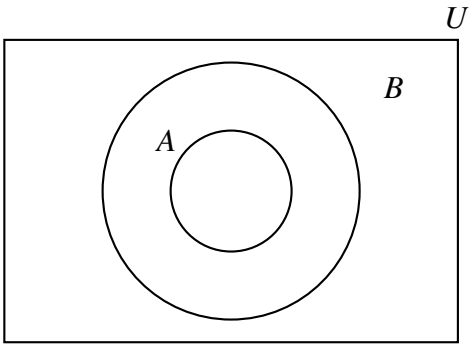


Figure 1.4: Diagrama de Venn para $A \subseteq B$

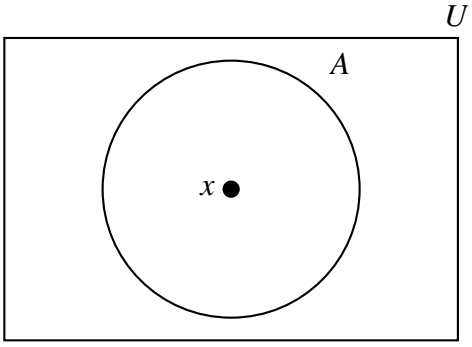


Figure 1.5: Diagrama de Venn para $x \in A$

Ejemplos:
$\wp(\emptyset) = \{\emptyset\}$
$\wp(a) = \{\emptyset, \{a\}\}$
$\wp(a,b) = \{\emptyset, \{a\}, \{b\}, \{a,b\}\}$
$\wp(\wp(\emptyset)) = \wp\{\{\emptyset\}\} = \{\emptyset, \{\emptyset\}\}$

Proposición. Sea A un conjunto finito arbitrario, entonces $|\wp(A)| = 2^{|A|}$.

Prueba. Sea A un conjunto con n elementos $\{a_1, a_2, ..., a_n\}$, considerar que $A' = \wp(\{a_1, a_2, ..., a_{n-1}\})$. Por hipótesis de inducción se supone que $|A'| = 2^{n-1}$. Se realiza la definición del conjunto A' como:

$A' = \{D : D = \{a_n\} \cup B, B \in A', A_n \in A\}$

A" está formado por todos los elementos de $A' \cup \{a_n\}$; por lo tanto para iniciar se cumple el caso base: $|\wp(\emptyset)| = 2^{|\emptyset|} = 2^0 = 1$

1. $|A'| = |A''|$
2. $A' \cap A'' = \emptyset$
3. $\wp(A) = A' \cup A''$, Puesto que los subconjuntos de A se pueden dividir en los que tienen (a_n) y los que no.

Tomando en cuenta que $|A| = n$ y la hipótesis de inducción $|A'| = 2^{n-1}$, se concluye que:

$|\wp(A)| = |A''| + |A^0| = 2|A'| = 22^{n-1} = 2^n$, finalmente, $|\wp(A)| = 2^n$■ que era el objetivo de la demostración.

Implementación con Listas

En el siguiente ejemplo se presenta la función conjunto_potencia la cual recibe como parámetro una lista de elementos, calcula y retorna una lista con todos los subconjuntos generados a partir de las lista inicial ingresada.

```
def conjunto_potencia(a):
    if len(a)==1:
        return [[],a]
    lst=[]
    b=conjunto_potencia(a[1:])
    for i in b:
```

```
        aux=[a[0]]+i
        lst=lst+[aux]
    return b+lst
```

Al realizar la invocación con la lista [1,2,3,4] y como se vió en el documento de fundamentos, el número de elementos del conjunto potencia corresponde a $2^{|A|}$, generando para la lista indicada, un total de 16 elementos (2^4).

```
conjunto_potencia([1,2,3,4])
```

```
[[],
 [4],
 [3],
 [3, 4],
 [2],
 [2, 4],
 [2, 3],
 [2, 3, 4],
 [1],
 [1, 4],
 [1, 3],
 [1, 3, 4],
 [1, 2],
 [1, 2, 4],
 [1, 2, 3],
 [1, 2, 3, 4]]
```

Implementación con Conjuntos

Luego de realizada la implementación con listas, ahora se realiza la implementación con la estructura de datos de conjunto (set) que maneja python. La implementación del código sería la siguiente:

```
def ConjuntoPotenciaSet(a):
    if len(a) == 0:
        return [set()]
    b = a.copy()
    for item in a:
        it1 = item
    b.discard(it1)
```

```
    r = ConjuntoPotenciaSet(b)
    Lista = r + [s | {it1} for s in r]
    return Lista
```

Y al realizar el llamado a la función con el mismo conjunto de datos de la llamada anterior:

```
ConjuntoPotenciaSet({1,2,3,4})
```

Se obtiene el siguiente resultado:

```
[set(),
 {1},
 {2},
 {1, 2},
 {3},
 {1, 3},
 {2, 3},
 {1, 2, 3},
 {4},
 {1, 4},
 {2, 4},
 {1, 2, 4},
 {3, 4},
 {1, 3, 4},
 {2, 3, 4},
 {1, 2, 3, 4}]
```

Donde se puede evidenciar que el resultado de la ejecución es el mismo, pero la forma de realizar la codificación para cada una de las estructuras de datos es diferente al momento de realizar la respectiva implementación.

2 Teoría de Números

2.1 Números enteros y naturales

2.1.1 Teoría elemental de números

El fundamento de la teoría de números es el estudio de la divisibilidad entre números naturales $\mathbb{N} = \{0,1,2,3,\cdots\}$ y su extensión los enteros $\mathbb{Z} = \{\cdots -1,0,1,\cdots\}$.

Definición. Decimos que un entero a divide b, también entero, si b es múltiplo de a, $b = ma$ para $m \in \mathbb{Z}$. En este caso escribimos $a \setminus b$.

Cuando esto no pasa podemos aproximar cualquier entero $b \in \mathbb{Z}$ con múltiplos $q \in \mathbb{Z}$ de un número natural $a \in \mathbb{N}$ y tenemos un residuo $r \in \mathbb{N}$, en general $0 \leq r < a$.

Nota. las desigualdades desde el. de vista algebraico se comportan como las igualdades, pero cuando se multiplican por un número negativo cambia el sentido de la desigualdad.

Veamos algunos resultados de desigualdades que se usarán más adelante: por supuesto las desigualdades son transitivas $a < b$ y $b < c$ implica $a < c$.

1. $a \leq b$ si y solo si $(b-a) \in \mathbb{N}$, es decir $0 \leq b-a$.

2. Si $a \leq b$ y $a \neq b$ decimos $a < b$.

3. $a \leq b$ y $b \leq a$ si y solo así $a = b$.

4. Si $a \setminus b$ y $0 \leq b < a$ entonces $b = 0$. Prueba. Como $a \setminus b$, $b = ma$ por tanto $ma < a$ pero m es entero y no puede ser $1 \leq m$ porque no se cumple la desigualdad, es decir $m = 0$ y $b = 0$, $a = 0$ ■.

5. Si $0 \leq a < b < c$ entonces $0 \leq (b-a) < c$. Prueba. Restando a de $b < c$ se obtiene $b-a < c-a$ y sumando c a ambos lados de $0 \leq a$, $c \leq c+a$, restando a, $c-a \leq c$. Finalmente $0 \leq (b-a) < c$ ■.

2.1.2 Principio del buen orden

Otra propiedad importante de los números naturales es que cualquier subconjunto no vacío tiene un primer elemento. En este caso decimos que el conjunto de números naturales está bien ordenado.

Proposición. El primer elemento de un subconjunto de números naturales es único.

Prueba. Sea $A \subseteq \mathbb{N}$ y $a \in A$ su primer elemento, supongamos que $a^{'}$ es otro primer elemento de A. Entonces $a \leq a^{'}$ por ser a primer elemento de A y $a^{'} \leq a$ por ser también $a^{'}$ primer elemento de A, de

las dos desigualdades anteriores $a \leq a^{'}$ y $a^{'} \leq a$ se concluye que $a = a^{'}$, es decir, el primer elemento es único ■.

El principio del buen orden trae como consecuencia que se cumple el siguiente método inductivo de demostración (inducción matemática).

Sea P una proposición que se anuncia para toda $n \in \mathbb{N}$.

Se cumple con:

1. $P(1)$ la proposición sobre $n = 1$, es verdadera (paso base).

2. $P(n-1)$ verdadera (hipótesis de inducción) implica $P(n)$ verdadera (paso inductivo).

Si ocurre en los dos pasos anteriores se concluye que P es cierta para toda $n \in \mathbb{N}$. Prueba. Sea $A \subseteq \mathbb{N}$ el conjunto donde P no se cumple. Si $A = 0$ y acabamos. En caso contrario sea n su primer elemento, entonces $P(n-1)$ es verdadera, por lo cual por el paso inductivo $P(n)$ también lo es. Por tanto, tenemos una contradicción, es decir, A debe ser vacío ■.

Proposición. (Teorema o algoritmo de la división). Sean $x \in \mathbb{Z}$ y $d \in \mathbb{N}$ entonces $x = qd + r$, $q \in \mathbb{Z}$, $q \in \mathbb{Z}$, $0 \leq r < d$.

Prueba. Sea $M = \{x - qd : q \in \mathbb{Z}\}$ se puede probar que $M \cap \mathbb{N} \neq \emptyset$ si este es el caso sea r su primer elemento. Por demostrar que $0 \leq r < d$, donde $x = qd + r$. Puesto que, en particular, $r \in \mathbb{N}$, esto implica $0 \leq r$, falta probar que $r < d$. Supongamos lo contrario, $d \leq r$ esto implica $0 \leq r - d \leq r$, pero $r - d = x - qd - d = x - (q+1)\,d$. Esto contradice que r es el primer elemento, por lo tanto $x = qd + r$, $0 \leq r < d$ ■.

Nota. Al residuo de un número entero x con respecto a un divisor d lo denotamos como $[x]_d$. Es decir, si $x = qd + r$, $r = [x]_d$.

Definición. (Relación de congruencia). Decimos que a es congruente con b módulo m $a \equiv b \mod m$, si y sólo si $m \setminus (b - a)$.

Proposición. Para $a, b, m \in \mathbb{Z}$, $m \setminus (b-a)$ si y sólo si $[a]_m = [b]_m$ (esta proposición nos da otro criterio para saber cuándo dos números son congruentes).

Prueba. Por el algoritmo de la división tenemos $b = q_b m + [b]_m$, $a = q_a m + [a]_m$ restando $b - a = (q_b - q_a)\,m + [b] - [a]$ por lo tanto $(b-a) + (q_a - q_b)\,m = [b] - [a]$ como $m \setminus (b-a)$ implica $m \setminus ([b] - [a])$, pero $0 \leq [a], [b] < m$ y la diferencia también es menor que m. Es decir, $[a] - [b]$ tiene que ser cero, $[a]_m = [b]_m$ ■. Por otro lado, si $b - a = (q_b - q_a)\,m + [b] - [a]$ y $[b]_m = [a]_m$ entonces $b - a = (q_b - q_a)\,m$, esto es $m \setminus (b-a)$ ■.

2.2 Divisibilidad y congruencia

2.2.1 Propiedades de las congruencias

Sean $a, b, c \in \mathbb{Z}$ y $m \in \mathbb{Z}$
Si
$a \equiv b \mod m$
$c \equiv d \mod m$
entonces $a+c \equiv b+d$, $a-c \equiv b-d$, $ac \equiv bc \mod m$. También $ac \equiv bd \mod m$, es decir, las congruencias se comportan como las ecuaciones, incluyendo la transitividad si $a \equiv b \mod m$ y $b \equiv c \mod m$, entonces $a \equiv c \mod m$.

Prueba. Si $a \equiv b \mod m$ y $c \equiv d \mod m$, tenemos $(b-a) = km$ y $(d-c) = lm$ sumando ambas ecuaciones $(b+d)-(a+c) = (k+l)m$, por tanto $a+c \equiv b+d \mod m$, lo mismo para la resta ■.
Para la propiedad transitiva, si $a \equiv b \mod m$ y $b \equiv c \mod m$, tenemos $b-a = km$ y $c-b = lm$, sumando tenemos $(c-b)+(b-a) = lm+km$, $c-a = (l+k)m$, es decir, $a \equiv c \mod m$■.

Ahora si $a \equiv b \mod m$, $(b-a) = km$, multiplicando por c ambos lados tenemos $c(b-a) = cb-ca = (ck)m$, por tanto $ca \equiv cb \mod m$■.
Por último, si $a \equiv b$ y $c \equiv d \mod m$, multiplicando por c la primera congruencia y por b la segunda tenemos, $ac \equiv bc \mod m$ y $bc \equiv bd \mod m$, aplicando la transitividad tenemos $ac \equiv bd \mod m$■.

2.2.2 Propiedades de los residuos

1. $a \equiv [a] \mod m$, 2. $[[a]] = [a]$
2. $[ab] = [[a][b]]$
3. $[a+b] = [[a]+[b]]$

Prueba 1. Del algoritmo de la división $a = qm + [a]$ ■ por tanto $qm = a - [a]$ ■.
Prueba 2. $[a] = 0m + [a]$, $[a]$ es su propio residuo ■.
Prueba 3. $a \equiv [a], b \equiv [b] \mod m$, multiplicando $ab \equiv [a][b] \mod m$, es decir, $[ab] = [[a][b]]$ ■.
Prueba 4. Lo mismo que en 3 pero sumando ■.

2.2.3 Mapeo residual y coordenadas residuales

Considere el conjunto $\mathbb{Z}/N = \{x : 0 \leq x < N\}$ dónde N es un número natural. Este conjunto es el conjunto de los residuos positivos de N.

Si $N = n_1 n_2 \cdots n_k$ es el producto de k números naturales. Entonces se puede considerar la siguiente función, llamada función residual r.

$$r : \mathbb{Z}/N \to \mathbb{Z}/n_1 \times \cdots \times \mathbb{Z}/n_k$$

Que tiene por dominio a los residuos de N y por contra dominio al producto cartesiano $\mathbb{Z}/n_1 x \cdots x \mathbb{Z}/n_k$ de los residuos de cada uno de los factores n_i de N.

La regla de correspondencia de la función residual es:

$$r : x \in \mathbb{Z}/N \to \left([x]_{n_1}, [x]_{n_2}, \cdots [x]_{n_k} \right)$$

k tupla formada por residuos.

Proposición. Si los factores de N son primos relativos dos a dos $\left(n_i, n_j\right) = 1,\ 1 \leq i, j \leq k,\ i \neq j$ entonces el mapeo residual es 1 a 1 y el domino cubre todo el contra dominio. Es decir, la función residual es biyectiva y la representación de los números en $\mathbb{Z}/N$ es única y cuando esto ocurre podemos hablar de coordenadas.

Prueba. Debido a que para dos conjuntos finitos A y B se cumple $|A \times B| = |A|\,|B|$ en este caso tenemos:
$|\mathbb{Z}/N| = N$ y $|(\mathbb{Z}/n_1) \times \cdots \times (\mathbb{Z}/n_k)| = |\mathbb{Z}/n_1| \cdots |\mathbb{Z}/n_k| = n_1 \cdots n_k = N$
Por lo cual el dominio y el contradominio tienen el mismo tamaño. Es este caso basta probar que el mapeo es 1 a 1 para tener biyección. Supongamos que $r(x) = r(y)$ para $x, y \in \mathbb{Z}/N$ entonces $[x]_{n_1} = [y]_{n_1}, \cdots, [x]_{n_k} = [y]_{n_k}$, es decir, $n_1/(y-x), \cdots, n_k/(y-x)$ por la proposición anterior tomando en cuenta $\left(n_i, n_j\right) = 1, i \neq j$, $\underbrace{(n_1, n_k)}_{N} \setminus (y-x)$. Por tanto $N \setminus (y-x)$ pero $0 \leq x, y < N$, o sea $(y-x) < N$, esto implica $y - x = 0$ por lo cual $x = y$ ■.

2.3 Máximo común divisor y algoritmos (Euclides, Euler, Chino, etc.)

2.3.1 Algoritmo de Euclides (forma escalar)

El algoritmo de Euclides nos permite encontrar el máximo común denominador entre dos números enteros. Para $a, b \in \mathbb{Z}$ este máximo común denominador lo denotamos como (a, b).

El algoritmo se basa en lo siguiente: si $b \setminus a$ entonces $(a, b) = b$ puesto que b forma parte de los divisores de a y es el mayor de los divisores de b. El otro hecho importante en el que se basa el algoritmo es que, si $a = qb + r$, entonces $(a, b) = (b, r)$.

Usando las dos observaciones anteriores tenemos: para $a > b$ (por ejemplo, números naturales).

1. Si $b \setminus a$ ya acabamos $(a,b) = b$.

2. En caso contrario $a = q_1 b + r_1$, $0 < r_1 < b$ hacemos $b = q_2 r_1 + r_2$, si $r_2 = 0$ entonces $(a,b) = (b,r_1) = r_1$.

3. Si aún no terminamos continuamos con

$$\begin{aligned} a &= q_1 b + r_1 \\ b &= q_2 r_1 + r_2 \\ r_1 &= q_3 r_2 + r_3 \\ &\vdots \\ r_{n-1} &= q_{n+1} r_n + r_{n+1} \end{aligned}$$

   $a > b > r_1 > r_2 > r_n$ secuencia decreciente de números positivos que cuando mucho converge a cero.
   Si $r_{n+1} = 0$ tenemos

$$(a,b) = (b,r_1) = (r_1,r_2) = \cdots = (r_{n-1},r_n) = r_n$$

   Por tanto $(a,b) = r_n$ con lo cual terminamos el procedimiento.

2.3.2 Algoritmo de Euclides (forma matricial)

Las iteraciones del algoritmo de Euclides se pueden ver en forma matricial de la siguiente manera. Para $b < a$ números enteros $s_0 = a$, $t_0 = b$.

$\begin{bmatrix} s_1 \\ t_1 \end{bmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & -Q_1 \end{pmatrix} \begin{pmatrix} s_0 \\ t_0 \end{pmatrix}$ y en general $\begin{bmatrix} s_n \\ t_n \end{bmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & -Q_n \end{pmatrix} \begin{pmatrix} s_{n-1} \\ t_{n-1} \end{pmatrix}$
con $Q_n = s_{n-1} // t_{n-1}$ (parte entera de dividir s_{n-1} entre t_{n-1})

Cómo puede verse t_n es el residuo de dividir s_{n-1} entre t_{n-1}, es decir, $s_{n-1} = Q_n t_{n-1} + t_n$.

En este enfoque matricial el algoritmo termina cuando $t_n = 0$ y el resultado (a,b) es s_n.

$$\begin{pmatrix} (a,b) \\ 0 \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & 1 \\ 1 & -Q_n \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & -Q_{n-1} \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ 1 & -Q_1 \end{pmatrix}}_{\begin{pmatrix} m & n \\ F & G \end{pmatrix}} \begin{pmatrix} a \\ b \end{pmatrix}$$

Multiplicando todas las matrices:

$\begin{pmatrix} (a,b) \\ 0 \end{pmatrix} = \begin{pmatrix} m & n \\ F & G \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$ es decir $ma + nb = (a,b)$.

Usando este enfoque no sólo se encuentra el máximo común divisor, también se encuentran los múltiplos ma y nb que sumados nos dan (a,b). Aquí los números importantes son m y n. A este algoritmo que no solo entrega (a,b) sino también m y n se le conoce como algoritmo extendido de Euclides.

2.3.3 La función φ de Euler

Considere el conjunto $(\mathbb{Z}/N)^* = \{x \in (\mathbb{Z}/N) : (x,N) = 1\}$ de los números naturales menores que N y primos relativos con él.

Definimos la función φ de Euler como el número de elementos de este conjunto $\left|(\mathbb{Z}/N)^*\right| = \varphi(n)$.

Proposición. Si $N = mn$ y $(m,n) = 1$ entonces $\varphi(mn) = \varphi(m)\varphi(n)$. Esto nos da en muchos casos una forma práctica de calcular φ.

Prueba. Consideremos el mapeo residual con $N = mn$ producto de factores primos. Sabemos que en este caso el mapeo $r : (\mathbb{Z}/N) \to (\mathbb{Z}/N) \times (\mathbb{Z}/M)$ es una biyección. Por lo tanto, si se logra hacer ver que $r\left((\mathbb{Z}/N)^*\right) = (\mathbb{Z}/n)^* \times (\mathbb{Z}/m)^*$ se tendría la prueba puesto que $\left|(\mathbb{Z}/N)^*\right| = \left|(\mathbb{Z}/n)^* \times (\mathbb{Z}/m)^*\right| = \left|(\mathbb{Z}/n)^*\right|\left|(\mathbb{Z}/m)^*\right|$, es decir, $\varphi(mn) = \varphi(m)\varphi(n)$.

Para ver lo anterior toma en cuenta que $(x,n) = (n,[x]_n) = 1$ y $(x,m) = (m,[x]_m) = 1$

Entonces si $x \in (\mathbb{Z}/n)^*$, $(x,nm) = 1$ implica $(x,n) = 1$ y $(x,m) = 1$ es decir $([x]_n,n) = 1$ y $([x]_m,m) = 1$ y viceversa. Si $([x]_n,n) = 1$, $(x,n) = 1$ y si $([x]_m,m) = 1$, $(x,m) = 1$ entonces $(x,nm) = 1$.

2.3.4 El teorema de Euler

Proposición. (Euler).
Si $(a,m) = 1$ para $a,w \in \mathbb{Z}$. Entonces $a^{\varphi(m)} \equiv 1 \mod m$, dónde φ es la función de Euler.

Demostración. Considere $A = \left\{r_1, r_2, \cdots, r_{\varphi(m)}\right\}$ el conjunto de residuos positivos primos relativos con m. Es decir, $A = (\mathbb{Z}/m)^*$, y por supuesto este conjunto tiene $\varphi(m)$ elementos. Considere la función $g : A \to A$ dada por $g(r) = [ar]_m$ para $r \in A$, afirmamos que g es 1 a 1.

Prueba. Sea $r_i r_j \in A$, $i \neq j$ y supongamos que $g(r_i) = g\left(r_j\right)$ o sea $[ar_i]_m = \left[ar_j\right]_m$ por tanto $ar_i \equiv ar_j \mod m$, como $(a,m) = 1$ la a se puede cancelar.

Tenemos entonces $r_i \equiv r_j \mod m$, o sea $[r_i] = \left[r_j\right]$ pero r_i y r_j ya eran residuos, por tanto $r_i = r_j$ ■.

De acuerdo a lo anterior $g(A) = \left\{\rho_1, \cdots, \rho_{\varphi(m)}\right\}$ es únicamente una permutación de los valores de $A = \left\{r_1, \cdots, r_{\varphi(m)}\right\}$. Entonces tenemos:

$$\begin{aligned} ar_1 &\equiv \rho_1 \mod m \\ &\vdots \\ ar_{\varphi(m)} &\equiv \rho_{\varphi(m)} \mod m \end{aligned}$$

Multiplicando todas las congruencias:

$$(ar_1)\cdots\left(ar_{\varphi(m)}\right) \equiv \rho_1 \cdots \rho_{\varphi(m)} \mod m$$

$$\underbrace{(a\cdots a)}_{\varphi(m)\ veces}\left(r_1\cdots r_{\varphi(m)}\right)\equiv \rho_1\cdots\rho_{\varphi(m)} \mod m$$

Puesto que cada r_i y ρ_i son primos relativos con m su producto también lo es, por lo tanto los podemos cancelar de la congruencia. Quedando finalmente $a^{\varphi(m)}\equiv 1 \mod m$ ■.

2.3.5 El teorema chino del residuo

Si el mapeo residual es biyectivo tiene sentido pensar en su mapeo inverso y de eso trata precisamente el teorema chino del residuo. De cómo obtener un número x a partir de sus residuos respecto a módulos primos 2 a 2.

Sean $m_1, m_2, \cdots, m_k \in \mathbb{N}$, números primos relativos 2 a 2 y sean $r_1, r_2, \cdots, r_k \in \mathbb{Z}$ residuos enteros tales que

$$\begin{aligned} x &\equiv r_1 \mod m_1 \\ &\vdots \\ x &\equiv r_k \mod m_k \end{aligned}$$

Para un $x \in \mathbb{Z}$ y el problema es precisamente encontrar la solución x.

Solución. (Teorema Chino)

1. $M = m_1, m_2, \cdots, m_k$, M es el producto de todos los módulos.
2. $N_i = M/m_i \;\; 1 \leq i \leq k$, N_i tiene a todos los módulos como factor menos al módulo m_i.
3. $y_i N_i \equiv 1 \mod m_i$, y_i es el inverso multiplicativo de N_i respecto a m_i. Este inverso existe debido a que $(N_i, m_i) = 1$.
4. Finalmente, $x = \sum_{i=1}^{k} r_i y_i N_i \mod M$

Prueba. Tómese el residuo respecto al módulo m_j, si el resultado es r_j hemos terminado.

$$[x]_{m_j} = \left[\sum_{i=1}^{k} r_i y_i N_i + lM\right] = \left[\left[\sum_{i=1}^{k} r_i y_i N_i\right] + [lM]\right] = \left[r_j y_j N_j\right]$$

Puesto que M tiene a todos los m_i como factor común $[lM]_{m_j} = 0$ y todos los $[r_i y_i N_i]_{m_j} = 0$ para $i \neq j$ por la misma razón. Pero $\left[r_j y_j N_j\right] = \left[\left[r_j\right]\left[y_j N_j\right]\right] = \left[r_j\right]$, es decir $x \equiv r_j \mod m_j$ ■.

2.3.6 Algoritmo de los cuadrados repetidos

Para obtener el residuo de un entero $a \in \mathbb{Z}$ elevado a una potencia $e \in \mathbb{N}$, respecto a un entero positivo N, $[a^e]_N$.

1. Obtener la representación binaria de e.

$$\imath = b_m \imath^m + b_{m-1}\imath^{m-1} + \cdots + b_1\imath^1 + b_0 = \sum_{i=0}^{m} b_i \imath^i$$

con $b_i e^i \in \{0,1\}$ siendo $[b_m, b_{m-1}, \cdots, b_1, b_0]$ las coordenadas o representación binaria de e.

2. Puesto que $a^{k+l} = a^k a^l,\ k,l \in \mathbb{Z}$

$$[a^e] = \left[a^{\sum_{i=1}^{m} b\imath^i}\right] = \left[a^{b_m\imath^m} \cdot a^{b_{m-1}\imath^{m-1}} \cdots a^{b_1\imath} \cdot a^{b_0}\right] = \left[\left[a^{b_m\imath^m}\right]\left[a^{b_{m-1}\imath^{m-1}}\right]\cdots\left[a^{b_1\imath}\right]\left[a^{b_0}\right]\right]$$

como b_i sólo puede ser 0 o 1 es suficiente calcular los residuos de a elevada a las potencias de 2 y en el caso que $b_i = 0$ estos factores son 1, puesto que $a^0 = 1$.

3. Calcular los $\left[a^{\imath^i}\right]_N,\ i = 1,2,\cdots,m$ usando el criterio de los cuadrados repetidos que se basa en:

$$\left[a^{\imath^{i+1}}\right] = \left[a^{\imath^i\imath^1}\right] = \left[a^{\imath^i+\imath^i}\right] = \left[a^{\imath^i}a^{\imath^i}\right] = \left[\left[a^{\imath^i}\right]\left[a^{\imath^i}\right]\right]$$

Es decir, para obtener el nuevo residuo $\left[a^{\imath^{i+1}}\right]$ se eleva al cuadrado el residuo anterior $\left[a^{\imath^i}\right]$ y luego se le saca el residuo.

$$\left[a^{\imath^{i+1}}\right] = \left[\left[a^{\imath^i}\right]\left[a^{\imath^i}\right]\right]$$

4. En el resultado final $[a^e] = \left[\left[a^{b_m\imath^m}\right]\cdots\left[a^{b_1\imath}\right]\left[a^{b_0}\right]\right]$ únicamente aparecen los factores con $b_i = 1$.

Definición. Decimos que $a,b \in \mathbb{Z}$ son primos relativos o coprimos si y sólo si $(a,b) = 1$.
Proposición. Para $a,b,c \in \mathbb{Z}$

1. Supongamos que $a \setminus bc$ y $(a,b) = 1$. Entonces $a \setminus c$.

2. Si a y b son primos relativos, $a \setminus c$ y $b \setminus c$ entonces $ab \setminus c$.

3. Si $(a,b) = 1$ y $(a,c) = 1$ entonces $(a,bc) = 1$.

4. Si $(a,bc) = 1$ entonces $(a,b) = 1$ y $(a,c) = 1$. Esto es trivial: $ma + nbc = 1$ de la definición.

Prueba.

1. Como $(a,b) = 1$ tenemos $ma + nb = 1$, multiplicando por c $mac + nbc = c$, como $bc = ka$ $mac + nka = c$, es decir, $(mc + nk)\,a = c$. Por lo tanto $ab \setminus c$ ■.

2. $ma + nb = 1$ (por ser primos relativos)
$mac + nbc = c$ (multiplicando por c)
$ma\,(kb) + nb\,(la) = c$ ($c = kb$ por $b \setminus c$ y $c = la$ por $a \setminus c$)
$(mk)\,ab + (nl)\,ab = c = ((mk) + (nl))\,ab = c$, es decir, $ab \setminus c$■.

3. $\dfrac{\begin{array}{c} ma+nb=1 \\ ka+lc=1 \end{array}}{mkaa+mlac+nkab+nlbc=1}$(multiplicando ecuaciones) $\underbrace{(mka+mlc+nkb)}_{f}a+\underbrace{(nl)}_{g}bc=1$, $fa+gbc=1$, $(a,b,c)=1$ ■

Definición. (Inverso multiplicativo congruencial). Si para $a \in \mathbb{Z}$ se tiene $a^{-1} \in \mathbb{Z}$ tal que $a^{-1}a \equiv 1 \mod m$, decimos que a^{-1} es el inverso multiplicativo de $a \mod m \in \mathbb{Z}$. Nota. Dado a y m se puede obtener a^{-1} si $(a,m)=1$, $a^{-1}a+km=1$, $m \setminus \left(1-a^{-1}a\right)$.

2.3.7 La identidad de Bézout

Si a y b son enteros positivos $a,b \in \mathbb{Z}^+$ y (a,b) es un máximo común denominador, entonces por el algoritmo de Euclides extendido sabemos que se tienen dos enteros $m,n \in \mathbb{Z}$, tales que $ma+nb=(a,b)$. Esta identidad es conocida como de Bézout. En general m y n no son únicos.

Proposición. Si $m,n \in \mathbb{Z}$ satisfacen $ma+nb=(a,b)$ entonces también $m'=m+kb$, $n'=n-ka$, lo hacen.

Prueba. $m'a+n'b=(m+kb)a+(n-ka)b=ma+kab+nb-kab=ma+nb=(a,b)$ por lo tanto $m'a+n'b=(a,b)$ ■.

Esta identidad es importante porque, a pesar de que m y n no son únicos, nos permite caracterizar cuando dos enteros son primos relativos.

Proposición. Si $a,b \in \mathbb{Z}$, entonces son primos relativos si y sólo si, existen $m,n \in \mathbb{Z}$ tales que $ma+nb=1$.

Prueba. Si $(a,b)=1$ entonces por el algoritmo de Euclides extendido podemos hallar $m,n \in \mathbb{Z}$ con $ma+nb=1$. Por otro lado, supongamos que $ma+nb=1$ para algún par $m,n \in \mathbb{Z}$. Tomemos $d=(a,b)$ máximo común divisor y por lo mismo divide a ambos a y b y por lo tanto divide a $ma+nb$ y por $ma+nb=1$ también divide al 1. Pero el único entero que divide a 1 es él mismo y por lo tanto $(a,b)=d=1$ ■.

En algunas ocasiones se requiere que m sea positivo, esto se puede lograr mediante el siguiente programa en Python:

```
def bezout(a,b,m,n):
    if m>0:
        return m,n
    else:
      k=0; m1=m;
        while m1<=0:
          k = k+1
          m1 = m + k*b
          n1 = n - k*a
```

```
10          return m1,n1
```

Code Listing 2.1: Código de la identidad de Bézout en Python

2.3.8 Ejercicios numéricos

Algoritmo de Euclides y Euclides extendido

Calcular el máximo común denominador de $a > b \in \mathbb{Z}$.

Ejemplo: $a = 18$, $b = 12$

a	b	q	r
18	12	1	6
12	⑥	2	0

Se usa una tabla y cada renglón representa una iteración del algoritmo $a = q_1 b + r_1$, $b = q_2 b + r_2$, etcétera. El algoritmo termina cuando $r = 0$ y en la columna de q se encuentra el resultado.

Aquí $m = 1$ para el algoritmo extendido de Euclides. $ma + nb = (a,b)$

Otro ejemplo: $a = 34$, $b = 13$

a	b	q	r
34	13	2	8
13	8	1	5
8	5	1	3
5	3	1	2
3	2	1	1
2	①	2	0

Para aplicar el algoritmo de Euclides extendido y obtener m y n tales que $34m + 13n = 1$, procedemos como sigue: de la tabla despejamos el residuo en cada renglón:

$$8 = 34 - 2 \cdot 13$$
$$5 = 13 - 8$$
$$3 = 8 - 5$$
$$2 = 5 - 3$$
$$1 = 3 - 2$$

Después se sustituyen de arriba a abajo:

$$5 = 13 - 34 + 2 \cdot 13 = 3 \cdot 13 - 34$$
$$3 = 8 - (3 \cdot 13 - 34) = 34 - 2 \cdot 13 - 3 \cdot 13 + 34 = -5 \cdot 13 + 2 \cdot 34$$
$$2 = 3 \cdot 13 - 34 - (-5 \cdot 13 + 2 \cdot 34) = 8 \cdot 13 - 3 \cdot 34$$
$$1 = -5 \cdot 13 + 2 \cdot 34 - (8 \cdot 13 - 3 \cdot 34) = -13 \cdot 13 + 5 \cdot 34$$

Entonces $m = 5$ y $n = 13$

Ejercicio. Para $a = 37$ y $b = 8$ calcular (a,b) y m,n tales que $ma + nb = (a,b)$.

Método de los cuadrados para calcular residuos

Ejemplo. Considere calcular $[13]_{17}$.

1. La representación binaria de 11 es 1011.
2. Usando $\left[13^{2^3+2+1}\right]_{17} = \left[[13^{2^3}][13^2][13]\right]$

$$[13]_{17} = 13,\ \left[13^2\right]_{17} = [169]_{17},\ [169]_{17} = 16$$

$$\left[13^{2^2}\right]_{17} = \left[16^2\right]_{17},\ \left[13^{2^2}\right]_{17} = 1,\ \left[13^{2^3}\right]_{17} = \left[1^2\right]_{17} = 1$$

Finalmente,

$$\left[13^{11}\right]_{17} = \left[13^{2^3+2+1}\right]_{17} = \left[\left[13^{2^3}\right]\left[13^2\right][13]\right]_{17} = [16\cdot 13]_{17} = [208]_{17}$$

$$\left[13^{11}\right]_{17} = 4$$

Haciéndolo en forma directa tenemos:

13^1	13
13^2	169
13^3	2197
13^4	28561
13^5	371293
13^{10}	137858491849
13^{11}	1792160394037
$13^{11}\%17$	4 que es el resultado correcto

Ejercicio. Calcular de forma convencional y por los cuadrados repetidos $\left[13^{15}\right]_{13}$.

Coordenadas congruenciales y el teorema chino del residuo

Sea $N = 3\cdot 5\cdot 7$ números primos relativos 2 a 2. $N = 105$ y $(\mathbb{Z}/N) = \{0,1,2,3,\cdots,104\}$ residuos positivos de 105.

Si $n \in (\mathbb{Z}/N)$ se puede representar de forma única como $(n)_r = ([n]_3, [n]_5, [n]_7)$.

Ejemplo. Encontrar las coordenadas residuales de $n = 20$

$$(20)_r = ([20]_3, [20]_5, [20]_7) = (2,0,6)$$

Ejercicio. Encontrar $(47)_r$.

Aplicando el algoritmo del teorema chino del residuo resolver el siguiente sistema de congruencias.

Ejemplo.

$$x \equiv 2 \mod 3$$
$$x \equiv 0 \mod 5$$
$$x \equiv 6 \mod 7$$

Solución

$$M = 3 \cdot 5 \cdot 7 = 105,\ N_1 = 35,\ N_2 = 21,\ N_3 = 15$$

$$y_1 N_1 \equiv 1 \mod 3,\ y_2 N_2 \equiv 1 \mod 5,\ y_3 N_3 \equiv 1 \mod 7$$

$$y_1 = 2,\ y_2 = 1,\ y_3 = 1$$

Aplicando el algoritmo de Euclides extendido.

$$x = r_1 y_1 N_1 + r_2 y_2 N_2 + r_3 y_3 N_3 + kM$$
$$x = 2 \cdot 2 \cdot 35 + 0 \cdot 1 \cdot 21 + 6 \cdot 1 \cdot 15 + k105$$
$$x = 140 + 90 + (-2) \cdot 105 = 230 - 210 = 20$$
$$x = 20$$

Se elige $k = -2$ para tener una solución x positiva y pequeña.

Ejercicio. Resolver

$$x \equiv 1 \mod 3$$
$$x \equiv 2 \mod 5$$
$$x \equiv 3 \mod 7$$

2.4 Números primos y factorización

2.4.1 Números primos

Definición. Un número $p \in \mathbb{P}$ es primo si únicamente es divisible por él mismo y la unidad.

Proposición. El divisor más pequeño de un número natural es primo.

Demostración. Sea $n \in \mathbb{N}$ y $a \in \mathbb{N}$ su divisor más pequeño. Supongamos que a no es primo, entonces existe $b \in \mathbb{N}$ con $b \setminus a$, $1 < b < a < n$ si $b \setminus a$ también $b \setminus n$ y esto no es posible porque a es el divisor más pequeño de n. Por lo tanto a es primo ■.

Proposición. Sea $n \in \mathbb{N}$ y a su divisor más pequeño entonces $a < \sqrt{n}$.

Demostración. Como $a < \sqrt{n}$ es el divisor más pequeño, tenemos $0 < a < b$ y $n = ab$, multiplicando la desigualdad por a, tenemos $a^2 < ab = n$ y sacando raíz $a < \sqrt{n}$.

Proposición. La cantidad de números primos es infinita.

Demostración. Supongamos que la cantidad es finita, es decir, $p_1, \cdots, p_n$ son todos los números primos. Considere el número $q = p_1 \cdot p_2 \cdots p_n + 1$ producto de todos los primos más 1.

1. Si q es primo, encontramos un primo que no está en la lista original de primos y por lo tanto la lista no es finita.

2. Si q no es primo, su divisor más pequeño p es primo y no aparece en la lista de primos porque ninguno de ellos divide a q. Por tanto, la lista no es finita ■.

2.4.2 Teorema fundamental de la aritmética

Todo número natural se puede representar como un producto de potencias de números primos, $n = p_1^{e_1} \dots p_k^{e_k}$ dónde $n \in \mathbb{N}$ y p_i es primo.

Prueba. Sea p_i el primer divisor de n, que pueda aparecer e_i veces como divisor. Sea $n_1 = n/p_1^{e_1}$, p_2 el primer divisor de n_1 que cabe e_2 veces, $n_2 = n_1/p_2^{e_2}$, etcétera. $n_1 = n/p_1^{e_1}, n_2 = n_1/p_2^{e_2}, \cdots, n_k = n_{k-1}/p_k^{e_k}, n_k = 1$

Por tanto $n = p_1^{e_1} \cdot p_2^{e_2} \cdots p_k^{e_k}$ ■.

2.4.3 La criba de Eratóstenes

Algoritmo para obtener los números primos menores a un número natural n dado. La idea es hacer una lista de 1 a n número dado. Por ejemplo, $n = 100$.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

Después empieza con el primer primo y se marcan todos sus múltiplos, en este caso el 2 y quedan marcados los pares. Se continúa con el siguiente que no está marcado y que resulta que es primo (porque si no fuera primo sería múltiplo de un primo más pequeño y estaría marcado) y así sucesivamente hasta llegar antes de $\sqrt{n}$ puesto que el primo más pequeño que divide a n es menor que $\sqrt{n}$, en este caso $\sqrt{100} = 10$.

Por último, nos quedamos con todos los números no marcados y esos son precisamente los números primos que andamos buscando: 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97.

1 2 3 ④ 5 ⑥ 7 ⑧ ⑨ ⑩ 11 ⑫ 13 ⑭ ⑮ ⑯ 17 ⑱ 19 ⑳ ㉑ ㉒ 23 ㉔ ㉕ ㉖ ㉗ ㉘ 29 ㉚ 31 ㉜ ㉝ ㉞ ㉟ ㊱ 37 ㊳ ㊴ ㊵ 41 ㊷ 43 ㊹ ㊺ ㊻ 47 ㊽ ㊾ ㊿ 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

2.4.4 Coordenadas primas

Usando la representación de $n \in \mathbb{N}$ en términos de potencia de números primos $n = p_1^{e_1} \cdots p_k^{e_k}$ y una tabla de números primos generada por el algoritmo de Eratóstenes, se pueden obtener los coeficientes $[e_1, \cdots, e_k]$ y a estos coeficientes les llamamos coordenadas primas de n. Es claro que si tenemos las coordenadas primas podemos recuperar a n usando $n = p_1^{e_1} \cdots p_k^{e_k}$. Por supuesto también tenemos que tener los generadores primos $p_1, \cdots, p_k$.

Aplicación de las coordenadas primas

Si $a \in \mathbb{N}$ sus coordenadas primas $[a_1, \ldots, a_k]$ están dadas por $a = p_1^{a_1} \ldots p_k^{a_k}$ dónde a_i son los exponentes en la representación de a como producto de potencias de números primos En estos términos podemos representar el conjunto de divisores de a, $\operatorname{div}(a)$ como:
$\operatorname{div}(a) = \left\{ p_1^{e_1} \cdots p_k^{e_k} : 0 \leq e_1 \leq a_1, \ldots, 0 \leq e_k \leq a_k \right\}$ puesto que
$a = \left(p_1^{(a_1 - e_1)} \cdots p_k^{(a_k - e_k)} \right) \left(p_1^{e_1} \cdots p_k^{e_k} \right)$, si $b = p_1^{b_1} \cdots p_k^{b_k}$
$\operatorname{div}(b) = \left\{ p_1^{f_1} \cdots p_k^{f_k} : 0 \leq f_1 \leq b_1, \ldots, 0 \leq f_k \leq b_k \right\}$ tenemos que

$$\operatorname{div}(a) \cap \operatorname{div}(b) = \left\{ p_1^{g_1} \cdots p_k^{g_k} : 0 \leq g_1 \leq \min(a_1, b_1), \cdots, 0 \leq g_k \leq \min(a_k, b_k) \right\}$$

esto debido a que si $0 \leq g_i \leq \min(a_i, b_i)$ por ejemplo $a_i < b_i$ se cumplen las dos desigualdades $0 \leq g_i \leq a_i$ y $0 \leq g_i \leq b_i$ y esto implica que el exponente g_i pertenece a un divisor común de a y b. Y esto permite obtener el mayor de los divisores comunes, como:

$$(a, b) = p_1^{\min(a_1, b_1)} \cdots p_k^{\min(a_k, b_k)}$$

siguiendo el mismo orden de ideas el conjunto de múltiplos de a se puede representar como:
$m \cup l(a) = \left\{ p_1^{e_1} \cdots p_k^{e_k} : a_1 \leq e_1, \ldots, a_k \leq e_k \right\}$ y los múltiplos de b:
$m \cup l(b) = \left\{ p_1^{f_1} \cdots p_k^{f_k} : b_1 \leq f_1, \ldots, b_k \leq f_k \right\}$ por tanto el mínimo común múltiplo se puede obtener como:
$[a, b] = p_1^{\max(a_1, b_1)} \cdots p_k^{\max(a_k, b_k)}$, donde se usan los paréntesis cuadrados $[,]$ para representar el mínimo común múltiplo.

2.4.5 Sistema criptográfico de llave pública RSA

En este sistema se usan algunos algoritmos vistos anteriormente como: algoritmo de Euclides extendido, de los cuadrados repetidos y el teorema de Euler.

En este esquema si alguien desea que le envíen información pública una llave que se considera pública (N,e) formada por un número N grande producto de 2 números primos $N = pq$ y el número e primo relativo con $\varphi(N)$, $(\varphi, \varphi(N)) = 1$.

El emisor transforma un bloque de información en un número grande $0 \leq x < N$ y lo envía encriptado como $[x^e]_N$, aquí se puede usar el algoritmo de los cuadrados repetidos para calcular este residuo.

El receptor posee una llave privada (d,N) que le permite recuperar el mensaje x, mediante una operación similar a la codificación. La d de la llave privada satisface $de - \alpha\varphi(N) = 1$, que se obtiene al aplicar el algoritmo de Euclides extendido a $(\varphi, \varphi(N)) = 1$. Tómese en cuenta también que puesto que $N = pq$, en general $(x,N) = 1$, dónde x es el dato a transmitir y por el teorema de Euler tenemos $x^{\varphi(N)} \equiv 1 \mod N$.

Proposición. $\left[[x^e]_N^d\right]_N = x$

Prueba. $\left[[x^e]^d\right] = \left[x^{ed}\right] = \left[x^{\alpha\varphi(N)+1}\right] = \left[\underbrace{\left[x^{\varphi(N)}\right]^{\alpha}}_{1}[x]\right] = [x] = x$ ■. Usando esta proposición se hace la codificación del mensaje.

2.4.6 Evaluación de la función φ de Euler usando la representación prima de un número

Sea $n \in \mathbb{N}$ y $n = p_1^{n_1} \cdots p_k^{n_k}$ su representación prima, sabemos que φ se comporta en forma multiplicativa cuando su argumento es un producto de números coprimos y en este sentido los $p_i^{n_i}$ son coprimos por tanto $\varphi(n) = \varphi\left(p_1^{n_1}\right) \cdots \varphi\left(p_k^{n_k}\right)$ entonces necesitamos calcular $\varphi(p^m)$ con p número primo y $m \in \mathbb{N}$

$\varphi(p^m) = |\{q : 1 \leq q < p^m \; y \; (q, p^m) = 1\}|$, es decir, el número de coprimos con p^m menores que p^m y mayores o iguales a 1.

Éstos los podemos calcular de la siguiente manera:

$$\varphi(p^m) = p^m - p^{m-1} = p^m\left(1-\frac{1}{p}\right)$$

Todos los números naturales de 1 a p^m menos los números que no son coprimos con p^m. Esto debido a que todos los múltiplos de p menores o iguales a p^m son no coprimos con p^m.

Es decir, los números de la forma kp, con $k = 1,2,\ldots,p^{m-1}$ son no coprimos con p^m y éstos son precisamente p^{m-1}. Por lo cual tenemos para $\varphi(n)$:

$$\varphi(n) = \varphi\left(p_1^{n_1}\right)\cdots\varphi\left(p_k^{n_k}\right) = p_1^{n_1}\left(1-\frac{1}{p_1}\right)\cdots p_k^{n_k}\left(1-\frac{1}{p_k}\right)$$

$$= \underbrace{p_1^{n_1}\cdots p_k^{n_k}}_{n}\prod_{i=1}^{k}\left(1-\frac{1}{p_i}\right) = n\prod_{i=1}^{k}\left(1-\frac{1}{p_i}\right)$$

Esto quiere decir que $\varphi(n)$ para cualquier número natural n es igual al número n multiplicado por una constante que depende de sus coordenadas primas.

2.4.7 Prueba de Wilson

La prueba de Wilson para saber si un número natural es primo o no. Esta prueba se basa en lo siguiente.

Proposición. Un número natural $n \in \mathbb{N}$ es primo si y sólo si $(n-1)! \equiv (n-1) \mod n$, esto también se escribe $(n-1)! \equiv -1 \mod n$.

Demostración. Supongamos que n es primo, entonces todos los números $1 \leq m < n$ son primos relativos con n, $(m,n) = 1$. Por lo tanto, cada uno de ellos tiene su inverso multiplicativo congruencial m', $mm' \equiv 1 \mod n$, que podemos ubicar también en $1 \leq m' < n$, es decir, $1 \leq m,m' < n$. Agrupando por pares cada m con su inverso congruencial m' y dado que en general tenemos $m \neq m'$ con la única excepción de $(n-1)$ que es su propio inverso congruencial. Tenemos

$$\begin{aligned} 1 \cdot 1 &\equiv \mod n \\ &\vdots \\ m \cdot m' &\equiv \mod n \end{aligned}$$

Multiplicando estas congruencias obtenemos $(n-2)! \equiv 1 \mod n$, y multiplicando por $(n-1)$ ambos lados de la congruencia $(n-1)! \equiv (n-1) \equiv -1 \mod n$ ■.

Por otro lado, si suponemos que $(n-1)! \equiv (n-1) \mod n$ si n no fuera primo entonces $(n-1)!$ no fuera primo relativo con n, pero como $(n-1)$ sí es primo relativo con n esto no puede ser y esto prueba que $(n-1)! \equiv (-1) \mod n$ implica n primo. De tarea pruebe que si $a \equiv b \mod n$ y $(a,n) = 1$ entonces $(b,n) = 1$.

2.5 Algoritmos para la factorización de primos (Pollard, Fermat-Kraitchik, etc.)

2.5.1 El método de factorización $(p-1)$ de Pollard

De acuerdo al teorema pequeño de Fermat tenemos $a^{p-1} \equiv 1 \mod p$, si $(a,p) = 1$ y esto ocurre si el número primo p no divide a. Multiplicando esta congruencia varias veces por sí misma se obtiene:

$a^{k(p-1)} \equiv 1 \mod p$, para un k arbitrario, haciendo $m = k(p-1)$

$a^m \equiv 1 \mod p$, con m múltiplo arbitrario de $(p-1)$ y $p \setminus (a^m - 1)$.

Si N es un número compuesto que se desea factorizar, se calcula $(N,(a^m-1)) = q$ el máximo común divisor entre N y (a^m-1). Si esto ocurre, q es divisor de N y posiblemente sea primo ya que es solución de $a^m \equiv 1 \mod q$, debido a que $q \setminus (a^m-1)$. Entonces el algoritmo hace una búsqueda sobre a y m hasta que $q = (N,(a^m-1))$ cumpla $1 < q < N$, si $q = 1$ se incrementa m y si $q = N$ se busca otra a.

De las propiedades del máximo común divisor se sabe que $((a^m-1),N) = (N,[a^m-1]_N)$ donde $(a^m-1) = lN + [a^m-1]_N$. Aprovechando esto se calcula $(N,[a^m-1]_N)$ que es más simple, además por propiedades de los residuos $[a^m-1]_N = [a^m]_N - 1$. Para la búsqueda de m una vez que se fija a, se usa una cota superior B tal que $p_1 < p_2 < \cdots < p_n < B$, para considerar únicamente en los primeros primos menores que B. Y la m se forma como la potencia de algunos de estos primos $m = p_1 \cdot p_2 \cdots p_l = \prod_{i=1}^{l} p_i^{e_i}$, $l \leq n$.

2.5.2 El algoritmo Ro de Pollard

Algoritmo para detectar un factor primo de un número compuesto (típicamente el más pequeño).

Sea $N \in \mathbb{N}$ un número compuesto y q un factor primo de él. Tomemos $f(a) = \left[a^2+1\right]_N$, una función $f : (\mathbb{Z}/N) \to (\mathbb{Z}/N)$. Si generamos la secuencia $a_n = f^n(a) = \underbrace{f(f\cdots f(a_0)}_{n\ veces}\cdots) = f\left(f^{n-1}(a_0)\right)$, ésta se comporta como un muestreo aleatorio de $(\mathbb{Z}/N) = \{0,1,2,\cdots,(N-1)\}$, cada instancia de la secuencia a_n es una muestra aleatoria de $(\mathbb{Z}/N)$. Para poder disponer de un subconjunto interesante de muestras se construye otra secuencia de muestras dadas por $b_n = \underbrace{f(f\cdots f(a_0)}_{2n\ veces}\cdots) = a_{2n}$ y la idea es considerar diferencias $(b_n - a_n)$.

El algoritmo procede de la siguiente manera:

1. Se elige un a_0 arbitrario (inicio de la secuencia de muestreo).
2. Se van formando en las secuencias $a_n = f(a_{n-1})$ y $b_n = a_{2n}$.
3. Se calcula $(b_n - a_n, N)$ el máximo común divisor entre $(b_n - a_n)$ y N. Si son primos relativos continúa la búsqueda. Si $(b_n - a_n, N) = q \neq 1,\ N$. Entonces q es un divisor de N, como su divisor más pequeño es primo, es probable que sea este. Por otro lado $q \setminus (b_n - a_n)$, es decir, $a_n \equiv b_n \mod q$. Si consideramos el espacio muestral $(\mathbb{Z}/q) = \{0,1,\cdots,q-1\}$ lo que tenemos es una repetición aleatoria de uno de sus valores $[a_n]_q = [b_n]_q$ y esto ocurre de acuerdo al problema del día de nacimiento en aproximadamente $\sqrt{q}$ ensayos y aquí es donde se consigue el ahorro puesto que el algoritmo del ensayo por división factoriza N en aproximadamente $\sqrt{N} > \sqrt{q}$.

2.5.3 Ejercicios

1. Haga un programa para probar que n es un número primo:
   (a) Calcule $(n-1)!$
   (b) Encuentre $[(n-1)!]_n$

(c) Verifique si $[(n-1)!]_n = (n-1)$ entonces n es primo.

2. Escriba un programa que por ensayo de división determine si un $n \in \mathbb{N}$ es primo o no.

3. Considere el número 5293 y encuentre sus factores primos a mano por los siguientes 3 métodos: Ro de Pollard, $(p-1)$ de Pollard, y diferencia de cuadrados de Fermat-Kraitchik.

4. El mismo ejercicio que el anterior, pero usando programas.

5. Verifique con los programas de división y criterio de Wilson si los números obtenidos en los dos incisos anteriores son primos.

6. Considere $n = 7$ y para $\{1,2,3,4,5,6\}$ sus residuos positivos diferentes de cero, encuentre las parejas de inversos multiplicativos módulo 7.

2.6 Anexo

Para $a,x,y,m \in \mathbb{Z}$

Proposición. Si $(a,m) = 1$ y $ax \equiv ay \mod m$ entonces $x \equiv y \mod m$.

Demostración. Por definición $ax \equiv ay \mod m$ implica $m \setminus (ay-ax)$, es decir, $m \setminus a(y-x)$, pero debido a que $(a,m) = 1$, $m \setminus (y-x)$ y esto implica $x \equiv y \mod m$ ■. Esta propiedad nos permite en algunos casos cancelar (dividir) en ambos lados de una congruencia.

Proposición. Si $a,b \in \mathbb{Z}$ su mínimo común múltiplo $[a,b]$ se puede calcular como $[a,b] = \frac{ab}{(a,b)}$, es decir, como el producto de los dos números a y b sobre el máximo común divisor (a,b).

Demostración. Para $a,b \in \mathbb{Z}$ sea M un múltiplo común de a y b, entonces M es de la forma $M = ka = lb$ por lo tanto $\frac{ka}{a} = l$ es un entero. Si $d = (a,b)$ entonces hagamos $a_1 = \frac{a}{d}$ y $b_1 = \frac{b}{d}$ como $(a,b) = d$, $\left(\frac{a}{d}, \frac{b}{d}\right) = 1$, $(a_1,b_1) = 1$; a_1 y b_1 son primos relativos y $\frac{ka}{b} = k\frac{a_1 d}{b_1 d} = k\frac{a_1}{b_1}$ pero como a_1 y b_1 son primos relativos $b_1 \setminus k$, es decir, $k = tb_1$ para

t número entero, pero $b_1 = \frac{b}{d}$, por tanto $M = t\frac{ab}{d}$. Y si se quiere el múltiplo común más pequeño se hace $t = 1$ y tenemos $[a,b] = \frac{ab}{(a,b)}$ ■.

Proposición. Para dos conjuntos finitos A y B: $|A \times B| = |A|\,|B|$.

Demostración. Sea $A = \{a_1, \cdots, a_n\}$ y $B = \{b_1, \cdots, b_n\}$:

$$A \times B = \bigcup_{i=1}^{n} \{(a_i, b) : b \in B\} = \sum_{i=1}^{n} \left| \underbrace{\{(a_i, b) : b \in B\}}_{m} \right| = nm = |A|\,|B| \ \blacksquare$$

Esto es debido a que tenemos una unión de conjuntos distintos.

3 Estructuras algebraicas

3.1 Grupos de clases de residuo módulo C

Considere la clase subconjunto de números enteros con el mismo residuo al dividirlos entre un entero donde $C \in \mathbb{Z}$.

Este subconjunto lo podemos escribir como $a+c \in \mathbb{Z} = [m : m = ax$ de manera equivalente podemos representar $a+c \in \mathbb{Z}$ como $X \in \mathbb{Z}]$.
$[a]_c$ + c $\in \mathbb{Z}$ puesto que a = kc + $[a]_c$ y cualquier a + C $\in \mathbb{Z}$ de la forma KC + $[a]_c + cx = [a]_c + c(k+x) \in [a]_c + c \in \mathbb{Z}$. En otras palabras siempre podemos representar cualquier $a+c \in \mathbb{Z}$ como $a+c \in \mathbb{Z}$, es decir $a+c \in \mathbb{Z} = [a]_c + c \in \mathbb{Z}$. A partir de esta consideración podemos hacer álgebra con estas clases residuales módulo c, formando un grupo de la siguiente manera:
Definición
Sean las clases congruenciales módulo c $[a]_c + c \in \mathbb{Z}$ representadas por sus residuos $[a]_c \neq [b]_c$. Entonces su suma está dada por $[a]_c + [b]_c = [[a]_c + [b]_c]_c$. Cuando el módulo está implícito es suficiente escribir $[a]$.
Nota*
Como los residuos satisfacen $0 \leq [a]_c < c, a \in \mathbb{Z}$. Las clases residuales se representarán como $[k]$, $k = 0, 1, 2, ..., (c-1)$ y el conjunto de clases residuales como proposición $(\mathbb{Z}/c\mathbb{Z}, +)$ es un grupo con $(\mathbb{Z}/c\mathbb{Z})$.
La suma definida como $[k] + [l] = [k+l], 0 \leq k, l < (c-1)$.
1. Por definición el conjunto es cerrado bajo la operación.
2.$[0]$ es el elemento neutro.
3. $[(c-k)]$ es el inverso de $[k]$.
4. La asociatividad depende de la propiedad $[k+l+m] = [(k+l)+(m)]$
Tarea: Hacer la tabla de composición de $(\mathbb{Z}/g\mathbb{Z}, ..+) = [[k] + [l+m]]$.

Subgrupos y clases laterales, el teorema de Lagrange.

Definición.

Un subconjunto $H \subseteq G$ de un grupo G es un subgrupo de G si es un grupo respecto a la operación de G. Para esto basta con que cumpla con:

1. Si $x, y \in H$ entonces $xy \in H, cerradura.$

2. El elemento neutro σdeG está en H, $\sigma \in H$.

3. Para cada $X \in H$, X^{-1} también está en H $X^{-1} \in H$.

Ejemplo: El núcleo $Ker(f) = x : f(x) = E$ de un Homomorfismo f $f : G- > k$ entre dos grupos G y K, es un subgrupo de G.

$E \in K$, neutro. Un Homomorfismo f es un mapeo entre dos grupos G y K, que preserva el producto en G, es decir $f(xy) = f(x) \cdot f(y), x, y \in G$.

Veamos que en efecto $Ker(f) \subseteq G$ es un subgrupo. Sean $\sigma \in G$ elemento neutro es G y E elemento neutro en K y F Homomorfismo.

1. $f(\sigma) = f(\sigma\sigma) = f(\sigma) \cdot f(\sigma)$, multiplicado por $f^{-1}(x) \in K, E = f(\sigma), f$ mapea al neutro de G en el neutro de K.

2. $E = f(\sigma) = f(xx^{-1}) = f(x) \cdot f($$x^{-1}$, multiplicado por $f^{-1}(x)$.

$f^{-1}(x) = f(x^{-1})$, f mapea el inverso de X en el inverso de f(x).

Para probar que $Ker(f) \subseteq G$ es subgrupo necesitamos:

1.$\sigma \in Ker(f) \sigma neutrodeG$. Esto es cierto $f(\sigma) = E$.

2. Si $X \in Ker(f)$, $E = f(\sigma) = f(xx^{-1} = f(x) \cdot f(x^{-1} = E \cdot f(x^{-1} = f(x^{-1}, f(x)E$. $f(x^{-1} = E$ y por tanto $X^{-1} \in Ker(f)$.

3. Si $x, y \in Ker(f), f(xy) = f(x) \cdot f(y) = E, E = E, xy \in Ker(f)$ como se cumplen 1,2 y 3 $Ker(f)$ es subgrupo de G.

Definición: Para un subgrupo $H \subseteq G$ de un grupo G y un elemento $g \in GdeG$, al subconjunto de G dado por $g \cdot H = [g \cdot H : K \in H]$ le llamamos clase lateral izquierda de H y a, $H \cdot g = [h \cdot g : h \in H]$ clase lateral derecha.

3.2 subgrupos

Es interesante considerarlos subconjuntos de un grupo que a su vez también tienen estructura de grupos.
Definición 13: Un subconjunto no vacío S de un grupo G es un subgrupo (de G) si, con respecto a la misma operación de G, S también es grupo.
Ejemplo 14 1): El conjunto 3 Z es un subgrupo de $(Z,+)$. 2) Sea (G,*) un grupo cualquiera; el conjunto $S = xGx = a...a = ap, pZ$ es un subgrupo de G, llamado subgrupo monógeno de generador a.
3) El conjunto del ejemplo 2 es un subgrupo del grupo del ejemplo 1.
4) Los conjuntos $S1 = f1, f2yS2 = f1, f4, f5$ son subgrupos de (F,o) (ejercicio 9) En todo grupo existen, al menos, dos subgrupos, , G, llamados subgrupos impropios; caso de que existan, los restantes subgrupos se denominan subgrupos propios.
Teorema 15: (de caracterización de subgrupos).- Un subconjunto S, no vacío, de un grupo G es un subgrupo si y solo si $\forall a, b \in S$, se verifica: $a \cdot b \in S$.

Demostrar $[\Rightarrow]$ Por ser S subgrupo, $\forall a \cdot b \in S$, se verifica que $b \in S$ y que, por ser operación estable en $S, a \cdot b \in S$. $[\Leftarrow]$ Hemos de probar que se verifican las propiedades 0), 1), 2), 3) de la definición.
Trivialmente, si x es asociativa en G también lo es en S. Tomando, en la hipótesis, $b = a$, resulta: $a \cdot a = e \in S(prop.2))$. Ahora, tomando e y a en la hipótesis, resulta: $ea = a \in S$ (prop. 3)). Finalmente, tomando a y b' en la hipótesis, sale: $a(b') = abS(prop.0))$.
Ejercicio 16: Probar que la intersección de dos subgrupos de un grupo es otro subgrupo y que la suma de dos subgrupos de un grupo con la adición como operación es otro subgrupo.

3.3 Clases o cogrupos

Unos subconjuntos de un grupo que tienen especial interés, ya que conducen a la definición del concepto de grupo cociente, son los cogrupos o clases de equivalencia asociadas a un subgrupo.

Teorema 17 *Sea S un subgrupo de un grupo $(G,*)$. Definimos la relación binaria $\mathcal{R}$ en G mediante*

$$a\mathcal{R}b \Longleftrightarrow a * b' \in S, a, b \in S$$

Entonces, se verifica: 1) $\mathcal{R}$ es una relación de equivalencia.
2) $[a] = S * a = \{x \in G | x = s * a, s \in S\}$

Demostr.:
1) reflexiva: $a\mathcal{R}a \Longleftrightarrow a * a' = e \in S$ (por ser S subgrupo, prop. 2)
simétrica: $a\mathcal{R}b \Longleftrightarrow a * b' \in S \Longleftrightarrow (a * b')' = b * a' \in S \Longleftrightarrow b\mathcal{R}a$.
transitiva: $a\mathcal{R}b$ y $b\mathcal{R}c \Longleftrightarrow a * b' \in S, b * c' \in S \Longrightarrow$ (prop. o)) $(a * b') * (b * c') = = a * c' \in S \Longleftrightarrow a\mathcal{R}c$.

2) Por definición de clase de equivalencia: $[a] = \{x \in G \,|\, x\mathcal{R}a\}$. Pero $x\mathcal{R}a \Longleftrightarrow x * a' \in S \Longleftrightarrow s \in S' x = s * a$, luego: $[a] = \{x \in G \,|\, x = s * a, s \in S\} = S * a$. ■

Análogamente podríamos estudiar la relación $\tilde{\mathcal{R}}$ en G:

$$a\tilde{\mathcal{R}}b \Longleftrightarrow b' * a \in S, a, b \in S$$

y ver que
1) $\tilde{\mathcal{R}}$ es una relación de equivalencia.
2) $[a]_{\tilde{\mathcal{R}}} = a * S = \{x \in G \,|\, x = a * s, s \in S\}$

En consecuencia, podemos dar la siguiente definición

Definición 18 *Sea $a \in G$ y S un subgrupo de G. Llamaremos clase a izquierda asociada al elemento a módulo S, al conjunto*
$a * S = \{x \in G \,|\, x = a * s, s \in S\}$
y clase a derecha asociada al elemento a módulo S, al conjunto
$S * a = \{x \in G \,|\, x = s * a, s \in S\}$

3.4 Grupo cociente

Definición 21 *Llamaremos subgrupo normal, invariante o distinguido a todo subgrupo N de un grupo G que verifique:*

$$a * N = N * a, \forall a \in G$$

Ejemplo 22 *1) Todo subgrupo de un grupo conmutativo es subgrupo normal; 3 Z es normal. 2) El subgrupo S_2 del ejercicio 19 es normal, en cambio, el S_1* **no** *lo es.*

Observemos que:

$$a * N = N * a, \forall a \in G \Longleftrightarrow a * N * a' = N, \forall a \in G$$

es decir,

$$\forall a \in G, \forall n_1 \in N \quad \exists n_2 \in N \text{ tal que } a * n_1 * a' = n_2 \in G$$

La importancia de los subgrupos invariantes deriva de la siguiente propiedad.

Teorema 23 *Si N es subgrupo invariante de G, la relación de equivalencia $\mathcal{R}$ asociada a N es compatible con la estructura de G.*

Demostr: Por definición, $\mathcal{R}$ es compatible con la estructura de grupo de G si y solo si

$$a\mathcal{R}b \text{ y } \tilde{a}\mathcal{R}\tilde{b} \Longrightarrow (a * \tilde{a})\mathcal{R}(b * \tilde{b}),$$

que es lo que debemos probar.

Pues bien, notemos que

$$(a * \tilde{a}) * (b * \tilde{b})' = a * \tilde{a} * \tilde{b}' * b' = a * (\tilde{a} * \tilde{b}') * (b' * a) * a' \in N$$

como se requería. ■

Estamos ahora en condiciones de definir coherentemente una operación en el conjunto cociente $G/\mathcal{R}$, que a partir de aquí denotaremos por G/N que, como sabemos, está constituido por todas las clases de equivalencia o cogrupos.

Definición 24 *En G/N, definimos la operación * mediante [a] * [b] = [a * b]*

Observemos que * está bien definida, pues, como consecuencia del teorema 23, el resultado no depende de los representantes de las clases [a], [b] elegidos.

3.5 Generación de grupos

Una pregunta que surge expontaneamente es la siguiente: ¿cuál es subgrupo S más pequeño que contiene a un subconjunto no vacío C de un grupo G?

Definición 30 *Al subgrupo S anterior se le denomina subgrupo engendrado por el conunto C.*

Para construir el subgrupo S engendrado por C, basta construir la tabla de Klein del conjunto C, ampliándola sucesivamente con el elemento neutro y con los elementos no considerados que vayan saliendo, debido a que S debe ser estable para la operación. Veamos varios ejemplos.

Ejemplo 31 *Sea* $G = \mathbb{Z}_8$ *y* $C = \{2,3\}$

	0	2	3	5	7	6	1	4
0	0	2	3	5	7	...		
2	2	4	5	7	1			
3	3	5	6	0	2			
5	5	7	0	2	4			
7	7	1	2	4	6			
⋮			...					

luego S = G

Ejemplo 32 *Sea* $G = \mathbb{Z}_8$ *y* $C = \{2,4\}$

	0	2	4	6
0	0	2	4	6
2	2	4	6	0
4	4	6	0	2
6	6	0	2	4

Ejemplo 33 *Sea* $G = \mathbb{Z}_8$ *y* $C = \{4\}$

	0	4
0	0	4
4	4	0

luego S = {0, 4}

3.6 Grupos monógenos y grupos finitos

Definición 34 *Se denomina grupo monógeno a todo grupo G engendrado por un solo elemento g; a este elemento g se le llama generador del grupo; se escribe*

$$G = (g) = \{g^k \mid k \in \mathbb{Z}\}$$

Ejemplo 35 *1)* $\mathbb{Z}$ *con la suma es un grupo engendrado por +1 y también por -1.*
2) $\mathbb{Z}_p$ *con la suma está engendrado por 1.*
3) $\{1, i, -1, -i\} = (i) = (-i)$

Teorema 36 *Todo grupo monógeno es isomorfo a* $\mathbb{Z}$ *o a* $\mathbb{Z}_p$

Dem.: Consultar, por ejemplo, Castellet [7]. ■

Teorema 37 *Todo subgrupo de un grupo monógeno es también monógeno*

Dem.: Consultar, por ejemplo, Castellet [7]. ■

Definición 38 *Se denomina orden de un grupo al número de sus elementos. Se denomina orden de un elemento de un grupo al orden del subgrupo engendrado por él.*

Ejemplo 39 *En el grupo S(3), el orden de* s_4 *y* s_5 *es 3; el orden de* s_2, s_3 *y* s_6 *es 2. No hay elementos de orden 6, luego S(3) no es un grupo cíclico.*

Teorema 40 *(de Lagrange) El orden de un subgrupo S de un grupo finito G es un divisor del orden de G.*

Dem.: Consultar, por ejemplo, Castellet [7]. ■

3.7 Anillos

Una estructura más ”fuerte” que la de grupo, con la que estamos familiarizados, es la que posee el conjunto $\mathbb{Z}$ de los números enteros respecto de la suma y el producto. Sirviéndonos ésta como ejemplo casi general, definimos axiomática mente el concepto de anillo.

Definición 42 *Si la multiplicación es conmutativa decimos que el anillo es conmutativo.*

Si existe elemento neutro (unidad) para la multiplicación se dice que el anillo es unitario (o con unidad).

Ejemplo 43 *1) ($\mathbb{Z}, +, \cdot$) es anillo conmutativo con unidad.*

2) $\mathbb{R}[x]$ (conjunto de todos los polinomios en una indeterminada x y coeficientes números reales) con respecto de la suma y el producto de polinomios es un anillo conmutativo con unidad.

3) El conjunto $A = \{(a,b) | \, a,b \in \mathbb{Z}\}$ con respecto a la suma definida por (a,b)+(a',b') = (a+a',b+b') y el producto (a,b)· (a',b') = (a a',b b') tiene estructura de anillo conmutativo unitario.

4) $M_2(\mathbb{R})$ (conjunto de matrices cuadradas 2x2 de elementos reales) con la suma y el producto de matrices habituales es anillo unitario no conmutativo.

5) $\mathbb{Z}_5$ con respecto a la suma y producto de clases de restos habituales es un anillo conmutativo.

Consecuencia 44 *1)* $a \cdot 0 = 0, \forall a \in A$

2) $0 \cdot a = 0, \forall a \in A$

3) $a \cdot b + a \cdot (-b) = 0, \forall a, b \in A$

4) $b \cdot a + (-b) \cdot a = 0, \forall a, b \in A$

5) $(-a) \cdot (-b) = a \cdot b, \forall a, b \in A$

Dem. 1): $\forall a \in A, a \cdot (b + 0) = a \cdot b + a \cdot 0 = a \cdot b$, por prop. distributiva y axioma 2), luego $a \cdot 0 = 0$.

Dem. 5): $(-a) \cdot (-b) = -(a \cdot (-b)) = -(-a \cdot b) = a \cdot b$ por consecuencias 4), 3) y axioma 3. ■

Notemos que como, en general, el producto no es conmutativo, el que se verifique una propiedad (por ejemplo consecuencia 1)) no implica que se cumpla la simétrica (por ejemplo, consecuencia 2)).

3.8 Subanillos

Los subconjuntos de un anillo que realmente tienen interés son los que también tienen estructura de anillo.

Definición 55 *Sea $S \subset A$ anillo. Diremos que S es un subanillo de A si con respecto a las mismas operaciones de A tiene también estructura de anillo.*

Teorema 56 *(de caracterización).- La condición necesaria y suficiente para que un subconjunto S de un anillo A sea un subanillo es que*

$$\forall a, b \in S : a - b \in S \text{ y } a.b \in S$$

Demostr.: se propone como un sencillo ejercicio. ■

Subanillos muy especiales de un anillo, porque permiten definir la estructura de anillo cociente, son los "ideales".

Definición 57 *Llamaremos ideal a izquierda I de un anillo A a todo subconjunto I de A que verifique:*

$$\forall a, b \in I : a - b \in I$$

$$\forall a \in I, \forall x \in A : \quad x.a \in I, \quad (A.I \subset I)$$

Análogamente se puede definir el concepto de ideal a derecha cambiando x.a por a.x. Entonces, un ideal bilátero es todo ideal a izquierda y a derecha.

Ejemplo 58 *1) $p\ \mathbb{Z}$ (conjunto de los enteros múltiplos de p) es un subanillo de $\mathbb{Z}$.*
2) $2\ \mathbb{Z}$ (conjunto de los enteros múltiplos de 2) es un ideal (bilátero) de $\mathbb{Z}$.

Definición 59 *Diremos que el ideal I del anillo unitario A es ideal principal si existe un elemento $i \in A$ tal que $\forall x \in I, \exists y \in A$ verificando $x = y.i$*
Al elemento i se le denomina base del ideal y entonces se escribe $I = (i)$.

Ejemplo 60 *El conjunto $5\ \mathbb{Z}$ de los enteros múltiplos de 5 es un ideal principal de $\mathbb{Z}$, y así, $5\ \mathbb{Z} = (5)$.*

3.9 Cuerpos

El conjunto de los números reales R representa un modelo de estructura algebraica con dos operaciones internas muy conocido, por lo que únicamente recordaremos la definición de la estructura abstracta de la que R es ejemplo y algunas de sus propiedades más sobresalientes.

Definición 61 *Llamaremos cuerpo, K, a todo anillo (K,+,.) unitario, conmutativo y tal que todo elemento distinto del cero posea inverso. Es decir, que verifique:*

**) (K,+) es grupo abeliano*

**) (K*,·) es grupo abeliano*

**) propiedad distributiva:* $\forall a, b, c \in A$: $a.(b+c) = a.b + a.c$

Ejemplo 62 *1)* $(\mathbb{R},+,.)$, $(Q,+,.)$, $(\mathbb{C},+,.)$ *son cuerpos (conmutativos).*

2) $\mathbb{Z}_5$ *es cuerpo, pero* $\mathbb{Z}_6$ *no lo es.*

Propiedad 63 *1) Un cuerpo no posee divisores de cero.*

2) Todo dominio dc intcgridad finito cs un cucrpo.

3) Los únicos ideales de un cuerpo K son 0 y K

4) En todo cuerpo K son válidas las reglas del cálculo con fracciones con denominadores no nulos.

(Otras cuestiones sobre esta estructura pueden consultarse en la bibliografía recomendada).

Ejercicio 64 *(El cuerpo de fracciones de un dominio de integridad) (En este ejercicio se muestra un procedimiento para construir un cuerpo a partir de un dominio de integridad, de la misma forma a como se construye el cuerpo de los números racionales a partir de los números enteros).*

Sea (A,+,.) un dominio de integridad. En el conjunto A x A definimos dos operaciones + y . mediante*

(a,b)+(c,d) = (a..d+b.c,b.d)

(a,b).(c,d) = (a.c,b.d)

1) Probar que $A \times A^*$ *es anillo unitario conmutativo.*

Se define ahora la relación binaria R en A x A mediante:*

(a,b) R (a',b') $\iff$ *a.b' = a'.b*

2) Probar que R es compatible con + y . de $A \times A^*$

3) Construir el conjunto cociente $A \times A^*/R$ *que denotaremos por K.*

Se definen en K otras operaciones + y . mediante:

[(a,b)] + [(c,d)] = [(a d + b c, b d)]

[(a,b)] . [(c,d)] = [(a c, b d)]

4) Probar que (K,+,.) es un cuerpo conmutativo.

Sea $i:(A,+,.) \longrightarrow (K,+,.)$, $a \longrightarrow i(a) = [(a,1)]$

5) Probar que i es un homomorfismo de anillos inyecctivo.

6) Probar que cualquier otro cuerpo K' para el que se pueda encontrar otro monomorfismo $m:A \longrightarrow K'$ *contiene a K.*

3.10 Permutaciones

El grupo simétrico Sn, permutaciones de n elemetos. $Sn = [r : [1,2,...,n] \to [1,2,...,n]]$, por ejemplo:
$S_3 = [\begin{pmatrix}1&2&3\\1&2&3\end{pmatrix}, \begin{pmatrix}1&2&3\\2&1&3\end{pmatrix}, \begin{pmatrix}1&2&3\\2&3&1\end{pmatrix}, \begin{pmatrix}1&2&3\\3&1&2\end{pmatrix}, \begin{pmatrix}1&2&3\\3&2&1\end{pmatrix}, \begin{pmatrix}1&2&3\\1&3&2\end{pmatrix}]$.

En este conjunto cada elemento tiene en el primer renglón el dominio y en el segundo los correspondientes a cada miembro del dominio. Este mismo conjunto se puede representar como $S_3 = [(1\,2\,3),(2\,1\,3),(2\,3\,1),(3\,1\,2),(3\,2\,1),(1\,3\,2)]$ donde el dominio se encuentra de manera implícita en la posición que ocupa cada miembro de la tupla.

Proposición: El número de permutaciones de n elementos es $n!$, $|S_n| = n!$.

Prueba: (Por inducción), caso base $n = 1, |S_1| = |[\binom{1}{1}]| = 1! = 1$.

Supongamos que para $|S_{n-1}| = (n-1)!$ y entonces n elementos.

Coloquemos el n-esimo elemento en la primera posición, para los otros $(n-1)$ elementos tenemos por hipótesis $(n-1)!$ posibilidades y esto por cada una de las n posiciones, por lo tanto en total se tienen $n(n-1)! = n!$ permutaciones.

Sn forma grupo con el producto dado por la composición de las permutaciones por ejemplo:
$$\begin{pmatrix}1&2&3\\3&1&2\end{pmatrix} = \begin{pmatrix}1&2&3\\3&2&1\end{pmatrix} \cdot \begin{pmatrix}1&2&3\\1&3&2\end{pmatrix}.$$

Tomando la asignación de derecha a izquierda $1 \to 1 \to 3, 2 \to 3 \to 1, 3 \to 2 \to 2$.

Este producto satisface:

1. Cerradura.
2. Asociatividad.
3.Existencia de elemento neutro.
4. Todo elemento tiene inverso.

Estas 4 propiedades hacen que Sn con esta operación sea un grupo.

3.11 COMBINATORIA: SIN REPETICIÓN

La Combinatoria es una herramienta que nos permite contar el número de situaciones que se pueden dar al someter a un conjunto finito a las acciones de ordenar y/o elegir entre sus elementos.

PERMUTACIONES de n elementos: posibles ordenaciones de un conjunto de n elementos distintos.
Acción: **ordenar**

Su número: $P_n = n! = n \cdot (n-1) \cdot (n-2) \cdots 3 \cdot 2 \cdot 1$ (se lee "factorial de n") . Por convenio 0!=1

En la calculadora: con la tecla $x!$ se calcula "factorial de x" , siendo x un número entero no negativo.

Modelo: ¿Cuántos números de 4 cifras distintas pueden escribirse con los dígitos 2, 3 , 5 y 8 ?
Solución: $4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$ números

VARIACIONES de n elementos tomados de r en r : posibles **muestras ordenadas** de r elementos distintos que se pueden extraer de un conjunto de n elementos ($r \leq n$) Acción: **elegir con orden**

Su número: $V_n^r = n \cdot (n-1) \cdot (n-2) \cdots (n-r+1)$ (r factores enteros consecutivos decrecientes a partir de n)

En la calculadora: con la tecla nPr se calcula V_n^r , siendo $r \leq n$

Notemos que $V_n^n = P_n = n!$

Modelo: En una carrera con 10 atletas, ¿de cuántas formas distintas podrían repartirse las medallas de oro, plata y bronce?

Solución: $V_{10}^3 = 10 \cdot 9 \cdot 8 = 720$ formas distintas

COMBINACIONES de n elementos tomados de r en r: posibles **muestras sin orden** de r elementos distintos que se pueden extraer de un conjunto de n elementos ($r \leq n$). Acción: **elegir sin orden**

Su número: $C_n^r = \binom{n}{r} = \frac{V_n^r}{P_r} = \frac{n \cdot (n-1) \cdot (n-2) \cdots (n-r+1)}{r!} = \frac{n!}{r! \cdot (n-r)!}$

En la calculadora: con la tecla nCr se calcula $\binom{n}{r}$ (que se lee "*n sobre r*")

Modelo: En una reunión de 10 personas debe nombrarse una comisión formada por tres de ellas. ¿Cuántas comisiones distintas podrían nombrarse?

Solución: $\binom{10}{3} = \frac{10 \cdot 9 \cdot 8}{3 \cdot 2 \cdot 1} = 120$ comisiones distintas.

3.12 COMBINATORIA: CON REPETICIÓN

PERMUTACIONES CON REPETICIÓN: posibles **ordenaciones** de una secuencia de n signos entre los que hay algunos repetidos (uno se repite α veces, oto β veces, otro γ veces… etc.).

Su número: $P_n^{\alpha,\beta,\gamma\cdots} = \frac{n!}{\alpha! \cdot \beta! \cdot \gamma! \cdots}$

Notemos que $P_n^{\alpha,n-\alpha} = \binom{n}{\alpha}$

Modelo: ¿Cuántos números distintos de 6 cifras se pueden escribir usando tres unos, dos cincos y un ocho?

Solución: $P_6^{3,2} = \frac{6!}{3! \cdot 2!} = 60$ números distintos.

VARIACIONES CON REPETICIÓN de *n* elementos tomados de *r* en *r*: posibles **muestras ordenadas** de *r* elementos no necesariamente distintos que se pueden extraer de un conjunto de *n* elementos.

Su número: $VR_n^r = n \cdot n \cdot n \cdots n = n^r$

Notemos que aquí puede ser r > n

Modelo: ¿Cuántos números distintos de 6 cifras se escriben usando solamente las cifras 1, 5 y 8 ?

Solución: $VR_3^6 = 3^6 = 729$ números distintos.

COMBINACIONES CON REPETICIÓN de *n* elementos tomados de *r* en *r*: posibles **muestras no ordenadas** de *r* elementos no necesariamente distintos que se pueden extraer de un conjunto de *n* elementos.

Su número: $CR_n^r = \binom{n+r-1}{r}$

Notemos que aquí puede ser r > n

Modelo: Un banco ofrece un regalo a elegir entre 5 posibles regalos por cada cartilla. Un señor que tiene tres cartillas en dicho banco ¿de cuántas formas puede elegir el lote de tres obsequios si no le importa repetir regalos?

Solución: $CR_5^3 = \binom{5+3-1}{3} = \binom{7}{3} = 35$ lotes distintos.

3.13 LOS NÚMEROS COMBINATORIOS

Los números combinatorios o coeficientes binomiales son los números de la forma $\binom{n}{r}$ siendo *r* y *n* enteros no negativos con r≤n. Se calculan, como ya se ha dicho mediante la fórmula:

$$\binom{n}{r} = \frac{n \cdot (n-1) \cdot (n-2) \cdots (n-r+1)}{r!} \quad , \quad \text{o también mediante} \quad \binom{n}{r} = \frac{n!}{r! \cdot (n-r)!}$$

Indican:

- El número de combinaciones sin repetición de n elementos tomados de r en r.
- El número de subconjuntos de r elementos que tiene un conjunto de n elementos.

Propiedades básicas:

1) $\binom{n}{r}=\binom{n}{n-r}$ Ejemplo: $\binom{11}{9}=\binom{11}{2}=55$

2) $\binom{n}{0}=\binom{n}{1}=1$

3) $\binom{n}{r}+\binom{n}{r+1}=\binom{n+1}{r+1}$ Ejemplo: $\binom{6}{4}+\binom{6}{5}=\binom{7}{5}$

4) EL TRIÁNGULO DE TARTAGLIA

```
          1   1
        1   2   1
      1   3   3   1
    1   4   6   4   1
.....................etc.
```

En cada fila:

- Los extremos son iguales a 1
- Cada elemento del interior es la suma de los dos que tiene encima
- Los elementos de la fila n-ésima son $\binom{n}{0},\binom{n}{1},\ldots,\binom{n}{n}$ en ese mismo orden.

5) BINOMIO DE NEWTON, nos da el desarrollo de la potencia n-ésima de un binomio.

$$(a+b)^n=\binom{n}{0}a^n+\binom{n}{1}a^{n-1}b+\binom{n}{2}a^{n-2}b^2+\cdots+\binom{n}{n-1}ab^{n-1}+\binom{n}{n}b^n$$

(Los coeficientes ordenados son la fila n-ésima del triángulo de Tartaglia)

3.14 OTRAS PROPIEDADES

En el caso de una diferencia los signos alternan:

$$(a-b)^n = \binom{n}{0}a^n - \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 - \cdots + (-1)^{n-1}\binom{n}{n-1}ab^{n-1} + (-1)^n\binom{n}{n}b^n$$

6) $\binom{n}{0}+\binom{n}{1}+\binom{n}{2}+\cdots+\binom{n}{n}=2^n$

(Un conjunto de n elementos tiene exactamente 2^n subconjuntos, contando el vacío y el conjunto total)

7) $\binom{n}{0}-\binom{n}{1}+\binom{n}{2}-\cdots+(-1)^n\binom{n}{n}=0$

Otras propiedades:

8) $\binom{r}{r}+\binom{r+1}{r}+\binom{r+2}{r}+\cdots\binom{n}{r}=\binom{n+1}{r+1}$

9) $\binom{n}{0}\binom{m}{r}+\binom{n}{1}\binom{m}{r-1}+\binom{n}{2}\binom{m}{r-2}+\cdots+\binom{n}{r}\binom{m}{0}=\binom{m+n}{r}$

10) $\binom{n}{0}^2+\binom{n}{1}^2+\binom{n}{2}^2+\cdots+\binom{n}{n}^2=\binom{2n}{n}$

4 Asignaciones y funciones

4.1 Funciones

Una función está representada por una regla de correspondencia y don conjuntos conocidos como dominio y contradominio. La regla toma elementos del dominio y los marea al contradominio. $f : A \to B$ en este caso el dominio es A, el contradominio B y la regla de correspondencia f.

Para que f sea función, a cada elemento $a \in A$ le debe corresponder un solo elemento de B. Es decir si $f(a) \neq f(b)$ entonces $a \neq b$.

Definición. La imagen de un subconjunto $A' \subseteq A$ del dominio A es $f(A') = \{y : y = f(x), x \in A'\}$.

Definición. La función inversa g de una función $f : A \to B$ cuando existe satisface $g : f(A) \subseteq B$ y $g(f(x)) = x$ para todo elemento $x \in A$ y $f(g(y)) = y$ para $y \in f(A)$. Cuando la inversa g exista la denotamos por $q = p^{-1}$.

Definición. Una función $f : A \to B$ es uno a uno, cuando cualquier elemento de la imagen de A le corresponde a un solo elemento del dominio A.

Por ejemplo la función constante $f(x) = C$, con C una constante no es 1 a 1, porque todos los elementos del dominio se corresponden con un solo elemento, en este caso C. La función identidad en $f : A \to A$, $f(x) = x$ si es $1a1\alpha$.

Proposición si una función $f : A \to B$ es la $1a1$ entonces tiene inversa.

Prueba. Sea $g : f(A) \subseteq B \to A$ dada por $g(y) = x$, con x el único elemento asociado a y por f. Por tanto se cumple $f(g(y)) = f(x) = y$, $g(f(x)) = g(y) = x$. Es decir $g = p^{-1}$

Definición. Sea una función $f : A \to B$, si $f(A) = B$ se dice que f es sobreyectiva y si es la 1 a 1 decimos que es invectiva. Si es ambos, biyectiva.

4.2 Logaritmos

4.2.1 Funciones Trigonométricas Circulares

Teorema de Pitágoras. En un triángulo rectángulo la suma de los cuadrados de los catetos es igual al cuadrado de la hipotenusa.
Prueba (gráfico pendiente)
El área del cuadrado grande es $(a+b)^2 = a^2+2ab+b^2$, pero también es igual a c^2+4, área del cuadrado pequeño más 4 veces el área del triángulo.
$a^2+2ab+b^2 = c^2+4(\frac{ab}{2}) = c^2+2ab$ finalmente $a^2+b^2 = c^2$.
Las funciones trigonométricas se calculan sobre un triángulo rectángulo.
$cos(\theta) = \frac{catetoadyacente}{hipotenusa} = \frac{a}{c}$, $sen(\theta) = \frac{b}{c}$, $tan(\theta) = \frac{b}{a} = \frac{cos(\theta)}{sen(\theta)}$
(GRÁFICO)
Del Teorema de Pitágoras
$a^2+b^2 = c^2$, dividiendo entre c^2 ambos lados de la ecuación $(\frac{a}{c})^2+(\frac{b}{c})^2 = 1$, $cos^2(\theta)+sen^2(\theta) = 1$ identidad fundamental. Desde un punto de vista práctico el $cosen(\theta)$ permite obtener la proyección horizontal de un segmento de recta y el $sen(\theta)$
La proyección vertical (gráfico) usando este hecho podemos calcular $cos(\alpha+\beta)$ y $sen(\alpha+\beta)$
(2 GRÁFICOS con fórmulas)
$cos(\alpha+\beta) = cos(\alpha)cos(\beta)-sen(\alpha)sen(\beta)$
Tarea: Encuentre $Tan(\alpha+\beta)$

4.2.2 Propiedades de los logaritmos

Un logaritmo base $a \geq 1$ es el inverso de una exponencial a^x, es decir $log_a(a^x) = x$. Usando esta relación y la propiedad de la exponencial $a^{x_1}a^{x_2} = a^{x_1+x_2}$ vamos a obtener las propiedades relevantes de los logaritmos.

1. Sea $y_1 = a^{x_1}$, $y_2 = a^{x_2}$, $log_a(y_1y_2) = log_a(a^{x_1}a^{x_2}) = log_a(a^{x_1+x_2}) = x_1+x_2$ pero $log_a(y_1) = log_a(a^{x_1}) = X_1$, $log_a(y_2) = x_2$ por tanto $log_a(y_1y_2) = log_a(y_1)+log_a(y_2)$, el logaritmo de un producto es la suma de los logaritmos.

2. Considere $log_a(y^2) = log_a(yy) = log_a(y) + log_a(y) = 2log_a(y)$ Tarea. Demuestre por inducción que $log_a(x^n) = nlog_a(x)$.

3. $log_a(1) = log_a(a^{\varnothing}) = \varnothing$, logaritmo de 1 para cualquier a es cero.

4. $\varnothing = log_a(\frac{x}{x}) = log(xx^{-1}) = log(x) + log(x^{-1}) = \varnothing$ Por tanto $log(x^{-1}) = -log(x)$, como si el exponente (-1) bajara fuera del logaritmo.

5. $log_a(\frac{x_1}{x_2} = log_a(x_1 {x_2}^{-1}) = log_a(x_1) + log_a({x_2}^{-1}) = log_a(x_1) - log_a(x_2)$, el logaritmo de una división (cociente) es igual a la resta de los logaritmos.

6. $log(x^{-n}) = log((x^n)^{-1}) = -log(x^n) = -(nlog(x)) = -nlog(x)$ en limpio $log(x^{-n}) = -nlog(x)$.

7. $log_b(a^{log_{a(x)}}) = log_a(x)log_b(a) = log_b(x))$, $log_a(x) = \frac{log_b(x)}{log_b(a)}$ Esto nos permite calcular un logaritmo con otro.

4.3 Cardinalidad y enumerabilidad

4.3.1 Cardinalidad

Se dice que un conjunto B tiene la misma potencia cardinal de un conjunto A si se puede establecer una correspondencia 1 a 1 entre ambos. En el caso de los conjuntos finitos, es común usar la palabra «cardinalidad» en lugar de potencia cardinal y se denota usando dos barras $|A|$ = Número de elementos del conjunto A.

Si un conjunto tiene la potencia cardinal de los números naturales $\mathbb{N}$, se dice que el conjunto es numerable o contable.

Proposición. Lo números naturales son contables.

Prueba. Para la demostración de este enunciado basta con establecer una correspondencia 1 a 1 entre los naturales y los números impares.

$f : \mathbb{N} \to Impares$

por ejemplo: $n \to 2n - 1$

Este mapeo da como asignaciones:
$1 \to 1$
$2 \to 3$
$3 \to 5$

Y así sucesivamente hasta abarcar con todos los números impares. ■

Proposición. Los números reales $\mathbb{R}$ tienen una potencia cardinal mayor que la potencia cardinal de los números naturales.

Prueba. Considerar la representación decimal de los números reales en el intervalo $(0,1)$, suponiendo que cada dígito decimal se puede ordenar (contar).

$0.a_{11}a_{12}...$
$0.a_{21}a_{22}a_{23}...$
$0.a_{31}a_{32}a_{33}a_{34}...$
...

Por lo que es posible mostrar un número $b \in (0,1)$ y que no esté en la lista $b = 0.b_1b_2...$ donde $b_i \neq a_{ii}$ debido a que b es diferente al i-ésimo elemento al menos en su cifra b_i. ■

4.3.2 Algunos resultados de conjuntos

Proposición.
Si A y B son conjuntos finitos y $A \cap B = \emptyset$ Entonces $|A \cup B| = |A| + |B|$.

Prueba. Por definición $A \cup B =$ x: x está en al menos A o B ; entonces todos los |A| elementos de A cumplen la definición y por lo tanto están en A ∪ B, pero además como $B \cap A = 0$ ningún elemento de B ha sido considerado y también todos cumplen con la definición para estar en $A \cup B$.

Por lo tanto, también todos los |B| elementos de B están en $A \cup B$ y son todos aquellos posibles elementos que cumplen con la definición para estar en $A \cup B$. Es decir, finalmente el número de elementos de $A \cup B$ es igual a $|A| + |B|$, $|A \cup B| = |A| + |B|$■

Definición. La resta del conjunto A menos el conjunto B, es denotada como $A \backslash B =$ { x: $x \in A$ y $x \notin B$}.

Proposición. $|A \backslash B| = |A| - |A \cap B|$

Prueba. De manera general se puede escribir que $A = (A \backslash B) \cup (A \cap B)$ y puesto que $(A \backslash B) \cap (A \cap B) = 0$; se puede utilizar el resultado anterior $|A| = |A \backslash B| + |A \cap B|$, y despejando se obtiene que $|A \backslash B| = |A| - |A \cup B|$■

Proposición. Para A y B conjuntos finitos arbitrarios se tiene que $|A \cup B| = |A| + |B| - |A \cap B|$

Prueba. Con la definición que $A = (A \backslash B) \cup (A \cap B)$ y $B = (B \backslash A) \cup (B \cap A)$, por lo tanto

$$A \cup B = [(A \backslash B) \cup (A \cap B)] \cup [(B \backslash A) \cup (B \cap A)] = (A \backslash B) \cup (B \backslash A) \cup (A \cap B)$$

asociando, y por la idempotencia de la unión. Como $(A \backslash B)$, $(B \backslash A)$ y $(A \cap B)$ son disjuntos, se tiene que:

$|A \cup B| = |A \backslash B| + |B \backslash A| + |A \cap B|$, usando que $|A \backslash B| = |A| - |A \cap B|$ y $|B \backslash A| =$

$|B| - |A \cap B|$, finalmente $|A \cup B| = |A| - |A \cap B| + |B| - |A \cap B| + |A \cap B|$

es decir, $|A \cup B| = |A| + |B| - |A \cap B|$■

4.4 Espacios Lineales funcionales

Una de las ideas más férvidas en matemáticas y sus aplicaciones es considerar a una función $f(x)$ como un vector. Esto se puede hacer porque se puede multiplicar una función por un escalar $cf(x)$ y el resultado sigue siendo una función y en general tiene sentido hablar de combinaciones lineales de funciones $c_1 f(x) + c_2 g(x)$ y con estas combinaciones lineales generamos lo que se conoce como espacios lineales o vectoriales de funciones.

Por ejemplo las funciones $cos(nx)$ y $sen(nx)$ general el espacio de las funciones periódicas continuas a trazos, por si esto puesto que $n \in \mathbb{N}$ este espacio es de dimensión infinita. Para estudiar este espacio con más precisión veamos lo siguiente.

Definition 4.1. *Una función f es por si $f(x) = f(-x)$ absorbe el signo y es impar si $f(-x) = -f(x)$ saca el signo.*

Cualquier función f se puede representar como la suma de una función por f_+ y una función impar f_-, $f(x) = f_+(x) + f_-(x)$ donde $f_+(x) = \frac{f(x)+f(-x)}{2}$ y $f_-(x) = \frac{f(x)-f(-x)}{2}$ la demostración se deja al lector.

Se puede verificar que $cos(x)$ es par y $sen(x)$ es impar. Entonces en una representación $f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n cos(nx) + bnsen(nx)$, los cosenos están representando la parte par de $f(x)$ y los senos la parte non. Y estos sub espacios son ortogonales respecto al producto interior $< f, g >= \int_a^b f(x)g(x)dx$. Entonces es de esperarse que $\int_{-\pi}^{\pi} cos(nx)sen(mx)dx = \varnothing$ como puede verificarse. En este caso todas las funciones tiene periodo $[-\pi, \pi]$.

También se puede probar que $\int_{-\pi}^{\pi} cos(nx)cos(mx)dx = \varnothing$, para $n \neq m$ y los mismo para el seno. Consideremos ahora $\int_{-\pi}^{\pi} cos^2(nx)dx$, como $cos^2(x) = \frac{cos(2x)+1}{2}$.
$\int_{-\pi}^{\pi} cos^2(nx)dx = \frac{1}{2}\int_{-\pi}^{\pi} cos(2nx)dx + \frac{1}{2}\int_{-\pi}^{\pi} dx$ pero de lo anterior
$\int_{-\pi}^{\pi} cos(2nx)dx = \varnothing$. Por lo que $\int_{-\pi}^{\pi} cos^2(nx)dx = \pi$, por un argumento similar $\int_{-\pi}^{\pi} sen^2(nx)dx = \pi$. A la representación de f en términos de $sen(nx)$ y $cos(nx)$ se le conoce como serie de Fourier o sumas trigonométricas.

$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n cos(nx) + b_n sen(nx)$ y los coeficientes a_n, b_n son los coeficientes de Fourier de $f(x)$, estos representan los componentes de frecuencia en f, puesto que nos dicen que tanto de las frecuencias de $cos(nx)$ y $sen(nx)$ hay en f. Para encontrar estos coeficientes basta usar las integrales anteriores por ejemplo para a_R multiplicamos ambos lados de la ecuación por $cos(Rx)$ e integramos de $-\pi$ a π.

$\int_{-\pi}^{\pi} cos(Rx) f(x) dx =$
$\int_{-\pi}^{\pi} \frac{a_0}{2} cos(Rx) dx + \sum_{n=1}^{\infty} \int_{-\pi}^{\pi} cos(Rx) cos(nx) dx + \int_{-\pi}^{\pi} cos(nx) sen(nx) dx.$
$\int_{-\pi}^{\pi} f(x) cos(Rx) dx = a_R \pi$, finalmente $a_R = \frac{i}{\pi} \int_{-\pi}^{\pi} f(x) cos(Rx) dx$ de manera análoga para $b_R = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) sen(Rx) dx$ y $a_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) dx$.

4.5 Optimización de funciones de una variable

Definición. Sea $f : D \to \mathbb{R}$ donde el dominio D puede ser por ejemplo $\mathbb{R}^4$ o $\mathbb{R}$. Decimos que f tiene un mínimo en X_0 si para X en una vecindad de X_0 , $f(x) \geq f(x_0)$. En este caso estamos hablando de un mínimo local y si la anterior desigualdad ocurre para cualquier x en el dominio de f, se trata de un mínimo global, lo mismo podemos decir de los máximos locales y globales, únicamente invirtiendo la desigualdad.

1. Usando un criterio geométrico es claro que en los máximos y mínimos locales la derivada vale cero

2. Para saber si es máximo o mínimo cuando $f'(x_0) = \phi$, usemos la representación de segundo orden de $f(x)$ en X_0. $f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2$ para x cercana a x_0, como $f'(x_0) = \varnothing$, tenemos $f(x) = f(x_0) + \frac{f''(x_0)}{2}(x - x_0)^2$, si x_0 es donde f toma el valor mínimo local, $\varnothing \geq f(x) - f(x_0) = \frac{f''(x_0)}{2}(x - x_0)^2$, como $(x - x_0)^2 \geq \varnothing$ entonces $f''(x0) \geq \varnothing$ es decir si $f'(x_0) = \varnothing$ y $f''(x_0) > \varnothing$ entonces x_0 es un mínimo local y se puede ver de manera análoga que si $f'(x_0) = \varnothing$ y $f''(x_0) < \varnothing$ entonces x_0 es un máximo, es decir donde $f(x)$ alcanza un máximo. Para indicar el valor máximo de $f(x)$ usamos $maxf(x)$. Aunque a veces se usa $orgmaxf(x)$ para ambos.

4.5.1 Optimización de funciones de varias variables

Para el caso de $f : D \to \mathbb{R}$ y $D = \mathbb{R}^n$ tenemos que f depende de varias variables en este caso en los máximos y mínimos locales el gradiente vale cero, $\bigtriangledown f(x_0) = \varnothing$ en este caso en (9 toma el máximo o el mínimo. Tomando la representación de segundo orden de f, tenemos $f(x) = f(x_0) + \bigtriangledown f(x_0) \cdot (x - x_0) + (x - x_0)^t H(x - x_0)$, donde x es un vector muy cercano a x_0 y H es el Hessiano de f, nuevamente si $\bigtriangledown f(x_0) = \varnothing$ y $f(x_0)$ es un valor mínimo, $f(x) - f(x_0) = (x - x_0)^x H(x - x_0) \geq \varnothing$ y esto implica que el Hessiano es positivo definido. Entonces si $\bigtriangledown f(x_0) = \varnothing$ y H es positivo definido tenemos un mínimo en x_0 y si $\bigtriangledown f(x_0) = \varnothing$ y el Hessiano es positivo negativo

tenemos un máximo. Muchas veces se tiene el problema de tener un criterio para aproximarse a un mínimo local y esto se logra al notar que
$df(x) = \bigtriangledown f(x) \cdot dx = |\bigtriangledown f(x)||dx|cos(\theta)$, esto implica si el desplazamiento dx lo hacemos en la dirección del $\bigtriangledown f(x)$ el crecimiento de la función $df(x)$ es máximo porque $\theta = cero$ y $cos(\theta) = 1$ y si dx es en la dirección de $\bigtriangledown f(x)$ pero en sentido contrario $\theta = 180$ y $cos(\theta) = -1$ entonces el decrecimiento es máximo.

En resumen el $\bigtriangledown f(x)$ apunta siempre en la dirección de máximo crecimiento de $f(x)$ y $-\bigtriangledown f(x)$ apunta en la dirección de máximo decrecimiento. Con este criterio heurístico podemos fácilmente implementar el algoritmo deseado.

5 Teoría de grafos

5.1 Definiciones

5.1.1 ¿Qué es un grafo?

Un grafo es un conjunto de objetos V=v1,v2,v3,... llamados vértices (también suelen ser llamados puntos o nodos) y otro set E=e1,e2,e3,... cuyos elementos son nombrados aristas o arcos.

Usualmente los grafos son denotados como G=(V,E)

Existen 2 tipos básicos de grafos: los dirigidos y los no dirigidos

5.1.2 Grafo dirigido

Sea V un conjunto finito no vacío y sea la relación no binaria $E \subseteq V \times V$. El par ordenado (V,E) es un grafo dirigido sobre V, o digrafo, donde V es ele conjunto de vértices o nodos y y E es su conjunto de aristas. Escribimos G = (V, E) para denotar tal digrafo. En el diagrama de un grafo dirigido, cada vértice e=(u,v) es representado por una flecha o una curva dirigida de un punto inicial u de e al punto terminal v. La siguiente figura es un ejemplo de un grafo dirigido.

Suponga e=(u,v) es un vértice dirigido en un digrafo, entonces:

- (i) u es llamado el vértice inicial de e y v es el vértice termimnal de e.
- (ii) se le conoce a e como incidente de u y de v
- (iii) u es adyacente a v y v es adyacente de u

5.1.3 Grafo no dirigido

Cuando no importa la dirección de las aristas, la estructura G = (V, E), donde E es ahora un conjunto de pares no ordenados sobre V, es decir el conjunto de aristas rep-

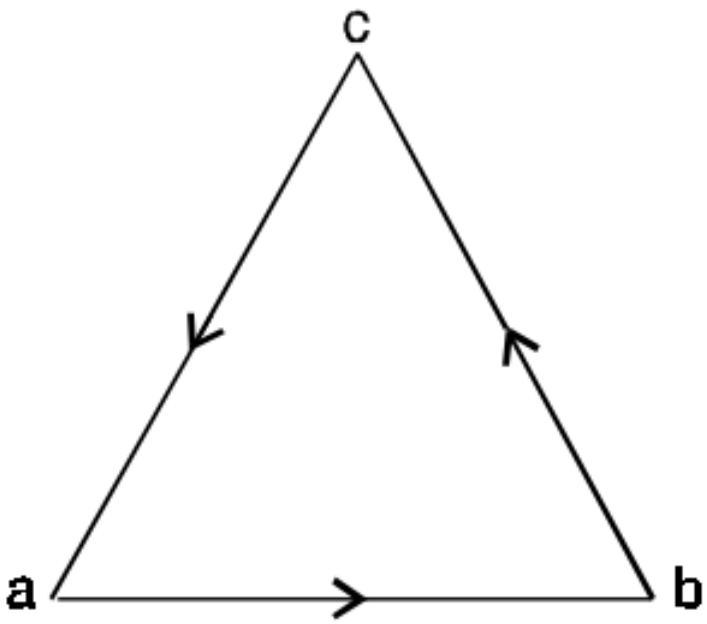


Figure 5.1: Grafo dirigido

resenta una relación simétrica binaria, donde si Vj y Vk son vértices cualesquiera del conjunto de vértices V de un grafo, $(V_j, V_k) \in E \rightarrow (V_k, V_j) \in E$ decimos que tenemos un grafo no dirigido. En otras palabras, si cada vértice del grafo G no tiene dirección entonces el grafo es nombrado como no dirigido, a continuación se muestra el ejemplo de un grafo no dirigido.

5.1.4 Subgrafos

Considere un grafo G = G(V, E). Un grafo $H = H(V^{'}, E^{'})$, se denomina subgrafo de G si los vértices y las aristas de H están contenidas en los vértices y en las aristas de G; es decir, si $V^{'} \subseteq V$ y $E^{'} \subseteq E$. En particular:

- i) Un subgrafo $H(V^{'}, E^{'})$ de G(V,E) se denomina subgrafo inducido por sus vértices $V^{'}$ si su conjunto de aristas $E^{'}$ contiene todas las aristas en G cuyos puntos extremos pertenecen a los vértices en H.
- ii) Si v es un vértice en V, entonces G-v es el subgrafo de G obtenida al eliminar v de G y al eliminar todas las aristas en G que contiene a v.

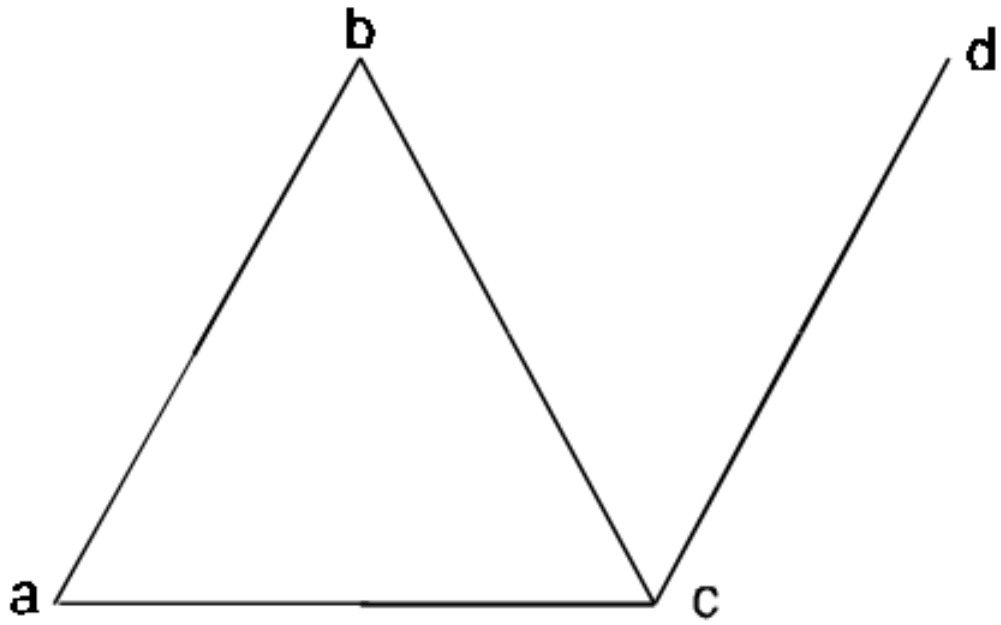


Figure 5.2: Grafo no dirigido

- iii) Si e es una aristas en G, entonces G-e es el subgrafo de G obtenido al eliminar la arista e de G.

5.1.5 Otros tipos de grafos

Además de los dos grafos básicos, dirigido y no dirigidos hay otros tipos de grafos que se denotan a continuación.

- **Grafo de cadena:** son un tipo de grafo híbrido formado por aristas dirigidas y no dirigidas
- **Grafo simple:** es un tipo de grafo el cual no incluye ciclos ni aristas paralelas
- **Multigrafo:** son grafos con dos o más aristas que pueden conectar a un mismo vértice
- **Grafo completo:** es un grafo con aristas entre cada par de vértices
- **Grafo bipartito:** son grafos que se pueden dividir en dos subconjuntos disjuntos de vértices, donde cada una de las aristas conecta un vértice del primer conjunto

con uno del segundo

- **Grafo pesado:** es un grafo que tiene pesos asociados a vértices y/o aristas

5.1.6 Caminos

Definición sean x, y vértices (no necesariamente diferentes entre sí) de un grafo no dirigido G=(V,E). Un camino x - y en G es una sucesión alternada finita (sin lazos) de vértices y aristas de G, que comienza en el vértice x y termina en el vértice y; que contiene las n aristas $e = x_{i-1}, x_i$ donde $1 \leq i \leq n$.

$$x = x_0, e_1, x_1, e_2, x_2, ..., e_{n-1}, x_{n-1}, e_n, x_n = y \tag{5.1}$$

La longitud n de un camino es el número de aristas que hay en el camino (Si n=0, no existen aristas, x=y, y el camino se denomina trivial).

5.1.7 Recorridos

Sean x,y vértices de un grafo G=(V,E). Si no se repite alguna arista en el camino x -y, entonces el camino es un recorrido x - y. Por ejemplo el camino simple mostrado en la siguiente figura, también es un recorrido.

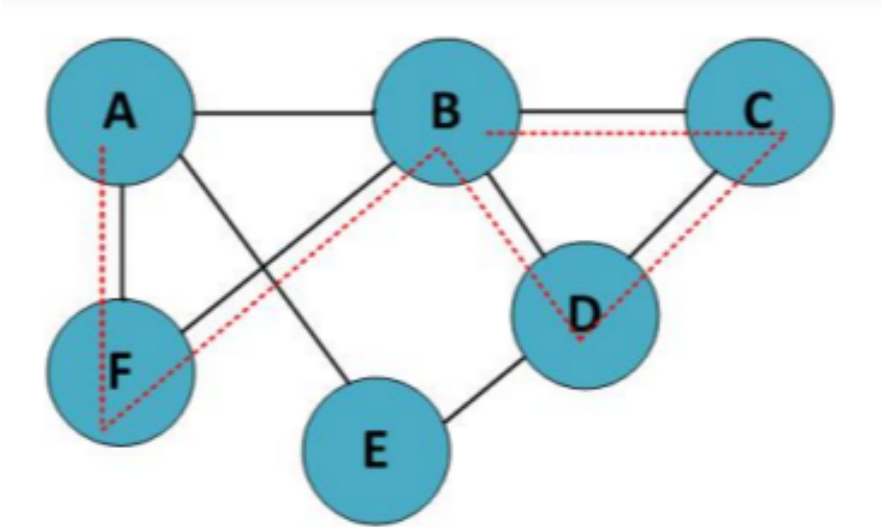


Figure 5.3: Camino

5.1.8 Circuitos

Sean x, y vértices de un grafo G=(V,E), donde x=y. Si no se repite alguna arista en el camino x - y, entonces existe un circuito x - y. Dicho de otra forma, se puede decir que un circuito es un recorrido donde el vértice inicial es también el final (recorrido cerrado). A continuación se muestra un circuito en el grafo.

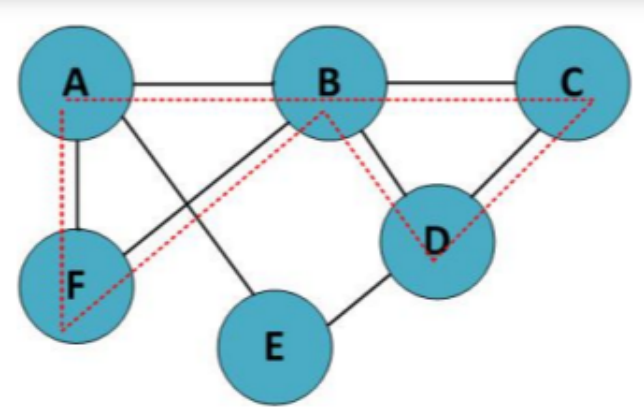


Figure 5.4: Circuitos

5.1.9 Conectividad

Un grafo G=(V,E) es conexo si para cualquier par de vértices u y v existe un camino en G que los une, dicho de otra forma, un camino con extremos u y v. En la siguiente figura se muestra un grafo G con vértices V=A,B,C,D,E,F,G,H y aristas $E = e_1, e_2, e_3, e_4, e_5, e_6, e_7$, del cual se pueden formar dos subgrafos W=A,B,C,D,E $\{e_1, e_2, e_3, e_4, e_5\}$ y Z=F,G,H$\{e_6, e_7\}$. Se puede afirmar que el grafo G no es conexo pues no existe un camino del vértice A con el vértice F, dado que si existe un solo par de vértices sin camino el grafo no es conexo. Sin embargo los subgrafos W y Z sí son conexos al existir un camino entre cualquier par de vértices que los forman. Un grafo G está conectado si hay una trayectoria entre cada par de vértices distintos en G. Si un grafo G no está conectado, cada parte si está conectada se llama componente conexo de G. Entre las diferentes trayectorias que podemos encontrar en un grafo se encuentran caminos, recorridos, circuitosy ciclos.

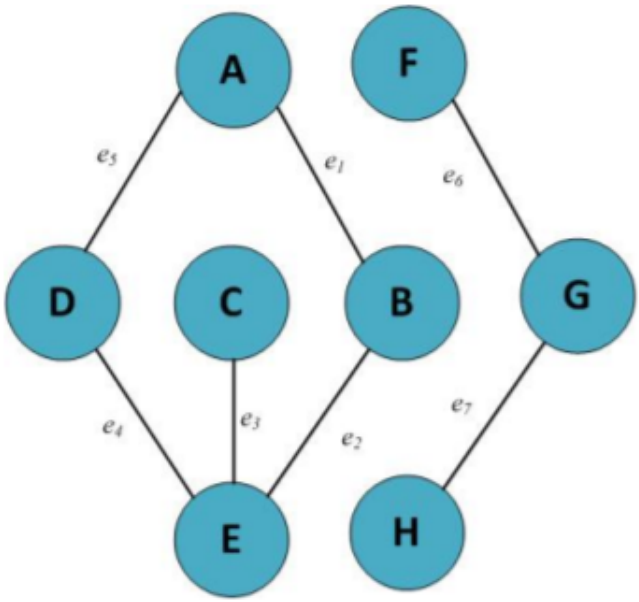


Figure 5.5: Grafo conexo

5.2 Árboles

Si un grafo es un árbol, escribimos T en vez de G para enfatizar dicha estructura. Algunas propiedades de los árboles en general son:

- Si a,b son vértices distintos de un árbol T=(V,E), entonces hay un único camino simple que conecta estos vértices
- En cualquier árbol T=(V,E) se cumple que el número de vértices es igual al número de aristas más uno $|V| = |E| + 1$
- Para cualquier árbol T=(V,E) si $|V| \leq 2$, entonces T tiene al menos dos vértices terminales, llamadas hojas

5.2.1 Árboles no dirigidos.

Sea G=(V,E) un grafo no dirigido sin lazos. El grafo G es un árbol no dirigido si G es conexo y no contiene lazos. En un árbol no dirigido existen dos tipos de vértices.

- Vértices terminales de grado 1

- Vértices internos de grado mayor a 1

Los vértices terminales tambipen se llaman hojas

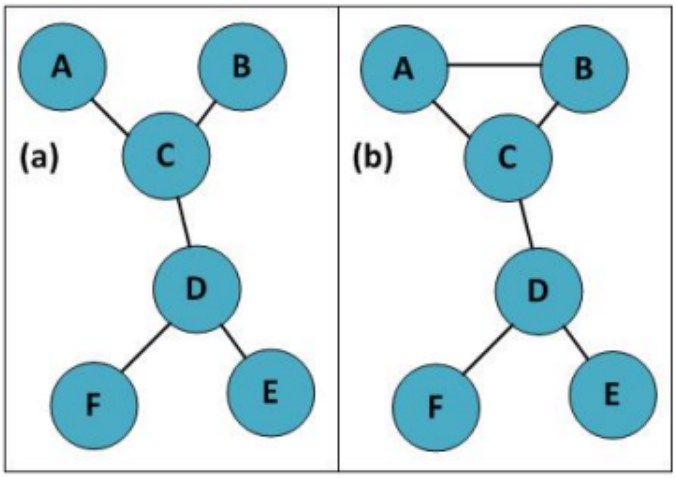


Figure 5.6: Árbol no dirigido

5.2.2 Árboles dirigidos.

Dado G un grafo dirigido, G es un árbol dirigido si el grafo no dirigido asociado es un árbol, existen dos tipos de árboles dirigidos: a) Árbol con una sola raíz o árbol dirigido simple y b) Poliárbol o árbol dirigido con múltiples raíces. G es un árbol simple si existe un único vértice r en G, tal que el grado de entrada de $r = grad_e(r) = 0$ y para todos los demás vértices v el grado de entrada de v es $grad_e(v) > 1$. G es un poliárbol si existe más de un vértice r, es decir que tenga más de una raíz. En la figura **??** se muestran los dos tipos de árboles.

- a) es un poliárbol dado que los vértices A y D son raíces ya que ambas tienen su grado de entrada igual a cero.
- b) es un árbol dirigido simple porque el vértice A es la única raíz dado que su grado de entrada es cero.

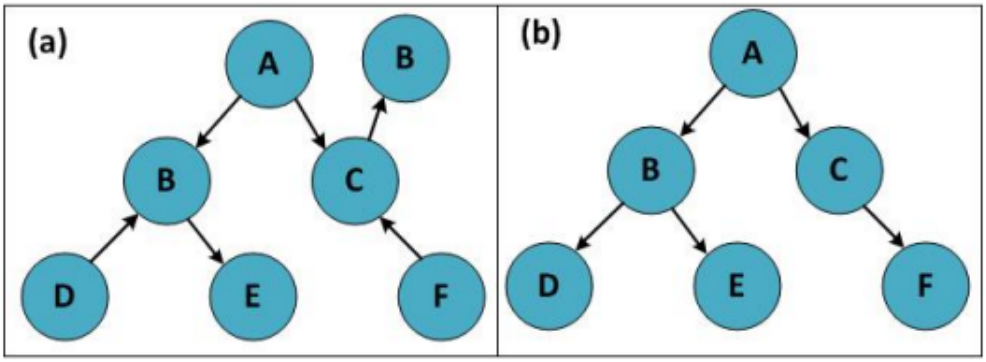


Figure 5.7: Árboles dirigidos

5.3 Grafos Eulerianos y Hamiltonianos

Sea G un grafo o multigrafo no dirigido. Para cualquier vértice v de G, el grado de v, que se denota como $grad(v)$, es el número de aristas en G que son incidentes con v. Cuando se tiene que $grad(v) = 0$ se dice que v es un vértice aislado. La sucesión de grados de un grafo se obtiene ordenando en forma creciente los grados de todos los vértices.

En todo Grafo G=(V,E) se cumple:

$$\sum_{v \in V} grad(v) = 2\,|E| \tag{5.2}$$

Caminos y circutios Eulerianos

Sea G=(V,E) un grafo o multigrafo no dirigido sin vértices aislados. Un circuito Euleriano es un circuito en G que recorra cada arista del grafo exactamente una vez. Si un recorrido abierto de A a B en G recorre cada arista de G exactamente una vez, se llamará camino Euleriano.

A continuación se muestra un camino Euleriano que va de A a B pasando a través de la siguiente ruta (A-B-C-D-A-B) y un circuito Euleriano, en donde se cierra una ruta exacta.

La determinación para conocer si un grafo contiene un recorrido Euleriano consiste en evaluar los siguientes puntos

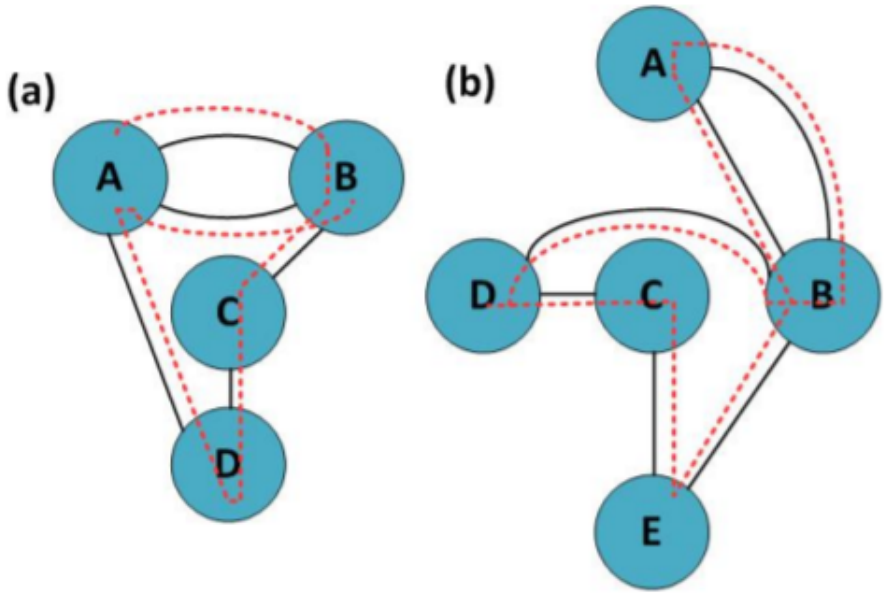


Figure 5.8: a) Camino Euleriano b) Circuito Euleriano

- Un grafo conexo tiene más de dos nodos con grado impar no contendrá un recorrido Euleriano
- Si existen exactamente dos vértices de grado impar, el grafo puede recorrer el recorrido Euleriano si éste empieza en uno de los dos vértices de grado impar y termina en el otro
- Si no existen vértices de grado impar, el grafo se puede recorrer, en este caso, el recorrido Euleriano será cerrado (es decir, se trata de un circuito Euleriano)

Grafo Hamiltoniano

En la sección anterior referente a los grafos Eulerianos se recalcaron las aristas recrridas, para el análisis de los grafos Hamiltonianos es preciso concentrarnos en los vértices. Se dice que un camino es Hamiltoniano en un grafo G si es aquel recorrido que visita todos los vértices de G exactamente una vez. Si dicho recorrido es cerrado se cuenta, entonces, con un circuito Hamiltoniano.

Debe enfatizarse que un recorrido Euleriano recorre cada arista exactamente una vez, aunque puede repetir vértices, mientras que un recorrido Hamiltoniano visita

exclusivamente cada vértices una sola vez, pero no puede repetir aristas y no tiene que pasar por todas necesariamente.

Dicho lo anterior se tiene la siguiente definición formal. Si G=(V,E) es un grafo o multigrafo con $|V| \geq 3$, decimos que G tiene un circuito Hamiltoniano si existe un circuito en G que contenga cada vértice de V y un camino Hamiltoniano es un camino simple de G que contiene todos los vértices.

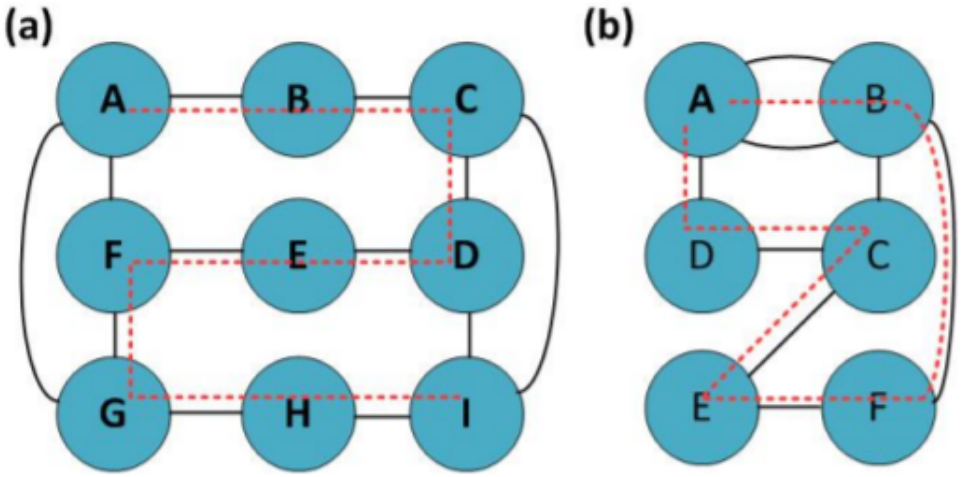


Figure 5.9: a) Camino Hamiltoniano b) Circuito Hamiltoniano

5.4 Grafos planares

Un grafo o multigrafo es plano si podemos dibujar G en el plano de modo que sus aristas se intersequen sólo en los vértices de G. Este dibujo de G se conoce como una inmersión de G en el plano.

La Figura 5.10 muestra grafos planos. El primero (figura a) es plano dado que sus aristas no se cruzan, excepto en los vértices. Por su parte, el grafo de la figura b es plano puesto que las aristas A,C y B,D se cruzan en un punto que no es un vértice, no obstante, es posible trazar nuevamente este grafo como se hizo en la figura c.

El grafo mostrado en la figura c es equivalente al grafo b, de c se observa que se cumple con las características de un grafo plano, por lo tanto b también lo es.

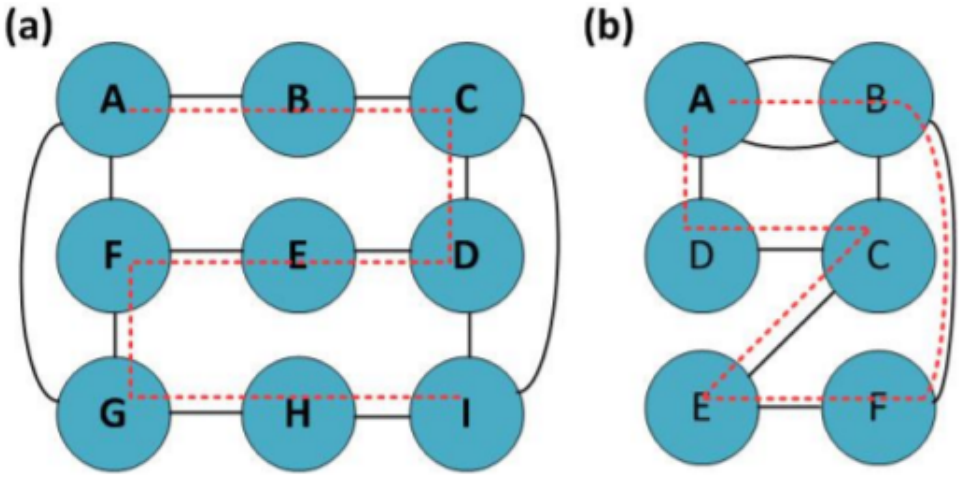


Figure 5.10: Grafos planares

5.5 Teorema de Oré

Si G es un grupo con $p \geq 3$ vértices tales que para todos los vértices no adyacentes u y v $deg(u) + deg(v) \geq p$ entonces G es Hamiltoniano.

Demostración. Sea k un entero que denote el número de vértices de G cuyo grado no excede n donde $1 \leq n \leq \frac{p}{2}$. Dichos k vértices inducen un subgrafo H el cual es completo, ya que si cualquiera de los dos vértices de H no fueran adyacentes, existirían dos vértices no adyacentes, la suma de los cuales es menor a p. Esto implica que $k \leq n+1$. No obstante sabemos que $k \neq n+1$ porque de ser así, cada vértices de H es adyacente solo a dos vértices de sí mismo, y si $u \in B(H)$ y $v \in V(G) - V(H)$ entonces tenemos que $deg(u) + deg(v) \leq n + (p-n-2) = p-2$ lo cual es una contradicción. Más aún, $k \neq n$; de otra manera cada vértice de H es adyacete a, máximo, un vértices de G que no esté en H.

No obstante, dado que $k = n < \frac{p}{2}$, existe un vértice $w \in V(G) - V(H)$ no adyacente a algún vértice de H. Entonces si $u \in V(H), deg(u) + deg(w) \leq n + (p-n-1) = p-1$ lo cual vuelve a ser una contradicción. Por lo tanto, $k < n$, lo cual implica que G satisface la condición, por lo tanto G es Hamiltoniano.

5.6 Problema del vendedor

El problema del vendedor se enuncia de la siguiente manera: "Un vendedor requiere visitar un número de ciudades durante un viaje. Dada la distancia entre las ciudades, ¿cuál es el orden en que debe viajar para poder visitar cada ciudad precisamente una vez y regresar a casa con la mínima distancia recorrida?" Para solucionar este problema es conveniente utilizar una representación en grafos, siendo las ciudades los nodos y la distancia los vértices. En este grafo, cada vértice está asociado a un número real (el cual representa la distancia recorrida entre los nodos a los cuales conecta), dicho grafo sera entonces un grafo con ponderado (también llamado grafo con pesos), representada dicha ponderación como $w(e_i)$ siendo w el peso del vértice e_i. Puesto en otras palabras, el vendedor quiere visitar cada una de las n ciudades exactamente una vez y regresar al punto inicial (su hogar), si cada una de las ciudades tiene un camino entre otra ciudad tenemos un grafo ponderado completo. Por ejemplo, supongamos que el vendedor quiere visitar cinco ciudades, nombradas como A,B,C,D y E, vea la figura 5.11.

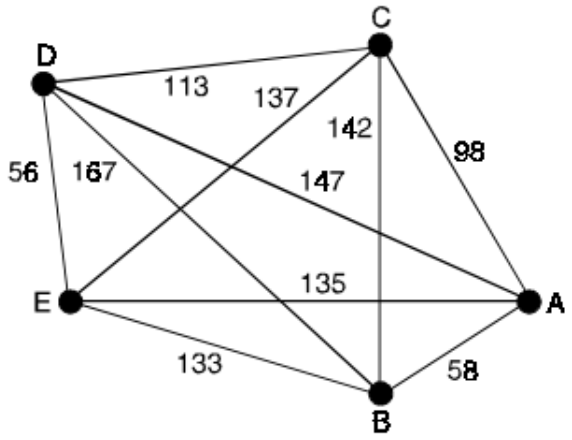


Figure 5.11: Grafo mostrando las distancias entre las ciudades

¿En qué orden debería visitar el vendedor las 5 ciudades para viajar la distancia mínima? Para resolver este problema podemos asumir que el vendedor comienza en A, dado que ella debe ser parte del circuito y examinar todas las posibles formas que tiene de viajar a través de las otras 4 ciudades y regresar a la ciudad A (empezar en otro

punto produciría los mismos circuitos). De tal forma que hay un total de 24 circuitos, al ser una permutación sin repetición de 4, $4! = 24$. Al viajar la misma distancia cuando el circuito se recorre en orden regresiva, basta únicamente con considerar 12 rutas para encontrar la mínima distancia, la cual se muestra en la imagen 5.12.

Ruta	Distancia
A – B – D – C – E – A	610
A – B – D – E – C – A	516
A – B – E – C – D – A	588
A – B – E – D – C – A	458
A – B – C – E – D – A	540
A – B – C – D – E – A	504
A – C – B – D – E – A	598
A – C – B – E – D – A	576
A – C – E – B – D – A	682
A – C – D – B – E – A	646
A – D – C – B – E – A	670
A – D – B – C – E – A	728

Figure 5.12: Distancia de los 12 circuitos posibles

En la imagen se observa que la distancia mínima corresponde al valor de 458 (se deja al lector la selección de unidades), siguiendo el circuito A-B-E-D-C-A o su reverso A-C-D-E-B-A.

El problema del vendedor requiere un circuito que sea el peso mínimo en un grafo completo no dirigido ponderado, que visite cada nodo exactamente una vez y regrese al punto inicial, lo cual es equivalente a buscar un circuito Hamiltoniano con un peso mínimo en un grafo completo, ya que cada vértice es visitado exactamente una sola vez a lo largo de dicho circuito. La forma más simple de resolver una instancia del problema del vendedor es examinar todos los circuitos Hamiltonianos posibles y seleccionar el que tenga el peso mínimo. ¿Cuántos circuitos tenemos que examinar para

resolver el problema si hay n vértices en el grafo?. Una vez habiendo seleccionado el punto de arranque hay $(n-1)!$ diferentes circuitos Hamiltonianos que examinar, dado que hay $n-1$ opciones en el segundo paso (partiendo del inicio al segundo vértice) y $n-2$ en el tercer paso y así de manera continua. Dado que el grafo Hamiltoniano puede ser recorrido en sentido inverso, la examinación se reduce de $(n-1)!$ a $\frac{(n-1)}{2}$.

Note que $\frac{(n-1)}{2}$ crece de manera extremadamente rápida, tratar de resolver el problema del vendedor de esta forma en donde hay, cuando menos algunas decenas de vértices es totalmente impráctico. Por ejemplo, para 25 vértices tendríamos un total de $\frac{24!}{2}$ lo cual se aproxima a 3.1×10^{23} circuitos Hamiltonianos diferentes que analizar.

Si tomara solo un nanosegundo 10^{-9} segundos para analizar cada circuito Hamiltoniano, entonces el resultado tardaría aproximadamente 10 millones de años para encontrar el resultado utilizando fuerza bruta.

5.7 Problema de los puentes de Königsberg

Había dos islas conectadas entre sí a las orillas del río Pregolia por 7 puentes como se muestra en la figura 5.13

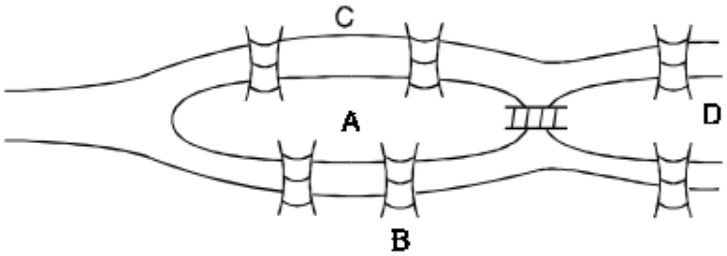


Figure 5.13: Problema de los puentes de Königsberg

El problema consiste en comenzar en cualquiera de las 4 áreas de tierra, caminar a través de cada puente una sola vez y regresar al punto inicial. Se podría pensar fácilmente en que existe una solución intuitiva, pero todos los intentos esultarían inútiles, al probar eso, el problema se volvería en un problema sin resolución. Euler reemplazó cada área terrestre por un vértice y cada puente por una arista que une los vértices correspondientes, finalizando con el grafo que se muestra en la figura 5.14, lo anterior se explica más a fondo en la siguiente sección.

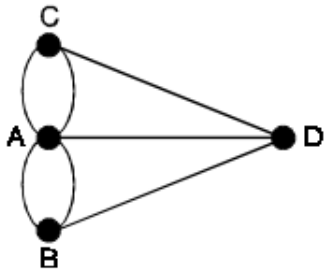


Figure 5.14: Grafo resultante del problema de los puentes de Königsberg

Efectivamente, existen justificaciones para aseverar que los grafos y la teoría de grafos podrían haberse originado en Europa durante la primera mitad del siglo XVIII. En ese entonces Leonhard Euler estudió y resolvió éste problema. Kónigsberg era el nombre de una ciudad en Prusia del Este. Se cuenta que los habitantes de Königsberg se entretenían tratando de encontrar una ruta alrededor de la ciudad que recorriese los siete puentes sin pasar por algún puente más de una vez. En 1736, Leonard Euler, uno de los matemáticos de primera línea de su época, publicó un artículo en el cual no solo soluciona este problema en particular, sino que además proporcionaba un método general para resolver otros problemas del mismo tipo. Puesto en palabras de Euler.

"(de este problema en particular) yo he formulado el problema general: cualquiera sea el arreglo y división del río en ramas, y sin importar cuantos puentes hay ¿puede uno determinar si es o no posible cruzar cada puente exactamente una vez? Mientras nuestro interés sea el problema de los siete puentes de Königsberg, éste ouede ser resuelto construyendo una lista exhaustiva de todas las rutas posibles para luego determinar si alguna de las rutas satisface las condiciones del problema. Debido al número de posibilidades, este método de solución sería demasiado difícil y laborios y en otros problemas con más puentes sería imposible...".

Posteriormente, el mismo Euler afirma "Todo mi método se basa en una conveniente y particular manera con la cual se puede representar el cruce de un puente... Si un pasante va de A a B sobre el puente a o b yo escribo AB".

A continuación, Euler observa que si el número m de puentes que conducen al área A es impar, entonces A debe aparecer m+12 veces en la representación correspondi-

ente a la ruta buscada, luego Euler concluye.

"En el caso de los puentes de Königsber por tanto, debe haber tres ocurrencias de la letra A en la represengación de la ruta dado que cinco puentes (a,b,c,d,e), conducen al área A. Luego, dado que tres puentes conducen a B, la letra B debe ocurrir dos veces; similarmente , D debe ocurrir dos veces y C también. Así en una serie de 8 letras representando el recorrido de los siete puentes, la letra A debe ocurrir 3 veces, y las letras B, C y D dos veces cada una, pero esto no puede suceder en una sucesión de 8 letras. Sigue de esto que tal travesía no puede ser tomada a lo largo de los siete puentes de Königsberg"

En los siguientes párrafos de su artículo, Euler se detiene a examinar la solución más general cuando el número de puentes que inciden en alguna de las áreas es par. En primer lugar, él generaliza su observación referente al número de ocurrencias de una letra en la representación de la ruta buscada.

"... Así en general, si el número de puentes (que conducen a la zona A) es par, entonces el número de ocurrencias de A será la mitad de este número si la travesía no se incia en A, y el número de ocurrencias será en una mitad mayor que la mitad del número de puentes si la travesía se incia en A".

En este momento, Euler se apresta a establecer una regla para el caso general, la cual será utilizada para determinar si puede o no existir un arreglo de letras que representa a la ruta buscada.

"Dado que en cualquier travesía uno puede partir desde una sola área, definiré... para cada área A, el número de ocurrencias de la letra A como igual a la mitad del número de puentes (que conducen a A) más uno, si este número de puentes es impar, y si este npumero de puentes es par, como igual a la mitad de este número".

Denotemos por $n(A)$ al número de ocurrencias de la letra A, según la definición dada por Euler. Sea p(A) el npumero de puentes que conducen al área A. Entonces la anterior definición se formula de la siguiente manera.

$$n(a) = \begin{cases} \frac{p(A)+1}{2} & si \quad p(A) \quad es \quad impar \\ \frac{p(A)}{2} & si \quad p(A) \quad es \quad par \end{cases} \tag{5.3}$$

Conviene definir la suma de Euler de la configuración de aguas y puentes que se esté considerando (así lo llama Euler) como la suma de los números de ocurrencias de todas las áreas de la configuración. Es decir, si E denota la suma Euler de una configuración, entonces.

$$E = \sum_{A} n(A) \tag{5.4}$$

Antres de enunciar y demostrar la regla encontrada por Euler, necesitamos el siguiente lema, cuya demostración se basa enteramente en las observaciones hechas por Euler con respecto al número de ocurrencias de una letra en la representación de la ruta buscada.

Lema. Dada una configuración de aguas y puentes, supongamos que existe una ruta R que recorre todos los puentes de la configuración, cada uno exactamente una vez

- Si a cada región llega un número par de puentes, entonces la ruta R se inicia y termina en la misma región.
- Si existe una región a la que llegan un número impar de puentes, entonces hay exactamente dos de estas regiones, y R se inicia en una de ellas y termina en la otra.

Demostración. Para demostrar el primer punto, supongamos que el número de puentes que conducen a cada región es par. Basta observar que si la ruta R se inicia en una región A pero no termina en A, entonces el npumero de puentes que conducen a A debe ser impar. De aquí se sigue el resultado. Para demostrar el segundo punto, observamos que si una letra no es ni inicial ni final en la representación de R, entonces el número de puentes que conducen a la región respectiva debe ser par. Por tanto, si existe una región A en la configuración a la que llegan un número impar de puentes,

entonces la ruta R comienza en A o termina en A. Por otro lado, si la representación de R comienza y termina en la misma letra, entonces el número de puentes que conducen a tal región debe ser par. Esto forzosamente implica que si existe una región a la cual llegan un número impar de puentes, entonces existen exactamente dos regiones en la configuración a las cuales llegan un número impar de puentes, siendo una de estas regiones la inicial y la otra la final de la ruta R. **Q.E.D.**

Teorema fundamental de Euler. En una configuración de aguas y puentes, denotemos por P al número total de puentes de la configuración. Entonces

$$E = \sum_A p(A) = 2P \tag{5.5}$$

Demostración. Cada puente está conectado exactamente dos veces en la sumatoria anterior.

Corolario. En una configuración de aguas y puentes, sea h el número de regiones a las cuales llega un número impar de puentes. Entonces:

$$E = \sum_A n(A) = P + h2 \tag{5.6}$$

Demostración.

$$\sum_A n(A) = \sum_A \frac{P(A)}{2} + \sum_A \frac{P(A)+1}{2} = P + \frac{h}{2} \tag{5.7}$$

Observe que del Corolario anterior, se concluye que el número h debe ser par. Esta conclusión, ahora parte fundamental de la teoría de grafos, también fue descubierta y probada por Euler.

Ahora ya tenemos todos los elementos para presentar y demostrar la regla encontrada por Euler, la cual permite decidir cuándo la ruta buscada existe y cuál es la naturaleza de dicha ruta.

5.7.1 Teorema de Euler

Dada una configuración de aguas y puentes, supongamos que existe una ruta R que recorre cada uno de los puentes exactamente una vez. Entonces, y solo entonces, una de las siguientes proposiciones es verdadera.

- La suma de Euler E de la configuración es igual al número total de puentes
- La suma de Euler E de la configuración es igual al número total de puentes más uno

En el primer caso, a cada región llegan un número par de puentes y la ruta R comienza y termina en una misma región. En el caso siguiente, existen exactamente dos regiones hacia las que llegan un número impar de puentes, debiendo la ruta R comenzar en una de estas regiones y terminar en la otra.

Demostración. Supongamos que para la configuración de aguas y puentes dada, existe una ruta R que recorre cada uno de los puentes exactamente una vez. Denotemos por E a la suma de Euler de la configuración dada y denotemos por P al número total de puentes que existen en la configuración. Si a cada una de las regiones llega un número par de puentes, entonces el corolario implica que:

$$E = P \tag{5.8}$$

Por el contrario, si existe una región a la cual llega un número impar de puentes, entonces el lema implica que hay exactamente dos de estas regiones. Aplicando el corolario, se obtiene.

$$E = P + 1 \tag{5.9}$$

A continuación, probaremos la suficiencia del primer o segundo punto del teorema de Eulera.

En primer lugar, supongamos que la primera idea es verdadera, entonces, del Corolario, se obtiene que a cada región llegan un número par de puentes. Sea A una región de la configuración. Construyamos una ruta R que parta de A y que, abandonando A, retorne a A. Para tal efecto, procedamos como sigue. Una vez que salimos de A, y cada ve que alcanzamos una región $X \neq A$ continuamos nuestra ruta abandonando la región X por alguno de los puentes no utilizados que conducen a X (la existencia de tal puente está garantizada, porque el número de puentes que conducen a X es par). Consecuentemente, la ruta R debe alguna vez retorna a A. Removamos mentalmente

los puentes recorridos por la ruta R y observemos que la configuración así obtenida satisface la condición primera.

De estas observaciones puede concluirse que es posible construir una familia de rutas cerradas que no comparten puente alguno en común y que contienen a todos los puentes de la configuración inicial. Notemos que si dos rutas de este tipo pasan por una región común, entonces uno fácilmente puede construir una ruta compuesta entre ellas.

En segunda instancia, supongamos que el segundo punto punto del teorema de Euler es verdadero. Entonces, del Corolario, se obtiene que hay exactamente dos regiones hacia las cuales llegan un número impar de puentes. Mentalmente construyamos un nuevo puente T entre estas dos regiones. Entonces, la nueva configuración de aguas y puentes así formada satisface el primer punto del teorema. Sea R una ruta que recorra todos los puentes de la nueva configuración y que pase exactamente una vez por cada puente. A partir de R, podemos construir una ruta $R^{'}$ que recorra la nueva configuración de tal modo que el último puente que R atraviese sea T. Esta ruta $R^{'}$ recorrida en la configuración inicial (sin el puente T) satisface las condiciones requeridas **Q.E.D.**

5.8 Recorrido de grafos

Dado un grafo $G(V,E)$ con vértices V y lasdos E, un recorrido puede tener entre otros, los siguientes propósitos

- Buscar un elemento entre los vértices del grafo G
- Generar, a partir de un nodo particular un árbol de expansión.
- Encontrar trayectorias de costo mínimo dentro del grafo

Para resolver los dos primeros problemas hay dos estrategias: búsqueda en amplitud y búsqueda en profundidad. Para el tercer problema se tiene el algoritmo de Dijkstra.

La búsqueda en amplitud se basa en lo siguiente

- Se empieza en un nodo inicial, raíz del árbol de expansión
- Se tiene un conjunto para incluir los nodos visitados para expandirlos, de manera que se procesen una sola vez
- Se tiene una lista de espera con política FIFO para colocar a los nodos que se van a procesar o expandir.

Para la búsqueda en profundidad todo es igual, excepto el tercer punto, donde la lista de espera se maneja con una política LIFO. En el caso de la búsqueda en profundidad la lista de espera se puede sustituir por llamadas recursivas. Por ejemplo para el ábol empezando el nodo A. el recorrido en amplitud es ABCDEFG es decir, primero los hijos de A y luego sus nietos. Para el caso en profundidad, tenemos que el recorrido es ABDECFG, es decir, primero se recorre la rama izquierda del árbol y luego la derecha.

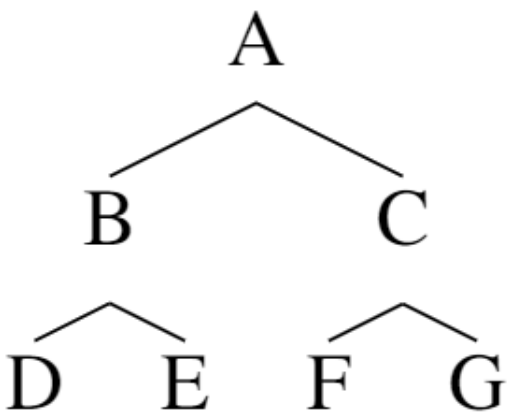


Figure 5.15: Recorrido de grafos

5.9 Algoritmo de Dijkstra

El algoritmo de Dijkstra para encontrar la ruta más corta entre dos nodos en un grafo conexo obedece a la siguiente explicación. Sea G(V,E) un grafo conexo y dos nodos $u, v \in V$ decimos que u es el vértice origen y v el nodo destino de la ruta, en general G es un grafo no dirigido con pesos, en caso de que no tenga pesos en forma explícita se asumen pesos unitarios.

La trayectoria de costo mínimo se va construyendo de manera iterativa asignando los costos de la forma siguiente:

- $costo(u) = 0$
- Si el costo $costo(x)$ del nodo x es mínimo, se calcula el costo mínimo de un nodo vecino y de x de la manera siguiente

$$costo(y) = min(costo(x) + distancia(x,y), costoprevio(y)) \tag{5.10}$$

  donde la distancia (x,y) es el costo de ir de x a su vecino "y" y este valor es fijo determinado de entrada en el grafo y costoprevio(y) es un costo mínimo provicional calculado anteriormente.

Después de aplicar iterativamente el punto 2, se llega al costo de v que es el $costo(v)$ mínimo requerido y una lista de nodos recorridos que se tiene que invertir para encontrar la trayectoria de costo mínimo desde un nodo inicial al nodo final v.

6 Otros temas

6.1 Binomios

6.1.1 Binomio de Newton

Una expresión de la forma $a+b$ es llamado un binomio. Aunque, en principio puede parecer sencillo elevar $a+b$ a cualquier potencia, elevar este binomio a una potencia muy alta, puede resultar una tarea tediosa. Existe una fórmula para la expansión de $(a+b)^n$ para cualquier número natural n. Para encontrar esta fórmula, se presentan los siguientes modelos para algunas potencias sucesivas de $a+b$.

$$
\begin{aligned}
(a+b)^1 &= a+b \\
(a+b)^2 &= a^2+2ab+b^2 \\
(a+b)^3 &= a^3+3a^2b+3ab^2+b^3 \\
(a+b)^4 &= a^4+4a^3b+6a^2b^2+4ab^3+b^4 \\
(a+b)^5 &= a^5+5a^4b+10a^3b^2+10a^2b^3+5ab^4+b^5
\end{aligned}
$$

Los coeficientes de cada polinomio resultante se rigen bajo la secuencia del triángulo de Pascal:

$$
\begin{array}{lccccccccc}
n=0 & & & & & 1 & & & & \\
n=1 & & & & 1 & & 1 & & & \\
n=2 & & & 1 & & 2 & & 1 & & \\
n=3 & & 1 & & 3 & & 3 & & 1 & \\
n=4 & 1 & & 4 & & 6 & & 4 & & 1 \\
n=5 & 1 & & 5 & & 10 & & 10 & & 5 \quad 1
\end{array}
$$

Para definir un número combinatorio es preciso conocer con anterioridad lo que es

el factorial de un número $(n!)$ que es definido como:

$0! = 1$ (caso base)
$n! = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 3 \cdot 2 \cdot 1$ (caso inductivo)

Un número combinatorio es un número natural de la forma $\binom{n}{k}$, donde $n \geq m$ y se lee "n sobre m". La formula es la siguiente:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Con base a la definición del numero combinatorio, el triángulo de Pascal cambia a la siguiente configuración:

$$\begin{array}{lccccccccccc}
n=0 & & & & & \binom{0}{0} & & & & & & \\
n=1 & & & & \binom{1}{0} & & \binom{1}{1} & & & & & \\
n=2 & & & \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} & & & & \\
n=3 & & \binom{3}{0} & & \binom{3}{1} & & \binom{3}{2} & & \binom{3}{3} & & & \\
n=4 & \binom{4}{0} & & \binom{4}{1} & & \binom{4}{2} & & \binom{4}{3} & & \binom{4}{4} & & \\
n=5 & \binom{5}{0} & & \binom{5}{1} & & \binom{5}{2} & & \binom{5}{3} & & \binom{5}{4} & & \binom{5}{5}
\end{array}$$

Y teniendo en cuenta lo anterior, se realiza la generalización el desarrollo de la potencia de un binomio a un exponente natural cualquiera, el cual es conocido como el Binomio de Newton

$$(a+b)^n = \binom{n}{0}a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \ldots + \binom{n}{n-2}a^2b^{n-2} + \binom{n}{n-1}ab^{n-1} + \binom{n}{n}b^n$$

Si se toma la función $f(x) = (x+b)^n$, donde $x, b \in \mathbb{R}, n \in \mathbb{N}$. Desarrollando en series de potencias se tiene, para $X_0 = 0$

$$f(x) = \sum_{R=0}^{\infty} \frac{f^{(R)}(0)}{R!} X^R$$

como $f(x)$ es un polinomio de orden n, su R-ésima derivada para R > n es la función constante cero, es decir, $f^{(R)}(x) = 0$ para $R < n$. entonces la suma infinita anterior se vuelve finita.

$$f(x) = \sum_{R=0}^{n} \frac{f^{(R)}(0)}{R!} X^R, pero f^{(1)}(x) = n(x+b)^{n-1}, f^1(x) = nb^{n-1}$$

$$f^{(2)}(x) = n(n-1)(x+b)^{n-2}, f^2(0) = n(n-1)b^{n-2} f^R(0) = n(n-1)...(n-(n+1))$$

en general

$$f^R(0) = n(n-1)...(n-R+1)$$

Y para el caso de $R = n$

$$f^{(n)}(0) = n(n-1)...(n-R+1)b^{n-n} = n!$$

Finalmente

$$(x+b)^n = \sum_{R=0}^{n} \frac{n(n-1)...(n-R+1)}{R!} X^R b^{n-R} = \sum_{R=0}^{n} \binom{n}{R} X^R b^{n-R}$$

Luego se realiza el reemplazo de x por a $x = a$

$$(a+b)^n = \sum_{R=0}^{n} \binom{n}{R} a^R b^{n-R}$$

que corresponde al binomio de Newton conocido si n=r, con $r \in \mathbb{R}$, por lo que la suma nunca termina.

6.1.2 Tarea de la sección

Realizando uso del binomio extendio, calcular $\sqrt{2}$, haciendo $r = 1/1$ y $a = b = 1$.

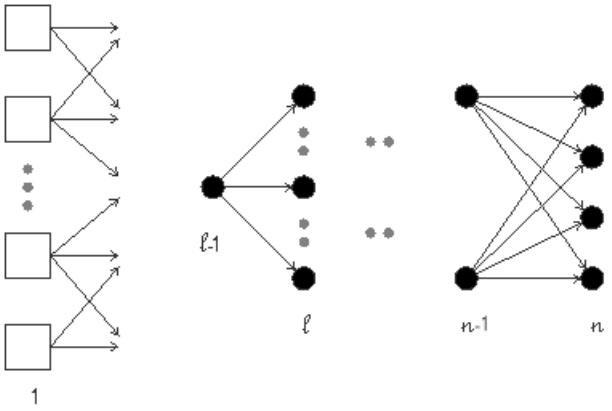
Realizar el procedimiento de manera manual y posteriormente implementarlo en Python.

6.2 Algoritmo de aprendizaje de propagación hacia atrás

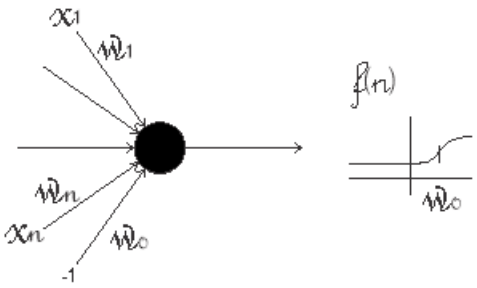
Considere un conjunto de datos

$$D = \left\{ \left(x^i,\ d^1 \right) :\ x^i \in \mathbb{R}^n \ patrón\ de\ entrada\ y\ d\ \mathbb{R}^m \ de\ salida\ deseada \right\}$$

Una red neuronal como la que sigue



En este grafo cada vértice es un perceptrón del tipo



$$y = p(n) \qquad u = \sum_{i=1}^{n} w_i x_i$$

La capa 1 es la capa de entrada y la capa n es la capa de salida, todas las demás capas se conocen como capas ocultas.

El algoritmo de entrenamiento consiste en calcular el gradiente del error $\frac{\partial E}{\partial w_{iJ}}$ respecto a los pesos sinápticos y actualizarlos en menos la dirección del gradiente para minimizar el error $\omega_{iJ} = \omega_{iJ} - \eta \frac{\partial E}{\partial \omega_{iJ}}$

Para el cálculo del gradiente se usa la regla de la cadena y el hecho de que $\frac{\partial E}{\partial \omega_{iJ}} =$

$\frac{\partial E}{\partial v_i} \bullet \frac{\partial u_i}{\partial \omega_{iJ}} = \frac{\partial E}{\partial v_i}$ haciendo $\delta = \frac{\partial E}{\partial u_i}$ gradiente local.

$\frac{\partial E}{\partial \omega_{iJ}} = \delta_i x_J$ es decir, teniendo el gradiente local, es fácil obtener las parciales del gradiente.

Considere una neurona J en la capa l-1 entonces por la regla de la cadena con las R neuronas en la capa l

$\frac{\partial E}{\partial u_J} = \delta_J = \sum_R \frac{\partial E}{\partial u_R} \frac{\partial u_R}{\partial u_J} = \sum_R \delta_R \frac{\partial u_R}{\partial y_J} \frac{\partial y_J}{\partial u_J} = f(u_j) \sum_R u_{RJ} \delta_R$ para una neurona J en la capa de n-1 tenemos

$$\frac{\partial E}{\partial u_J} = \frac{\partial}{\partial u_J} \left(\sum_R \frac{1}{2} (d_R - y_R)^2 \right) = -(d_J - y_J) \frac{\partial y_J}{\partial v_J} \frac{\partial E}{\partial v_J} = -(d_J - y_J) f'(v_J) = \delta_J^{\circ} f'(v_J) \text{ con } \delta_J^{\circ} = -(d_J - y_J)$$

Esta última ecuación nos permite iniciar los gradientes locales y la anterior $\delta_J = f'(R) \sum_R \omega_{RJ} S_R$ calcularlos de manera recurrente y a partir de ellos calcular el gradiente del error, para cada pareja (x, d) se actualizan todos los pesos sinápticos $_{iJ}$ hasta llegar a un óptimo local, típicamente se inicializan los pesos sinápticos con valores aleatorios pequeños.

6.3 Redes neuronales y aprendizaje profundo

La evolución de las matemáticas se puede revisar a partir de la evolución de la representación y el concepto de función, desde las funciones elementales dadas por la aplicación de operaciones aritméticas básicas, junto con trigonométricas, exponenciales y sus inversas.

En la actualidad se cuenta con una definición bastante general de una función donde lo más importante es la regla de la correspondencia dada entre dos conjuntos llamados dominio y contra dominio.

En este contexto las redes neuronales son una forma de construir funciones, que biológicamente representan trabajos cognitivos. En el caso de las redes neuronales, las funciones no se programan, se entrenan en forma supervisada a partir de a partir de datos en forma de pareja $(x_i; y_i)$ de entrada, salida.

El entrenamiento se lleva a cabo mediante un proceso de aproximación, partiendo de un valor inicial aleatorio, de los parámetros de la red neuronal. Se usa el gradiente de una función de error, como un criterio para aproximar loa solución, que en este caso busca ser la función buscada.

En el aprendizaje de maquina el descriptor del objetivo o señal a reconocer o predecir, se calculaba aparte y luego se le aplicaba la función cognitiva. En el caso del aprendizaje profundo el cálculo del descriptor a patrón se calcula dentro del mismo proceso de entrenamiento de la red neuronal, logrando mejores desempeños de las funciones cognitivas.

Tanto el caso biológico como tecnológico, las funciones cognitivas agrupan y organizan la información que capturan los sensores y sentidos, en conceptos y clases.

Si bien las funciones cognitivas no se programan, si se programa el proceso de entrenamiento.

6.3.1 Programación vs Aprendizaje Automático

Hay una clara diferencia entre programación tradicional y Machine Learning, mientras en la programación tradicional se trata de programar cada aspecto y tener bien definida tanto la solución como el método de solución, en el caso del machine learning, se enfoca más en la solución y los datos de entrada, dejando que la computadora se encargue de encontrar la función que resuelva el problema.

Si bien hay procesos que son bastante faciales de programar, debido a la sencillez del problema hay muchos otros que son bastante más difíciles o incluso imposible de programar de forma tradicional y aun así en las universidades se sigue enseñando ese método que consiste en "Decirle a la computadora lo que debe buscar, y que debe hacer, debes enumerar cada condición y definir cada acción" (Perrotta, 2020, P.4).

Y esto puede ser muy difícil, hay que estar atentos en cada detalle, cada posible error, cada etapa y secuencia del funcionamiento del programa, esto claro si ya se sabe a la perfección que es lo que hay que hacer, pero si a lo anterior hay que sumarle que debemos resolver un nuevo problema del cual no sabemos como resolverlo, por diversas razones o porque su complejidad es demasiado alta para ser programado, entonces no parece tan mala idea dejar que la computadora haga el trabajo difícil.

El siguiente ejemplo es un programa sencillo que servirá tanto para demostrar como funciona la red neuronal, como para indicar en que consiste cada parte de la red neuronal.

Problema: Crear un programa que convierta de grados Centígrados a grados Fahrenheit

Suponiendo que se desconoce la fórmula para convertir de grados Centígrados a Fahrenheit, puesto que no soy médico, ni biólogo, ni químico y vivo debajo de una piedra donde no llega el internet, no se como convertir de una unidad a otra, pero si sabemos programar, veamos como quedaría el programa.

Librerías.

Hay muchas librerías que se pueden ocupar y depende del programa que se quiera realizar, en este caso se ocuparán

```
import tensorflow as tf
import numpy as np
import matplotlib.pyplot as plt
```

Tensorflow: sirve para para crear las redes neuronales (hay otras librerías).
Numpay: maneja arrays, hace que el programa sea más eficiente
Matplotlib: es para crear gráficos.

Datos de entrenamiento.

Es fundamental contar con suficientes datos reales para entrenar la red neuronal, entre más datos se tengan mejor, hay maneras de solventar esto, pero es mejor contar con la mayor cantidad de datos reales posibles.

```
C=np.array  ([-40,  0, 15,   21,    38, 100, 45,     71,   119,  50],dtype=float)
F=np.array  ([-40, 32, 59, 69.8, 100.4, 212, 113, 159.8, 246.2, 122],dtype=float)
```

Aquí se ocuparon extremadamente pocos datos, por lo que puede ser un problema, pero se mitigara esta falta de datos más adelante.

Red neuronal, modelo y función de activación.

En este paso se crea la red neuronal, para este ejemplo es una neurona de entrada y una de salida, el modelo va a ser secuencial, útil cuando tienes una única entrada y una única salida, para este ejemplo no se ocupará función de activación (por su sencillez

de), pero casi siempre se ocupará alguna función de activación.

```
capa=tf.keras.layers.Dense(units=1, input_shape=[1])
modelo=tf.keras.Sequential([capa])
```

“Dense” significa que todas las neuronas estan conectadas con todas, “units” es el número de neuronas para datos de entrada, “input shape” es el número de neuronas para datos de salida, dentro del argumento de “tf.keras.Sequential”, se ponen las capas ocultas y su función de activación.

Compilar.

En esta etapa le decimos a la red neuronal como debe buscar la solución, establecemos los parámetros del descenso del gradiente, para encontrar el mínimo local y la función para minimizar el error.

```
modelo.compile(
               optimizer=tf.keras.optimizers.Adam(0.1),
               loss="mean_squared_error"
             )
```

El optimizador ocupado es Adam, que de hecho es la mezcla de otros dos optimizadores y la función de error es “mean squared error” que en términos más mundanos quiere decir que es mejor tener algunos errores pequeños que pocos errores muy grandes.

Entrenamiento.

En esta etapa comienza el entrenamiento de la red neuronal, como parámetros de entrada tenemos los datos de entrenamiento que para este caso son los grados Centígrados y los grados Fahrenheit, como el array de entrada tenía muy pocos valores, usamos “epochs” para que vuelva a revisar el array de entrada.

```
entrenado=modelo.fit(C,F,epochs=1500)
```

Para este ejemplo se contaba con sólo 10 datos de entrenamiento, un número extremadamente bajo, pero gracias a “epochs” se va a repetir ese array de entrada la cantidad que deseemos para simular que tenemos más datos y el entrenamiento se haga de la mejor manera.

Probar la red neuronal

Para finalizar se hace una prueba de la red neuronal para verificar que funciona correctamente.

```
resultado = modelo.predict([17])
```

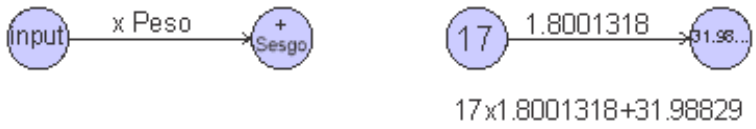
En este caso se ocupa un valor que no esté dentro de nuestro array de entrada y corroboramos.
Pero que fue lo Hace el programa de forma interna, como se puede estar seguro de que funciona correctamente, se va a comparar los valores iniciales de la red neuronal con los valores finales y con la fórmula para calcular los grados Fahrenheit.

Se obtendrán valores del peso y del sesgo, para diferentes epochs, y se compararan con el valor real y la fórmula para convertir a grados Fahrenheit, y así poder visualizar cual es el comportamiento de la red neuronal.

```
print (capa.get_weights())
```

Épocas	Peso	Sesgo	Error	Resultado para 17˚C
1	0.29470003	0.10000001	14976.2441	5.1099005
500	1.8812783	24.785013	27.8823	56.766747
1000	1.8019698	31.82491	0.0166	62.458397
1500	1.8001318	31.98829	0.00000741	62.59053

Como podemos ver el error es muy grande al principio, y el peso y sesgo que son asignados van cambiando a lo largo del programa, pero que son realmente esos valores y para qué sirven, lo que hace nuestra neurona es tomar el valor que va a convertir, lo multiplica por el peso, después suma el sesgo y entrega el resultado.



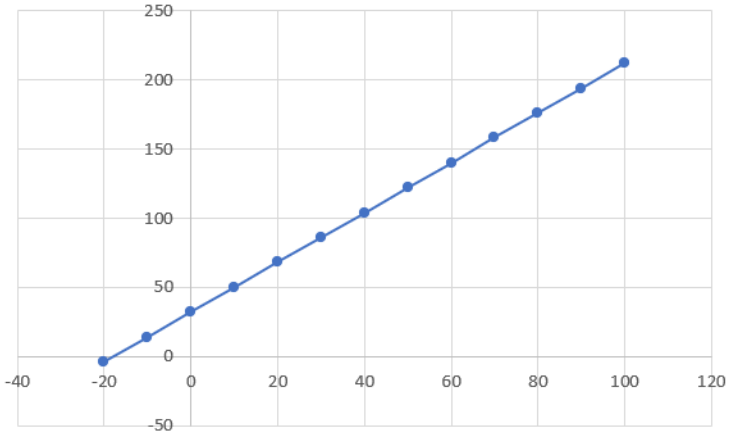
Entonces tenemos que

$$℃ \bullet peso + sesgo = ℉$$

Y ahora veamos la fórmula para convertir entre unidades $℉ = ℃ \cdot 1.8 + 32$ como se puede observar el peso y el sesgo, son prácticamente iguales.

Este problema podrá parecer muy sencillo e incluso se podrá pensar (acertadamente) que es mucho más fácil y rápido programar la formula y que entregue un resultado exacto, pero hay que tener en cuenta que muchos de los problemas a los que nos enfrentaremos, desconocemos como es su comportamiento y en muchos de los casos no sabremos si existe un modelo matemático que encaje a la perfección con el problema que debemos resolver, de ahí la importancia de aprender estas técnicas.

Analicemos ahora su comportamiento en una gráfica.

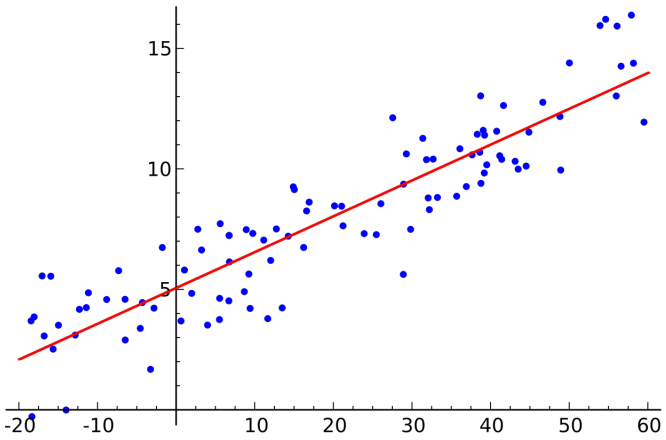


Como se observar en la gráfica de grados Celsius y de grados Fahrenheit se com-

porta de forma lineal, es por eso que este tipo de técnicas es muy buena para fenómenos que se comportan de forma lineal, ya que el modelo matemático de una neurona es similar a la ecuación de la recta $y = mx + b$ donde m es el peso y b es el sesgo, pero que pasa cuando un problema no se comporta de manera lineal, cuando no se puede ajustar una línea recta a una curva, para eso hay otras técnicas que se suman a esta primer neurona.

En el caso de los para convertir grados es fácil ver su comportamiento, pero no siempre es así, hay que ver los datos del problema que queremos resolver y de ser posible hacer una gráfica, para ver su comportamiento casi siempre sus gráficos tendrán varios puntos dispersos, nuestra misión será ver que tipo de recta se acerca mas a esos datos, por poner algunos ejemplos existen varios tipos de regresiones lineales, como regresión lineal.

Esta es la función más sencilla, ajusta los datos a una línea recta, también es importante recordada que difícilmente nos encontraremos con este tipo de problemas.



Para ocupar regresión lineal se ocupa el siguiente algoritmo, no olviden que el tamaño importa y que entre mayor sea su base de datos, mejor.

```
from sklearn import linear_model
x_entrenamiento = variablesIndependientes
y_entrenamiento = VariablesDependientes
algoritmo=linear_model.LinearRegression()
algoritmo.fit(x_entrenamiento,y_entrenamiento)
algoritmo.predict(x_prueba)
```

Hay otros tipos de regresiones, son las regresiones no lineales, una muy utilizada es la regresión logística, pero deben ocupar la que mas se parezca a sus datos, ahora hay que tomar en cuenta que las neuronas resuelven muy bien los problemas que son de tipo lineal, para otro tipo de problemas donde los datos tienen otro tipo de forma va a ser casi imposible que se puedan ajustar a una línea recta, para ello se ocupan las funciones de activación.

6.4 Modelos Lineales.

Un modelo es una representación de un objeto o fenómeno, existen dos tipos de modelos, estáticos y dinámicos. Un modelo estático solo representa la forma pero no el funcionamiento del modelado. El modelo dinámico por lo contrario algunos de los elementos que intervienen en la modelización no permanecen invariables, sino que se consideran como funciones del tiempo, describiendo trayectorias temporales. En el caso de los problemas científicos o tecnológicos, es aún más importante representar la función que la forma, aunque es deseable tener las dos representaciones, o una forma equivalente.

El caso que nos interesa son los modelos matemáticos que representan la forma externa del fenómeno, su estructura interna, la forma en que se relaciona y dependen de sus variables descriptivas.

Algo interesante de este caso, con el conocimiento de estas variables se pueden visualizar aspectos de forma del fenómeno. Los modelos matemáticos pueden ser continuos, discretos, lineales, no lineales o probabilístivos y normalmente corresponden al planteamiento de un problema, donde la solución consiste de aspectos del fenómeno que se está investigando.

Los modelos matemáticos continuos corresponden usualmente a :

- Ecuaciones Diferenciales Ordinarias.
- Ecuaciones Diferenciales Parciales.

Los modelos discretos se relacionan con :

- Sistema de Ecuaciones.
- Ecuaciones en Diferencias.

En los modelos probabilíticos se usan en:

- Técnicas de Simulación.
- Métodos de Montecarlo.

6.4.1 Modelo Matemático.

El modelo matemático se construye, tomando los aspectos escenciales del fenómeno a considerar y haciendo a un lado lo que no es de interes o que tiene poca relevancia para lo que se pretende estudiar.

Para construir un modelo matemático se necesita primero un marco conceptual formal del área a la que pertenece. Una vez que se tiene el modelo se obtiene los algoritmos relacionados con su análisis y solución. Finalmente estos algoritmos se programan con algún lenguaje de programación.

Problema.

Encontrar el máximo común divisor (a,b) de dos números enteros $b < a$ y los enteros m y n tales que $ma+nb = (a,b)$.

1. Partimos del siguiente echo si b divide a entonces $(a,b) = b$.
2. Si b no divide a, entonces $a = q_1$, $b = n_1$ y $(a,b) = (b,r_1)$. Entonces si r_1 divide b terminamos con $(a,b) = (b,y_1) = y_1$.
3. Si y_1 no divide b entonces esto se puede repetir, $b = q_2y_1 + y_2$, $(a,b) = (b_1,y_1) = (y_1,y_2) = (y_{n-1},y_n) = y_r$, hasta que para algún y_r divida a y_{r-1} en cuyo caso $(a,b) = y_r$.

Estos 3 puntos nos dan el algoritmo de Euclides, que por primera vez aparece aproximadamente en el año 300 a.c en el libro *XIII* de los elementos de Euclides.

Código en Python.

Edgar Hernan Rosas Espinosa

```
def euclides(a,b): #b<a son los enteros iniciales
  if a%b==0: # si b divide a
    return b #regresa b
  else: # caso contrario
    return euclides(b,a%b) # se aplica la recursividad e intenda con b
```

Para hayar los enteros m y n de $ma+nb=(a,b)$, vamos a usar el modelo matricial de la iteración de Euclides, llamando al algoritmo que nos permite obtener m y n *Euclides extendido*.

Sea $b<a$ dos números enteros, hacemos $a=s_0$ y $b=t_0$. La primera iteración de Euclides la podemos representar como:

$$\begin{pmatrix} s_1 \\ t_1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & -Q_1 \end{pmatrix} * \begin{pmatrix} s_0 \\ t_0 \end{pmatrix}, \quad Q_1 = s_0//t_0, \quad s_0 = Q_1 t_0 + y_0. \tag{6.1}$$

$y_0 = s_0 - Q_1 t_0$, en este caso $t_1 = Y_0$, de esta misma forma podemos seguir iterando hasta que último residuo sea igual a cero $t_n = 0$, si esto ocurre $s_n = (a,b)$.

$$\begin{pmatrix} (a,b) \\ 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & -Q_n \end{pmatrix} * \begin{pmatrix} 0 & 1 \\ 1 & -Q_{n-1} \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ 1 & -Q_1 \end{pmatrix} * \begin{pmatrix} a \\ b \end{pmatrix} \tag{6.2}$$

Multiplicando todas las matrices podemos agruparlas en una sola

$$\begin{pmatrix} (a,b) \\ 0 \end{pmatrix} = \begin{pmatrix} m & n \\ f & g \end{pmatrix} * \begin{pmatrix} a \\ b \end{pmatrix}, \quad \text{en particular} (a,b) = ma+nb. \tag{6.3}$$

y esto último es precisamente lo que andábamos buscando.

Para encontrar m y n tenemos que multiplicar todas las matrices, esto lo podemos desarrollar de manera recursiva procediendo de izquierda a derecha y tomando en cuenta que solo nos interesa el primer renglón de la matriz producto. Para ejecutar el producto recursivamente procedemos lo siguiente:

1. En el caso base tenemos la última matriz y regresamos el primer renglón $[0,1]$.
2. En el paso inductivo 1, se recibe el reglón $[A,B]$ y con este multiplicando por ña matriz de la iteración 1 de Euclides, se calcula el nuevo primero renglón.

$$\begin{bmatrix} A & B \end{bmatrix} * \begin{pmatrix} 0 & 1 \\ 1 & -Q_1 \end{pmatrix} = \begin{bmatrix} B, A - Q_1 B \end{bmatrix} \tag{6.4}$$

3. Este paso inductivo se aplica de manera recursiva hasta terminar con la primera iteración de Euclides.

Esto en código python queda de la siguiente manera:

```
def euclidex(a,b): # b<a son los enteros de entrada.
  if a%b==0: # caso base
    return 0,1,b # Regresa (a,b)
  else:
    A,B,r=euclidex(b,a%b) #Paso Inductivo
    return B,A-(a//b)*B,r #Retorno del valor correspondiente.
```

6.5 Espacios Vectoriales Fundamentos de Modelos Lineales.

Desde un punto de vista físico, un vector es un objeto matemático con tres atributos: magnitud, dirección, sentido. Y los representamos con una flecha o un segmento de recta o vectorial. Desde el punto de vista algebraico lo más relevante son los vectores que se pueden sumar y se pueden multiplicar por un número o escalar. En resumen si denotamos el espacio vectorial por V y los escalares que son números reales son el simbolo $\mathbb{R}$, tenemos $a,b \in V$ y $\alpha,\beta \in \mathbb{R}$.

1. $\alpha a \in V \rightarrow$ Producto de un escalar por un vector.
2. $a+b \in V \rightarrow$ Suma de dos vectores.
3. $a+0=0+a=a \rightarrow$ Existencia del vector nulo.

4. $\alpha(a+b) = \alpha a + \alpha b, (\alpha+\beta)a = \alpha a + \beta a \rightarrow$ Ley distributiva.

5. $a + (-a) = 0 \rightarrow$ Existencia del inverso aditivo.

Definición.

Un conjunto de vectores $ai : i = 1, 2, \cdots, n = A$ es linealmente independiente si $\sum_{i=1}^{n} \alpha i ai = 0$, implica $\alpha_1 = \alpha_2 \cdots \alpha_m = 0$

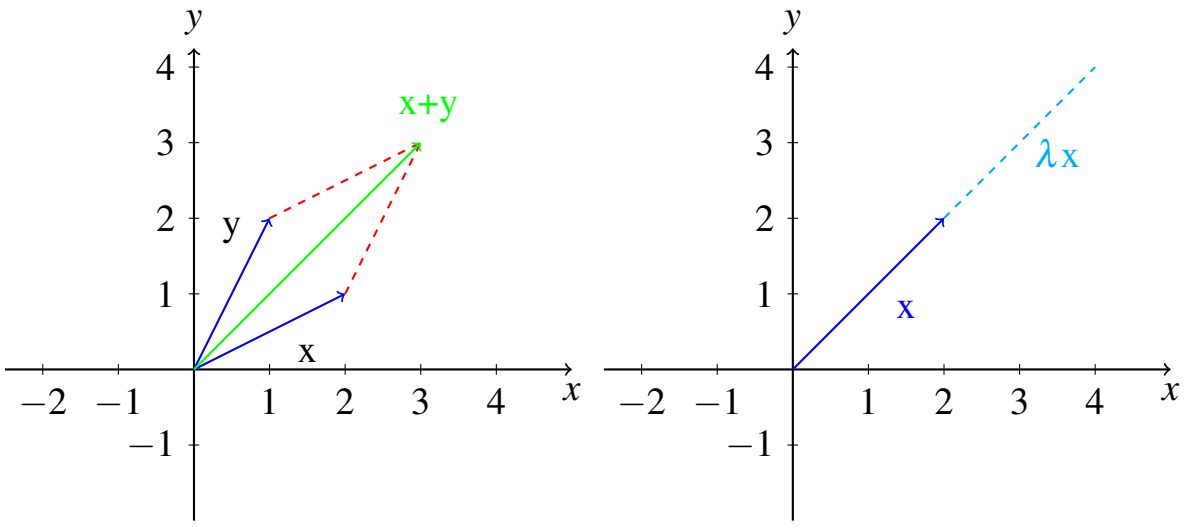
Esto quiere decir que ningún vector *ai* se puede representar como una combinación lineal de los demas.

Definición.

La dimensión de un espacio vectorial para el caso finito es igual al número máximo de vectores linealmente independiente que puede tener.

Un ejemplo importante de espacio vectorial es cuando se tiene un conjunto de tuplas $\mathbb{R}^n$ con n números reales. $\mathbb{R}^n$ en el espacio vectorial de n tuplas de números reales, conta de elementos como $x = (x_1, \cdots, x_n)$, $y = (y_1, \cdots, y_n)$ sumándose término a término.

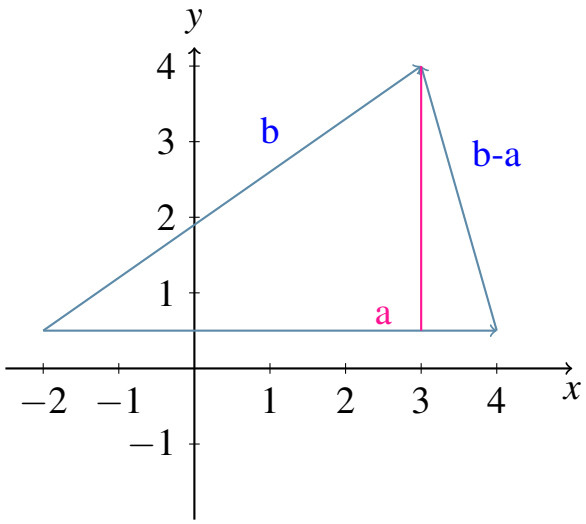
$x + y = (x_1 + y_1, \cdots, x_n + y_n)$ y la multiplicación de un número λ por un vector $x = (x_1. \cdots, x_n)$ esta dada por $\lambda x = (\lambda x_1, \cdots, \lambda x_n)$. Con estas dos operaciones se puede probar que $\mathbb{R}^n$ es un espacio vectorial y su representación geométrica de las operaciones son:



La suma es la diagonal principal del paralelogramo formado por los vectores (x,y) y λx es múltiplo de x_n. Generalizando el teorema de Pitagoras, la longitud de tamaño x es $|x^2| = \sum Xi^2$. Otra operación en particular de $\mathbb{R}^2$ es el producto de puntos o prodcutos interior.

Ejemplo.

Tomemos $a = (a_1, a_2, a_3)$, $b = (b_1, b_2, b_3)$ vectores en el espacio $\mathbb{R}^3$.



Por el teorema de Pitagoras.

$$
\begin{aligned}
|b-a|^2 &= (|b|sen\theta)^2 + (|a| - |b|cos\theta)^2. \\
|b-a|^2 &= |b|^2 sen^2\theta + |a|^2 + |b|^2 cos^2\theta - 2|a||b|Cos\theta. \\
|b-a|^2 &= |(b_1-a_1, b_2-a_2, b_3-a_3)|. \\
&= (b_1-a_1)^2 + (b_2-a_2)^2 + (b_3-a_3)^2. \\
&= b_1^2 + a_1^2 - 2a_1b_1 + b_2^2 + a_2^2 - 2a_2b_2 + b_3^2 + a_3^2 - 2a_3b_3. \\
&= |a|^2 + |b|^2 - 2(a_1b_1 + a_2b_2 + a_3b_3) \\
&= |a|^2 + |b|^2(Cos^2\theta + Sen^2\theta) - 2|a||b|Cos\theta.
\end{aligned} \tag{6.5}
$$

Finalmente.

$$(a_1b_1 + a_2b_2 + a_3b_3) = |a||b|Cos\theta$$

Y denotamos el producto punto de a y b como $a \cdot b = a_1b_1 + a_2b_2 + a_3b_3$.

6.5.1 Producto Vectorial.

El producto cruz o producto vectorial a x b es considerado un producto importante de vectores en el espacio $\mathbb{R}^3$, usando como referencia la figura anterior, su área es:

$$
\begin{aligned}
A^2 &= |a|^2|b|^2 Sen^2\theta. \\
&= |a|^2|b|^2(1 - Cos^2\theta). \\
&= |a|^2|b|^2 - |a|^2|b|^2 Cos^2\theta. \\
&= |a|^2|b|^2 - (a \cdot b)^2. \\
&= (a_1^2 + a_2^2 + a_3^2)(b_1^2 + b_2^2 + b_3^2) - (a_1b_1 + a_2b_2 + a_3b_3). \\
&= a_1^2b_1^2 + a_1^2b_2^2 + a_1^2b_3^2 + a_2^2b_1^2 + a_2^2b_2^2 + a_2^2b_3^2 + a_3^2b_1^2 + a_3^2b_2^2 + a_3^2b_3^2 - (a_1b_1 + a_2b_2 + a_3b_3)^2 \\
&- a_1^2b_1^2 - a_1a_2b_1b_2 - a_1a_3b_1b_3 - a_1a_2b_1b_2 - a_2^2b_2^2 - a_2a_3b_2b_3 - a_1a_3b_1b_3 - a_2a_3b_2b_3 - a_3^2b_3^2.
\end{aligned} \tag{6.6}
$$

Agrupando y eliminando terminos obtenemos:

$$A^2 = (a_2b_3 - b_2a_3)^2 + (a_1b_3 - b_1a_3)^2 + (a_1b_2 - b_1a_2)^2. \tag{6.7}$$

La ecuación anterior se puede definir como a x $b = ([a_2b_3 - b_2a_3], [a_1b_3 - b_1a_3], [a_1b_2 - b_1a_2], [b_1a_3 - a_1b_3])$, por lo tanto podemos definir A^2 como:

$$\begin{aligned} A^2 &= |a \ x \ b|^2. \\ A &= |a \, x \, b|. \end{aligned} \tag{6.8}$$

Otra opción de calcular $a \, x \, b$ es desarrollando el determinante.

$$\begin{aligned} a \, x \, b &= \begin{bmatrix} i & j & k \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{bmatrix} \\ &= \begin{bmatrix} a_2 & a_3 \\ b_2 & b_3 \end{bmatrix} i - \begin{bmatrix} a_1 & a_3 \\ b_1 & b_3 \end{bmatrix} j + \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix} k \\ &= \begin{bmatrix} a_2 & a_3 \\ b_2 & b_3 \end{bmatrix}, \begin{bmatrix} a_1 & a_3 \\ b_1 & b_3 \end{bmatrix}, \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix} \\ &= ([a_2b_3 - b_2a_3], [b_1b_3 - a_1b_3], [a_1b_2 - b_1a_2]). \end{aligned} \tag{6.9}$$

En conclusión la operación $a \, x \, b$ proporciona un vector perpendicular al plano generado por a y b cuya magnitud $|a \, x \, b|$ es igual al área del paralelo al plano designado por a y b.

6.5.2 Espacios Vectoriales.

Definición.

Sean V y W dos espacios vectoriales de un mapeo A entre ellos $A : V \to W$ es lineal si, $A(\alpha v + \beta u) = \alpha Av + \beta Au$, para $\alpha, \beta \in \mathbb{R}$ y $v, u \in V$.

Si $B = B_1, \cdots, B_k$ es un conjunto de vectores linealmente, entonces sus combinaciones lineales son únicos.

Supongamos que el vector b es generado de dos formas diferentes por los vectores del conjunto B, es decir $b = \sum_{i=1}^{k} b_i B^i$ y $b = \sum_{i=1}^{k} b_i' B_i$, restando $\sum_{i=1}^{k} b_i B^i - \sum b_i' B_i = 0$,

factorizando $\sum_{i=1}^{k}(b_i^{'} - b_i)B_i = 0$, pero los B_i son linealmente independiente por lo tanto $b_i^i = b_i$. Debidp a que $b_i \in \mathbb{R}$ son únicos les llamamos las coordenas de b respecto a B.

Definición.

Decimos que $B = B_1, \cdots, B_n$ es una base de V en el espacio conjunto de vectores linealmente independiente vectorial.

- Los elementos se pueden representar como una combinación lineal de elementos.
- En un caso finito, la dimensión de un espacio es igual a el número de elementos de cualquier de sus bases.

Definición.

Si un vector A es una transformación lineal $A : V \to W$, y $B = B_1, \cdots, B_n$ es base de V entonces $a \in V$.

Si $a = a_1B_1 + \cdots + a_nB_n$ y $Aa = a_1AB_1 + a_nAB_n \to (AB_1, \cdots, AB_n)(a_1, \cdots, a_n)$. La matriz $AB_1, \cdots, AB_n$ se le conoce como matriz de la transformación A respecto a la base B. En otras palabras a cada transformación lineal se le puede asociar una matriz que la represente en terminos de una base dada.

6.6 Matrices

Definición.

Una matriz es un arreglo rectangular de números reales.

$$A_{ij} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{pmatrix} \tag{6.10}$$

- Los números que están agrupados por renglones se denotan como *Ai*.
- Las columnas se denotan como *aj*.
- El elemento conjunto en el renglón *i* columna *j* se indica como **Aij**.
- La matriz de *m* renglones y *n* columnas se denota como *Amxn*.

Definición.

Si $\alpha \in \mathbb{R}$ es un número y A una matriz, el producto de αA es la matriz con elementos αAij.

$$2 \cdot \begin{pmatrix} 2 & 0 & 1 \\ 3 & 0 & 0 \\ 5 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 2\cdot 2 & 2\cdot 0 & 2\cdot 1 \\ 2\cdot 3 & 2\cdot 0 & 2\cdot 0 \\ 2\cdot 5 & 2\cdot 1 & 2\cdot 1 \end{pmatrix} = \begin{pmatrix} 4 & 0 & 2 \\ 6 & 0 & 0 \\ 10 & 2 & 2 \end{pmatrix} \tag{6.11}$$

- α multipla a cada elemnto de A.

Definición.

Si A y B son matrices del mismo tamaño, entonces $[A+B]_{ij} = a_{ij} + b_{ij}$.

$$\begin{aligned} A+B &= \begin{pmatrix} 3 & 7 & -1 \\ 2 & -5 & 5 \\ 7 & -8 & 1 \end{pmatrix} + \begin{pmatrix} -7 & 2 & -5 \\ -1 & 2 & 6 \\ -4 & 3 & 2 \end{pmatrix} \\ &= \begin{pmatrix} 3-7 & 7+2 & -1-5 \\ 2-1 & -5+2 & 5+6 \\ 7-4 & -8+3 & 1+2 \end{pmatrix} \\ &= \begin{pmatrix} -4 & 9 & -6 \\ 1 & -3 & 11 \\ 3 & -5 & 3 \end{pmatrix} \end{aligned} \tag{6.12}$$

- Si las matrices son iguales se suman elemento a elemento.

Definición.

Si una matriz A esta compuesta de números complejos, la matriz conjugada $\overline{A}$ tiene como elementos $\overline{a_{ij}}$.

$$A = \begin{pmatrix} 1+i & -2i \\ 3-5i & 4 \end{pmatrix} \rightarrow \overline{A} = \begin{pmatrix} 1-i & 2i \\ 3+5i & 4 \end{pmatrix} \tag{6.13}$$

- En una matriz conjugada se conjuga elemento a elemento.

Definición.

La transposición de una matriz A_{mxn} se denota como A^t y su tamaño de mxn se modifica, tal que $A^t_{ij} = A_{ji}$

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \rightarrow A^T = \begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix} \tag{6.14}$$

- Las filas de una matriz A^t va a ser igual a las columnas de la matriz A, mientras que las columnas de la matriz A^t son las filas de la matriz A.
- Si $A = A^t$ decimos que la matriz A es simétrica y cuadrada.
- También podemos demostrar que $(AB)^t = B^t A^t$.

Definición.

El producto de una fila $a^t = [a_1, \cdots, a_n]$ y una columna $b = [b_1, \cdots, b_n]$, esta dado por el producto punto $a^t b = a_1 b_1 + \cdots + a_n b_n$.

$$
\begin{aligned}
A * B^T &= \begin{pmatrix} 5 & 12 & -3 \\ 0 & -13 & 21 \\ 4 & 72 & -17 \end{pmatrix} * \begin{pmatrix} 7 & -10 & 3 \\ 15 & 0 & 42 \\ 31 & -14 & 3 \end{pmatrix} \\
&= \begin{pmatrix} 5 \cdot 7 & 12 \cdot (-10) & (-3) \cdot 3 \\ 0 \cdot 15 & (-13) \cdot 0 & 21 \cdot 42 \\ 4 \cdot 31 & 72 \cdot (-14) & (-17) \cdot (3) \end{pmatrix} \\
&= \begin{pmatrix} 35 & -120 & -9 \\ 0 & 0 & 882 \\ 124 & -1008 & -51 \end{pmatrix}
\end{aligned} \tag{6.15}
$$

6.7 Sistema de Ecuaciones.

Definición.

Un sistema de m ecuaciones y n incógnitas, es un conjunto finito de ecuaciones en las que se buscan soluciones comunes a resolver.

Una sistema de ecuaciones se puede representar de forma escalar, donde x_i son las incógnitas y (a_{ij},b_i) los datos como lo muestra la siguiente ecuación.

$$
\begin{aligned}
a_{11}x_1 + \cdots + a_{in}x_n &= b_i \\
&\vdots \\
a_{m1}x_1 + \cdots + a_{mn}x_n &= b_n
\end{aligned} \tag{6.16}
$$

También se puede representar de forma vectorial, como la combinación lineal de vectores columna.

$$
X_1 \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} \tag{6.17}
$$

Una representación común de un sistema de ecuaciones es de forma matricial:

$$Ax = b, \;\; A = (A_1, \cdots, A_n), \;\; X = (x_1, \cdots, x_n)^t \tag{6.18}$$

En el caso de la forma matricial A respresenta una transformación lineal y se busca el vector x que bajo A se mapea en b. Si las columnas de A son linealmente independientes y b se encuentra en el sub espacio generado por las columnas de A, existiendo una solución única.

En el caso de que b no este en el espacio generado por las columnas de A, se busca obtener b_*, que es la proyección ortogonal de b sobre s, $(b - b_*/s)$. Esto da origen a un problema de optimización, conocido como el problema de las mínimos cuadrados $x = min(|b - Ax|)$, en forma álgebraica usando $(b - b_*/s)$. $A^t(b - Ax) = 0$, $A^t b - A^t Ax = 0$, $A^t Ax = A^t b$, debido a que $A^t A$ es una matriz cuadrada, su inverso es $x = (A^t A)^{-1} A^t b$ y la matriz $[(A^t A)^{-1} A^t]$ se le denomina como matriz pseudo inversa.

6.8 Determinantes.

Definición.

El determinante de una matriz cuadrada A_{nxn} está definida por $|A| = \sum_{\Gamma \in S_n} sgn(\Gamma) a_1 \Gamma(1) \cdots a_n \Gamma(n)$.

- Los elementos de la matriz $A_{nxn} \rightarrow a_{ij}, 1 \leq ij \leq n$.
- S_n es el grupo de permutaciones de n números $1, 2, \cdots, n$.
- sgn(Γ) es el signo de las permutaciones Γ.
  - Donde, +1 es un signo par.
  - Donde, -1 es signo impar.

En una matriz escalar a_{1x1}, el resultado del determinante de la matriz va se igual a $|a| = \sum_{\Gamma \in S_n} sgn(\Gamma) a_1 \Gamma(1) \cdots a_n \Gamma(n) = a$. En un caso de una matriz de 2 x 2, siendo este la forma tradicional de calcular un determinante es igual a:

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = (a_{11}a_{22}) - (a_{12}a_{22}). \tag{6.19}$$

6.8.1 Relación recurrente para calcular el determinante.

En un conjunto S_n de permutaciones de $(1,2,\cdots,n)$ con $|S_n| = n!$ permutaciones. Se puede obtener una partición de $S_n = U_{i=1}S_{ni}$, con $S_{ni} \cap S_{nj} = 0.$ para $i \not= j$ donde S_{ni} son todas las permutaciones de S_n que tienen un 1 en la *i-esima* posición es decir si $\Gamma \in S_{ni}$ entonces $\Gamma(i) = 1$. Tomando en cuenta lo anterior el determinante de una matriz A se define de la siguiente manera:

$$\begin{aligned} |A| &= \sum_{\Gamma \in S_n} sgn(\Gamma)\ a_{\Gamma(1)1}, \cdots, a_{\Gamma(n)n}. \\ &= \sum_{i=1}^{n} \sum_{\Gamma \in S_{ni}} sgn(\Gamma)\ a_{\sigma(1)1}, \cdots, a_{\sigma(n)n}. \end{aligned} \tag{6.20}$$

Con una identificación adecuada cada $\Gamma \in S_{ni}$ se puede ver como el producto de una permutación de S_{n-1} y una permutación que por medio de $(i-1)$ transposiciones lleva el 1 a la i *-esima* posición $i = (1,2),(2,3),\cdots(i-1,i)$ por tanto tenemos $sgn(\Gamma_i) = (-1)**(i-1)$, lo que genera:

$$\begin{aligned} |A| &= \sum_{i=1}^{n} a_{1i} \sum_{\sigma \in S_{n-1}} sgn(\sigma_i, \sigma)\ a_{\sigma(1)1}, \cdots, a_{\sigma(n-1)n-1}. \\ &= \sum_{i=1}^{n} a_{1i}(-1)^{1+i}|A_{1i}|. \end{aligned} \tag{6.21}$$

a la matriz A_{1i} se le conoce como un menor $1,i$ de A y se obtiene de eliminar el primer renglón y la columna i de A. La fórmula recurrente para calcular un determinante, en general se puede desarrollar por cualquier renglón o columna.

$$|A| = \sum_{j=1}^{n} (-1)^{i+j} a_{ij}|A_{ij}|. \tag{6.22}$$

La ecuación anterior determina el desarrollo por columnas a partir del reglón i, por lo que los determinantes com signo se les conoce como cofactores $C_{ij} = cof(i,j) = (-1)^{i+j}|A_{ij}|$.

Una vez que se tiene una relación recurrente para calcular el determinante de A, ya no es necesario recurrir a las permutaciones.

$$|A| = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} |A_{ij}|. \tag{6.23}$$

Para poder efectuar la relación recurrente, basta con saber calcular el caso más sencillo, es decir cuando la matriz A es un escalar, $A = a \in \mathbb{R}$ en este caso $|a| = A$.

Código

Este código muestra, la solución de matrices implementando un lenguaje de programación en python.

```
def delk(a,k):
  n=len(a);b=zeros((n-1,n-1))
  for i in range(1,n):
    for j in range(n):
      if j<k:
        b[i-1,j]=a[i,j]
      if j>k:
        b[i-1,j-1]=a[i,j]
  return b

def det(a):
  n=len(a)
  d=0
  if n==1:
    return a[0,0]
  for j in range(n):
```

```
    if j%2:
      s=-1.0
    else:
      s=1.0
    d=d+s*a[0,j]*det(delk(a,j))
  return d
```

6.9 Transformación Lineal

Definición.

Sea A una transformación lineal $A : \mathbb{R} \to \mathbb{R}^n$, se dice que $v \in \mathbb{R}^n$ es vector propio de A si $Av = \lambda v$ para $\lambda \in \mathbb{R}$ y en el caso de λ es valor propio dde A correspondiente a v.

Para obtener los valores y vectores propios de una matriz A se desarrolla lo siguiente:

- Se calculan los valores propios de A, encontrando las raices del polinomio característico $P(\lambda)$, dado por $P(\lambda) = |A - \lambda I|$, donde $I = \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix}_{nxn}$, es la matriz identendidad $Av = \lambda v$, $Av - \lambda v = 0$, $Av - \lambda Iv = 0$, $(A - \lambda I)v = 0$, si $v \neq 0$, la unica forma de que $(A - \lambda I) = 0$ es que $|(A - \lambda I)| = 0$.
- En general $P(x)$ es un polinomio de grado n con n racies, tal vez con números complejos o repetidos. Para cada valor de λ se obtenie un sistema de ecuaciones, con varias soluciones. De la solución del sistema de ecuaciones propuesto se obtiene el vector propio correspondiente a $a\lambda$, esto para cada λ raiz de $P(\lambda)$.

Ejemplo.

Sea $A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, entonces $P(\lambda) = |\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}|$.

Desarrollo de $P(\lambda)$.

$$\begin{aligned} P(\lambda) &= |\begin{pmatrix} 1-\lambda & 0 \\ 0 & -(1+\lambda) \end{pmatrix}|. \\ &= -(1+\lambda)(1-\lambda). \\ &= -(1-\lambda^2) \end{aligned} \tag{6.24}$$

La racices de $P(\lambda)$ son $\lambda_1 = 1$ y $\lambda_2 = -1$ para encontrar el vector v_1 resolvemos $\begin{pmatrix} 0 & 0 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. Todos los vectores con $y = 0$ son vectores propios de $\lambda_1 = 1$ y todos los vectores con $x = 0$ para $\lambda_2 = -1$.

6.9.1 Matriz Cuadrada con Vectores Propios.

Si una matriz A cuadrada tiene vectores propios $v_1, \cdots, v_n$ con sus correspondientes valores propios $\lambda_1, \cdots, \lambda_n$. Es decir $Av_i = \lambda_i v_i$, $i = 1, 2, \cdots, n$ entonces A se puede diagonalizar, se puede llevar mediante operaciones matriciales a una matrix diagonal. $A = VDV^{-1}$, $D = V^1AV$, donde $V = (v_1, \cdots, v_n)$ tiene en las columnas a los vectores propios de A y $D = diag(x_1, \cdots, x_n)$ es una matriz diagonal con los valores propios de A.

Demostración.

$AV = (AV_1, \cdots, AV_n) = (\lambda_1 V_1, \cdots, \lambda_n V_n) = (v_1, \cdots, v_n) \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_n \end{pmatrix} = VD$, multiplicando por la inversa de V, V^{-1} por la derecha AVV^{-1}, Finalmente $A = VDV^{-1}$.

Es importante notar que en el caso que las columnas de V sea ortonormales dos a dos $V^t = V^{-1}$, puesto que $V^t = \begin{pmatrix} v_1^t \\ v_n^t \end{pmatrix}$, $V^tV = \begin{pmatrix} v_1^t v_1 & v_1^t v_1 & \cdots & v_1^t v_n \\ \vdots & v_2^t v_2 & \ddots & \vdots \\ v_i^t v_n & \vdots & \cdots & v_n^t v_n \end{pmatrix} = \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix}$, en este caso A se puede denotar como $A = VDV^t$.

6.9.2 Matriz Cuadrada Simétrica.

Si una matriz A cuadrada es simétrica entonces sus valores propios reales y sus vectores propios respectivos a valores propios diferentes son ortogonales. Para probar la primera aseveración, definamos como un número complejo a es real si es igual a su conjugación $a = \overline{a}$, tomando en cuenta que para un vector v con componentes complejos es $\overline{v}^t v = |v|^2 \in \mathbb{R}$.

1. Sea $Au = \lambda u$, valor λ y vector u propios de una matriz simétrica A, donde A contiene valores reales $\overline{A} = A$.

$$\begin{aligned} (\overline{Au})^t Au &= \overline{u}^t \overline{A^t} Au. \\ &= \overline{u}^t AAu. \\ &= \overline{u}^t \lambda \lambda u. \\ &= \lambda \lambda \overline{u}^t u. \end{aligned} \tag{6.25}$$

   Por otro lado.

$$\begin{aligned} \overline{Au}^t AU &= \overline{\lambda} \overline{u}^t Au. \\ &= \overline{\lambda} \overline{u}^t \lambda u. \\ &= \overline{\lambda} \lambda \overline{u}^t u. \end{aligned} \tag{6.26}$$

   Por lo tanto $\lambda \in \mathbb{R}$.

2. Sea U_i, U_j, vectores propios de A con sus correspondientes valores proprios λ_1 / λ_s.

$$\begin{aligned} u_i^t Au_j &= (A^t u_i)^t u_j. \\ &= (Au_i)^t u_j. \\ &= (\lambda_i u_i)^t u_j. \\ &= \lambda_i (u_i^t u_j). \end{aligned} \tag{6.27}$$

pero $u_i^t A u_j = u_i^t(\lambda_j u_j)^t = \lambda_j u_i^t u_j$, es decir $\lambda_i(u_i^t u_j) = \lambda_j u_i^t u_s$. $(\lambda_i - \lambda_j)u_i^t u_s = 0$, como $\lambda_i \not\lambda_s$, $\lambda_i - \lambda_j \not\emptyset$ y por lo tanto $u_1^t u_s = 0$.

Sea A_{mxn}, $n < m$ una matriz con valores reales, rectangulares. Donde $A_{mxn} = U_{nxm}, \Sigma_{nxm}, V_{nxm}$, una factorización *SVD* de valores singulares. Donde U tiene por columnas los vectores propios izquierdos de A y V en sus columnas los vectores propios de lado derecho. La matriz Σ es diagonal y sus elemento σ_i se conocen como valores singulares de A.

Para demostrar lo anterior, tenemos una matriz $A_{nxm}A_{mxn}$ cuadrada y simétrica de tamaño $n \rightarrow (A^tA)^t = A^t(A^t)^t = A^tA$, es decir queda invariante bajo la transposición por lo tanto asumimos que tiene vectores propios v_i ortonormales con valores propios λ_i, donde $A^tAv_i = \lambda_i v_i$, $i = 1, 2, \cdots, n$. Los mismo podemos intuir de (AA^t) simétrica de tamaño m, donde los vectores propios de u_i, donde $AA^tu_i = M_i u_i$ ortonormales. Vemos que $u_i = \dfrac{Au_i}{|Au_i|}$ y que $M_i = \lambda_i$, y también se desarrolla $\sigma_i = |Au_i|$.

$AA^tu_i = AA^t\dfrac{Au_i}{\sigma_i} = \dfrac{1}{\sigma_i}A\lambda_i u_i = \dfrac{\lambda_i}{\sigma_i}Au_i = \lambda_i(\dfrac{Au_i}{\sigma_i}) = \lambda_i u_i$. Es decir los vectores u_i spn los vectores propios de AA^t con valores propios λ_i y como AA^t es simétrica los vectores u_i son ortogonales entre si, Por lo tanto $AV = (Av_i, \cdots, Av_n) = (\sigma_1 u_1, \cdots, \sigma_n u_n) = U\Sigma$, multiplicando por $V^t \rightarrow$ obtenemos $A = U\Sigma V^t$, por lo tanto podemos comprobar que $\sigma_i = \sqrt{\lambda_i}$.

6.10 Bowbelli

A una expresión de la forma:

$$a_0 + \frac{b_1}{a_1 + \frac{b2}{a_2 + \ldots}} \tag{6.28}$$

Ejemplo:

Si $b < a$ enteros positivos usando el algoritmo de Euclides

$$a = q_1 b + y_1, \qquad b = q_2 y_1 + y_2, \ldots \tag{6.29}$$

Escrito de otra forma:

$$\frac{a}{b} = q_1 + \frac{y_1}{b}, \qquad \frac{b}{y_1} = q_2 + \frac{y_2}{y_1}, \tag{6.30}$$

$$\frac{y_1}{y_2} = q_0\frac{y_3}{y_2} + \frac{y_3}{y_2}\cdots\frac{y_1}{y_3} = q_3 + \frac{y_3}{y_2} \tag{6.31}$$

$$\frac{a}{b} = q_1 + \frac{1}{\frac{b}{y_1}}, \qquad \frac{a}{b} = q_1 + \frac{1}{q_2 + \frac{1}{\cdots + \frac{1}{q_{n-1} + \frac{1}{q_n}}}} \tag{6.32}$$

Lo interesante es que las fracciones continuas son finitas para el caso de los números racionales infinitos para el caso de los números irracionales, como se puede ver enseguida para el caso de los raices cuadradas no exactas que como se sabe representan números irracionales. Método de Bowbelli para calcular $\sqrt{N}$ donde N no es cuadrado perfecto.

Sea $a^2 < N$ y $N = a^2 + y$, $y > 0$. Hagamos $\sqrt{a^2+y} = a + x$ elevando al cuadrado ambos lados:

$$a^2 + y = (a+x)^2 = a^2 + 2ax + x^2, y = 2ax + x^2 = x(2a+x) \tag{6.33}$$

$$x = \frac{y}{2a+x} \tag{6.34}$$

$$\sqrt{N} = a + \frac{y}{2a + \frac{y}{2a+\cdots}} \tag{6.35}$$

por lo cual la raiz cuadrada que se busca se representa como una fracción continua infinita.

6.11 Conversión de base

Sea:

$n, b \in N$, la representación de n en la base b esta dada por:

$n = n_R b^R + \cdots + n_1 b + n_0$siendo los números $0 \leq b$ la representación de n en la base b. Factorizando a b tenemos:

$$n = b(n_R b^{R-1} + \cdots + n_1) + n_0 \tag{6.36}$$

Se puede ver como:

$$\frac{b}{n} = n_R b^{R-1} + \cdots + n_1 \tag{6.37}$$

En el residuo obtenemos el digito de la representación n_0.
De esta manera si dividimos el cociente entre b y nos quedamos con el residuo obtenemos n_1 y haci sucesivamente hasta obtener n_R.
Para reconstruir n apartir de los n_i, basta con evaluar el polinomio $n_R b^R + \cdots + n_1 b + n_0$, esto lo podemos hacer por ejemplo usando el método de Horney.

6.11.1 Método de Horney

Para evaluar un polinomio de la forma $P(x) = a_3x^3 + a_2x^2 + a_1x + a_0$ tenderiamos $(((((a_4x) + a_3)x) + a_2)x + a_1)x + a_0$.

6.12 Regla de Cramer

La regla de Cramer es un teorema del álgebra lineal que da la solución de un sistema lineal de ecuaciones en términos de determinantes. Recibe este nombre en honor a Gabriel Cramer (1704-1752), quien publicó la regla en su Introduction à l'analyse des lignes courbes algébriques de 1750, aunque Colin Maclaurin también publicó el método en su Treatise of Geometry de 1748 (y probablemente sabía del método desde 1729).

La regla de Cramer es de importancia teórica porque da una expresión explícita para la solución del sistema. Sin embargo, para sistemas de ecuaciones lineales de más de tres ecuaciones su aplicación para la resolución del mismo resulta excesivamente costosa: computacionalmente, es ineficiente para grandes matrices y por ello no es usado en aplicaciones prácticas que pueden implicar muchas ecuaciones.

Si $Ax = b$ es un sistema de ecuaciones, A es la matriz de coefifientes del sistema, $x = (x_1 \ldots x_n)$ es el vector columna de las incógnitas, y b es el vector columna de los

términos independientes, entonces la solución al sistema se presenta así:

$$x_j = \frac{det(A_j)}{del(A)}. \tag{6.38}$$

Donde A_j es la matriz resultante de reemplazar la j-ésima columna de A por el vector columna b. Hágase notar que para que el sistema sea compatible determinado, el determinante de la matriz A ha de ser **no nulo**.
antes de proceder a obtener el método de cramer, veamos la siguiente proposición auxiliar.

Propoción
Si una matriz A cuadra de 2 columnas son iguales entonces su determinante $\mid A \mid = 0$.
Demostración:
Sea:

$$A = (A_1, a_i, A_j, \ldots, A_n) \qquad A_i = A_j. \tag{6.39}$$

Intercambiamos A_i con A_j por la propiedad de la antisimetrica del determinante esto genera un cambio de signo, pero como $A_i = A_j$, la matriz Aqueda igual.

$$\mid A \mid = - \mid A \mid, \quad \mid A \mid + \mid A \mid = 0, \quad 2 \mid A \mid = 0 \tag{6.40}$$

esto implica que:

$$\mid A \mid = 0 \tag{6.41}$$

Considera la matriz A' que es igual a la matriz A en todas las columnas, diferentes de i y b es su columna i, $A' = (A_1, \ldots, A_{i-1}, b, \ldots, A_n)$
Entonces $\mid A' \mid = \mid A_i \ldots A_{i-1}, \sum_{R=1}^{n} x_i A_i, \cdots, A_{i+1}, A_n \mid$ porque $b = \sum_{b=1}^{n} x_i A_R$
Por la multinealidad del determinante $\mid A' \mid = \sum_{R=1} x_R \mid A_1, \cdots, A_i, A_R, A$ pero por la proposición anterior todas las A_R se repiten en los determinantes de la suma, excepto $r = i$ por tanto $\mid A' \mid = x_i \mid A \mid$, se despeja x_i.
Solución:

$$x_i = \mid A_1, \ldots, A_{i-1}, b, A_{i+1}, \ldots, A_n \mid \quad i = 1, 2, 3, \ldots, n \tag{6.42}$$

6.13 Matriz Inversa y Cofactores

Se puede probar a partir de la definición del determinante$|A| = \sum sgn(j_n)a_{1j(1)} \cdots a_n j_(n)$ que este se puede calcular recursivamente, usando el concepto de menor $|A_{ij}|$. Donde A_{ij} es la sub matriz $(n-1)x(n-1)$ que se obtine al eliminar el renglón i y la coluna j de A.

$$|A| = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} |A_{ij}| \tag{6.43}$$

a_{ij} elemento de A, aqui el desarrollo es sobre la columna j. También se puede desarrollar sobre cualquier renglón.

Ejemplo:

$$A^{-1} = \frac{cof(A)^t}{|A|} \tag{6.44}$$

Donde:

$cof(A)$ es la matriz de cofactores de A.

Demostración:

Sea:

$x = (x_1, \ldots, x_n)$ una matriz incognita, con $Ax = j$ es decir x pretende ser A^{-1} la inversa A. Si $j = (e_1, \ldots, e_n)$, e_i es un vector de ceros excepto en la posición i donde vale 1. Entonces tenemos n sistemas de ecuaciones uno por cada columna x.

$Ax_1 = e_1, \ldots Ax_n = e_n$ para la primera columna de x_1, tenemos que, usando el metodo de cramer.

$$x_{i1} = \frac{|(A_1, \ldots, A_{i-1}, e_1, \ldots, A_n)|}{|A|} \tag{6.45}$$

Desarolando sobre la columna i, $x_{i1} = \frac{cof_{(j,i)}}{|A|}$, donde x_{i1} es elemento iesimo de A. En general para cualquier x_{ij} de la inversa de A.

$x_{ij} = \frac{cof(j,i)}{|A|}$ pero $cof(j,i)$ es el elemento i, j de $cof(A)^t$, esto implica:

$$A^{-1} = \frac{cof(A)^t}{|A|}. \tag{6.46}$$

6.14 Determinante

La relación recurrente para calcular el determinante considere el conjunto S_n de permutaciones de $(1,2,\ldots,n)$ con $\mid S_n \mid = n!$ permutaciones. Sepuede obtener una partición de $s_n = \cup_{i=1}^{n} s_{ni}$ con $s_{ni} \cap_{nj} = 0$ para $i \neq j$ donde s_{ni} son todas las permutaciones de s_n que tienen un 1 en la $i-esima$ posición es decir si $r \in s_{ni}$ entonces $r(i) = 1$.

Tomando en cuenta lo anterior el determinante de una matriz, A se ve:

$$\mid A \mid = \sum_{r \in s_n} sgn(r) a_{r(1)1} \cdots a_{r(n)n} = \sum_{i=1}^{n} \sum_{r \in s_{ni}} agn(r) a_{11} \cdots a_{(1)2} \cdots a_{r(n)n} \tag{6.47}$$

$$= \sum_{i=1}^{n} a_i \sum_{r \in s_{n-1}} sgn(r,-r) a_{r(1)1} \cdots a_{r(n-1)n-1} = \sum_{i=1}^{n} a_i (-1)^{1+i} \mid A_{ji} \mid \tag{6.48}$$

A la matriz A_{ij} se le conoce como un menor ij de A y se obtiene de eliminar el renglon 1 y columna i de A. y la anterior es precisamente la formularecurrente para calcular un determinante, en general se puede desarrollar por cualquier renglon o columna en terminos de miembros.

También a los determinantes $\mid A_{ij} \mid$ se les dice menores.

$$\mid A \mid = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \mid A_{ij} \mid \tag{6.49}$$

en este caso se esta desarrollando por columnas a partir del renglon i.

A los menores signo se les conoce como cofactores $c_{ij} = cof(i,j) = (-1)^{i+j} \mid A_{ij} \mid$
Una vez que se tiene una relación recurrente para calcular un determinante ya no es necesario recurrir a las permutaciones.

$$\mid A \mid = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \mid A_{ij} \mid \tag{6.50}$$

Para poder efectuar la relación recurrente, basta con saber calcular el caso más sencillo, es decir, cuando la matriz A es un escalar, $A = a \in \mathbb{R}$ en este caso $\mid A \mid = a$. Veamos como se implementa este cálculo recursivo del determinante en Python, desarrollando en el primer renglón, lo primero es tener una función de f(a,j) que genere los menores $A + j$. Para posteriormente usando los menores calcular es forma recursiva el determinante de la matriz A.

6.15 Geometría

La matemática griega era principalmente geometría y fue Euclides en su obra fundamental los elementos donde por primera vez de una manera contundente se expone el método axiomático que es l forma actual en que se presentan las matemáticas y puntual de la ciencia moderna.

Este método axiomático consiste en definiciones, axiomas y postulados. Las definiciones caracterizan a los entes con los que sea va a trabajar, los axiomas establecen los hechos o proposiciones auto evidentes, que no requieren demostraciones con respecto a estos los postulados son las proposiciones primitivas, que son la base de todo el cuerpo de conocimiento que se construye después mediante demostraciones.Por lo que en relidad el método debería nombrarse postumatico en lugar axiomático.

En el caso de los elementos de Euclides que se escribió aproximadamente 300 antes de Cristo y consta de 132 definiciones, 5 postulados y 5 nociones comunes o axiomas. Durante mucho tiempo esta obra parte de ser uno de los libros ,más leído consideró como el referente básico para el entrenamiento del uso de la razón y puerta obligada de la matemática superior.

A continuación se vera una pequeña muestra de lo anterior.

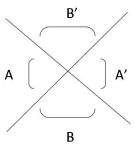
Postulados

1. Se puede trazar una recta entre dos puntos cualesquiera.
2. Se puede prolongar un segmento de recta.
3. Construir un circulo con cualquier radio y centro.
4. Todos los ángulos rectos son iguales entre si.
5. Dado una recta y un punto fuera de ella se puede trazar una única paralela a la recta dada.

Estos 5 postulados excepto el 4 presentan operaciones básicas, a partir de los cuales Euclides construyo la geometría plana.

Proposición

Los ángulos opuestos por el vértice son iguales.



$A+B=180^o$, $A'+B=180^o$, $A+B=A'+B$ $A=A'$ del dibujo se puede apreciar que $A*B=A'+B$ y de ahí se desprende la des mostración.

Definición 5.1 Dos figuras son congruentes si son la misma figura colocada en diferentes posiciones. En el caso por ejemplo de triángulos, la congruencia implica que sus ángulos y lados son del mismo tamaño y además de sus áreas son iguales.

Proposición.

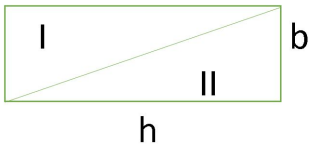
Si dos triángulos tienen un lado y su ángulo opuesto iguales entonces son congruentes. Sean los triángulos ABC y $A'B'C'$ supongamos que los lados $AByA'B'$ son iguales y que el ángulo en C y C' también son iguales. Consideremos dos pequeños segmentos de recta que forman al ángulo en C y C'. Colocamos al lado AB sobre $A'B'$ como son iguales coinciden.

Como los ángulos en C y C' también coinciden por hipótesis si hacemos coincidir $CconC'$ los segmentos de recta que forman los ángulos coinciden y si los prolongamos por consecuencia los lados $AC, A'C'$ y $BC, B'C'$ también coinciden cumpliéndose la congruencia.

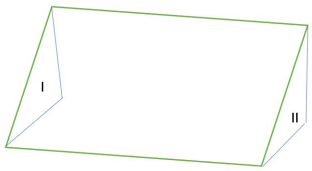
También se puede probar que si en dos triángulos un lado y los ángulos se puede en los extremos son iguales, entonces los triángulos son congruentes. Con la noción de congruencia se puede obtener la formula general para el calculo de área de un triángulos.

Proposición.

Para un triángulo rectángulo su área se puede calcular como $\frac{bh}{2}$. Por un criterio de congruencia los triángulos I y II son iguales, como el área del rectángulo que los contiene es *bh* entonces el área del triángulo rectángulo por ejemplo II es $\frac{bh}{2}$



Proposición.
El área de un paralelogramo es igual al área de su rectángulo equivalente.

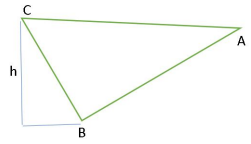


Por criterio de congruencia los triángulos I y II son congruentes y por tanto su área es la misma.

Para formar el triángulo equivalente se quita el triángulo I de la izquierda del paralelogramo y se pone en la posición del triángulo II, el resultado final de esta operación es el rectángulo equivalente que por construcción tiene la misma área del paralelogramo original.

Definición 5.2 Una altura *h* en un triángulo es el tamaño del segmento de recta que va en forma perpendicular de un vértice al lado opuesto.Por lo anterior en todo triángulo hay 3 alturas.
Proposición.

Dado un triángulos cualquiera su área se puede calcular como base por altura entre dos. Donde la base es cualquiera de sus lados y la altura h la perpendicular que va de ese lado o su prolongación al vértice opuesto.



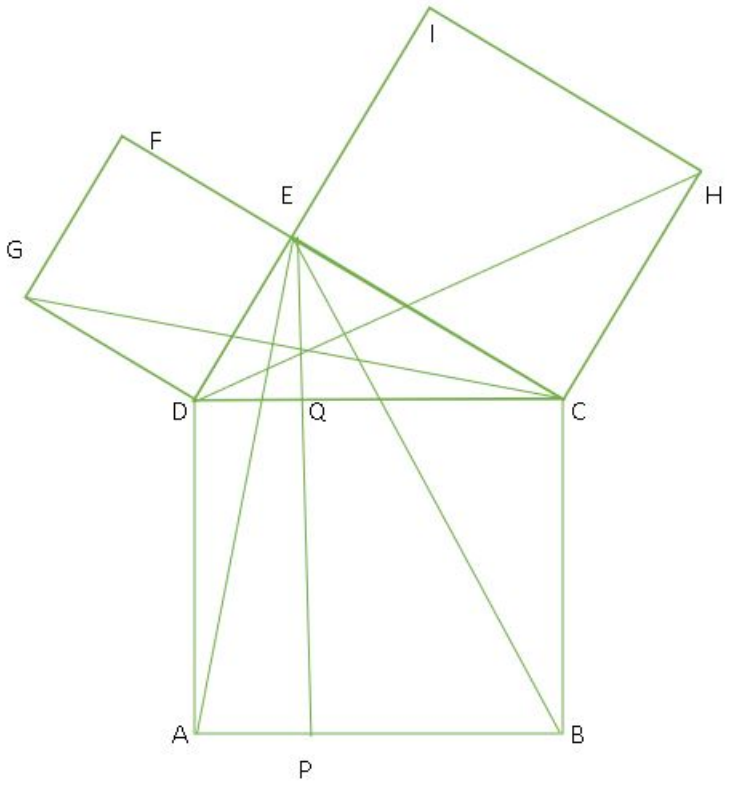
Consideremos la base $b = \mid AB \mid$ triángulo ABC la altura h es el tamaño perpendicular que va del vértice C a la prolongación del lado AB. bh es el área del rectángulo equivalente que es igual al área del paralelogramo determinado por el triángulo ABC. como el área del triángulo ABC es la mitad del paralelogramo entonces también es la mitad del área del rectángulo. Por lo tanto area $(ABC) = \frac{bh}{2}$.

6.16 El teorema de Pitagóricas.

Sin duda uno de los resultados más interesantes de los elementos de Elucides es la demostración del teorema de Pitagóricas, esta se basa es esencialmente en la congruencia y áreas de triángulos. La relevancia de este teorema se debe a que es el fundamento de todas las aplicaciones que tienen que ver con distancias y medidas de similitud y magnitudes.

Proposición.

En un triángulo rectángulo el cuadrado de la hipotenusa es igual a la suma de los cuadrados de los catetos; Pitagóricas de Samos.



En la figura los catetos son los lados DE y CE y la hipotenusa DC. El cuadrado de la hipotenusa es le área del cuadrado $ABCD$, que es la suma de los rectángulos $APQD$ y $PBCQ$. Por otro lado, los triángulos GDC y DAE son congruentes, lo mismo los triángulos BCE y DCH.

La congruencia anterior bajo el criterio de dos lados y el ángulo entre ellos iguales. Ahora bien el área del cuadrado $GDEF$ es el doble de un cateto. Pero el área del cuadrado $GDEF$ es el doble del área del triángulo $DAPQ$ es el doble del área del tri ánulo DAE. Por tanto el área del cuadrado $GDEF$ es igual al área del rectángulo $DAPQ$ y por un argumento similar y el área del cuadrado $BCHJ$ es igual al área del rectángulo $PBCQ$

6.17 ICA

Supongamos que se tiene una colección de señales x resultado de las mediciones de algún experimento.

Podemos considerar que estas señales x son la mezcla de algunas señales fuentes S, mediante la matriz de mezcla A. $x = As$, por supuesto no conocemos las señales s y posiblemente tampoco la matriz A. Entonces nuestro problema es encontrar una matriz des mezcladoras que permita obtener las señales s a partir de las señales x.
$Wx = s$, de tal manera que las señales s sean independientes entre si, en términos probabilísticos.
Este problema se puede atacar tomando en cuenta que de acuerdo l teorema del limite central, entre menos gausiana sea una señal es más independiente y tomando en cuanta el criterio de entropía negativa la medida $E[G \mid Z]$ para G una función no cuadrática y E el operador promedio, es una medida de independencia siendo 0 para Z gausiana y en general Z de media 0 y varianza 1.
En nuestro caso necesitamos maximizar $E[G(w^t x)]$ para encontrar una señal $s_i = w^t x$ independientemente. $s_i = maxE[G(w^t x)]$ y de esta forma se va consiguiendo la matriz W tal que $S = Wx$, donde la matriz S tiene en sus renglones a todas las señales independientes. Para resolver este problema de optimización se requiere que las señales medidas que están en los renglones de la matriz X, tengan media 0 y estén blanqueadas, es decir estén colacionadas y tengan varianza 1. Estas dos propiedades anteriores constituyen la etapa de preprocesa miento.
Sea E el operador lineal en el caso de que Z sea una variable discreta con instancias $z_1, \ldots, z_n$ cada una con probabilidades $P_1, \ldots, P_n$. Este operador E actúa sobre la variable aleatoria discreta como:

$$E[Z] = \int_{-\infty}^{\infty} zP(z)dz \tag{6.51}$$

En cualquiera de los dos casos E es lineal, es decir:

$$E[\alpha\gamma + \beta z] = dE[\gamma] + \beta E[z] \tag{6.52}$$

6.17.1 Preprocesa miento en ICA

1.- Si $E[\gamma] = m$ entonces $E[\gamma - m] = 0$, puesto que:

$$E[\gamma - m] = E[\gamma] - E[m] = m - m = 0. \tag{6.53}$$

Notese que si únicamente se tienen datos de una variable aleatoria z su promedio $E[z]$ se puede estimar con $E[z] \approx \frac{1}{N}\sum_{R=1}^{N} z_R$, donde z_R son instancias de Z

2.- Una vez que todas las señales muestreadas tienen media cero se blanquean mediante el cambio de variable:

$$x' = \sqrt{0^{-1}}V^t X, \tag{6.54}$$

Donde V es la matriz de vectores propios de $E[xx^t]$ matriz covarianza de las señales de entrada y D es la matriz diagonal de los valores propios de $E[xx^t]$. Consideremos la matriz de covarianza de las señales transformadas.

$$E[x'x'^t] = E[(\sqrt{D^{-1}}v^t x)(\sqrt{D^{-1}}v^t x)^t] = E[\sqrt{D^{-1}}v^t xx^t v\sqrt{D^{-1}}], \tag{6.55}$$

por linealidad = $\sqrt{D^{-1}}v^t E[xx^t]v\sqrt{D^{-1}}$, como v son vectores propios de:

$$E[xx^t] = \sqrt{D^{-1}}v^t vD\sqrt{D^{-1}} \tag{6.56}$$

v^t y v son inversas $= \sqrt{D^{-1}}D\sqrt{D^{-1}} = DD^{-1} = I$ Es decir la matriz de covarianza de las señales transformadas es la matriz identidad, por lo tanto están descorrelacionadas y tienen varianza 1.

Código:

Una vez hecho el pre procesamiento, lo que sigue es resolver el problema de maximización, para cada renglón de la matriz de des acoplamiento W

$$maxE[w^t x], \quad con \quad restriccion \quad E[(w^t x)^2] = \mid w \mid^2 = 1 \tag{6.57}$$

Este problema se resuelve con multiplicadores de Lagrange la solución se encuentra donde los gradientes son colineales

$$\bigtriangledown_w E[G(w^t x)] - \alpha \bigtriangledown (w \cdot w) = 0 \tag{6.58}$$

$$\bigtriangledown_w E[G(w^t x)] = E[\bigtriangledown_w G'(w^t x)] = E[G(w^t x) \bigtriangledown_w W^t x] = E[G'(w^t x)x] \tag{6.59}$$

$\bigtriangledown_w(w \cdot w) = 2w$, haciendo $\beta = 2\alpha$ y $G' = g$, tenemos $E[g(w^t x)x] - \beta w = 0$, es decir para encontrar a w tenemos que resolver este sistema de ecuaciones no lineales.

6.18 Método de Newton Raphson

Método de Newton Raphson para resolver un sistema de ecuaciones no lineas.
Sea una ecuación vectorial de la forma $F(x) = 0$, donde $F = (f_1(x), \ldots, f_m(x))^t$ y cada f_i es una función $f_i : \mathbb{R}^n \longrightarrow \mathbb{R}$. Es decir se tiene un sistema de m ecuaciones no lineales:

$$f_1(x) = 0$$
$$.$$
$$.$$
$$f_n(x) = 0$$

El método de Newton permite en contar una secuencia $x \in \mathbb{R}^n$ que en general converge a la solución x, tal que $F(x) = 0$. Esto en base a la siguiente iteración

$$(y_{R+1} - y_R) = JF(x_R)(x_{R+1} - x_R) \tag{6.60}$$

JF es la derivada de F su Jacobiano

$$0 = y_{R+1} = F(x_R) + JF(x_R)(x_{R+1} - x_R) \tag{6.61}$$

con $y_R = F(x_R)$ y x_{R+1} el valor donde se anula $F(x)$

$$JF(x_R)(x_{R+1} - x_R) = -F(x_R),$$
$$(x_{R+1} - x_R) = JF^{-1}(x_R)(-F(x_R))$$

Finalmente $x_{R+1} = x_R - JF^{-1}(x_R)F(x_R)$, donde:

$JF(x) = (\bigtriangledown f(x) \cdot \cdot \bigtriangledown f_n(x))$ Esto se puede ver como $x_{R+1} = H(x_R)$ donde H es la formula de iteración y se cumple que $x_* = \lim_{R \longrightarrow \infty} x_R$ es un punto fijo para H, veamos:

$$x_* = \lim_{R \longrightarrow \infty} x_{R+1} = \lim_{R \longrightarrow \infty} H(x_R) = H(\lim_{R \longrightarrow \infty} x_R) = H(x_R) \tag{6.62}$$

En efecto si H es continua y en general lo es, podemos buscar a x_* como el punto fijo de H.

Sea $F(w) = E[xg(w^t x)] - \beta w = 0$ calculando $JF(w)$ el Jacobiano respecto a w

$$J_w F(w) = J_w E[xg(w^t x)] - \beta J_w w \tag{6.63}$$

$$= E[xJ_w g(w^t x)] - \beta I = E[xg'(w^t x)J_w(x^t w) - \beta I \tag{6.64}$$

$$= E[xg(w^t x)x^t J_w(w)] - \beta I = E[xx^t g(w^t x)I] - \beta I \tag{6.65}$$

$$\approx E[xx^t]E[g'(w^t x)] = IE[g'(w^t x)] \tag{6.66}$$

$$J_w F(w) = E[g'(w^t x)] \cdot I - \beta I \tag{6.67}$$

Es una matriz diagonal con $E[g'(w^t x)] - \beta$ es la diagonal por tanto $JF(w)$ también es una diagonal con $\frac{1}{E[g'(w^t x)]-\beta}$ en la diagonal. Entonces la formula de iteración del punto fijo queda como:

$$H(w) = \frac{w - (E[xg(w^t x)] - \beta w)}{(E[g'(w^t x) - \beta])} \tag{6.68}$$

en otras palabras $w^t = H(w)$, donde w^t es el valor actualizado de la iteración multiplicando ambos lados de la ecuación de atención por $(\beta - E[g'(w^t x)])$ tenemos:

$$(\beta - E[g'(w^t)x])w^t = \beta\bar{w} - wE[g'(w^t x)] + E[xg(w^t x)] - \beta\bar{w} \tag{6.69}$$

como después de cada iteración se normaliza w_j el factor $(\beta - E[g'(w^t x)])$ no le afecta.

$$w^t = E[xg(w^t x)] - wE[g'(w^t x)] \tag{6.70}$$

y la convergencia se da cuando $w \cdot w^t = cos(\theta) \approx 1$ puesto que en este caso son prácticamente el mismo.

Típicamente se toma a $g(\cdot)$ como la tangente hiperbólica, $g(z) = tanh(z)$.

6.19 La Serie de Leibniz

Calculo de π usando series de Fourier consideremos la función identidad $f(x) = x$ en $[-\pi, \pi]$ como $f(x) = x$ es una función impar sus componentes a_n en su desarrollo de Fourier son ceros, por tanto

$$x = \sum_{n=1}^{\infty} b_n sen(nx) \tag{6.71}$$

$$b_n = \frac{1}{n}\int_{-\pi}^{\pi} xsen(nx)dx \tag{6.72}$$

pero,

$$\int_{-\pi}^{\pi} xsen(nx)dx \quad u = x,\ du = dx,\ v = \frac{-1}{n}cos(nx),\ dv = sen(nx) \tag{6.73}$$

Se tiene

$$\int_{-\pi}^{\pi} xsen(nx)dx = \left(\frac{-1}{n}\right)[xcos(nx)]_{-\pi}^{\pi} + \frac{1}{n}\int_{-\pi}^{\pi} cos(nx)dx \tag{6.74}$$

tomando en cuenta que

$$\int_{-\pi}^{\pi} cos(nx)dx = 0 \tag{6.75}$$

0

$$\int_{-\pi}^{\pi} sen(nx)dx = -\left(\frac{2\pi}{n}\right)cos(n\pi) \tag{6.76}$$

Finalmente

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} xsen(nx)dx = \frac{2(-1)^{n+1}}{n} \tag{6.77}$$

$$x = 2\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} sen(nx) \tag{6.78}$$

Evaluando en $x = \frac{\pi}{2}$ y tomando en cuenta que $sen(n\frac{\pi}{2}) = 0$ si n es par

$$\frac{\pi}{2} = a\sum_{R=0}^{\infty} \frac{(-i)^{2R+1+1}}{2R+1} sen((2R+1)\frac{\pi}{2}) \tag{6.79}$$

de esta forma $2R+1$ recorre solo valores nones.

Observando que el exponente de (-1), $2R+1+1$ es par, y $sen((2R+1)\frac{\pi}{2}) = (-1)$ tenemos al final

$$\pi = 4\left(\sum_{R=0}^{\infty} \frac{(-1)^R}{2R+1}\right) \tag{6.80}$$

6.20 Algoritmos y Métodos

Un problema central en ciencia aplicada es la solución de un sistema de ecuaciones

$$\begin{aligned} A_{11}x_1 + A_{12}x_2 + \cdots + A_{1n}x_n &= b_1 \\ &\cdots \\ &\cdots \\ A_{n1}x_1 + A_{n2}x_2 + \cdots + A_{nn}x_n &= b_n \end{aligned}$$

$$\begin{bmatrix} A_{11} \ldots A_{1n} \\ \cdot \\ \cdot \\ A_{n1} \ldots A_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \cdot \\ \cdot \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \cdot \\ \cdot \\ b_n \end{bmatrix} \tag{6.81}$$

Donde A_{ij} son los coeficientes del sistema x_i las incógnitas y b_i el objetivo. En forma simbólica
$Ax = b$, donde A es la matriz del sistema, x el vector incógnita $x = (x_1, \ldots, x_n)^t$ y $b = (b_1, \ldots, b_n)^t$ el vector objetivo.

Uno de los métodos más importante para la solución de este problema es el de eliminación Gaussiana, que consiste primero en transformar el sistema en triangular superior y luego en aplicar sustitución hacia atrás para encontrar la solución.
Por ejemplo:

$$\begin{aligned} 4x_1 - 2x_2 + x_3 &= 11 \\ -2x_1 + 4x_2 - 2x_3 &= -16 \\ x_1 - 2x_2 + 4x_3 &= 17 \end{aligned} \tag{6.82}$$

Tomando únicamente los coeficientes:

$$\begin{bmatrix} 4-2+1 \\ -2+4-2 \\ 1-2+4 \end{bmatrix} \begin{bmatrix} 11 \\ -16 \\ 17 \end{bmatrix} \tag{6.83}$$

Multiplicando la primera ecuación por $\frac{2}{4}$ y sumándola a la segunda obtenemos el renglón ampliado $[0,3,-1.5,-10.5]$

Multiplicando la primera ecuación por $\frac{-1}{4}$ y sumándola a la tercera obtenemos $[0,-1.5,3.75,14.2$
Por lo que el sistema se ve como:

$$\begin{bmatrix} 4-2+1 \\ 0+3-1.5 \\ 0-1.5+3.75 \end{bmatrix} \begin{bmatrix} 11 \\ -10.5 \\ 14.25 \end{bmatrix} \tag{6.84}$$

Procesando de igual manera con la sub matriz

$$\begin{bmatrix} 3-1.5 \\ -1.5+3.75 \end{bmatrix} \begin{bmatrix} -10.5 \\ 14.25 \end{bmatrix} \tag{6.85}$$

Finalmente obtenemos los coeficientes de un sistema equivalente al sistema de ecuaciones inicial, pero triangular superior, que es más fácil de resolver.

$$\begin{bmatrix} 4-2+1 \\ 0+3-1.5 \\ 0+0+3 \end{bmatrix} \begin{bmatrix} 11 \\ -10.5 \\ 9 \end{bmatrix} \tag{6.86}$$

y como sistema de ecuaciones

$$\begin{bmatrix} 4x_1-2x_2+x_3 \\ 0+3x_2-1.5x_3 \\ 0+0+3x_3 \end{bmatrix} \begin{bmatrix} 11 \\ -10.5 \\ 9 \end{bmatrix} \tag{6.87}$$

Aplicando sustitución hacia atrás $x_3 = 3, x_2 = -2, x_1 = 1$ y se puede ver sustituyendo que estos valores satisfacen el sistema original.

6.21 PCA

A los atributos componentes x_i de un patrón $x \in \mathbb{R}^n$ los podemos ver como variables aleatorias y por tanto obtener su promedio $E[x_i]$, que cuando únicamente se tienen muestras discretas de x_i se tiene $E[x_i] \approx \frac{1}{n}\sum muestras(x_i)$ donde n es el número de muestras. Otro hecho importante es que E un operador lineal es decir $E[\alpha a + \beta b]$ = $\alpha E[a] + \beta E[b]$ para ayb variables aleatorias. En lo que sigue consideramos variables aleatorias con media cero, porque para este caso la correlación adquiere una forma sencilla y simple restándole su promedio se puede transformar cualquier variable aleatoria en una variable de promedio o media cero.

Consideremos al vector columna $x = (x_1, \ldots, x_n)^t$ formado por atributos $x_i \in \mathbb{R}$ a cada uno de estos atributos los consideramos variables aleatorias. Entonces $A = E(xx^t)$ es una matriz simétrica con componentes $[A]_{ij} = E[x_i x_j]$. por ser simétrica como se sabe esta matriz se puede diagonal izar usando su matriz V de vectores propios ortogonales $D = V^t AV$,

Donde: D = es diagonal con los valores propios de A

Ahora notemos que la auto correlación nos mide el grado del promedio de variabilidad de una variable aleatoria y por lo tanto su representatividad.

Sea $Z = v^t x$ un cambio de coordenadas. Entonces

$$E[zz^t] = E[v^t x(v^t x)^t] = E[v^t xx^t v] = v^t E[xx^t]v = v^t Av = D \tag{6.88}$$

Es decir en las nuevas coordenadas Z las variables aleatorias están descorrelacionados y su auto correlación mide los valores propios de A. La idea del algoritmo PCA es reducir la dimensionalidad del problema que dándonos con los z_i más representativos.

Algoritmo PCA

1. Calcular los promedios restarlos valores de los atributos
2. Calcular A aproximándola con proyecto
3. En contar los valores propios de A y quedarnos con los más representativos
4. Obtener los vectores propios correspondientes para obtener la transformación

6.22 Método de Euler

Método de Euler para resolver ecuaciones diferenciales ordinarias

Dada una ecuación diferencial ordinaria $y' = f(x,y)$, $y_0 = y(x_0)$ se busca una función $yy = g(x)$ tal que cumpla la condición inicial $y_o = g(x_0)$ y satisfaga la ecuación $y' = f(x, g(x))$.

La función f(x,y) se puede interpretar como un campo de direcciones establecido en cada punto (x,y).

El método de Euler consiste en ir construyendo la solución de manera iterativa a partir de la condición inicial, mediante la siguiente ecuación:

$$y_{R+1} = Y_R + f(x_R, y_R)h \tag{6.89}$$

esto porque

$$y_{R+1} - y_R \approx y_R h \tag{6.90}$$

y como

$$y'_R = f(x_R, y_R) \tag{6.91}$$

entonces con un pequeño error podemos considerar, si h es pequeña

$$y_{R+1} = y_R + f(x_R, y_R)h \tag{6.92}$$

Suponemos que queremos aproximar la solución en un intervalo $[a,b]$, entonces haciendo $x_0 = a$, $h = \frac{b-a}{n}$ donde n es el número de puntos donde queremos aproximar la solución.

Código:

En el caso vectorial tenemos que el campo $f(y_1,\dots y_n)$ es un vector en $\mathbb{R}^n$ y cada una de las y_i es una función de un parámetro t. Entonces tenemos el siguiente sistema de ecuaciones:

$$\begin{aligned} y_1' &= f_1(y_1,\dots,y_n) \\ &\vdots \\ y_n' &= f_n(y_1,\dots,y_n) \end{aligned} \tag{6.93}$$

Donde el campo esta dado por $f = (f_1,\dots,f_n)$ y la solución es una curva $g(t) = (y_1(t),\dots,y_n(A))$ tal que en cada valor de t su derivada coincide con el valor del campo

$$\frac{d}{dt}y(t) = f(t) = (f_1(t),\dots,f_n(t)) \tag{6.94}$$

para ejemplificar veamos el caso $n = 2$, tenemos que la solución es una curva en el plano.

$$\begin{aligned} y_1' &= f_1(y_1,y_2) \\ y_2' &= f_2(y_1,y_2) \end{aligned} \tag{6.95}$$

Aquí el método de Euler se ve como sigue

$$\begin{aligned} y_{1R+1} &= y_{1R} + hf_1(y_{1R},f_{2R}) \\ y_{2R+1} &= y_{2R} + hf_2(y_{1R},f_{2R}) \end{aligned} \tag{6.96}$$

Cuando tenemos una ecuación diferencial de orden $f^n = f(x,y,y',\dots,y^{n-1})$ la podemos transformar en un sistema de ecuaciones mediante el siguiente cambio de variable

$$y_0 = y, y_1 = y', \ldots, y_{n-1} = y^{n-1} \tag{6.97}$$

y ahora tenemos el siguiente sistema de ecuaciones diferenciales ordinarias

$$y_0 = y, y_1 = y', \ldots, y_{n-1} = y^{n-1} = f(x, y_0, y_1, \ldots, y_{n-1}) \tag{6.98}$$

6.23 Método de Ronge Kutta

La idea de Ronge Kutta es mejorar la aproximación de Euler tomando otros valores del campo de direcciones.Para el caso del Ronge Kutta de orden dos basta con que se tome otro valor adecuado a la siguiente expresión.

$$y(x+h) = y(x) + C_0 F(x,y)h + C_1 F[x+ph, y+qhF(x,h)]h \tag{6.99}$$

Donde los parámetros C_0, c_1, pyq se tienen que determinar.

Desarrollando

$$F[x+ph, y+qhF(x,y)] = F(x,y) + \frac{\partial F}{\partial x}ph + \frac{\partial F}{\partial y}qh \tag{6.100}$$

Sustituyendo en la primera expresión

$$y(x+h) = y(x) + (C_0 + C_1)F(x,y)h + \frac{\partial F}{\partial x}C_1 ph^2 + \frac{\partial F}{\partial y}C_1 qh^2 \tag{6.101}$$

Por otro lado desarrollando en serie de Taylor hasta el segundo termino

$$y(x+h) = y(x) + y'(x)h + \frac{1}{2}y''(x)h^2 \tag{6.102}$$

en términos de $F(x,y) = y'(x)$

$$y(x+h) = y(x) + F(x,y)h + \frac{1}{2}F'(x,y)h^2 \tag{6.103}$$

pero $F'(x,y) = \frac{\partial F}{\partial x} + \frac{\partial F}{\partial y}y'$

$$y(x+h) = y(x) + F(x,y)h + \frac{1}{2}\left(\frac{\partial F}{\partial x} + \frac{\partial F}{\partial y}y'\right)h^2 = y(x) + F(x,y)h + \frac{1}{2}h^2\frac{\partial F}{\partial x} + \frac{1}{2}h^2\frac{\partial F}{\partial y}y'$$

$$y(x) + (C_0 + C_!)F(x,y)h + \frac{\partial F}{\partial x}C_! p h^2 + \frac{\partial F}{\partial y}C_1 q h^2 = \tag{6.104}$$

$$y(x) + F(x,y)h + \frac{1}{2}h^2\frac{\partial F}{\partial x} + \frac{1}{2}h^2\frac{\partial F}{\partial y} \tag{6.105}$$

Por tanto $C_0 = 0, C_! = 1, p = q = \frac{1}{2}$ es una solución.

De manera similar se puede obtener un esquema Ronge Kutta de orden 4 que mejora aún más la solución. Con constantes

$$K_0 = hF(x,y), K_! = hF\left(x + \frac{h}{2}, y + \frac{k_0}{2}\right) \tag{6.106}$$

$$K_2 = hF\left(x + \frac{h}{2}, y + \frac{K_!}{2}\right) \quad k_3 = hF(x+h, y+K_") \tag{6.107}$$

$$y(x+h) = y(x) + \frac{1}{6}(K_0 + 2K_1 + 2K_2 + K_3) \tag{6.108}$$

Aplicación del Método de Ronge Kutta

Consideremos la siguiente ecuación diferencial ordinaria de primer orden $y' = -\frac{x}{y}$, en este caso $f(x,y)$ le asociamos un pequeño segmento de recta con pendiente $-\frac{x}{y}$ obtenemos la descripción geométrica del campo

Como puede verse las curvas solución para esta ecuación diferencial son círculos concéntricos respecto al origen. Si resolvemos analíticamente la ecuación mediante separación de variables eso es precisamente lo que obtenemos

$$\frac{dy}{dx} = -\frac{x}{y}, \int ydy = -\int xdx, \tag{6.109}$$

$$\frac{1}{2}y^2 = -\frac{1}{2}x^2 + c, x^2 + y^2 = k = 2c \tag{6.110}$$

que es la ecuación de círculos concéntricos.

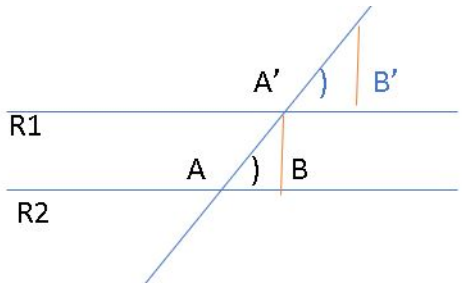
Si queremos usar la función integrante para generar el circulo, podríamos hacer solo un segmento del circulo puesto que cuando $y = 0$, tendríamos problemas para evaluar el valor del campo $-\frac{x}{y}$. Por lo que en este caso es mejor usar integrante que implementa un sistema de ecuaciones y aquí ya no hay problema con $y = 0$, puesto que en cada punto (x, y) el campo es un vector $(y, -x)$. En python la invocación de integrar $(F_!, F_2, -1.0, 0.0, 8.0, 0.01)(F_!, F_2)$ es el vector del campo y $(-1.0, 0.0)$ es el punto inicial con paso $n = 0.01$

6.24 El teorema de tales

Este teorema establece la base para manejar la propiedad de semejanza entre triángulos que junto con la congruencia permite manejar varios aspectos geométrico esenciales.

Proposición: Si una recta corta dos paralelas el ángulo que forma con estas es el mismo.

Sea las rectas paralelas R_1, R_2 y una recta R que corta.



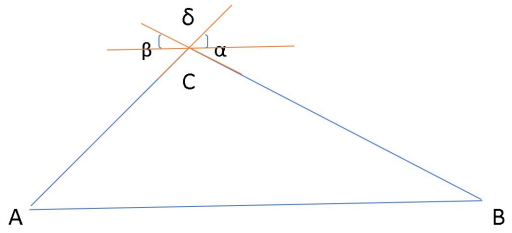
En A' sobre R_2 tómese B', tal que el segmento $A'B'$ es del mismo tamaño que el segmento AB. En B' levantar una perpendicular tal que $B'C'$ sea del mismo tamaño que

BA' y usando el criterio de congruencia.

Lados y ángulo que forman iguales. Tenemos que el triángulo ABA' es congruente con el triángulo $A'B'C'$ y por lo tanto el ángulo en A es igual al ángulo en A'.

6.25 La suma de los ángulos internos de un triángulo suman 180^o.

Sea el triángulo ABC arbitrario. Hagamos pasar a C prolongaremos los lados AC y BC. Por la posición anterior α es igual al ángulo de triángulo en A, β es igual al ángulo en C por opuestos por el vértice.



Pero por construcción $\alpha+\beta+\gamma=180^o$ que es igual a la suma de todos los ángulos del triángulo.

6.25.1 Proposición. Tales de Mileto

Si dos rectas concurrentes son cortadas por dos rectas paralelas, los segmentos que estas determinan sobre aquellas son proporcionales.

Sea las rectas AB y $A'B'$ concurrentes en el punto O y $A'A$ y $B'B$ dos rectas paralelas que las cortan. Observemos lo siguiente:

1. Los triángulos $AA'B$ y $AA'B'$ tienen la misma área porque comparten la misma base y altura h.
2. El área del triángulo $AA'B$ también se puede calcular como $\frac{|AB|x}{2}$ puesto que x es perpendicular a AB y el área de $AA'B'$ como $\frac{|A'B'x'|}{2}$ porque x' es perpendicular a $A'B'$ y como el área de $AA'B$ y $AA'B'$ son iguales

$$\frac{|\,AB\,|\,x}{2} = \frac{|\,A'B'\,|\,x'}{2} = \frac{x}{x'} = \frac{|\,A'B'\,|}{|\,AB\,|} \tag{6.111}$$

3. El área del triángulo OAA' se puede calcular de dos formas diferentes:

$$\frac{|\,OA\,|\,x}{2} = \frac{|\,OA'\,|\,x'}{2} \tag{6.112}$$

$$\frac{x}{x'}\frac{|\,OA'\,|\,x}{|\,OA\,|} \tag{6.113}$$

$$\frac{|\,A'B'\,|}{|\,AB\,|} = \frac{|\,OA'\,|}{|\,OA\,|} \tag{6.114}$$

4. En general si $\frac{a}{b} = \frac{c}{d}$ entonces se tiene $\frac{a}{b} = \frac{c}{d} = \frac{a+c}{b+d}$ para cualesquiera números reales a,b,c,d. Esto es así por que $\frac{a}{b} = \frac{c}{d}$ implica $ad = bc$, sumando cd en ambos lados $ad + cd = bc + cd$, factorizando $d(a+c) = c(b+d)$, $\frac{c}{d} = \frac{a+c}{b+d}$ en nuestra caso

$$\frac{|\,OA'\,|}{|\,OA\,|} = \frac{|\,A'B'\,|}{|\,AB\,|} = \frac{|\,OA'\,| + |\,A'B'\,|}{|\,OA\,| + |\,AB\,|} \tag{6.115}$$

$$\frac{|\,OA\,|}{|\,OB\,|} = \frac{|\,OA'\,|}{|\,OB'\,|} \tag{6.116}$$

5. Si trazamos una paralela a OB por A' y tomando B' en vez de O y repetimos el razonamiento anterior.

$$\frac{|\,B'O\,|}{|\,B'B\,|} = \frac{|\,A'O\,|}{|\,CB\,|} \tag{6.117}$$

pero, $|\,CB\,| = |\,A'A\,|$

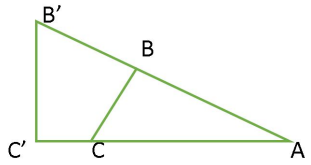
$$\frac{|\,B'O\,|}{|\,B'B\,|} = \frac{|\,A'O\,|}{|\,A'A\,|} \tag{6.118}$$

$$\frac{|\,AA'\,|}{|\,BB'\,|} = \frac{|\,OA'\,|}{|\,OB'\,|} \tag{6.119}$$

Definición: Son triángulos son semejantes si sus ángulos son iguales dos a dos. A los lados opuestos a los ángulos iguales se le llama correspondientes.

Proposición Si dos triángulos son semejantes entonces sus lados correspondientes son proporcionales.

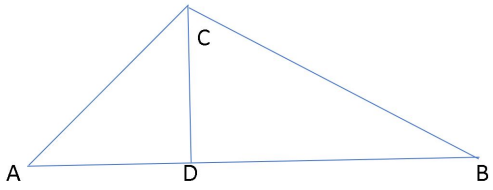
Si dos triángulos tienen sus tres lados iguales dos a dos, podemos hacer coincidir dos de sus ángulos iguales en un vértice A como se muestra en la figura. Como triángulos ABC y $AB'C'$ son semejantes, los ángulos en B y B' son iguales, lo mismo los ángulos en C y C'. Esto quiere decir que la recta BC y $B'C'$ son paralelas y por tanto podemos aplicar el teorema de Tales, es decir los lados correspondientes son proporcionales



$$\frac{|AB|}{|AB'|} = \frac{|AC|}{|AC'|} = \frac{|BC|}{|BC'|} = \lambda \tag{6.120}$$

Proposición: Si dos triángulos tienen dos ángulos iguales, entonces son semejantes. Esto se debe a que el tercero suma 180^o, con los dos primeros.

Problema: Usando semejanza probar el teorema de Pitágoras.



Sea ABC el triángulo original.Si C trazamos una perpendicular al lado AB que lo toque en D, Entonces los triángulos DBC y ABC, ADC y ABC son semejantes y

se cumple

$$\frac{|\,BC\,|}{|\,AB\,|} = \frac{|\,DB\,|}{|\,BC\,|} \tag{6.121}$$

$$|\,BC\,|^2 = |\,AB\,||\,DB\,| \tag{6.122}$$

$$\frac{|\,AC\,|}{|\,AB\,|} = \frac{|\,AD\,|}{|\,AC\,|} \tag{6.123}$$

$$|\,AC\,|^2 = |\,AB\,||\,AD\,| \tag{6.124}$$

$$|\,AC\,|^2 + |\,BC\,|^2 = |\,AB\,|\,(|\,DB\,| + |\,AD\,|) = |\,AB\,||\,AB\,| = |\,AB\,|^2$$

6.26 Magnitudes conmesurables

Los griegos creían que todas las magnitudes eran conmensurables entre si. Es decir dadas dos magnitudes A y B estas son conmensurables si existe una unidad U que mide a ambas, $A = pu, B = qu$ donde p y q son enteros positivos. Lo anterior implica que el cociente $\frac{A}{B} = \frac{p}{q}$ se puede representar mediante un número racional en este caso $\frac{p}{q}$. La intuición les decía y nos dice que esta u siempre existe, aunque sea muy pequeña.
Proposición $\sqrt{2}$ y 1 no son conmensurables.
Prueba (por contradicción)
Supongamos que $\frac{\sqrt{2}}{1} = \frac{p}{q}$ (suponemos lo contrario de lo que queremos probar) como parte de la hipótesis tomamos a p/q irreducible $(p,q) = 1$.
Elevando al cuadrado $2 = \frac{p^2}{q^2}$, $p^2 = 2q^2$ esto implica que p^2 es par y si el cuadrado de un número es par, el mismo número es par, es decir $p = 2m$, $4m^2 = 2q^2$, $2m^2 = q^2$. Por lo que q^2 es par y por lo tanto q también es par. Y esto es una contradicción con la hipótesis, puesto que si p y q son pares $\frac{p}{q}$ no puede ser irreducible, por lo tanto y en conclusión $\sqrt{2}$ no es un número racional.
Si a cada magnitud le asignamos un punto en la recta numérica y a cada punto un número, a $\sqrt{2}$ no le corresponde un número racional entonces decimos que $\sqrt{2}$ es un número irracional. Finalmente se concluye que en la recta numérica hay números racionales e irracionales y a todos juntos los nombramos números reales $\mathbb{R}$.

6.27 Expansión decimal de un número racional

El conjunto de números racionales $Q = \{\frac{p}{q} : p,q \in \mathbb{Z}\}$ representan las magnitudes que podemos expresar como el cociente de dos números enteros. En términos de su representación decimal, la parte fraccionaria puede ser finita como $\frac{1}{8} = 0.125$ a periódica $\frac{1}{3} = 0.333...$, aquí el periodo es de una cifra y lo indicamos con una barra encima del periodo $\frac{1}{3} = 0.\overline{3}$.

Proposición. Si un número tiene una representación decimal donde su parte fraccionaria es finita o infinita periódica, entonces es un número racional.

Prueba. Supongamos que la parte fraccionaria del número $X = a \bullet X_1..X_n$ es finita, donde a es la parte entera expresada en cifras decimales. Entonces $X = \frac{aX_1..X_n}{10^n}$ y esto es un número racional de la forma $\frac{p}{q}$. Supongamos ahora que $X = a \bullet \overline{X_1..X_n}$ tiene parte fraccionaria periódica. Entonces $10^n X - X = aX_1..X_n \bullet X_1..X_n - a \bullet X_1..X_n$ en esta resta se cancelan las partes fraccionarias periódicas y nos queda $(10^n - 1)X = aX_1..X_n - a$ es decir $X = \frac{aX_1..X_n - a}{(10^n - 1)}$ que es racional.

Ejemplo $X = 0.\overline{3}$, $10X - X = 3.\overline{3} - 0.\overline{3} = 3$, $9X = 3$, $X = \frac{1}{3}$ corolario. Si un número esta representado por cifras decimales y su parte fraccionaria no es finita ni infinita periódica. Entonces es un número irracional.

Ejemplo. $\sqrt{2} = 1.41...$

6.28 Probabilidad

Definición. Una probabilidad o función de probabilidad. Es una función que mapa un espacio de eventos E al intervalo $[0,1]$ de números reales. $P : E \to [0,1]$, E mismo es un evento $P(E) = 1$ y cualquier subconjunto de E, $A \subseteq E$. También es un evento, en particular $P(\phi) = 0$ donde ϕ es el conjunto vacío y si $A \cap B = \phi$ (eventos disjuntos) entonces $P(A \cup B) = P(A) + P(B)$ (esto es dado si se interpreta, la probabilidad como área).

Si $A \cap B \neq \phi$, podemos hablar de la probabilidad condicional de A dado B $P(A \mid B) = \frac{P(A,B)}{P(B)}$. Si visualizamos la probabilidad como área, esta probabilidad condicional es la proporción de la intersección $P(A,B)$ entre el área representada por la probabilidad de B.

De manera similar podemos expresar $P(B \mid A) = \frac{P(A,B)}{P(B)}$, $P(A,B) = P(B \mid A)P(A)$, pero también $P(A,B) = P(A \mid B)P(B)$. Por lo tanto $P(B \mid A)P(A) = P(A \mid B)P(B)$ y por ejemplo $P(A \mid B) = \frac{P(B|A)P(A)}{P(B)}$ (fórmula de Bayes) que permite expresar una probabilidad condicional en términos de su dual. También nótese que si el evento A no depende del B, entonces $P(A \mid B) = P(A)$ y en este caso tenemos $P(A,B) = P(A \mid B)P(B)$. Es decir cuando dos eventos son independientes su probabilidad conjunta es igual al producto de sus probabilidades. Cuando se tiene un experimento y se quiere estimar la probabilidad de un evento esto se puede hacer como casos favorables / casos posibles.

6.28.1 Variables aleatorias

Dado un experimento a los resultados y a los conjuntos de resultados les llamamos eventos, a los cuales les podemos asignar una probabilidad. Si a este espacio de eventos le llamamos E, entonces una variable aleatoria es una función $X : E \to \mathbb{R}$. Si X es una variable aleatoria discreta $P(X = X_0 \in \mathbb{R} = P(X^{-1}(a))$, donde $X^{-1}(a)$ representa la imagen inversa de X respecto a. Por ejemplo si el experimento consiste en arrojar 2 dados los resultados son parejas (i,j), $1 \leq i,j \leq 6$ y podemos considerar la variable aleatoria $(i,j) \xrightarrow{X} i+j$. Con esta variable aleatoria podemos preguntarnos por $P(X = 3) = P(X^{-1}(3)) = P(\{(1,2)\} \cup \{(2,1)\}) = P(\{(1,2)\}) + P(\{(2,1)\} = \frac{1}{36} + \frac{1}{36} = \frac{2}{36}$.

En el caso que la variable aleatoria X sea continua se busca $P(X > X_0)$ o $P(X_o < X < X_1)$, si se conoce la función de densidad de probabilidad fx esto queda como $P(X < X_0) = \int_{-\infty}^{X_0} fx(u)du$ por aditividad $P(X < X_0) + P(X_0 < X < X_1) = P(X < X_1)$, por tanto $P(X_0 < X < X_1) = P(X < X_1) - P(X - X_0) = \int_{X_0}^{X_1} fx(u)du$ si $dF_x(u) = f_x(u)du, \int_{X_0}^{X_1} fx(u)du = \int_{X_0}^{X} dFx(u)$, a $F_x(\bullet)$ se le conoce como la función de distribución de probabilidad de X. Entonces $dF_x(X)$ es la probabilidad de que la variable aleatoria X caiga en un intervalo diferencial dx. Como $P(x < +\infty) = 1$, su función de densidad de probabilidad debe cumplir $\int_{-\infty}^{\infty} fx(u)du = 1$.

6.29 Cálculo discreto

Considere una secuencia a_n de números reales, es decir una función de los naturales $\mathbb{N}$ a los reales $\mathbb{R}$. $n \in \mathbb{N} \to a_n \in \mathbb{R}$, el cálculo discreto se establece a partir de dos operadores: el operador suma Σ y el operador diferencia, que activan sobre secuencias y son lineales.

$$\sum_{i=1}^{n} a_i = a_1 + ... + a_n, \ \Delta a_R = a_R - a_{R-1}, \text{ así activan.}$$

La linealidad consiste en que abren sumas y sacan constantes.

1. $\sum_{i=1}^{n}[ca_i + db_i] = [ca_1 + .. + ca_n] + [da_1 + .. + da_n] = c[a_1 + .. + a_n] + d[b_1 + .. + b_n] = c\sum_{i=1}^{n} a_i + d\sum_{i=1}^{n} b_i$, esto prueba la linalidad para el operador Σ

2. $\Delta(ca_R + db_R) = (ca_R + cb_R) - (ca_{R-1} + db_{R-1}) = (ca_R - ca_{R-1}) + (cb_R - db_R) = c(a_R - a_{R-1}) + d(b_R - bR - 1) = c\Delta a_R + d\Delta b_R$, esto prueba la linealidad para Δ.

6.29.1 Teorema fundamental del cálculo discreto.

El resultado más importante del cálculo discreto, tiene que ver con la aplicación uno después del otro de los operadores Δ y Σ.

1. $\sum_{R=1}^{n} \Delta a_R = \sum_{R=1}^{n}(a_R - a_{R-1}) = (a_1 - a_0) + (a_2 - a_1) + .. + (a_{R-1} - a_{R-2}) + (a_n - a_{n-1}) = a_n - a_0)$, después de un proceso de cancelación, si $a_0 = \varnothing \sum_{R=1}^{n} \Delta a_R = a_n$. Como si Σ y Δ fueran inversos.

2. $\Delta\sum_{R=1}^{n} a_R = \sum_{R=1}^{n} a_R - \sum_{R=1}^{n-1} a_R = a_n + \sum_{R=1}^{n-1} a_R - \sum_{R=1}^{n-1} a_R = a_n$. Nuevamente los efectos de Δ y Σ se cancelan.

6.29.2 Aplicación del cálculo discreto a la suma de los primeros números naturales y sus potencias.

Considere el problema de encontrar la suma de $1 + 2 + 3 + ... + n = \sum_{R=1}^{n} R$, para encontrar la fórmula que nos permita calcular esta suma, sea.

$\Delta R^2 = R^2 - (R-1)^R = R^2 - R^2 + 2R - 1 = 2R - 1$, sumando de 1 a n
$n^2 = \sum_{R=1}^{n} \Delta R^2 = \sum_{R=1}^{n} (2R-1) = 2\sum_{R=1}^{n} R - \sum_{R=1}^{n} 1$, usando el Teorema fundamental y la literalidad de Σ

$n^2 = 2\sum_{R=1}^{n} R - n$, $2\sum_{R=1}^{n} R = n^2 + n$, finalmente $\sum_{R=1}^{n} R = \frac{n^2+n}{2} = \frac{n(n+1)}{2}$

Si queremos encontrar una fórmula en términos de n para $\sum_{R=1}^{n} R^2$ procedemos de la misma manera solo que ahora empezamos con la diferencia de R^3.

$\Delta R^3 = R^3 - (R-1)^3 = R^3 - (R^3 - 3R^2 + 3R - 1) = 3R^2 - 3R + 1$, sumando en ambos lados

$n^3 = \sum_{R=1}^{n} \Delta R^3 = \sum_{R=1}^{n} (3R^2 - 3R + 1) = 3\sum_{R=1}^{n} R^2 - 3\sum_{R=1}^{n} R + \sum_{R=1}^{n} 1 =$
$3\sum_{R=1}^{n} R^2 - \frac{3n(n+1)}{2} + n$, $3\sum_{R=1}^{n} R^2 = n^3 + \frac{3n(n+1)}{2} - n = \frac{2n^3+3n^2+3n-2n}{2} = \frac{2n^3+3n^2+n}{2}$, finalmente $\sum_{R=1}^{n} R^2 = \frac{2n^3+3n^2+n}{6}$, que es la fórmula que andábamos buscando.

Tarea. Encontrar fórmulas para $\sum_{R=1}^{n} R^3$, $\sum_{R=1}^{n} R^4$ y $\sum_{R=1}^{n} R^5$ y demostrar su validez por inducción. También probar por inducción $\sum_{R=1}^{n} R = \frac{n(n+1)}{2}$ y $\sum_{R=1}^{n} R^2 = \frac{2n^3+3n^2+n}{2}$

6.29.3 El Triángulo de Pascal

Considere el problema de encontrar los coeficientes en el polinomio de grado n, dado por la expresión $(a+b)^n$, para hacerlo en general, veamos que ocurre en algunos casos particulares. $(a+b)^0 = 1$, $(a+b)^1 = a+b$, $(a+b)^2 = a^2 + 2ab + b^2$, para ver que ocurre para $n = 3, 4, etc.$ acomodemos estas expresiones en forma vertical y tomemos en cuenta que $(a+b)^n = (a+b)^{n-1}(a+b)$, notemos que cada término de $(a+b)^{n-1}$ es multiplicado para $a+b$ y por b y luego sumado y esto se repite para la siguiente potencia. Esto lo podemos ir haciendo como lo muestra el diagrama. Las flechas izquierdas representan multiplicación por a y las derechas por b.

(DIAGRAMA)

$(a+b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4$ Si únicamente nos fijamos en el triángulo formado por los coeficientes, tenemos lo siguiente. De acuerdo con esto tenemos

unas coordenadas triangulares dadas por $\binom{n}{R}$ donde n es el renglón en el triángulo y R la columna. Con estos coeficientes dados por las coordenadas triangulares podemos resolver nuestro problema inicial. $(a+b)^n = \sum_{R=0}^{n} \binom{n}{R} a^{n-b} b^R$ donde los $\binom{n}{R}$ son los coeficientes que andamos buscando.

6.30 Cálculo vectorial

A la extensión de las ideas y operadores del cálculo, de funciones $f:\mathbb{R}\to\mathbb{R}$, vectores y matrices se le conoce como cálculo vectorial.

1. Si $f:\mathbb{R}\to\mathbb{R}^m$ tenemos una curva parametrizada por ejemplo por un parámetro t. $x(t)=(x_1(t),..,X_n(t))$ y su derivada $\frac{d}{dt}x=(\frac{d}{dt}x_1(t),..,\frac{d}{dt}x_n(t)$ representa un vector tangente a la curva representada por $x(t)$.

2. $f:\mathbb{R}^n\to\mathbb{R}$ su gráfica representa una híper superficie en $\mathbb{R}^n$. En este caso f depende de n variables $x_1,..,x_n$ y se puede derivar haciendo variar a una sola de ellas por ejemplo x_i y manteniendo constantes a los demás. A esta variación de f que depende de la variación parcial ∂x de (x_i, la llamamos diferencial parcial de f y la denotamos por (∂f, por tanto la derivada parcial de f respecto a x_i, es $\frac{\partial f}{\partial x_i}$. La diferencial total de f que en general depende de los incrementos de todas las x_i, se calcula como una suma ponderada de todas las diferenciales de las variables independientes x_i y los pesos de ponderación son las derivadas parciales. $df=\sum_{R=1}^{n}\frac{\partial f}{\partial x}dx_R$ El vector formado por las derivadas parciales se le llama gradiente $\bigtriangledown f=(\frac{\partial f}{\partial x_1},..,\frac{\partial f}{\partial x_n})^t$, entonces $df=\bigtriangledown f\cdot dx$ la diferencial de f es el producto punto entre el gradiente de f y el vector formado por las diferenciales de las $x_i, dx=(dx_1,..,dx_n)^t$.

3. $f:\mathbb{R}^n\to\mathbb{R}^n$ estas funciones se conocen como campos vectoriales. Si $x=(x_1,..,x_n)$ su imagen f es un vector con componentes $f=(f_1(x),..f_m(x))$ y cada componente $f_1:\mathbb{R}^n\to\mathbb{R}$ es una función de $\mathbb{R}^4$ a $\mathbb{R}$. Por tanto, $df=(df_1,..,df_m)^t$. $df=Jdx$, donde J es una matriz de derivadas parciales $[J]_{ij}=\frac{\partial f_i}{\partial x_j}$.

Para calcular la segunda derivada de $f:\mathbb{R}^n\to$, tenemos
$d\bigtriangledown f=(d\frac{\partial f}{\partial x_1},..,d\frac{\partial f}{\partial x_n})=(\sum_{R=1}^{n}\frac{\partial}{\partial x_R}[\frac{\partial f}{\partial x_1}]dx_R,..,\sum_{R=1}^{n}\frac{\partial}{\partial x_R}[\frac{\partial f}{\partial x_n}]dx_R)$.
$d\bigtriangledown f=(\sum_{R=1}^{n}\frac{\partial^2 f}{\partial x_R\partial x_1}dx_R,..,\sum_{R=1}^{n}\frac{\partial^2 f}{\partial x_n\partial x_n}dx_R)$ esto en forma matricial

$$\begin{pmatrix}\frac{\partial^2 f}{\partial x_1\partial x_1} & \ddots & \frac{\partial^2 f}{\partial x_n\partial x_1}\\ \frac{\partial^2 f}{\partial x_1\partial x_n} & \ddots & \frac{\partial^2 f}{\partial x_n\partial x_n}\end{pmatrix}\begin{pmatrix}dx_1\\ \vdots\\ dx_n\end{pmatrix}$$

, donde el gradiente de $\bigtriangledown f$ es la primer derivada de f.

La matriz de arriba es la segunda derivada de f y es conocida como el Hessiano, como en general las parciales conmutan. $[H]_{iS} = \frac{\partial^2 f}{\partial x_i dx_j}$, con H matriz Hessiano.

Para calcular un volumen irregular, por ejemplo el volumen bajo una superficie $f(x,y)$ (gráfico) Aquí el elemento de volumen lo podemos representar como $[\int_a^b f(x \cdot y)dy]dx$, en este caso $\int_a^b f(x,y)dy$ nos da () para una x dada una base irregular y por dx () obtenemos una pequeña rebanada de volumen que al integrar respecto de x nos da todo el volumen V. $\int_c^d \int_a^b f(x,y)dydx = V$ y esto se puede generalizar para $\mathbb{R}^4$.

6.31 Los polinomios de Bernoulli

A finales del siglo XVII Jacob Bernoulli notó que la expresión para evaluar la suma de R ponencias de números enteros era un polinomio de orden $R+1$, que se podía representar como la integral de ϕ a n de un polinomio de grado R, conocido actualmente como polinomio de Bernoulli.

$1^R + 2^R + .. + (n-1)^R = \int_{\phi}^{n} B_R(x)dx$, esta expresión resulta interesante porque vincula el dominio discreto con el continuo. De la fórmula anterior tenemos
$1^R + 2^R + .. + (n-1)^R + n^R = \int_{\phi}^{n+1} B_R(x)dx$, restando
$[1^R + 2^R + .. + (n-1)^R + n^R] - [1^R + 2^R + .. + (n-1)^R] =$
$\int_{\phi}^{n+1} B_R(x)dx - \int_{\phi}^{n} B_R(x)dx = \int_{n}^{n+1} B_R(x)dx$ En limpio $\int_{n}^{n+1} B_R(x)dx = n^R$, haciendo variar a n en forma continua
$\frac{d}{dn}\int_{n}^{n+1} B_R(x)dx = \int_{n}^{n+1} \frac{d}{du} B_R(x)dx = [B_R(x)]^{n+1} = B_R(n+1) - B_R(n) = Rn^{R-1}$,
tomando la expresión $B_R(n+1) - B_R(n) = Rn^{R-1}$ para $0, 1, .., n-1$ y sumando.

$B_R(1) - B_R(0) + B_R(2) - B_R(1) + .. + B_R(n) - B_R(m) =$
$R[\varnothing^{R+1} + 1^{R+1} + .. + (n-1)^{R+1}] = \int_{\varnothing}^{n} B_R(x)dx$. $B_R(n) - B_R(\varnothing) = R\int_{\varnothing}^{n} B_r(x)dx$
Jacob Bernoulli encuentra una fórmula recurrente para calcular los polinomios de Bernoulli. $B_R(n) = R\int_{\varnothing}^{n} B_R(x)dx + B_R(\phi)$, a las constantes $B_R(\phi) = B_R$ se les conoce como número de Bernoulli y son necesarios en la fórmula de reclusión.

Tarea. Hacer un programa en Python o Sympy que calcule los números de Bernoulli. Y con los números de Bernoulli encontrar las fórmulas para la suma de los cuadrados y cubos.

6.32 Cálculo continuo

Tomemos en cuenta 3 tipos de cantidades: grandes, muy grandes y muy pequeñas. Cada una de ellas representada por sus respectivos números. Las cantidades grandes son las cantidades convencionales que asociamos a la percepción de nuestros sentidos. Las cantidades muy pequeñas se encuentran cercanas al cero y las muy grandes al infinito. La relación entre las cantidades convencionales y las muy pequeñas está dada por las siguientes reglas.

1. La combinación lineal de cantidades convencionales, es una cantidad convencional (cerradura).
2. La combinación lineal de cantidades muy pequeñas es una cantidad muy pequeña (cerradura).
3. La suma de una cantidad convencional A y una cantidad muy pequeña a, es la cantidad convencional A y $A+a=A$, propiedad de absorción o simplificación.

También las cantidades muy grandes son cerradas bajo combinaciones lineales y tienen una ley de absorción respecto a las cantidades convencionales.

6.32.1 La diferencial y sus propiedades

Sea x una variable que puede tomar cualquier número real, considere x' muy cercana a x'. De tal manera que $x'-x=a$ sea un valor muy pequeño, y a este valor muy pequeño lo llamamos la diferencial o pequeño incremento de x, y lo denotamos con una d seguida por la variable $dx=x'-x$ o $x'=x+dx$. Para el caso de una función $f(x)$ su diferencial esta dada por $df(x)=f(x+dx)-f(x)$.

Propiedades de la diferencial de una función: La diferencial d aplicada a una función $f(x)$, nos da su crecimiento (pequeño) en x.

1. La diferencial es un operador lineal $d[C_1f(x)+C_2g(x)]=C_1df(x)+C_2g(x)$, C_1 y C_2 constantes. Prueba: Por definición $d[C_1f(x)+C_2g(x)]=[C_1f(x+dx)+$

$C_2g(x+dx]-[C_1f(x)+C_2g(x)]=C_1[f(x+dx)-f(x)]+C_2[g(x+dx)-g(x)]=C_1df(x)+C_2dg(x)$

2. 2. La diferencial satisface la regla de Leibniz o del producto. $d[f(x)g(x)]=f(x)dg(x)+g(x)df(x)$ Prueba: Como $df(x)=f(x+dx)-f(x)yf(x+dx)=f(x)+df(x)d[f(x)g(x)]=[f(x+dx)g(x+dx)]-[f(x)g(x)]=[f(x)+df(x)][g(x)+dg(x)]-f(x)g(x)=f(x)g(x)+f(x)dg(x)+g(x)df(x)+df(x)dg(x)-f(x)g(x)=f(x)dg(x)+g(x)df(x)$ Aplicando la regla de simplificación.

6.32.2 Cálculo de algunas diferenciales

1. $dc=c(x+dx)-c(x)=c-c=\varnothing$, pero la constante c vista como función siempre vale c independientemente de su argumento.

2. Diferencial de la función identidad $f(x)=x$. $df(x)=f(x+dx)-f(x)=x+dx-x=dx$, por ser $f(x)$ la función identidad.

3. Diferencial de x^2.

   (a) Usando la regla de Leibniz. $d(x^2)=d(x\cdot x)=x\cdot dx+x\cdot dx=2xdx$, *Tomando* $f(x)=x\ y\ g(x)=x$

   (b) Usando únicamente la definición $d(x^2)=(x+dx)^2-x^2=x^2+2xdx+dx^2-x^2=2xdx+d^2x=(2x+dx)dx$, aplicando la regla de simplificación a $2x+dx=2x$ esto porque $2x$ es una cantidad finita x, dx es una cantidad muy pequeña.

4. $d\sqrt{x}$. Escribimos $x=\sqrt{x}\cdot\sqrt{x}$; $d(\sqrt{x}\cdot\sqrt{x})=dx$; $\sqrt{x}d\sqrt{x}+\sqrt{x}d\sqrt{x}=dx$; $2\sqrt{x}d\sqrt{x}=dx$; $d\sqrt{x}=\frac{dx}{2\sqrt{x}}$ resultado final.

6.32.3 La derivada de una función y sus propiedades

En algunos cálculos de la diferencial de una función $f(x)$, que hemos efectuado. Podemos observar que para obtener $df(x)$, tenemos que multiplicar dx por una expresión que depende de x, es decir una función $f(x)$. Esta función $f(x)$ establece

una relación proporcional (lineal) entre la diferencial de la variable independiente dx y la diferencial de la función $df(x)$, $df(x) = f(x)dx$ y a este factor de proporcionalidad le llamamos la derivada de $f(x)$ respecto a x, la denotamos como $f'(x)$ y nos mide que tanto crece localmente $f(x)$ en términos del crecimiento pequeño de x.

Si $df(x) = f'(x)dx$, entonces $f'(x) = \frac{df(x)}{dx}$ y entonces podemos definir el operador derivada como un cociente de diferenciales $\frac{d}{dx}f(x) = \frac{df(x)}{dx}$, lo interesante del asunto es que el operador derivada hereda las propiedades del operador diferencial.

6.32.4 Propiedades del operador derivada

1. Linealidad: $\frac{d}{dx}[c_1 f(x) + c_2 g(x)] = \frac{d[c_1 f(x)+c_2 g(x)]}{dx} = \frac{c_1 df(x)+c_2 dg(x)}{dx} = c_1 \frac{df(x)}{dx} + c_2 \frac{dg(x)}{dx} = c_1 \frac{d}{dx} f(x) + c_2 \frac{d}{dx} g(x)$
2. Regla de Leibniz: $\frac{d}{dx}[f(x)g(x)] = \frac{d[f(x)g(x)]}{dx} = \frac{f(x)dg(x)+g(x)df(x)}{dx} = f(x)\frac{dg(x)}{dx} + g(x)\frac{df(x)}{dx} = f(x)\frac{d}{dx}g(x) + g(x)\frac{d}{dx}f(x)$

6.32.5 Calculo de alguna derivadas

1. $\frac{d}{dx}c = \frac{dc}{dx} = \frac{\varnothing}{dx} = \varnothing$, c constante
2. $\frac{d}{dx}(x) = \frac{dx}{dx} = 1$
3. $\frac{d}{dx}(x^2) = \frac{d}{dx}(xx) = x\frac{d}{dx}(x) + x(\frac{d}{dx}(x)) = 2x$
4. $\frac{d}{dx}(x^3) = \frac{d}{dx}(xx^2) = x\frac{d}{dx}x^2 + x^2\frac{d}{dx}x = x(2x) + x^2 = 3x^2$
5. Demuestre por inducción que $\frac{d}{dx}x^n = nx^{n-1}$ (Tarea)
6. $\frac{d}{dx}\sqrt{x}$, $\frac{d}{dx}(\sqrt{x}\sqrt{x}) = \frac{d}{dx}x = 1 = \sqrt{x}\frac{d}{dx}\sqrt{x} + \sqrt{x}\frac{d}{dx}\sqrt{x}$; $2\sqrt{x}\frac{d}{dx}\sqrt{x} = 1$; $\frac{d}{dx}\sqrt{x} = \frac{1}{2\sqrt{x}}$
7. $\frac{d}{dx}[\frac{1}{x}]$; $1 = (x(\frac{1}{x}))$; $x\frac{d}{dx}(\frac{1}{x}) + \frac{1}{x}\frac{d}{dx}x = \frac{d}{dx}(1) = \varnothing$; $x\frac{d}{dx}(\frac{1}{x}) = -\frac{1}{x}$; finalmente $\frac{d}{dx}(\frac{1}{x}) = -\frac{1}{x^2}$
8. Derivadas del $sen(x)$ y $cos(x)$. Para obtener estas derivadas primero vamos a calcular sus diferenciales. $dcos(x) = cos(x + dx) - cos(x) = cos(x)cos(dx) -$

$sen(x)sen(dx) - cos(x) = [cos(dx) - 1]cos(x) - sen(x)sen(dx)$, tomando en cuenta que para dx muy pequeña $cos(dx) = 1$ y $sen(dx) = dx$, tenemos finalmente $dcos(x) = -sen(x) \cdot dx$ y haciendo algo parecido para $sen(x)$, tenemos $dsen(x) = cos(x)dx$ y para sus derivadas $\frac{d}{dx}cos(x) = -sen(x)$ y $\frac{d}{dx}sen(x) = cos(x)$

6.32.6 La regla de la cadena

Supongamos que una función f depende de otra g, es decir tenemos $f(g(x))$. En este caso tenemos dos dependencias, la de g respecto de x y la de f respecto de g, a esto se le llama una composición de funciones y usamos la regla de la cadena para calcular la derivada de f respecto de x $\frac{df(g(x))}{dx}$. Para lograrlo hacemos lo siguiente.

Sea $u = g(x)$, entonces $df(u) = \frac{d}{du}f(u)du$, pero $du = dg(x) = \frac{d}{dx}g(x)dx$, por lo tanto $df(u) = \frac{d}{du}f(g(x)) \cdot \frac{d}{dx}g(x)dx$ dividiendo ambos lados de la ecuación por dx, obtenemos $\frac{d}{dx}f(g(x)) = \frac{df(g(x))}{dx} = \frac{d}{du}f(g(x))\frac{d}{dx}g(x)$.

Regla de la cadena. La derivada de una composición de funciones es igual al producto de las derivadas de cada una de las funciones que intervienen en la composición.

Cálculo de la inversa de una función usando la regla de la cadena. La inversa $f(u)$ de una función $f(x)$, satisface $f^{-1}(f(x)) = x$, por lo tanto haciendo $u = f(x)$ y aplicando la regla de la cadena tenemos $x = f^{-1}(u)$. $\frac{d}{dx}f^{-u}(u) = \frac{d}{dx}x = 1$, $\frac{d}{du}f^{-1}(u)\frac{d}{dx}f(x) = 1$, $\frac{d}{du}f^{-1}(u) = \frac{1}{\frac{d}{dx}f(f^{-1}(u)}$

Uso de la regla de la cadena y de la fórmula de la inversa para calcular algunas derivadas.

1. $\frac{d}{dx}x^{-n}$, x^{-n} la podemos ver como $(x^n)^{-1}$, entonces $\frac{d}{dx}(x^4)^{-1} = -\frac{1}{(x^n)^2}\frac{d}{dx}x^4 = 3 - \frac{1}{x^{2n}}(\frac{1}{n}c^{n-1}) = -\frac{1}{n}x = -\frac{1}{n}x$, entonces $\frac{d}{dx}x^{-n} = -\frac{1}{n}x^{-n-1}$, usando $\frac{d}{du}u^{-1} = -\frac{1}{u^2}$ y $\frac{d}{dx}x^n = nx^{n-1}$

2. La derivada de una división de funciones $\frac{d}{dx}[\frac{f(x)}{g(x)}] = \frac{d}{dx}[f(x)[\frac{1}{g(x)}]] = f(x)\frac{d}{dx}[\frac{1}{g(x)}] + [\frac{1}{g(x)}]\frac{f(x)}{dx} = f(x)[-\frac{1}{g(x)^2}\frac{dg(x)}{dx}] + [\frac{1}{g(x)}]\frac{d}{dx}f(x)[\frac{g(x)}{g(x)}] = \frac{g(x)\frac{d}{dx}f(x) - f(x)\frac{dg(x)}{dx}}{g^2(x)}$. En palabras: La derivada de la división de dos funciones $\frac{f(x)}{g(x)} = \frac{g(x)\frac{d}{dx}f(x) - f(x)\frac{d}{dx}g(x)}{g(x)}$, es

igual al de abajo $g(x)$ por la derivada del de arriba $\frac{d}{dx}f(x)$ menos el de arriba $f(x)$ por la derivada del de abajo $\frac{d}{dx}g(x)$, entre el de abajo al cuadrado $g^2(x)$. Tarea. Siguiendo la idea de 3. Calcular $\frac{d}{du}arccos(u)$ y $\frac{d}{du}arctan(u)$.

3. Tarea. Use la fórmula del cociente para calcular la derivada de la tangente $\frac{d}{dx}tan(x)$

4. Derivada del arco seno. $\frac{d}{du}arcsen(u) = \frac{1}{\frac{d}{dx}sen(arcsen(u))} = \frac{1}{cos(arcsen(u))}$ Donde $u = sen(x)$ $\frac{d}{du}arcsen(u) = \frac{1}{cos(arcsen(u))} = \frac{1}{\sqrt{1-u^2}}$, usando $cos(\theta) = \sqrt{1-sen^2(\theta)}$