



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

DEPARTAMENTO
DE INFORMÁTICA

Universidad Técnica Federico Santa María

Departamento de Informática

Draft
Apuntes de
Computación Científica

2020-v0.552

Profesor: Claudio E. Torres
y un gran equipo, ver capítulo [10](#).

Índice general

1. Introducción a Álgebra Lineal	5
1.1. Operaciones con vectores y matrices	5
1.1.1. Vector	5
1.1.2. Matriz	5
1.1.3. Producto matriz-vector	5
1.1.4. Producto matriz-matriz	6
1.2. Espacios Vectoriales y aplicaciones lineales	6
1.2.1. Espacio Vectorial	6
1.2.2. Transformación Lineal	8
1.2.3. Combinación lineal	9
1.3. Dependencia lineal y bases generadoras	9
1.3.1. Dependencia lineal	9
1.3.2. Independencia lineal	9
1.3.3. Ejemplo	10
1.4. Propiedades de las matrices	13
1.4.1. Rango	13
1.4.2. Nullspace	13
1.4.3. Rank	13
1.4.4. Full Rank	13
1.4.5. Inversa	13
1.4.6. Transpuesta	14
1.4.7. Adjoint	14
1.4.8. En resumen	14
1.5. Operatoria Vectorial	15
1.5.1. Producto Interno	15
1.5.2. Vectores ortogonales	15
1.5.3. Componentes de un vector	15
1.5.4. Matrices unitarias	16
1.5.5. Normas Vectoriales	16
1.5.6. Normas Matriciales inducidas por normas vectoriales	18
1.6. Ejercicios Propuestos	20
2. Aritmética de punto flotante y pérdida de precisión	21
2.1. Fundamentos	21
2.2. Números Binarios	22
2.2.1. Estándar de "Punto flotante" en base 2 de Números Reales (R):	24
2.2.2. Machine Epsilon	25
2.2.3. Error de Redondeo	26

2.3. Representación de Máquina	28
2.3.1. Casos Especiales del Exponente:	29
2.3.2. Pérdida de Significancia	30
2.4. Ejercicios Propuestos	30
3. Raíces en 1D	31
3.1. Método de la Bisección	32
3.1.1. Algoritmo	32
3.1.2. ¿Qué tan exacto y qué tan rápido es el método de la bisección?	32
3.2. Iteración de Punto Fijo	33
3.2.1. Algoritmo	33
3.2.2. Convergencia	34
3.2.3. Geometría de la iteración de Punto Fijo (cobweb diagram)	35
3.2.4. Convergencia lineal de iteración de punto fijo	36
3.2.5. Método localmente convergente	37
3.3. Teoremas útiles	38
3.3.1. Teorema de los límites continuos	38
3.3.2. Teorema del valor intermedio	38
3.3.3. Teorema del Valor Medio	38
3.3.4. Teorema de Rolle	38
3.3.5. Teorema de Taylor (con residuo)	38
3.4. Método de Newton	38
3.4.1. Introducción	38
3.4.2. Definición	39
3.4.3. Error de iteración	39
3.4.4. Convergencia local y cuadrática	40
3.4.5. Convergencia Cuadrática	40
3.4.6. Convergencia Lineal del método de Newton	40
3.4.7. Convergencia local	41
3.4.8. Método de Newton Modificado	41
3.5. Método de la Secante	41
3.6. Criterios de parada	42
3.7. Un breve comentario para entender la relación entre Backward-error y Forward-error	42
3.8. The Wilkinson polynomial (Wilkinson 1994)	43
3.9. Sensibilidad en la búsqueda de raíces en 1D	43
3.9.1. Condicionamiento (“¡primer encuentro!”)	44
4. Sistemas de Ecuaciones Lineales	45
4.1. Métodos conocidos de resolución	46
4.2. Eliminación Gaussiana Simple	47
4.2.1. Algoritmo Gaussian Elimination (G.E.)	48
4.2.2. Cantidad de operaciones en G.E	49
4.3. Backward Substitution	49
4.3.1. Algoritmo	50
4.3.2. Número de Operaciones elementales	50
4.4. Factorización LU de A	51
4.4.1. Utilización de la factorización LU	52
4.4.2. Origen de los errores	54
4.4.3. Número de condición	55
4.5. Factorización PA = LU	56

4.5.1. Matriz de Permutación	56
4.5.2. Ejemplo	57
4.6. La Factorización de Cholesky	59
4.6.1. Submatriz	59
4.6.2. Propiedades de una matriz definida positiva y simétrica	59
4.6.3. Descomposición de matrices simétricas y definida positiva	60
4.7. Métodos Iterativos	63
4.7.1. Método de Jacobi	63
4.7.2. Método de Gauss-Seidel	64
4.7.3. Successive Over-Relaxation (SOR(w))	64
4.7.4. Tabla Resumen	65
4.7.5. Convergencia	65
4.7.6. Comentarios varios	66
4.8. Método del Gradiente Descendente	67
4.8.1. Minimizar una función cuadrática convexa	68
4.8.2. ¿Qué es y cómo se obtiene α sub k	68
4.8.3. Convergencia en matrices simétricas y definida positiva	70
4.9. Método del Gradiente Conjugado	71
4.9.1. Derivación del Método del Gradiente Conjugado	72
4.9.2. Análisis de la A-ortogonalidad	74
4.9.3. Gradiente Conjugado - Interpretación Gráfica	77
4.9.4. Convergencia del Algoritmo del Gradiente Conjugado para una matriz simétrica y definida positiva	77
4.9.5. Gradiente Conjugado - Versión 2	77
4.10. Método de Newton en R elevado a n	79
4.10.1. Algoritmo	81
5. Interpolación (1D)	82
5.1. Matriz de Vandermonde	83
5.2. Interpolación de Lagrange	83
5.2.1. Ejemplo	84
5.2.2. Teorema de la interpolación polinomial (Unicidad)	85
5.3. Diferencias divididas de Newton	86
5.3.1. Ejemplo	86
5.3.2. Algoritmicamente	87
5.3.3. Ejemplo	87
5.4. Error de Interpolación	88
5.4.1. Fenómeno de Runge	89
5.5. Interpolación de Chebyshev	90
5.5.1. Motivación	90
5.5.2. Teorema de Chebyshev	91
5.5.3. Interpretación Gráfica	92
5.5.4. Cambio de Intervalo	94
6. Splines	95
6.1. Propiedades de Spline Cúbica	96
6.1.1. Propiedad 1 (Continuidad)	96
6.1.2. Propiedad 2 (Diferenciabilidad)	97
6.1.3. Propiedad 3 (Continuidad en la segunda derivada)	97
6.2. Interpretación de las 3 propiedades	97

6.3. Cantidad de ecuaciones a satisfacer	97
6.4. Tipos de condiciones de borde para splines cúbicas	98
6.4.1. Spline Natural	98
6.4.2. Spline con curvatura ajustada	98
6.4.3. Spline Terminada parabólicamente	98
6.4.4. Spline cúbica sin nodo (Not-a-knot cubic spline)	99
6.5. Unicidad de la Spline Cúbica	99
6.5.1. Ejemplo	99
6.6. Ejercicios Propuestos	100
7. Mínimos Cuadrados	101
7.1. Generalización	103
7.1.1. Ejemplo	104
7.2. Tipos de modelos	106
7.2.1. Modelo lineal	106
7.2.2. Modelo cuadrático	106
7.2.3. Modelo periódico	107
7.2.4. Modelo exponencial	107
7.2.5. Ley de potencia	108
8. Descomposición QR	109
8.1. Ortogonalización de Gram-Schmidt y mínimos cuadrados	109
8.1.1. Matriz ortonormal	110
8.1.2. Descomposición QR reducida y full	110
8.1.3. Algoritmo Clásico de Gram-Schmidt para obtener la descomposición QR	113
8.1.4. ¿Cómo usamos la descomposición QR en problemas de mínimos cuadrados?	113
8.1.5. Ejemplo (4.14) del libro guía	114
8.1.6. Interpretación gráfica del proceso de ortonormalización	114
9. Método del residuo mínimo generalizado (GMRes)	115
9.1. Derivación de GMRes	115
9.2. Algorítmicamente	120
10. Agradecimientos	121

Capítulo 1

Introducción a Álgebra Lineal

1.1. Operaciones con vectores y matrices

1.1.1. Vector

Un vector es un arreglo unidimensional de números, donde podemos definir un vector columna como:

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \in \mathbb{R}^n,$$

o un vector fila:

$$\mathbf{u} = [u_1 \quad u_2 \quad \cdots \quad u_n].$$

Notar que la transpuesta de un vector columna es un vector fila y viceversa.

1.1.2. Matriz

Una matriz $A \in \mathbb{R}^{m \times n}$ es un arreglo bidimensional de números, con una cantidad m de filas y n columnas,

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

1.1.3. Producto matriz-vector

¿Existe alguna compatibilidad requerida?

$$\Rightarrow A \cdot \mathbf{u} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{\square} \end{bmatrix} = \begin{bmatrix} a_{11} \cdot u_1 + a_{12} \cdot u_2 + \cdots + a_{1n} \cdot u_{\square} \\ \vdots \\ a_{m1} \cdot u_1 + a_{m2} \cdot u_2 + \cdots + a_{mn} \cdot u_{\square} \end{bmatrix}.$$

¿Qué es $A\mathbf{u}$? ¿Vector o matriz?¹

¿Qué tal si vemos A como una colección de vectores columna?

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = [\mathbf{a}_1 \mid \mathbf{a}_2 \mid \cdots \mid \mathbf{a}_n],$$

donde $\mathbf{a}_k = \begin{bmatrix} a_{1k} \\ \vdots \\ a_{mk} \end{bmatrix},$

$$\Rightarrow A\mathbf{u} = u_1 \mathbf{a}_1 + u_2 \mathbf{a}_2 + \cdots + u_n \mathbf{a}_n = \sum_{i=1}^n u_i \mathbf{a}_i \iff \text{combinación lineal de las columnas de } A$$

¿Es esto importante?

1.1.4. Producto matriz-matriz

Sea $A \in \mathbb{R}^{m \times n}$ y $B \in \mathbb{R}^{n \times l}$,

$$AB = [\mathbf{a}_1 \mid \mathbf{a}_2 \mid \cdots \mid \mathbf{a}_n] [\mathbf{b}_1 \mid \mathbf{b}_2 \mid \cdots \mid \mathbf{b}_i \mid \dots \mid \mathbf{b}_l] = C \in \mathbb{R}^{m \times l}.$$

$$\Rightarrow \mathbf{c}_i = \sum_{k=1}^n b_{k,i} \mathbf{a}_k = A\mathbf{b}_i, \text{ para } i = 1 : n \text{ y donde } b_{k,i} \Rightarrow \text{el } k\text{-ésimo elemento del vector } \mathbf{b}_i.$$

1.2. Espacios Vectoriales y aplicaciones lineales

1.2.1. Espacio Vectorial

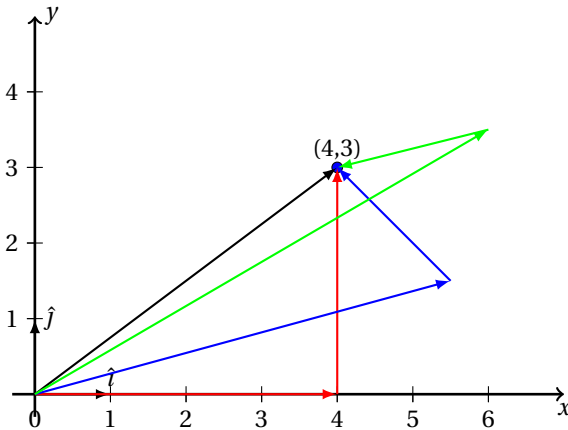
Un espacio vectorial es un conjunto no vacío V de objetos, llamados vectores, en el cual hay 2 operaciones definidas, llamadas “adición” y “multiplicación por escalar” (números reales), restringido a los siguientes axiomas. Los axiomas deben ser válidos para todos los vectores $\mathbf{u}, \mathbf{v}, \mathbf{w}$ de V y para todos los escalares c y d .

1. La suma de \mathbf{u} y \mathbf{v} , denotada como $\mathbf{u} + \mathbf{v}$, es en V
2. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ (conmutatividad)
3. $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$ (asociatividad de adición)
4. Hay un vector nulo $\mathbf{0}$ en V de tal forma $\mathbf{u} + \mathbf{0} = \mathbf{u}$
5. Para cada \mathbf{u} en V , hay un vector $-\mathbf{u}$ en V de tal forma que $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$

¹totalmente

6. El múltiplo escalar de \mathbf{u} por c , denotado por $c \cdot \mathbf{u}$, está en V
7. $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$
8. $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$
9. $c(d\mathbf{u}) = (cd)\mathbf{u}$
10. $1\mathbf{u} = \mathbf{u}$

Por ejemplo: El espacio \mathbb{R}^n consiste en todos los vectores columnas con “n” componentes, en particular:



Descomposición del vector $\langle 4, 3 \rangle$ en la base canónica (en rojo): $\langle 4, 3 \rangle = 4 \cdot \hat{i} + 3 \cdot \hat{j} = 4 \cdot \langle 1, 0 \rangle + 3 \cdot \langle 0, 1 \rangle$.

En la gráfica anterior se puede visualizar además la descomposición de $\langle 4, 3 \rangle$ en otras bases, identificados con los vectores verdes y azules.

1.2.1.1. ¿Qué significa hacer un cambio de base?

En el ejemplo anterior, se observa que:

$$4 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 3 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}.$$

Esto mismo puede hacerse con las otras 2 bases, supongamos que los 2 vectores azules o los 2 vectores verdes pueden escribirse de la siguiente forma: $\mathbf{b}_1 = \langle b_{1,1}, b_{1,2} \rangle$ y $\mathbf{b}_2 = \langle b_{2,1}, b_{2,2} \rangle$, respectivamente. Sin embargo en ese caso, no conocemos los escalamientos que debemos hacer de cada uno de esos vectores para poder obtener $\langle 4, 3 \rangle$, lo cual lo podemos escribir en la siguiente ecuación vectorial:

$$\alpha \begin{bmatrix} b_{1,1} \\ b_{1,2} \end{bmatrix} + \beta \begin{bmatrix} b_{2,1} \\ b_{2,2} \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}.$$

La cual se convierte en el siguiente sistema de ecuaciones lineales:

$$\underbrace{\begin{bmatrix} b_{1,1} & b_{2,1} \\ b_{1,2} & b_{2,2} \end{bmatrix}}_A \underbrace{\begin{bmatrix} \alpha \\ \beta \end{bmatrix}}_{\mathbf{b}} = \underbrace{\begin{bmatrix} 4 \\ 3 \end{bmatrix}}_{\mathbf{b}}.$$

En este caso, el problema se convirtió en resolver un sistema de ecuaciones lineales.

Desde el punto de vista Matemático, esto se resuelve de la siguiente forma:

$$A\mathbf{x} = \mathbf{b} \Rightarrow \mathbf{x} = A^{-1}\mathbf{b},$$

donde A^{-1} es denominada la inversa de A . Lo cual puede inducir a pensar que es necesario obtener la matriz inversa de A (i.e. A^{-1}), sin embargo veremos en el curso que eso no es realmente necesario.

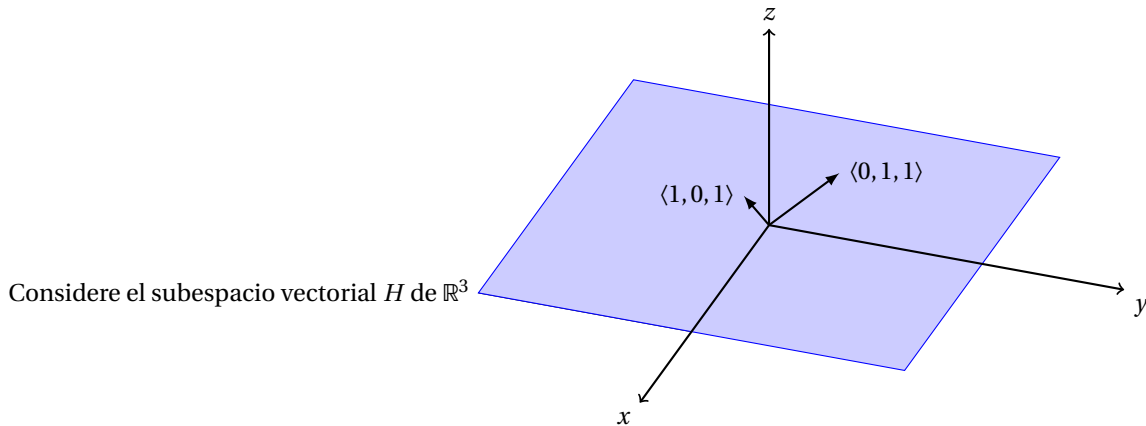
Un muy particular de sistema de ecuaciones lineales muy simple es el siguiente: $I\mathbf{x} = \mathbf{c} \Rightarrow \mathbf{x} = \mathbf{c}$, donde la matriz A del sistema de ecuaciones lineales es la matriz identidad I . La cual se define como:

$$I = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix}$$

1.2.1.2. Subespacio Vectorial

Un “subespacio vectorial H ” de un espacio vectorial V es un sub-conjunto de V que tiene 3 propiedades:

- El vector nulo de V está en H .
- H es cerrado bajo la adición. Esto es, para cada \mathbf{u} y \mathbf{v} en H , la suma $\mathbf{u} + \mathbf{v}$ está en H .
- H es cerrado bajo la multiplicación escalar. Esto es, para cada \mathbf{u} en H y cada escalar c , el vector $c\mathbf{u}$ está en H .



$$H = a \langle 0, 1, 1 \rangle + b \langle 1, 0, 1 \rangle = \text{span}(\langle 0, 1, 1 \rangle, \langle 1, 0, 1 \rangle)$$

¿Pertenece $(1, 1, 0)$ a H ?

1.2.2. Transformación Lineal

Se denomina Transformación Lineal a una aplicación tal que su dominio y codominio sean espacios vectoriales y que cumplan lo siguiente

$$\text{Sea } A \in \mathbb{R}^{n \times n}, \mathbf{x}, \mathbf{y} \in \mathbb{R}^n \text{ y } \alpha \in \mathbb{R} \Rightarrow A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y} \text{ y } A(\alpha\mathbf{x}) = \alpha A\mathbf{x}$$

1.2.3. Combinación lineal

Una Combinación Lineal de dos o más vectores es el vector que se obtiene al sumar estos vectores multiplicados por escalares.

$$\text{Sea } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \alpha \in \mathbb{R}, \alpha \neq 0 \text{ y } \mathbf{x} \neq \beta \mathbf{y}$$

$$\text{Ej: } \mathbf{v}_1 = \mathbf{x} + \mathbf{y}$$

$$\mathbf{v}_2 = \alpha \mathbf{x}$$

$$\mathbf{v}_3 = \frac{1}{\alpha} \mathbf{y}$$

En el ejemplo anterior:

¿Son \mathbf{v}_2 y \mathbf{x} linealmente independientes?

¿Son \mathbf{v}_1 , \mathbf{x} e \mathbf{y} linealmente independientes?

¿Son \mathbf{v}_2 y \mathbf{v}_3 linealmente independientes?

1.3. Dependencia lineal y bases generadoras

1.3.1. Dependencia lineal

Se dice que ciertos vectores son Linealmente Dependientes si hay una combinación lineal de ellos que es igual al vector cero, sin que todos coeficiente de la combinación lineal sea cero:

$$\sum_{i=1}^m \lambda_i \mathbf{x}_i = \mathbf{0} \Rightarrow \exists \lambda_i \neq 0.$$

1.3.2. Independencia lineal

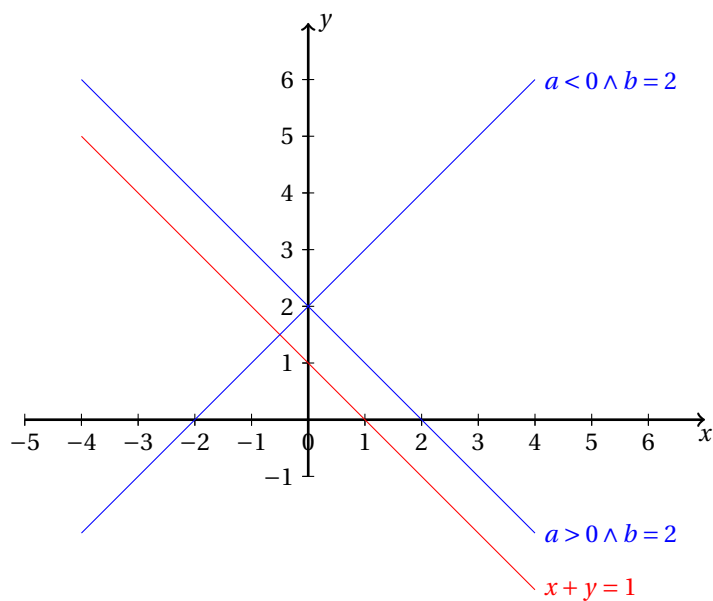
Se dice que ciertos vectores son Linealmente Independientes si ninguno de ellos puede ser escrito con una combinación lineal de los restantes.

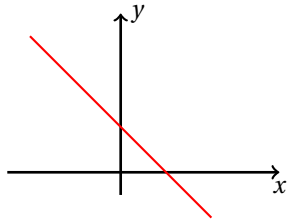
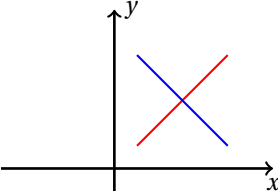
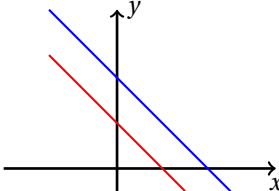
$$\sum_{i=1}^m \lambda_i \mathbf{x}_i = \mathbf{0} \Rightarrow \lambda_i = 0, \quad \forall i \in \{1, \dots, m\}.$$

1.3.3. EjemploResolver $x + y = 1$, $ax + y = b$

$$ax + y = b \Rightarrow y = b - ax$$

$$x + y = 1 \Rightarrow y = 1 - x$$

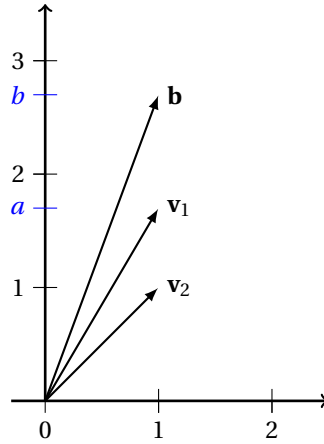


Caso 1	Caso 2	Caso 3
$a = 1$	$a \neq 1$	$a = 1$
$b = 1$		$b \neq 1$
$y = 1 - x$	$x = \frac{b-1}{a-1} \wedge y = \frac{a-b}{a-1}$	
∞ soluciones	1 solución	No existe solución
		
Existe sobreposición de rectas	Las rectas intersecan en un solo punto	Las rectas son paralelas

Otra forma de verlo:

$$x + y = 1, ax + y = b \Rightarrow \begin{pmatrix} 1 & 1 \\ a & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ b \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ b \end{pmatrix} = 1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + b \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

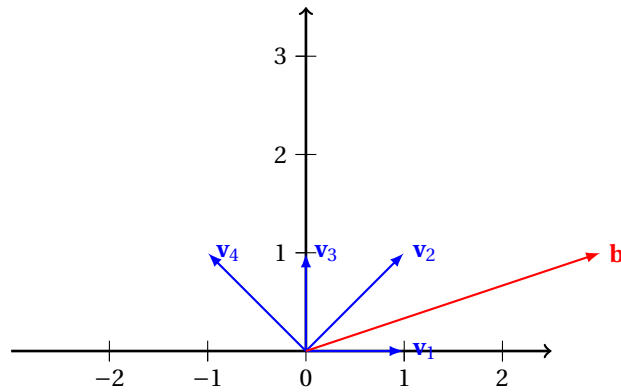
$$\Rightarrow x \underbrace{\begin{pmatrix} 1 \\ a \end{pmatrix}}_{\mathbf{v}_1} + y \underbrace{\begin{pmatrix} 1 \\ 1 \end{pmatrix}}_{\mathbf{v}_2} = \underbrace{\begin{pmatrix} 1 \\ b \end{pmatrix}}_{\mathbf{b}}$$



¿Podemos encontrar las coordenadas x e y para construir $\mathbf{b} = \begin{pmatrix} 1 \\ b \end{pmatrix}$
 como una combinación lineal entre $\mathbf{v}_1 = \begin{pmatrix} 1 \\ a \end{pmatrix}$ y $\mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$?

Por ejemplo:

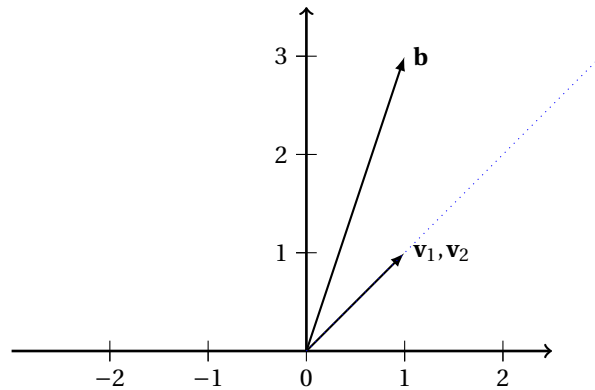
$$\begin{aligned} & \text{¿} \mathbf{b} = x \mathbf{v}_1 + y \mathbf{v}_2 ? \\ \Rightarrow & \text{si "a" es } \neq 1 \Rightarrow \mathbf{v}_1 \text{ y } \mathbf{v}_2 \text{ son linealmente independientes} \\ \Rightarrow & \mathbf{v}_1 \text{ y } \mathbf{v}_2 \text{ forman una base para } \mathbb{R}^2 \end{aligned}$$

Otro ejemplo

¿Cuál sería una mejor base para encontrar las coordenadas x e y ?

por ejemplo: $\mathbf{b} = x\mathbf{v}_3 + y\mathbf{v}_4$

Volviendo al caso anterior y considerando $a = 1$:



Considerando el gráfico superior

¿Cuándo no hay solución?

¿Cuándo hay solución? ¿Interpretación gráfica?

1.4. Propiedades de las matrices

1.4.1. Rango

El "Rango" de una matriz A , $\text{Range}(A)$ es el espacio vectorial "spanned" (generado) por las columnas de A .

1.4.2. Nullspace

El Nullspace de A , $\text{null}(A)$, es el espacio vectorial tal que todos sus elementos \mathbf{x} no nulos satisfacen la ecuación $A\mathbf{x} = \mathbf{0}$.

Por ejemplo: considere $\mathbf{a}_1 \neq \mathbf{0}$, $A\mathbf{x} = [\mathbf{a}_1 \mid \mathbf{0}] \cdot \begin{bmatrix} 0 \\ c \end{bmatrix} = \mathbf{0}$, en este caso $\text{null}(A) = \text{span}(\langle 0, 1 \rangle)$

1.4.3. Rank

- Column Rank: es la dimensión del column space, es decir, el número de columnas linealmente independientes.
- Row Rank: es la dimensión del row space, es decir, el número de filas linealmente independientes.

1.4.4. Full Rank

La matriz $A \in \mathbb{R}^{m \times n}$ con $m \geq n$ tiene full rank si y solo si no mapea 2 vectores distintos al mismo vector.

1.4.5. Inversa

Una matriz es invertible si es cuadrada y full rank.

$$\Rightarrow AZ = ZA = \underbrace{I}_{\text{Identidad}} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

$Z = A^{-1}$, es la inversa de A

Recordar: Si $Z = A^{-1} \Rightarrow AZ = I$

\mathbf{b} : coeficientes (coordenadas) de la expansión de \mathbf{b} en las bases canónicas (vector de 0, con 1 en la posición i) $\mathbf{e}_i = \langle 0, 0, \dots, 0, 1, 0, \dots, 0 \rangle$.

$A^{-1}\mathbf{b}$: coordenadas de la expansión de \mathbf{b} en la base $\langle \mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n \rangle$.

$$A \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \leftrightarrow A\mathbf{x} = \mathbf{b} \leftrightarrow A \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \leftrightarrow \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = A^{-1} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

$$A \cdot \mathbf{x} = I \cdot \mathbf{b} \leftrightarrow x_1 \cdot \mathbf{a}_1 + x_2 \cdot \mathbf{a}_2 \dots + x_n \cdot \mathbf{a}_n = b_1 \cdot \mathbf{e}_1 + b_2 \cdot \mathbf{e}_2 + \dots + b_n \cdot \mathbf{e}_n$$

1.4.6. Transpuesta

La Transpuesta de una matriz A es la matriz A^T y se obtiene de la siguiente forma:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \Rightarrow A^T = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \end{bmatrix}$$

1.4.7. Adjoint

La matriz transpuesta conjugada, operador adjoint de una matriz A es la matriz A^* que cumple lo siguiente:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \Rightarrow A^* = \begin{bmatrix} \overline{a_{11}} & \overline{a_{21}} & \overline{a_{31}} \\ \overline{a_{12}} & \overline{a_{22}} & \overline{a_{32}} \end{bmatrix}$$

donde $z = a + ib \Rightarrow \bar{z} = a - ib$

Si $A \in \mathbb{R}^{m \times n} \Rightarrow A^* = A^T$

Si $A = A^* \Rightarrow A$ es hermitiana

Ej:

$$A = \begin{bmatrix} a & a - ic \\ a + ic & b \end{bmatrix} \Rightarrow A^* = \begin{bmatrix} a & a - ic \\ a + ic & b \end{bmatrix}$$

Observación: Hermitiana \neq Simétrica.

1.4.8. En resumen

Para $A \in \mathbb{R}^{m \times m}$, las siguientes condiciones son equivalentes:

1. A tiene inversa A^{-1}
2. $\text{Rank}(A) = m$
3. $\text{Range}(A) = \mathbb{R}^m$
4. $\text{null}(A) = \emptyset$
5. 0 no es un valor propio de A
6. 0 no es un valor singular (singular value) de A
7. $\det(A) \neq 0$

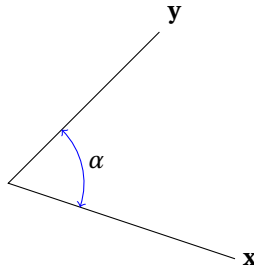
1.5. Operatoria Vectorial

1.5.1. Producto Interno

El Producto Interno de dos vectores columnas, es el producto del adjunto del primer vector por el segundo.

$$\text{Sea } \mathbf{x}, \mathbf{y} \in \mathbb{C}^m \Rightarrow \mathbf{x}^* \mathbf{y} = \sum_{i=1}^m \overline{x_i} y_i \Rightarrow \text{Euclidean length} \Rightarrow \|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^* \mathbf{x}} = \left(\sum_{i=1}^m |x_i|^2 \right)^{1/2}$$

También recuerde $\mathbf{x}^* \mathbf{y} = \|\mathbf{x}\|_2 \|\mathbf{y}\|_2 \cos(\alpha)$, donde α es el ángulo entre los vectores \mathbf{x} y \mathbf{y} .



Nota: $(AB)^* = B^* A^*$ y $(AB)^{-1} = B^{-1} A^{-1}$

1.5.2. Vectores ortogonales

- Dos vectores $\mathbf{x}, \mathbf{y} \neq \mathbf{0}$ son ortogonales si $\mathbf{x}^* \mathbf{y} = 0$

$$\text{Ej: } \mathbf{x} = \langle 1, 0 \rangle, \mathbf{y} = \langle 0, 1 \rangle \Rightarrow \mathbf{x} \cdot \mathbf{y} = 1 \cdot 0 + 0 \cdot 1 = 0$$

- Un conjunto de vectores son ortonormales si son ortogonales y la norma de cada uno es igual a uno.

1.5.3. Componentes de un vector

Sea $Q = \langle \mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_m \rangle$ un conjunto ortonormal, donde $\mathbf{q}_i \in \mathbb{R}^n$ y sea \mathbf{v} un vector arbitrario.

$$\Rightarrow \mathbf{v} = \mathbf{r} + \sum_{i=1}^m a_i \mathbf{q}_i$$

en donde \mathbf{r} es el residuo ortogonal a Q ,

$$\Rightarrow \mathbf{q}_k^* \mathbf{v} = \mathbf{q}_k^* \mathbf{r} + \sum_{i=1}^m a_i \mathbf{q}_k^* \mathbf{q}_i$$

$$\Rightarrow \mathbf{q}_k^* \mathbf{v} = 0 + a_k$$

$$\Rightarrow a_k = \mathbf{q}_k^* \mathbf{v} \Rightarrow \mathbf{v} = \mathbf{r} + \sum_{i=1}^m (\mathbf{q}_i^* \mathbf{v}) \mathbf{q}_i$$

¿Qué pasa si $m = n$?

$$\Rightarrow \mathbf{v} =$$

$$\Rightarrow \mathbf{v} = \sum_{i=1}^n (\mathbf{q}_i^* \mathbf{v}) \mathbf{q}_i = \sum_{i=1}^n \mathbf{q}_i (\mathbf{q}_i^* \mathbf{v}) = \sum_{i=1}^n (\mathbf{q}_i \mathbf{q}_i^*) \mathbf{v}$$

1.5.4. Matrices unitarias

Sea Q una matriz cuadrada $\in \mathbb{C}^{m \times m}$, $\Rightarrow Q^* = Q^{-1}$, $Q^* Q = I$

$$\begin{bmatrix} \mathbf{q}_1^* \\ \mathbf{q}_2^* \\ \vdots \\ \mathbf{q}_n^* \end{bmatrix} \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

¿Es útil tener matrices unitarias?

Ej: $A\mathbf{x} = \mathbf{b} \rightarrow Q\mathbf{x} = \mathbf{b} \rightarrow \mathbf{x} = Q^*\mathbf{b}$

¡Bastante más simple que calcular $A^{-1}\mathbf{b}$ en general!

Nota: En general no es recomendado calcular A^{-1} explícitamente. Es más “fácil” y más “rápido” obtener el vector $A^{-1}\mathbf{b}$.

1.5.5. Normas Vectoriales

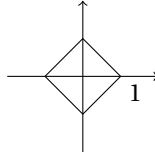
Las normas son la noción esencial del tamaño o distancia en un espacio vectorial. Una norma es una función $\|\bullet\| : \mathbb{C}^m \rightarrow \mathbb{R}$ que asigna un valor real (largo, tamaño) a cada vector. Formalmente, una norma debe satisfacer las siguientes 3 condiciones, para todo vector \mathbf{x} y \mathbf{y} , y escalar $\alpha \in \mathbb{C}$

- (1) $\|\mathbf{x}\| \geq 0$, y $\|\mathbf{x}\| = 0 \Rightarrow \mathbf{x} = \mathbf{0}$
- (2) $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ Desigualdad triangular
- (3) $\|\alpha\mathbf{x}\| = |\alpha| \cdot \|\mathbf{x}\|$

1.5.5.1. Las normas más FAMOSAS:

(las graficas corresponden a los isocontornos de las siguientes normas)

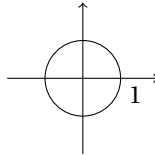
$$\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$$



$$\|\mathbf{x}\|_1 = 1$$

$$|x_1| + |x_2| = 1$$

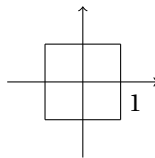
$$\|\mathbf{x}\|_2 = \left(\sum_{i=1}^n |x_i|^2 \right)^{\frac{1}{2}} = \sqrt{\mathbf{x}^* \mathbf{x}}$$



$$\|\mathbf{x}\|_2 = 1$$

$$\sqrt{x_1^2 + x_2^2} = 1$$

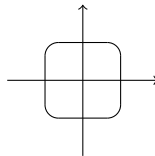
$$\|\mathbf{x}\|_\infty = \max_{i=1}^n |x_i|$$



$$\|\mathbf{x}\|_\infty = 1$$

$$\max\{|x_1|, |x_2|\} = 1$$

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}, \quad (1 \leq p \leq \infty)$$



$$\|\mathbf{x}\|_p = 1$$

(las ilustraciones corresponden a los casos de $n = 2$)

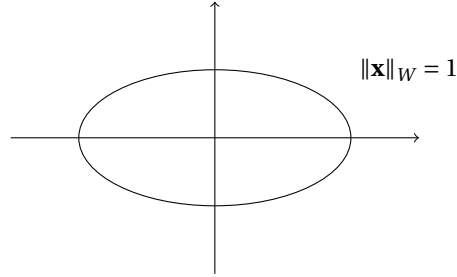
W-p-norms (Weighted p-norms) Una norma útil es la W-p-norms donde cada una de las coordenadas del espacio vectorial está dado por su propia ponderación. En general, dada una norma $\|\bullet\|$, una norma ponderada pueda ser escrita como

$$\|\mathbf{x}\|_W = \|W\mathbf{x}\| = \left\| \begin{bmatrix} w_1 & & & 0 \\ & w_2 & & \\ & & \ddots & \\ 0 & & & w_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right\| \quad \square \rightarrow \text{Alguna norma}$$

donde $w_i \neq 0$.

Ej:

$$\|\mathbf{x}\|_W = \left(\sum_{i=1}^n |w_i \cdot x_i|^2 \right)^{\frac{1}{2}} \quad \text{Los pesos } W_i \text{ hacen que se vea como una elipse}$$



1.5.5.2. Cauchy-Schwarz e Inegualdad de Hölder

Sea p y q que se satisfacen $\frac{1}{p} + \frac{1}{q} = 1$, donde $1 \leq p, q \leq \infty$, entonces la desigualdad de Hölder para cualesquier vectores \mathbf{x} e \mathbf{y}

$$|\mathbf{x}^* \mathbf{y}| \leq \|\mathbf{x}\|_p \cdot \|\mathbf{y}\|_q$$

La desigualdad de Cauchy-Schwarz es el caso especial de $p = q = 2$

$$|\mathbf{x}^* \mathbf{y}| \leq \|\mathbf{x}\|_2 \cdot \|\mathbf{y}\|_2$$

1.5.6. Normas Matriciales inducidas por normas vectoriales

Dada las normas vectoriales $\|\cdot\|_{(n)}$ y $\|\cdot\|_{(m)}$ en el dominio y rango de $A \in \mathbb{C}^{m \times n}$, \Rightarrow la norma matricial inducida $\|A\|_{(m,n)}$ es el número "C" más pequeño posible que hace válida la siguiente igualdad para todo $\mathbf{x} \in \mathbb{C}^n$:

$$\|A \cdot X\|_{(m)} \leq C \cdot \|X\|_{(n)}$$

1.5.6.1. Normas generales de matrices

Nótese que las normas matriciales no necesariamente deben ser inducidas de normas vectoriales, pero deben seguir las siguientes 3 condiciones:

$$(1) \|A\| \geq 0, \text{ y } \|A\| = 0 \Rightarrow A = 0$$

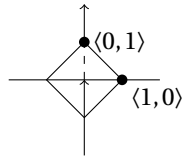
$$(2) \|A + B\| \leq \|A\| + \|B\|$$

$$(3) \|\alpha \cdot A\| = |\alpha| \cdot \|A\|$$

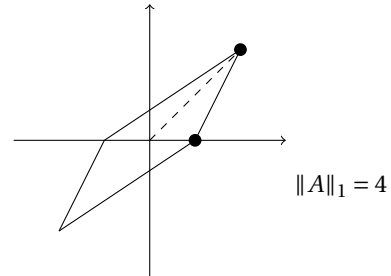
Ejemplo:

$$\text{Sea } A = \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}$$

1-norm:



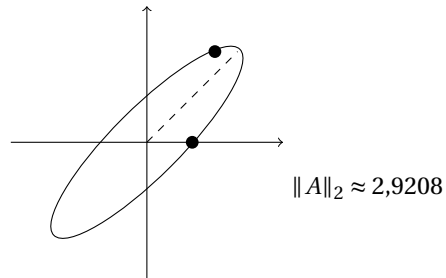
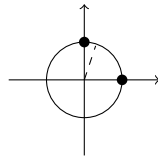
\Rightarrow



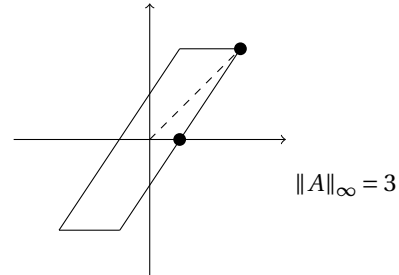
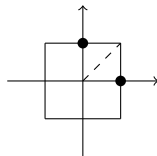
$$\|A\|_1, \|A \cdot \mathbf{x}\|_1 \leq C \cdot \|\mathbf{x}\|_1$$

$$\left\| \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right\|_1 \leq C \cdot \left\| \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right\|_1$$

2-norm:



∞ -norm:



1-Norm de una matriz \Rightarrow "Máxima suma **absoluta** de columnas"

∞ -Norm de una matriz \Rightarrow "Máxima suma **absoluta** de filas"

¡Se debe sumar el valor absoluto de cada elemento!

1.5.6.2. Norma de Frobenius

$$\begin{aligned}
\|A\|_F &= \left(\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \right)^{\frac{1}{2}} \\
&= \left(\sum_{j=1}^n \|a_j\|_2^2 \right)^{\frac{1}{2}} \\
&= \sqrt{\text{tr}(A^* A)} = \sqrt{\text{tr}(A A^*)}
\end{aligned}$$

donde $\text{tr}(B)$ denota la traza, i.e. la suma de los elementos de la diagonal de la matriz B .

Si la matriz tiene traza nula NO quiere decir que tiene determinante 0.
Como observación:

$$\begin{aligned}
\text{tr}(A) &= \text{tr}(B) \not\Rightarrow A = B \\
A = B &\Rightarrow \text{tr}(A) = \text{tr}(B)
\end{aligned}$$

Teorema: Para cualquier $A \in \mathbb{C}^{m \times n}$ y $Q \in \mathbb{C}^{m \times m}$ unitaria, tenemos:

$$\|QA\|_2 = \|A\|_2, \quad \|QA\|_F = \|A\|_F$$

Demostración:

$$\begin{aligned}
\|Q \cdot A\|_F &= \sqrt{\text{tr}((QA)^* QA)} = \sqrt{\text{tr}(A^* Q^* QA)} \\
&= Q^* = Q^{-1} \\
&= \sqrt{\text{tr}(A^* A)} = \|A\|_F
\end{aligned}$$

1.6. Ejercicios Propuestos

“El teorema de Pitágoras” asegura que para un conjunto de n vectores ortogonales \mathbf{x}_i ,

$$\left\| \sum_{i=1}^n \mathbf{x}_i \right\|^2 = \sum_{i=1}^n \|\mathbf{x}_i\|^2$$

1. Demuestre esto para el caso $n = 2$ por computación explícita de $\|\mathbf{x}_1 + \mathbf{x}_2\|$.
2. Demuestre lo anterior para el caso general
3. Sea $A \in \mathbb{C}^{m \times m}$ Hermitiana. Un eigenvector m (vector propio) de A es un vector no-nulo $\mathbf{x} \in \mathbb{C}$ de tal forma que $A \cdot \mathbf{x} = \lambda \cdot \mathbf{x}$ para algún $\lambda \in \mathbb{C}$, donde λ es el valor propio correspondiente (eigenvalue).
 - Demuestre que todos los eigenvalues de A son reales.
 - Demuestre que si \mathbf{x} y \mathbf{y} son eigenvectors correspondientes a eigenvalues diferentes $\Rightarrow \mathbf{x}$ y \mathbf{y} son ortogonales.

Capítulo 2

Aritmética de punto flotante y pérdida de precisión

2.1. Fundamentos

Ideas Útiles:

¿Cuál es la mejor forma de evaluar el polinomio $P(x)$?

$$P(x) = c_5 x^4 + c_4 x^3 + c_3 x^2 + c_2 x + c_1$$

Forma directa Simplemente multiplicamos y sumamos:

$$\begin{array}{rcl} c_5 x^4 & = & 4 \text{ multiplicaciones} \\ c_4 x^3 & = & 3 \text{ multiplicaciones y 1 suma} \\ c_3 x^2 & = & 2 \text{ multiplicaciones y 1 suma} \\ c_2 x & = & 1 \text{ multiplicación y 1 suma} \\ c_1 & = & 1 \text{ suma} \end{array}$$

Se necesitan 10 multiplicaciones y 4 sumas, un total de 14 operaciones elementales.

Forma directa pre-calculada Precalculamos y guardamos:

$$\begin{array}{l} a = x^2, \\ b = xa, \\ d = xb, \end{array}$$

Al momento de evaluar el polinomio tenemos:

$$\begin{array}{rcl} c_5 d (x \cdot x \cdot x^2) & = & 3 \text{ multiplicaciones} \\ c_4 b (x \cdot x^2) & = & 2 \text{ multiplicaciones y 1 suma} \\ c_3 a (x^2) & = & 1 \text{ multiplicación y 1 suma} \\ c_2 x & = & 1 \text{ multiplicación y 1 suma} \\ c_1 & = & 1 \text{ suma} \end{array}$$

Se necesitan 7 multiplicaciones y 4 sumas, un total de 11 operaciones elementales.

Multiplicación anidada o Metodo de Horner

$$P(x) = c_1 + x(c_2 + x(c_3 + x(c_4 + c_5x)))$$

$$\begin{array}{llll} f_1 & = & (c_4 + c_5x) & = \text{1 multiplicación y 1 suma} \\ f_2 & = & (c_3 + xf_1) & = \text{1 multiplicación y 1 suma} \\ f_3 & = & (c_2 + xf_2) & = \text{1 multiplicación y 1 suma} \\ & = & (c_1 + xf_3) & = \text{1 multiplicación y 1 suma} \end{array}$$

Se necesitan 4 multiplicaciones y 4 sumas, un total de 8 operaciones elementales.

2.2. Números Binarios

$$B = \dots b_3 b_2 b_1 \cdot b_{-1} b_{-2} b_{-3} \dots$$

donde cada dígito binario, o bit, es 0 o 1. Notar que existe una separación de la parte entera con la decimal hecha por el punto.

Ejemplo 1 Convertir el siguiente número binario en su forma decimal.

$$(1001.01)_2$$

$$\Rightarrow \begin{array}{ccccccc} 1 \cdot 2^3 & + & 0 \cdot 2^2 & + & 0 \cdot 2^1 & + & 1 \cdot 2^0 & + & 0 \cdot 2^{-1} & + & 1 \cdot 2^{-2} \\ 8 & + & 0 & + & 0 & + & 1 & + & 0 & + & \frac{1}{4} \end{array}$$

$$(9.25)_{10}$$

Ejemplo 2 Convertir el siguiente numero decimal a binario.

$$(28.7)_{10} = (28)_{10} + (0.7)_{10}$$

Comenzaremos con la parte entera:

$$28 \div 2 = 14 \text{ con resto } 0$$

$$14 \div 2 = 7 \text{ con resto } 0$$

$$7 \div 2 = 3 \text{ con resto } 1$$

$$3 \div 2 = 1 \text{ con resto } 1$$

$$1 \div 2 = 0 \text{ con resto } 1$$

Por lo tanto la parte entera se expresa como $(11100)_2$. Ahora su parte decimal:

$$,7 \times 2 = ,4 + 1$$

$$,4 \times 2 = ,8 + 0$$

$$,8 \times 2 = ,6 + 1$$

$$,6 \times 2 = ,2 + 1$$

$$,2 \times 2 = ,4 + 0$$

$$,4 \times 2 = ,8 + 0$$

.

.

.

Por lo tanto la parte decimal se expresa como $(.1011001100110\dots)_2 = (.1\overline{0110})_2$.

$$\therefore (28,7)_{10} = (11100.1\overline{0110})_2.$$

Pregunta

¿Qué número representa $(0.\overline{10})_2$ en base 10?

$$(0.101010\dots)_2$$

$$(0.\underbrace{101010\dots}_{2^{-1}+2^{-3}+2^{-5}\dots})_2 = \sum_{i=0}^{\infty} 2^{-(2i+1)} = \frac{1}{2} \cdot \sum_{i=0}^{\infty} (2^{-2})^i = \frac{1}{2} \cdot \frac{1}{1-\frac{1}{4}} = \frac{1}{2} \cdot \frac{4}{3} = \frac{2}{3}$$

Recuerde la serie geométrica:

$$1 + r + r^2 + r^3 + \dots = \sum_{i=0}^{\infty} r^i = \frac{1}{1-r}$$

$$|r| < 1$$

2.2.1. Estándar de “Punto flotante” en base 2 de Números Reales (\mathbb{R}):

(IEEE 754 Floating point standard)

$$\underbrace{\pm}_{\text{signo siempre 1}} \underbrace{1}_{\text{}} \cdot \underbrace{bbbbbb...}_{\text{mantisa}} \cdot 2^{p-\text{exponente}} \rightarrow \text{palabra (word)}$$

Ejemplo

$$\begin{array}{rclcl} 9 & = & (1001.)_2 & = & +1.001 \cdot 2^3 \Rightarrow 1 \cdot 2^3 + 1 \cdot 2^0 = (1 + 1 \cdot 2^{-3}) \cdot 2^3 = +1.001 \cdot 2^3 \\ 2 & = & (10.)_2 & = & +1.00 \cdot 2^1 \\ 0.5 & = & (0.1)_2 & = & +1.00 \cdot 2^{-1} \\ 0.75 & = & \frac{1}{2} + \frac{1}{4} = (0.11)_2 & = & +1.10 \cdot 2^{-1} \end{array}$$

Formatos de ejemplo: (bits usados)

Precisión	Signo	Exponente	Mantisa	Total
Single	1	8	23	32
Double	1	11	52	64
Long Double	1	15	64	80

Formato Normalizado: $\pm 1.bbb \dots \cdot 2^p$

(Justificado a la Izquierda - "left justified")

Ejemplo

- ¿Qué es “1” en “Double Precision”?

Signo : 1

Exponente: 11

Mantisa: 52

$$1 + \underbrace{1.\text{000}}_{\text{mantisa } (52 \rightarrow 0' s)} \cdot 2^0$$

- ¿Cuál es el siguiente número representable mayor a 1 en double precision?

$$1 + "2^{-52}" = +1.00\dots\dots$$

			1
--	--	--	---

 $\cdot 2^0$

2.2.2. Machine Epsilon

El número “Machine Epsilon”, ϵ_{mach} , es la distancia entre 1 y el menor número representable mayor a 1. Para IEEE double precision FPS.

$$\epsilon_{mach} = 2^{-52} \approx 2 \cdot 10^{-16}$$

¿Cuáles son los problemas?

“Truncamiento v/s Redondeo”

Ejemplo

$$\begin{aligned}
 (9.4)_{10} &= (1001.\overline{0110})_2 \\
 &= \boxed{+1.001 \ 0110 \ 0110 \ \dots \ 0110 \ 0110 \ 0} \underbrace{110 \dots}_{\downarrow} \cdot \boxed{2^3}
 \end{aligned}$$

Esto no se puede almacenar

¿Qué hacemos?

?

Truncamiento \Rightarrow

Redondeo \Rightarrow

2.2.2.1. Regla IEEE de redondeo al más cercano

Para “double precision”, si el 53vo bit a la derecha del punto binario es 0, redondear hacia abajo (truncar después del bit 52). Si el 53vo bit es 1, entonces redondear hacia arriba (agregar 1 al bit 52), a menos que todos los bits “conocidos” a la derecha de 1 son 0, en este caso 1 es agregado al bit 52 si y solo si el bit 52 es 1.

$$+1. \text{-----} \boxed{} | \boxed{} \boxed{} \boxed{}$$

52 53

(la línea que separa el bit 52 del 53 indica la frontera de lo que si y lo que no se puede almacenar)

¿Cómo se aplica la regla de redondeo para los siguientes casos?

$$1.- \quad (9.4)_{10} = +1.0010110 \dots 01100 | 110 \dots \cdot 2^3$$

⇓ Aplicando “IEEE” regla de redondeo

$$= +1.0010110 \dots 01101 \cdot 2^3 = \text{fl}((9.4)_{10})$$

↪ Esto es lo que “llamamos” $(9.4)_{10}$

$$2.- \quad = -1.0001 \dots 0001 | 1001 \dots \cdot 2^8$$

⇓

$$= \boxed{-1.0001 \dots \text{-----}} \cdot 2^8$$

$$3.- \quad = +1.0001 \dots 0001 | 1000 \dots \cdot 2^5$$

⇓

$$= \boxed{+1.0001 \dots \text{-----}} \cdot 2^5$$

$$4.- \quad = +1.1001 \dots 0100 | 1000 \dots \cdot 2^3$$

⇓

$$= \boxed{+1.1001 \dots \text{-----}} \cdot 2^3$$

2.2.3. Error de Redondeo

2.2.3.1. Error Absoluto

Sea x_c la cantidad computada y x la cantidad exacta. Entonces:

$$\text{Error Absoluto} = |x_c - x|$$

2.2.3.2. Error Relativo

$$\text{Error Relativo} = \frac{|x_c - x|}{|x|}, |x| \neq 0$$

2.2.3.3. Interpretación de “Machine Epsilon”

Se denota $\text{fl}(x)$ al número realmente almacenado utilizando el formato de punto flotante y la “regla IEEE de redondeo al más cercano”, cuando se quiere almacenar “ x ”.

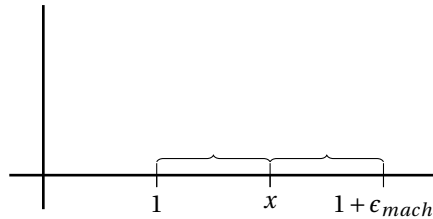
En el modelo IEEE de aritmética de máquina, el error relativo de redondeo de $\text{fl}(x)$ no es más que la mitad de ϵ_{mach} :

$$\frac{|\text{fl}(x) - x|}{|x|} \leq \frac{1}{2} \cdot \epsilon_{mach}$$

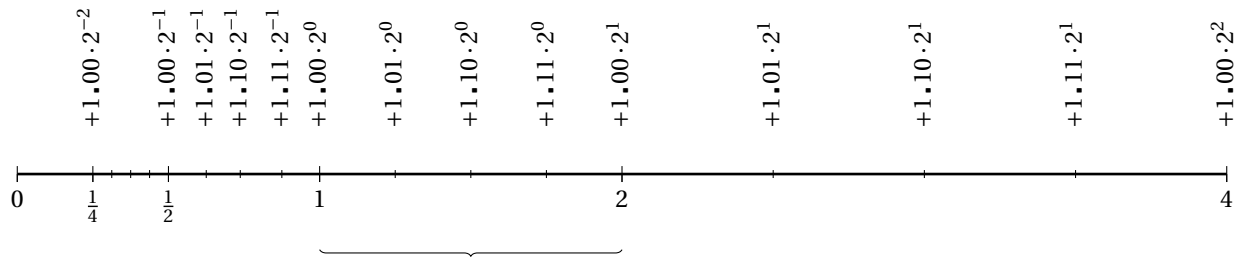
O mejor:

$$|\text{fl}(x) - x| \leq \frac{1}{2} \cdot \epsilon_{mach} \cdot |x|$$

i.e. el error de representar un número en FPS es proporcional al tamaño del número original.



Así se ve la representación de punto flotante en la recta numérica. En este caso particular se considera una mantisa de 2 bits.



El número de $\text{fl}(x)$ entre potencias de 2 es constante.

Ejercicio ¿Cuál es el primero entero no representable para este caso?

Considere que hay disponible solo 2 bits en la mantisa y 3 bits para el exponente.

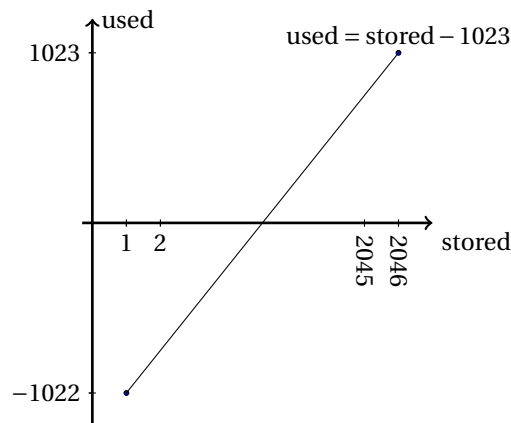
2.3. Representación de Máquina

Double precision : 8-byte word o 64 bits

$$\Rightarrow \begin{array}{|c|c|c|} \hline s & e_1 e_2 \dots e_{10} e_{11} & b_1 b_2 \dots b_{51} b_{52} \\ \hline 1\text{- bit} & 11 \text{ bits} & 52 \text{ bits} \\ \hline \end{array}$$

$$s: \begin{cases} 0, \text{ si el número es positivo} \\ 1, \text{ si es negativo} \end{cases}$$

$$e_1 \dots e_{11}: \quad 0 - 2047 \Leftrightarrow \begin{array}{ccc} -1022 & a & 1023 \\ \downarrow & & \downarrow \\ 1 & - & 2046 \end{array}$$



0 & 2047 son "diferentes"

Siendo "stored" lo que se almacena y "used" como se interpreta. La recta representa el "shift", es decir, como se interpreta matemáticamente lo que se almacena. Recuerde que solo se almacenan los "enteros" $\{0, 1, 2, \dots, 2047\}$.

$$\begin{aligned} 1023 &= 2^{10} - 1 && \text{: "Exponent bias"} \\ &= 2^{11-1} - 1 && \hookrightarrow \text{Convierte exponentes negativos en exponentes positivos para ser usado en su almacenamiento.} \end{aligned}$$

b_i : "Los bits de la representación de punto Flotante".

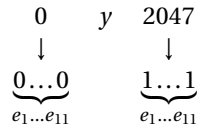
$$\pm 1 \cdot b_1 b_2 \dots b_{52} \cdot 2^p = \pm (2^0 + b_1 \cdot 2^{-1} + b_2 \cdot 2^{-2} + \dots + b_{52} \cdot 2^{-52}) \cdot 2^p$$

$$\boxed{\underbrace{\pm}_{\text{signo}} \left(1 + \underbrace{\sum_{i=1}^{52} b_i \cdot 2^{-i}}_{\text{mantisa}} \right) \cdot 2^p \rightarrow \text{exponente}}$$

Donde: $p = e_1 \cdot 2^{10} + e_2 \cdot 2^9 + e_3 \cdot 2^8 + \dots + e_{11} \cdot 2^0 - 1023$

$$= \left(\sum_{i=1}^{11} e_{12-i} \cdot 2^{i-1} \right) - 1023$$

2.3.1. Casos Especiales del Exponente:



2.3.1.1. 1111111111

$$2047 : \boxed{1111111111} = \boxed{e_1 e_2 e_3 \dots e_{11}}$$

S	e_1	e_2	e_3	...	e_{11}	b_1	b_2	...	b_{52}	lo que representa	
0	1	1	1	...	1	0	0	...	0	$+\infty$	1/0
1	1	1	1	...	1	0	0	...	0	$-\infty$	-1/0
1	1	1	1	...	1	x	x	...	x	<u>NaN</u>	0/0

2.3.1.2. 0000000000

$$0: 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 = e_1 \ e_2 \ \dots \ e_{11}$$

\Rightarrow "Non-normalized FP number"

$$\Rightarrow$$

S	0	0	0	0	0	0	0	0	0	0	0	b_1	b_2	...	b_{52}
	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	e_9	e_{10}	e_{11}				

$$\Rightarrow \pm 0.b_1 b_2 b_3 \dots b_{52} \cdot 2^{-1022}$$

\hookrightarrow Sub-normal FP numbers

\Rightarrow El número más pequeño representable es

0	0	...	0	0	0	...	0	1
S	e_1	...	e_{11}	b_1	b_{51}	b_{52}

$$\begin{aligned}
 &\Rightarrow +(0.0\dots 01) \cdot 2^{-1022} \\
 &= (b_1 \cdot 2^{-1} + b_2 \cdot 2^{-2} + \dots + b_{52} \cdot 2^{-52}) \cdot 2^{-1022} \\
 &= 1 \cdot 2^{-52} \cdot 2^{-1022} \\
 &= 2^{-1074} \approx 4,94 \cdot 10^{-324}
 \end{aligned}$$

Además tenemos 2 formas de representar 0

0	00...0	0...0	+0
1	00...0	0...0	-0

2.3.2. Pérdida de Significancia

Calcula “ $\sqrt{9.01} - 3$ ” en un computador con aritmética de 3-dígitos-decimales.

Ayuda: $\sqrt{9.01} = 3.0016$

$$\sqrt{9.01} - 3 = 0$$

Resultado “exacto”:

$1.6 \cdot 10^{-3}$

¿Qué podemos hacer?

2.4. Ejercicios Propuestos

1. $E_1(x) = \frac{1 - \cos(x)}{\sin^2(x)}$, $E_2(x) = \frac{1}{1 + \cos(x)}$ ¿Son $E_1(x)$ y $E_2(x)$ iguales? ¿Qué ocurre al evaluarlas cerca de 0?

2. Encuentre las raíces de $x^2 + 9^{12} \cdot x = 3$

3. ¿Cuándo habría problemas al evaluar $f(x)$ y $g(x)$?

$$f(x) = \frac{1 - (1-x)^3}{x}, g(x) = \frac{1}{1+x} - \frac{1}{1-x}$$

4. Diseñe un algoritmo para calcular las raíces de, $x^2 + b \cdot x - 10^{-12} = 0$, $b \geq 100$, lo más “exactamente” posible.

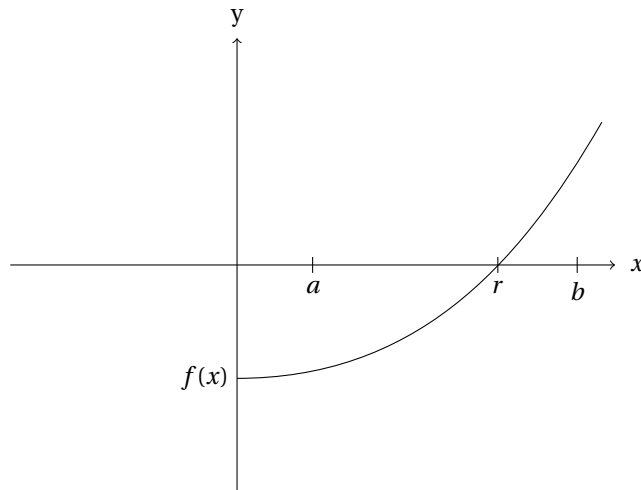
Capítulo 3

Raíces en 1D

Def 1. La función $f(x)$ tiene una raíz en $x = r$ si $f(r) = 0$.

Thm 1. Sea f una función continua en $[a, b]$, satisfaciendo $f(a) \cdot f(b) < 0$. Entonces f tiene una raíz entre a y b ; es decir, existe un número r que satisface $a < r < b$ y $f(r) = 0$.

Ej: $f(x) = x^3 + x - 1$



$$f(r) = 0,$$

¿Cómo obtengo r ?

Digamos $a = 0$, $f(a) = -1$, $b = 1$, $f(b) = 1$ y f es continua.

$$\Rightarrow f(a) \cdot f(b) = (-1) \cdot 1 = -1 < 0$$

\therefore Existe una raíz entre $a < r < b$.

3.1. Método de la Bisección

3.1.1. Algoritmo

Dado un intervalo inicial $[a, b]$ y $f(a) \cdot f(b) < 0$, el algoritmo para el método de la bisección es:

```

1  def biseccion(a, b, f, TOL) :
2      while (  $\frac{b-a}{2} > \text{TOL}$  ) :
3           $c = \frac{a+b}{2}$ 
4          if (  $f(c) = 0$  ) :
5              break
6          if (  $f(a) \cdot f(c) < 0$  ) :
7               $b = c$ 
8          else :
9               $a = c$ 
10         return  $\frac{a+b}{2}$ 

```

- El intervalo final $[a, b]$ contiene una raíz.
- La aproximación a la raíz es $\frac{a+b}{2}$

3.1.2. ¿Qué tan exacto y qué tan rápido es el método de la bisección?

Bisección:

- Largo intervalo inicial $(b - a)$
- Largo intervalo 1era iteración $\frac{(b-a)}{2}$
- \vdots
- Largo intervalo tras n -ésima iteración $\frac{(b-a)}{2^n}$

3.1.2.1. Error de la solución

Luego de n pasos tenemos que el error de la solución será: $|x_c - r| \leq \frac{(b-a)}{2^{n+1}}$, donde $x_c = \frac{a_n + b_n}{2}$ y requiere $n+2$ evaluaciones de f .

3.1.2.2. Precisión

Una buena forma para evaluar la eficiencia del método, es preguntarse cuanta precisión puede obtenerse por cada evaluación de la función. En cada evaluación de la función se reduce la incertidumbre en la raíz por un factor de dos.

Se definirá como una solución es correcta en p decimales si el error es menor que $0,5 \cdot 10^{-p}$

3.1.2.3. Ejemplo:

Use el método de la Bisección para encontrar una raíz de $f(x) = \cos x - x$ en el intervalo $[0, 1]$ correcta en 6 decimales.

$$\begin{aligned} \Rightarrow |x_c - r| &< \frac{1-0}{2^{n+1}} < 0,5 \cdot 10^{-6} \\ \Rightarrow n &> \frac{6}{\log_{10} 2} \approx 19,9 \\ \therefore n &= 20. \end{aligned}$$

3.2. Iteración de Punto Fijo

Aplice la función “cos” en una calculadora (¡asegúrese de que este en radianes!) repetidas veces con un número inicial arbitrario, es decir aplicar la función cos al número de inicio, luego al resultado y luego al nuevo resultado y así sucesivamente. Después de cierto número de iteraciones notara que el número no cambia, es decir converge a un número en específico. En esta sección se dará una explicación a este fenómeno, el cual es un ejemplo de iteración de punto fijo.

Def 2. El número real “ r ” es un punto fijo de la función $g(x)$ si $g(r) = r$.

3.2.1. Algoritmo

$$\begin{aligned} i = 0 : x_0 &= \text{“initial guess”} \\ i = 1 : x_1 &= g(x_0) \\ i = 2 : x_2 &= g(x_1) \\ &\vdots \\ i = n : x_n &= g(x_{n-1}) \end{aligned}$$

Por lo tanto:

```

1  x0 = Initial Guess
2  for i in range(1,n):
3      xi = g(xi-1)

```

3.2.2. Convergencia

Thm 2. Sea f una función continua en un vecindario de x_0 , y asume que $\lim_{n \rightarrow \infty} x_n = x_0$. Entonces:

$$\lim_{n \rightarrow \infty} f(x_n) = f\left(\lim_{n \rightarrow \infty} x_n\right) = f(x_0)$$

La secuencia x_i puede converger o no a medida que el número de pasos tiende a infinito. Sin embargo si g es continua y las x_i convergen a un número r , entonces r es un punto fijo.

Demostración.

$$\Rightarrow g(r) = g\left(\lim_{i \rightarrow \infty} x_i\right) = \lim_{i \rightarrow \infty} g(x_i) = \lim_{i \rightarrow \infty} x_{i+1} = r$$

□

Por lo tanto para nuestro ejemplo inicial de la función coseno, tenemos que la iteración de punto fijo quedara definida como:

$$x_{i+1} = \cos(x_i) \Rightarrow g(x) = \cos(x)$$

Equivalentemente estamos encontrando una raíz de $f(x)$ donde $f(x) = x - \cos(x)$.

Ejemplo: Considere que interesa buscar una raíz de $f(x) = x^3 + x - 1$ utilizando una iteración de punto fijo, considere las siguientes 3 opciones:

①

$$x = 1 - x^3 = g_1(x)$$

②

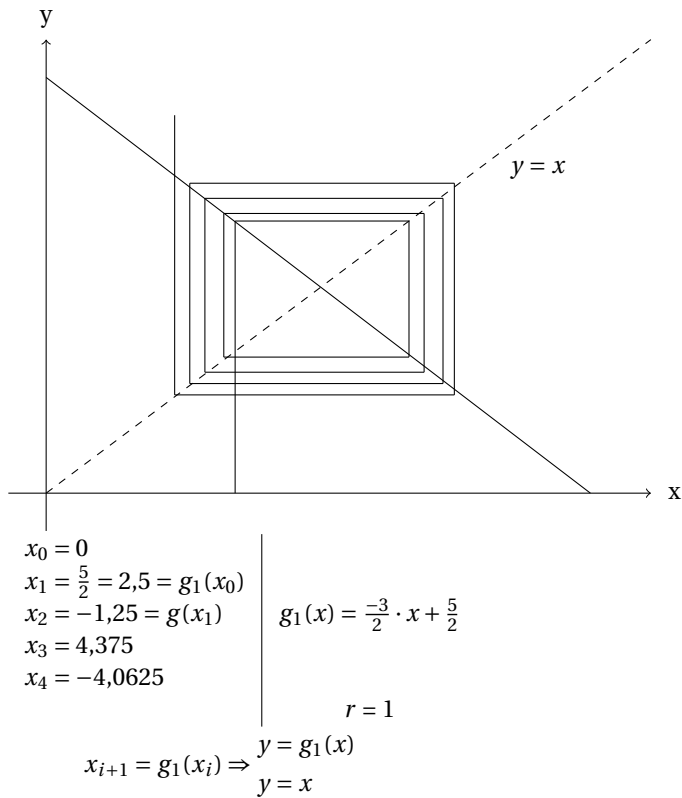
$$x = \sqrt[3]{1-x} = g_2(x)$$

③

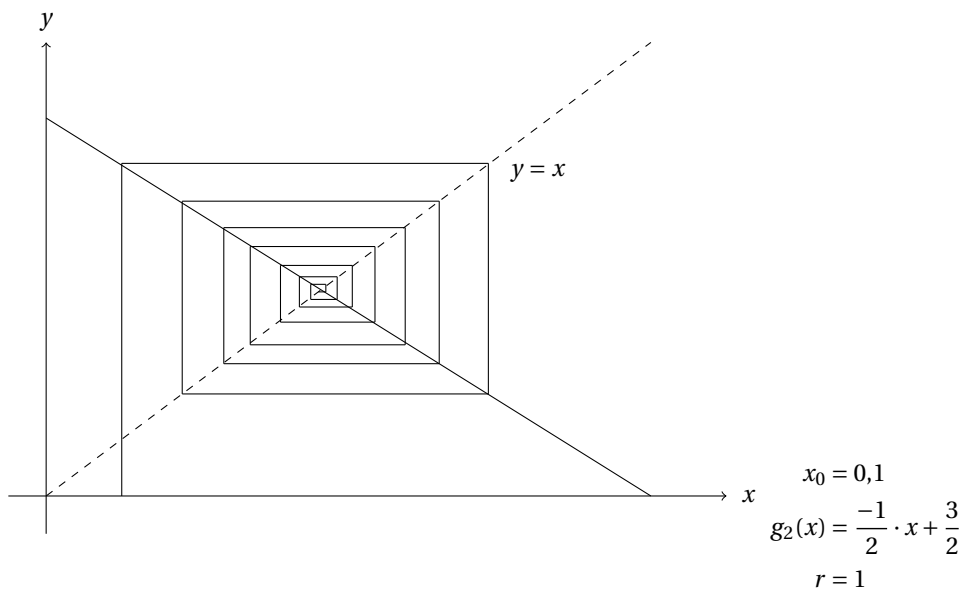
$$\begin{aligned} x^3 + x - 1 &= 0 & / + 1 \\ x^3 + x &= 1 & / + 2x^3 \\ 3x^3 + x &= 1 + 2x^3 \\ x(3x^2 + 1) &= 1 + 2x^3 \\ x &= \frac{1 + 2x^3}{1 + 3x^2} = g_3(x) \end{aligned}$$

3.2.3. Geometría de la iteración de Punto Fijo (cobweb diagram)

3.2.3.1. Divergencia de la iteración de punto fijo



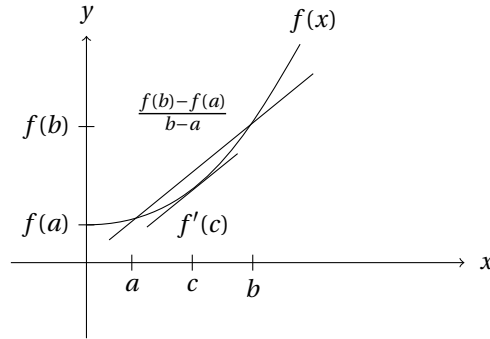
3.2.3.2. Convergencia de la iteración de punto fijo



3.2.4. Convergencia lineal de iteración de punto fijo

3.2.4.1. “Teorema del valor medio”

Thm 3. Sea f una función continua y diferenciable en el intervalo $[a, b]$.
Entonces existe un número “ c ”, entre a y b de tal forma que $f'(c) = \frac{f(b) - f(a)}{(b - a)}$.



Def 3. Sea $e_i = |x_i - r|$ el error del paso i de un método iterativo. Si:

$$\lim_{i \rightarrow \infty} \frac{e_{i+1}}{e_i} = S < 1.$$

Se dice que el método obedece “convergencia lineal” con tasa S .

Thm 4. Asuma que $g(x)$ es continua y diferenciable, que $g(r) = r$, y que $S = |g'(r)| < 1$. Entonces la iteración de Punto Fijo converge linealmente con tasa S al punto fijo r para “initial guesses” lo suficientemente cerca de r .

Demostración. Sea x_i el resultado de la iteración i . De acuerdo al teorema del valor medio $\left(f'(c) = \frac{(f(b)-f(a))}{(b-a)}\right)$

$$g'(c_i) = \frac{(g(x_i) - g(r))}{x_i - r}$$

$$a = r$$

$$b = x_i$$

y recuerde $x_{i+1} = g(x_i)$ y $r = g(r)$

$$\Rightarrow g'(c_i) = \frac{(x_{i+1} - r)}{(x_i - r)}$$

$$\Rightarrow x_{i+1} - r = g'(c_i) \cdot (x_i - r)$$

$$\Rightarrow e_{i+1} = |g'(c_i)| \cdot e_i$$

Si $S = |g'(r)| \Rightarrow$ Existe un pequeño vecindario alrededor de r tal que para $x \in [r - \epsilon, r + \epsilon] \Rightarrow |g'(x)| \leq \frac{S+1}{2}$

$$\Rightarrow e_{i+1} = |g'(c_i)| \cdot e_i < \frac{S+1}{2} \cdot e_i$$

$$\Rightarrow \lim_{i \rightarrow \infty} \frac{e_{i+1}}{e_i} = \lim_{i \rightarrow \infty} |g'(c_i)| = |g'(r)| = S \quad \square$$

$$\Rightarrow e_{i+1} \approx S \cdot e_i$$

□

3.2.5. Método localmente convergente

Un método iterativo es llamado localmente convergente a “ r ” si el método converge a r para “initial guesses” suficientemente cercanos a r .

En otras palabras, el método converge localmente a la raíz r , si existe una vecindad $(r - \epsilon, r + \epsilon)$, para $\epsilon > 0$, de tal manera que se de la convergencia a r a partir de “initial guesses” que esten en la vecindad $(r - \epsilon, r + \epsilon)$, i.e. $x_0 \in (r - \epsilon, r + \epsilon)$.

3.3. Teoremas útiles

3.3.1. Teorema de los límites continuos

Thm 5. Sea f una función continua en las cercanías de x_0 , y suponga que $\lim_{n \rightarrow \infty} x_n = x_0$. Entonces:

$$\lim_{n \rightarrow \infty} f(x_n) = f\left(\lim_{n \rightarrow \infty} x_n\right) = f(x_0)$$

En otras palabras los límites pueden trasladarse al interior de funciones continuas.

3.3.2. Teorema del valor intermedio

Thm 6. Sea f una función continua en un intervalo $[a, b]$, entonces f reconoce cada valor entre $f(a)$ y $f(b)$. Es decir, si y es un número tal que $f(a) < y < f(b)$, existe un número c de forma que $a < c < b \Rightarrow f(c) = y$

3.3.3. Teorema del Valor Medio

Thm 7. Sea f una función continua y diferenciable en un intervalo $[a, b]$. Entonces existe un número c , entre a y b de tal forma que $f'(c) = \frac{f(b) - f(a)}{b - a}$.

3.3.4. Teorema de Rolle

Thm 8. Sea f una función continua y diferenciable en un intervalo $[a, b]$ y asuma que $f(a) = f(b)$. Entonces existe un número c , entre a y b de tal forma que $f'(c) = 0$.

3.3.5. Teorema de Taylor (con residuo)

Thm 9. Sea " x " y " x_0 " números reales, y " f " " $k+1$ " - veces continua y diferenciable en el intervalo entre " x " y " x_0 ", entonces existe un número " c " entre " x " y " x_0 " tal que:

$$f(x) = f(x_0) + f'(x_0) \cdot (x - x_0) + \frac{f''(x_0)}{2!} \cdot (x - x_0)^2 + \dots + \frac{f^{(k)}(x_0)}{k!} \cdot (x - x_0)^k + \frac{f^{(k+1)}(c)}{(k+1)!} \cdot (x - x_0)^{k+1}$$

3.4. Método de Newton

3.4.1. Introducción

$$f(x) = f(x_0) + f'(x_0) \cdot (x - x_0) + \frac{f''(c)}{2!} \cdot (x - x_0)^2$$

Considere si $x = r$, la raíz.

$$f(r) = f(x_0) + f'(x_0) \cdot (r - x_0) + \frac{f''(c)}{2!} \cdot (r - x_0)^2$$

$$x = r \Rightarrow f(r) = 0$$

$$0 = f(x_0) + f'(x_0) \cdot (r - x_0) + \frac{f''(c)}{2!} \cdot (r - x_0)^2$$

¿Qué podemos hacer?

$$\begin{aligned}
 -f'(x_0) \cdot (r - x_0) &= f(x_0) + \frac{f''(c)}{2!} \cdot (r - x_0)^2 \\
 r - x_0 &= -(f'(x_0))^{-1} \cdot f(x_0) - (f'(x_0))^{-1} \cdot \frac{f''(c)}{2} \cdot (r - x_0)^2 \\
 \therefore r &= x_0 - (f'(x_0))^{-1} \cdot f(x_0) - (f'(x_0))^{-1} (f'(x_0))^{-1} \\
 \Rightarrow x_{i+1} &= x_i - \frac{f(x_i)}{f'(x_i)} \\
 &= g_n(x_i)
 \end{aligned}$$

3.4.2. Definición

El método de Newton también conocido como Newton-Raphson, por lo general converge mucho más rápido que los dos métodos vistos anteriormente. Para encontrar una raíz $f(x) = 0$, se da una estimación inicial x_0 y se traza la recta tangente a la función f en x_0 . La recta tangente seguirá en forma aproximada a la función hasta el eje x hacia la raíz. El punto de intersección de la línea con el eje x es una raíz aproximada, pero probablemente no es exacta si f no es lineal. Por lo tanto en este paso se itera.

3.4.2.1. Algoritmo

```

1   $x_0 = \text{InitalGuess}$ 
2  for  $i$  in range( $n$ ):
3       $x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$ 

```

¿Recuerda? $f(x) = x^3 + x - 1$
 $f'(x) = 3x^2 + 1$

$$\begin{aligned}
 \Rightarrow x_{i+1} &= x_i - \frac{x_i^3 + x - 1}{3x_i^2 + 1} \\
 x_{i+1} &= \frac{2x_i^3 + 1}{3x_i^2 + 1}
 \end{aligned}$$

i	x_i	e_i	$\frac{e_i}{e_{i-1}^2}$
0	-0.7	1.38	-
1	0.12	0.55	0.2906
2	0.95	0.27	0.8933
3	0.73	0.052	0.6924
4	0.6845	0.0022	0.8214
5	0.682332	0.00000437	0.8527
6	0.68232780	0.00000000	0.8541
7	0.68232780	0.00000000	---

3.4.3. Error de iteración

Sea e_i el error de la iteración " i ". La iteración converge cuadráticamente si:

$$M = \lim_{i \rightarrow \infty} \frac{e_{i+1}}{e_i^2} < \infty$$

3.4.4. Convergencia local y cuadrática

Thm 10. Sea “ f ” dos veces continuamente diferenciable y $f(r) = 0$. Si $f'(r) \neq 0$, entonces el método de Newton es local y cuadráticamente convergente a “ r ”. El error e_i en el paso i es:

$$\lim_{i \rightarrow \infty} \frac{e_{i+1}}{e_i^2} = M$$

donde

$$M = \frac{|f''(r)|}{2|f'(r)|}$$

Demostración. FPI con $g(x) = x - \frac{f(x)}{f'(x)}$

$$\begin{aligned} \Rightarrow g'(x) &= 1 - \frac{f'(x)^2 - f(x) \cdot f''(x)}{f'(x)^2} = \frac{f(x) \cdot f''(x)}{f'(x)^2} \\ &\text{pero } f(r) = 0 \quad \text{y} \quad f'(r) \neq 0 \\ \Rightarrow g'(r) &= \frac{f(r) \cdot f''(r)}{(f'(r))^2} = 0 \end{aligned}$$

□

3.4.5. Convergencia Cuadrática

$$\begin{aligned} f(r) &= f(x_i) + f'(x_i) \cdot (r - x_i) + \frac{f''(c_i)}{2} \cdot (r - x_i)^2 \\ 0 &= f(x_i) + (r - x_i) \cdot f'(x_i) + (r - x_i)^2 \cdot \frac{1}{2} \cdot f''(c_i) \\ \frac{-f(x_i)}{f'(x_i)} &= r - x_i + (r - x_i)^2 \cdot \frac{1}{2} \cdot \frac{f''(c_i)}{f'(x_i)} \\ x_{i+1} - r &= (r - x_i)^2 \cdot \frac{1}{2} \cdot \frac{f''(c_i)}{f'(x_i)} \\ \Rightarrow \frac{e_{i+1}}{e_i^2} &= \frac{1}{2} \cdot \frac{|f''(c_i)|}{|f'(x_i)|} \\ \therefore \lim_{i \rightarrow \infty} \frac{e_{i+1}}{e_i^2} &= \left| 2^{-1} \cdot \frac{f''(r)}{f'(r)} \right| \end{aligned}$$

Asuma que r es una raíz de f y que f es diferenciable, esto es, asuma que $f(r) = 0$. Entonces si $0 = f(r) = f'(r) = \dots = f^{(m-1)}(r)$, pero $f^{(m)}(r) \neq 0$, decimos que f tiene una raíz de multiplicidad “ m ” en r . Decimos que f tiene una raíz múltiple si $m > 1$, en otro caso ($m = 1$), la raíz se dice que es simple.

3.4.6. Convergencia Lineal del método de Newton

Ex: $f(x) = x^2, \quad f(r) = 0 \quad \Rightarrow \quad r = 0$

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}, \quad f'(x) = 2x$$

$$\Rightarrow x_{i+1} = x_i - \frac{x_i^2}{2 \cdot x_i}$$

$$= \frac{x_i}{2}$$

Ejemplo: $f(x) = x^m, f'(x) = m \cdot x^{m-1}$

$$\Rightarrow x_{i+1} = x_i - \frac{x_i^m}{m \cdot x_i^{m-1}} = \frac{m-1}{m} \cdot x_i \quad \therefore \text{Convergencia lineal...}'$$

i	x_i	e_i	$\frac{e_i}{e_{i-1}}$
0	1	1	-
1	$\frac{1}{2}$	0.5	0.5
2	0.25	0.25	0.5
3	0.125	0.125	0.5

3.4.7. Convergencia local

Thm 11. Asuma que f es una función $(m+1)$ - veces continua y diferenciable en $[a, b]$ y tiene una multiplicidad “ m ” en la raíz “ r ”. Entonces el método de Newton es linealmente convergente a “ r ” y el error e_i es:

$$\lim_{i \rightarrow \infty} \frac{e_{i+1}}{e_i} = S = \frac{m-1}{m} \neq 0$$

3.4.8. Método de Newton Modificado

Thm 12. Si f es $(m+1)$ - veces continua y diferenciable en $[a, b]$, donde hay una raíz “ r ” de multiplicidad $m > 1$, entonces el método modificado de Newton

$$x_{i+1} = x_i - m \cdot \frac{f(x_i)}{f'(x_i)}$$

Converge local y cuadráticamente a “ r ”.

3.5. Método de la Secante

Como se vio anteriormente, el método de Newton converge cuadráticamente pero requiere el uso de la derivada de “ f ”. Sin embargo, en muchas ocasiones uno no tiene acceso a esta. Por lo cual no es posible utilizarlo, para resolver este problema se propone aproximarla.

Por ejemplo se propone la siguiente aproximación:

$$f'(x) \approx \frac{f(x) - f(x - \Delta x)}{x - (x - \Delta x)} = \frac{f(x) - f(x - \Delta x)}{\Delta x}$$

Luego usando nuestra nueva aproximación en el método de Newton

$$x_0 = \text{“initial guess”}$$

método de Newton:

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$$

$$x_0, x_1 = \text{“initial guesses”}, \quad x_0 \neq x_1$$

método de la secante:

$$x_{i+1} = x_i - \frac{f(x_i) \cdot (x_i - x_{i-1})}{f(x_i) - f(x_{i-1})}$$

Es posible demostrar que el método de la secante bajo el supuesto de que converge a r y si $f'(r) \neq 0$, se cumple la relación aproximada del error:

$$\begin{aligned} e_{i+1} &\approx \left| \frac{f''(r)}{2 \cdot f'(r)} \right| \cdot e_i \cdot e_{i-1} \\ &= \left| \frac{f''(r)}{2 \cdot f'(r)} \right|^{\alpha-1} \cdot e_i^\alpha \end{aligned}$$

Donde, $\alpha = \frac{1+\sqrt{5}}{2} \approx 1,62$. La convergencia de este método se denomina convergencia super-lineal, lo cual significa que está entre la convergencia de los métodos linealmente convergentes y los cuadráticamente convergentes.

3.6. Criterios de parada

- ① $|x_{i+1} - x_i| < TOL$
- ② $\frac{|x_{i+1} - x_i|}{|x_{i+1}|} < TOL$
- ③ $\frac{|x_{i+1} - x_i|}{\max(|x_{i+1}|, \Theta)} < TOL$, por ejemplo $\Theta = 10^{-10}$

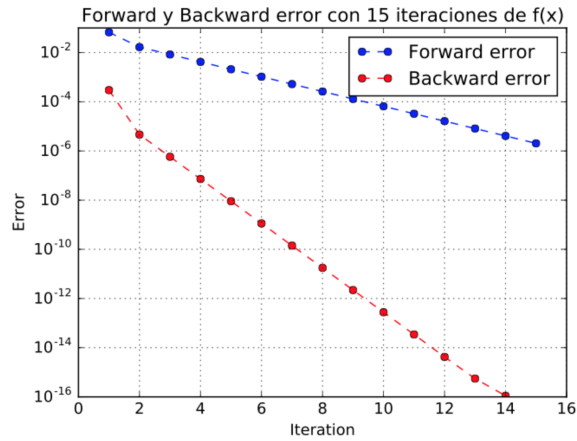
3.7. Un breve comentario para entender la relación entre Backward-error y Forward-error

Asuma que f es una función y que r es una raíz, i. e. $f(r) = 0$. Asuma que x_a es una aproximación a r . Para el problema de encontrar una raíz, el backward-error de la aproximación x_a es $|f(x_a) - 0|$ y el forward error es $|x_a - r|$.

$$\begin{aligned} f(r) &= f(x_a) + f'(c) \cdot (r - x_a) \\ \Rightarrow \underbrace{|f(r) - f(x_a)|}_{\text{Backward - error}} &= |f'(c)| \cdot \underbrace{|r - x_a|}_{\text{Forward - error}} \end{aligned}$$

Si usamos algún método conocido para poder encontrar raíces (por ejemplo, bisección), al encontrar valores aproximados cercanos a la raíz x_a por cada iteración, y al evaluarla en la función ($f(x_a)$), podemos encontrar los backward y forward error correspondientes.

Al realizar un gráfico de backward y forward error a medida que hacemos iteraciones por bisección, encontramos lo siguiente:



Nos damos cuenta en el gráfico que el backward error disminuye de forma mucho mas rápida que el forward error, es decir, si el backward error es pequeño, no implica que necesariamente el forward error es pequeño.

Lo que se quiere en búsqueda de ceros es que el forward error sea pequeño pero para poder medirlo necesitamos conocer la solución (¡que es precisamente lo que andamos buscando!) por lo cual lo que se hace es obtener al backward error y concluir sobre el forward error.

Es **¡¡¡MUY IMPORTANTE!!!** que se entienda este punto, **backward error pequeño no necesariamente implica forward error pequeño**, solo es válido cuando el problema es “bien condicionado”.

$$f(x_a) = f(r) + f'(r) \cdot (r - x_a) + \dots + \frac{f^{(m+1)}(c)}{(m+1)!} \cdot (r - x_a)^{m+1}$$

$$\Rightarrow |f(r) - f(x_a)| = \left| \frac{f^{(m+1)}(c)}{(m+1)!} \right| \cdot |r - x_a|^m$$

3.8. The Wilkinson polynomial (Wilkinson 1994)

$$W(x) = \prod_{i=1}^{20} (x - i), \text{ roots} = 1, 2, \dots, 20$$

$$W(x) = x^{20} - 210 \cdot x^{19} + \dots + 1206647803780373360 \cdot x^6 + \dots + 243 \dots 000 \text{ (19 dígitos).}$$

Se debe expandir el polinomio de Wilkinson (¡SymPy!) y luego buscar sus raíces.

$$\text{root}(W(x), 16) \Rightarrow 16,002294682053055 \leftarrow \text{lo que se obtiene}$$

$$16,000000000000000 \leftarrow \text{la raíz real}$$

$$|x - x_a| \approx 10^{-3} \leftarrow \text{el error}$$

¿Por qué pasa esto si estamos trabajando con números enteros?

3.9. Sensibilidad en la búsqueda de raíces en 1D

Problema: Encontrar r , dado $f(r) = 0$, para pequeñas perturbaciones agregadas $\epsilon \cdot g(x)$, $\epsilon \ll 1$
 $\Rightarrow f(r + \Delta r) + \epsilon \cdot g(r + \Delta r) = 0$

Expandiendo:

$$f(r) + \Delta r \cdot f'(r) + \epsilon \cdot g(r) + \epsilon \Delta r \cdot g'(r) + O(\Delta r^2) = 0$$

$$\Rightarrow \Delta r \approx \frac{-\epsilon \cdot g(r)}{f'(r) + \epsilon \cdot g'(r)} \quad \text{y asumiendo } \epsilon g'(r) \ll f'(r)$$

$$\Delta r \approx -\epsilon \cdot \frac{g(r)}{f'(r)}$$

Ex: Estime la sensibilidad de $r = 16$ para cambios en x^{15}
donde:

$$W(x) = \prod_{i=1}^{20} (x - i) = x^{20} + \dots - 1672280820 \cdot x^{15} + \dots$$

$$\Rightarrow W_\epsilon(x) = W(x) + \epsilon \cdot g(x), \quad W'(16) = 15!4!$$

$$\Rightarrow \Delta r = \frac{-\epsilon \cdot g(r)}{f'(r)} \approx \frac{16^{15} \cdot 1672280820}{15!4!} \cdot \epsilon \approx 6,14 \cdot 10^{13} \cdot \epsilon$$

$$\Rightarrow \Delta r \approx 6,14 \cdot 10^{13} \cdot (\pm \text{Emach})$$

$$\approx 6,14 \cdot 10^{13} \cdot (\pm 2,22 \cdot 10^{-16}) \approx \pm 0,0136$$

$$\approx |x - x_a|$$

3.9.1. Condicionamiento (“¡primer encuentro!”)

“Magnificación del error por parte del problema teórico”

$$K = \frac{\text{relative Forward error}}{\text{relative Backward error}} = \frac{\frac{|x_a - r|}{|r|}}{\frac{|f(x_a) - f(r)|}{|f(r)|}} \Rightarrow \frac{\frac{|\Delta r|}{|r|}}{\left| \frac{\Delta f}{f} \right|} = \frac{|f|}{\left| \frac{\Delta f}{\Delta r} \right| \cdot |r|} = \frac{\epsilon \cdot |g(r)|}{|r| \cdot |f'(r)|}$$

Ex:

$$\begin{aligned} x^2 - 2 \cdot x + 1 &= 0 & \Leftrightarrow & (x - 1)^2 = 0 \\ \Rightarrow x^2 - 2x + 1 - \epsilon &= 0 \\ \Rightarrow x_1 &= 1 - \sqrt{\epsilon} \\ \Rightarrow x_2 &= 1 + \sqrt{\epsilon} \\ \text{Si } \epsilon &\approx 10^{-8} \\ \Rightarrow x^2 - 2x + 1 - 10^{-8} &= 0 \\ x^2 - 2x + 0,99999999 &= 0 \\ \Rightarrow x_1 &= 1 - 10^{-4} = 0,9999 \\ x_2 &= 1 + 10^{-4} = 1,0001 \end{aligned}$$

Volviendo a $W(x)$

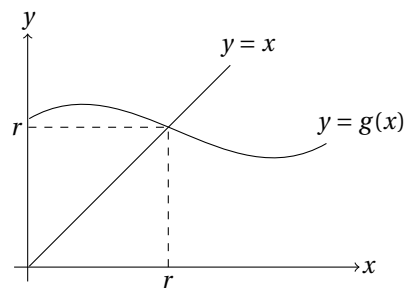
$$K = \frac{|g(r)|}{|r| |f'(r)|} = \frac{16^{15} \cdot 1672280820}{15! 4! 16} \approx 3,8 \cdot 10^{12}$$

\Rightarrow Se espera perder “12” dígitos de “accuracy”.
recuerda $16 \approx 16,014 \dots$

Capítulo 4

Sistemas de Ecuaciones Lineales

Recuerde: Búsqueda de ceros mediante Iteración de Punto Fijo



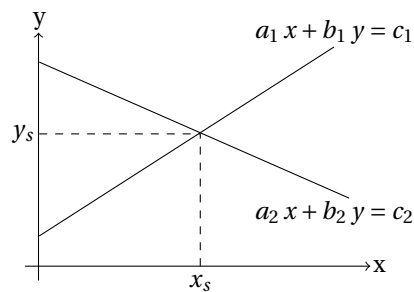
$$\Rightarrow r = g(r)$$

o más general

$$\begin{array}{l} y - x = 0 \\ y - g(x) = 0 \end{array}$$

Encontrar raíz r es equivalente a resolver este sistema de ecuaciones no lineales.

Ahora: “algo más simple... por ahora”



$$\begin{array}{l} (1) \quad a_1 x + b_1 y = c_1 \\ (2) \quad a_2 x + b_2 y = c_2 \end{array} \quad \Leftrightarrow \quad \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$

$$\uparrow$$

2 ecuaciones y
2 incógnitas

$$A \quad \mathbf{x} = \mathbf{c}$$

$A\mathbf{x} = \mathbf{c}$ ← Sistema de ecuaciones lineales

donde: $\mathbf{x} = A^{-1}\mathbf{c}$

4.1. Métodos conocidos de resolución

Ⓐ Por sustitución

A partir de (1) despejamos $x \Rightarrow x = \frac{c_1 - b_1 y}{a_1}$ (3).

Utilizamos (3) en (2) $\Rightarrow a_2 \left(\frac{c_1 - b_1 y}{a_1} \right) + b_2 y = c_2$ (1 ecuación y 1 incógnita)

Despejamos y , $y = \frac{c_2 - a_2 \frac{c_1}{a_1}}{b_2 - a_2 \frac{b_1}{a_1}} = \frac{a_1 c_2 - a_2 c_1}{a_1 b_2 - a_2 b_1}$

Ⓑ Usando Regla de Cramer

$$y = \frac{\begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}} = \frac{a_1 c_2 - a_2 c_1}{a_1 b_2 - a_2 b_1}$$

$$\begin{aligned} \text{¿Cómo obtengo } x? &\rightarrow x = \frac{c_1}{a_1} - \frac{b_1}{a_1} y \\ &= \frac{\begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}} \end{aligned}$$

De lo anterior se concluye lo siguiente:

- Ⓐ Simple pero un poco compleja para “algoritmizar”
- Ⓑ Muy simple de explicar pero MUY costosa computacionalmente. ¿Cuál es el costo computacional?

¿Cuál es un sistema de ecuaciones lineales simple pero no trivial de resolver?

$$\begin{array}{l} \underline{a}x + \underline{b}y = \underline{d} \\ \underline{0}x + \underline{c}y = \underline{e} \end{array} \Leftrightarrow \begin{bmatrix} \underline{\quad} & \underline{\quad} \\ \underline{\quad} & \underline{\quad} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \underline{\quad} \\ \underline{\quad} \end{bmatrix}$$

4.2. Eliminación Gaussiana Simple

Ejemplo, para 2 ecuaciones y 2 incógnitas:

$$\begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \Leftrightarrow \begin{array}{l} a_1 x + b_1 y = c_1 \\ a_2 x + b_2 y = c_2 \end{array}$$

$$\Rightarrow \text{tableau} \rightarrow \left[\begin{array}{cc|c} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{array} \right] \quad R_2 = R_2 - \left(\frac{a_2}{a_1} \right) R_1$$

coeficiente necesario para eliminar a_2 de la segunda fila

$$\Rightarrow \left[\begin{array}{cc|c} a_1 & b_1 & c_1 \\ 0 & b_2 - \frac{a_2}{a_1} b_1 & c_2 - \frac{a_2}{a_1} c_1 \end{array} \right]$$

$$\begin{array}{l} a_1 x + b_1 y = c_1 \\ 0x + \left(b_2 - \frac{a_2}{a_1} b_1 \right) y = c_2 - \frac{a_2}{a_1} c_1 \end{array}$$

↑
Forma "simple"

$$\Rightarrow y = \frac{c_2 - \frac{a_2}{a_1} c_1}{b_2 - \frac{a_2}{a_1} b_1}$$

↙

$$\Rightarrow x = \frac{c_1 - b_1 y}{a_1}$$

Algoritmo

- ① $A\mathbf{x} = \mathbf{c} \rightarrow$ construir tableau
- ② Aplicar operaciones fila para dejar la matriz triangular superior y modificar $\mathbf{c} \Rightarrow U\mathbf{x} = \hat{\mathbf{c}}$
- ③ Resolver $U\mathbf{x} = \hat{\mathbf{c}}$ usando backward substitution.

En forma general: $A\mathbf{x} = \mathbf{b}$

$$\begin{array}{c}
 \left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & & a_{2n} & b_2 \\ \vdots & & \ddots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & b_n \end{array} \right] = \left[\begin{array}{c|c} & \\ A & b \end{array} \right] \\
 \uparrow \\
 \text{tableau general} \\
 A\mathbf{x} = \mathbf{b} \quad \xrightarrow{\text{G.E.}} \quad U\mathbf{x} = \hat{\mathbf{b}} \quad \xrightarrow{\text{B.S.}} \quad \mathbf{x}
 \end{array}$$

4.2.1. Algoritmo Gaussian Elimination (G.E.)

```

1  for j in range(n):
2      for i in range(j+1, n):
3          mult =  $\frac{B_{ij}}{B_{jj}}$ 
4           $B_{i,j+1:n+1} = B_{i,j+1:n+1} - \text{mult} \cdot B_{j,j+1:n+1}$ 

```

¿Cuántas operaciones elementales se necesitan en G.E.?

Previo:

- $1 + 1 + 1 + \cdots + 1 = \sum_{i=1}^n 1 = n$
- $1 + 2 + 3 + \cdots + n = \sum_{i=1}^n i = \frac{n(n+1)}{2}$
- $1^2 + 2^2 + 3^2 + \cdots + n^2 = \sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}$

4.2.2. Cantidad de operaciones en G.E

Contemos cuántas operaciones se realizan en G.E.

línea	# operaciones
(3)	1
(4)	2 ("n + 1 - (j + 1 - 1)"-veces)

$$\begin{aligned}
 \text{Total} &= \underbrace{\sum_{j=1}^{n-1}}_{(1)} \underbrace{\sum_{i=j+1}^n}_{(2)} \left(\underbrace{1}_{(3)} + \underbrace{2(n-j+1)}_{(4)} \right) \\
 &= \sum_{j=1}^{n-1} \sum_{k=1}^{n-j} (1 + 2(n-j+1)) \\
 &= \sum_{j=1}^{n-1} (1 + 2(n-j+1))(n-j) \\
 &= \sum_{j=1}^{n-1} (3n - 3j + 2n^2 - 4nj + 2j^2) \\
 &= (3n + 2n^2) \left[\sum_{j=1}^{n-1} 1 \right] - 3 \left(\sum_{j=1}^{n-1} j \right) - 4n \left(\sum_{j=1}^{n-1} j \right) + 2 \sum_{j=1}^{n-1} j^2 \\
 &= (3n + 2n^2)(n-1) - 3 \frac{(n-1)n}{2} - 4n \frac{(n-1)n}{n} + 2 \frac{(n-1)n(2n-1)}{6} \\
 &= \frac{2}{3}n^3 + \frac{1}{2}n^2 - \frac{7}{6}n \\
 &\sim \frac{2}{3}n^3
 \end{aligned}$$

4.3. Backward Substitution

$$Ux = \hat{b} \quad \Leftrightarrow \quad \begin{bmatrix} u_{11} & u_{12} & \cdots & & u_{1n} & \vdots & \tilde{b}_1 \\ 0 & u_{22} & & & u_{2n} & \vdots & \tilde{b}_2 \\ \vdots & 0 & \ddots & & & \vdots & \vdots \\ & & & \ddots & u_{n-1,n-1} & u_{n-1,n} & \vdots \\ 0 & & \cdots & 0 & u_{nn} & & \tilde{b}_n \end{bmatrix}$$

\uparrow
 triangular superior

$$\textcircled{1} \quad \begin{array}{l} \text{última ecuación} \\ \text{(ecuación } n) \end{array} \quad : \quad \begin{array}{l} u_{nn} x_n = \tilde{b}_n \\ \Rightarrow x_n = \frac{\tilde{b}_n}{u_{nn}} \end{array}$$

$$\textcircled{2} \quad \text{ecuación } (n-1) \quad : \quad u_{n-1,n-1} x_{n-1} + u_{n-1,n} x_n = \tilde{b}_{n-1}$$

$$\Rightarrow x_{n-1} = \frac{\tilde{b}_{n-1} - u_{n-1,n} x_n}{u_{n-1,n-1}}$$

$$\textcircled{i} \quad i\text{-ésima ecuación} \quad : \quad u_{i,i} x_i + \sum_{j=i+1}^n u_{i,j} x_j = \tilde{b}_i$$

$$\Rightarrow x_i = \frac{\tilde{b}_i - \sum_{j=i+1}^n u_{i,j} x_j}{u_{i,i}}$$

4.3.1. Algoritmo

```

1 for i in range(n,1,-1):
2     for j in range(i+1,n):
3          $b_i = b_i - u_{i,j} \cdot x_j$ 
4          $x_i = \frac{b_i}{u_{ii}}$ 

```

4.3.2. Número de Operaciones elementales

$$\begin{aligned} &\Rightarrow \sum_{i=n}^1 \left[\left(\sum_{j=i+1}^n 2 \right) + 1 \right] \\ &= \sum_{i=1}^n \left[2 \left(\sum_{j=i+1}^n 1 \right) + 1 \right] \\ &= \sum_{i=1}^n [2(n-i+1-1) + 1] \\ &= 2n \left(\sum_{i=1}^n 1 \right) - 2 \left(\sum_{i=1}^n i \right) + \sum_{i=1}^n 1 \\ &= 2nn - 2 \frac{n(n+1)}{2} + n \\ &= 2n^2 - n^2 - n + n \\ &= n^2 \square \end{aligned}$$

4.4. Factorización LU de A

- Una matriz $L \in \mathbb{R}^{m \times n}$ es **triangular inferior** si sus coeficientes satisfacen $l_{ij} = 0$ para $i < j$.
- Una matriz $U \in \mathbb{R}^{m \times n}$ es **triangular superior** si sus coeficientes satisfacen $u_{ij} = 0$ para $i > j$.

Buscamos factorizar la matriz de coeficientes A en la multiplicación de dos matrices triangulares L y U. Esto se basa en los siguientes tres hechos.

1. Sea $L_{ij}(-c)$ la matriz triangular inferior la cual tiene coeficientes no-cero en su diagonal y la posición (i, j) .

La diagonal son 1's y en la posición (i, j) es $(-c)$. Entonces $L_{ij}(-c) \cdot A$ representa la operación fila "restar la fila j multiplicada por $-c$ a la fila i ".

Ejemplo: $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, $L_{21}(-c) \cdot A = \begin{bmatrix} 1 & 0 \\ -c & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} - c a_{11} & a_{22} - c a_{21} \end{bmatrix}$

2. $L_{ij}(-c)^{-1} = L_{ij}(c)$, recuerde $A \cdot A^{-1} = I$

Ejemplo: $\begin{bmatrix} 1 & 0 \\ -c & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 \\ -c & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
 $\begin{bmatrix} 1 & 0 \\ c & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -c & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ c - c & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

3. La siguiente ecuación matricial es cierta:

$$\begin{bmatrix} 1 & 0 & 0 \\ c_1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ c_2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & c_3 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ c_1 & 1 & 0 \\ c_2 & c_3 & 1 \end{bmatrix}$$

Ejemplo:

$$\begin{matrix} L_3 & L_2 & L_1 & \leftarrow \text{fact 1} & \uparrow A \end{matrix} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{7}{3} & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & -1 \\ 2 & 1 & -2 \\ -3 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & -1 \\ 0 & -3 & 0 \\ 0 & 0 & -2 \end{bmatrix} = U$$

Es decir: $L_3 L_2 L_1 A = U$

$$\begin{aligned} A &= L_1^{-1} L_2^{-1} L_3^{-1} U && \leftrightarrow \text{Fact 2} \\ &= \tilde{L}_1 \tilde{L}_2 \tilde{L}_3 U && \leftrightarrow \text{Fact 3} \\ &= LU \end{aligned}$$

4.4.1. Utilización de la factorización LU

¿Cómo se utiliza la factorización LU de A para resolver $A\mathbf{x} = \mathbf{b}$?

$$\begin{array}{ccc}
 A\mathbf{x} = \mathbf{b} & & \text{incógnita} \\
 & & \downarrow \\
 L \underbrace{U\mathbf{x}}_{\mathbf{c}} = \mathbf{b} & \rightarrow \text{Resolver} & L\mathbf{c} = \mathbf{b} \\
 & & \uparrow \quad \uparrow \\
 & & L \text{ y } \mathbf{b} \text{ conocidas} \\
 & & \mathbf{c} = L^{-1}\mathbf{b} \\
 & & \uparrow \\
 & & \text{NO calcule la inversa} \\
 & & \text{utilice Forward Substitution}
 \end{array}$$

Recuerde: $L\mathbf{c} = \mathbf{b} \Rightarrow$

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ l_{21} & 1 & \ddots & 0 \\ l_{31} & \cdots & \ddots & 0 \\ \vdots & & & \\ l_{n1} & & l_{n,n-1} & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

Ok, ahora conocemos \mathbf{c} , ¿y de qué sirve?

$$\begin{array}{ccc}
 & & \text{incógnita} \\
 & & \downarrow \\
 L\mathbf{c} = \mathbf{b} & \Rightarrow & U\mathbf{x} = \mathbf{c} \\
 \uparrow \uparrow \uparrow & & \uparrow \quad \uparrow \\
 \text{conocido} & & \text{conocido} \\
 \therefore \mathbf{x} = U^{-1}\mathbf{c} & \Rightarrow & \text{"Usar Backward Substitution!"}
 \end{array}$$

$$\begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & \\ 0 & 0 & \ddots & \vdots \\ \vdots & & & \\ 0 & 0 & \cdots & u_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \Rightarrow \mathbf{x}$$



Operaciones elementales para
calcular $A = LU$: $\frac{2}{3}n^3$

4.4.1.1. ¿Cuándo es útil?

- ① Secuencia de sistemas de ecuaciones lineales

$$\begin{array}{lcl}
 A\mathbf{x}_1 = \mathbf{b}_1 & & \\
 A\mathbf{x}_2 = \mathbf{b}_2 & \Leftrightarrow & A_{n \times n} X_{n \times m} = B_{n \times m}, \text{ donde } X_{n \times m} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \\
 \vdots & & \\
 A\mathbf{x}_m = \mathbf{b}_m & &
 \end{array}$$

$AX + XA = B \Rightarrow \tilde{A}\mathbf{x} = \tilde{\mathbf{b}}$
 ¿Cómo resolvemos $AX + XA = B$?

(i) Calcular $LU = A$

(ii) m Backward Substitutions $\Rightarrow \text{Costo} = \frac{2}{3}n^3 + m \cdot 2n^2$

(iii) m Forward Substitutions

y si lo hubiéramos resuelto para cada uno independientemente: $\frac{2}{3}m \cdot n^3 + 2m \cdot n^2$

por ej: si $m \gg n$.

$$\begin{array}{lcl}
 \textcircled{2} \quad A^l \mathbf{x} = \mathbf{b} & & A \underbrace{A^{l-1} \mathbf{x}}_{c_{l-1}} = \mathbf{b} \\
 \downarrow & & \\
 \underbrace{(A A A \cdots A)}_{l\text{-veces}} \mathbf{x} = \mathbf{b} & & A \underbrace{A^{l-2} \mathbf{x}}_{c_{l-2}} = c_{l-1} \\
 & & \\
 A c_1 = b_1 & & \\
 A c_2 = b_2 & & \\
 \vdots & & \\
 A c_l = c_{l-1} & \Rightarrow & \mathbf{c}_l = \mathbf{x} \quad \square
 \end{array}$$

Entonces: Cuando decimos $\mathbf{x} = A^{-1} \mathbf{b}$ pensamos lo siguiente:

$$\begin{array}{lcl}
 A\mathbf{x} = \mathbf{b} & \leftrightarrow & LU \\
 LU\mathbf{x} = \mathbf{b} & & \\
 UX = \underbrace{L^{-1} \mathbf{b}}_{\text{E.S.}} & & \\
 \mathbf{x} = \underbrace{U^{-1} (L^{-1} \mathbf{b})}_{\text{B.S.}} & &
 \end{array}$$

Recuerde: $A^{-1} = (LU)^{-1} = U^{-1} L^{-1}$

4.4.2. Origen de los errores

Sea \mathbf{x}_a una solución aproximada al sistema de ecuaciones lineales $A\mathbf{x} = \mathbf{b}$. El vector residual es definido como $\mathbf{r} = \mathbf{b} - A\mathbf{x}_a$. El Backward-Error es $\|\mathbf{b} - A\mathbf{x}_a\|$ y el Forward-Error es $\|\mathbf{x} - \mathbf{x}_a\|$.

Ejemplo: Ejemplo: $\begin{bmatrix} 1 & 1 \\ 3 & -4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$, $\mathbf{x} = [2, 1]^T$
 $\mathbf{x}_a = [1, 1]^T$

$$\underbrace{\|\mathbf{b} - A\mathbf{x}_a\|_\infty}_{\text{Backward-Error}} = \left\| \begin{bmatrix} 3 \\ 2 \end{bmatrix} - \begin{bmatrix} 2 \\ -1 \end{bmatrix} \right\|_\infty = \left\| \begin{bmatrix} 1 \\ 3 \end{bmatrix} \right\|_\infty = 3.$$

$$\underbrace{\|\mathbf{x} - \mathbf{x}_a\|_\infty}_{\text{Forward-Error}} = \left\| \begin{bmatrix} 2 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\|_\infty = \left\| \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\|_\infty = 1.$$

Ex: "Muy importante"

$$\begin{bmatrix} 1 & 1 \\ 1+\epsilon & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 2+\epsilon \end{bmatrix}$$

$$\mathbf{x} = [1, 1]^T, \quad \mathbf{x}_a = [-1, 3+\epsilon]^T$$

$$\begin{aligned} \underbrace{\|\mathbf{b} - A\mathbf{x}_a\|_\infty}_{\text{Backward-Error}} &= \left\| \begin{bmatrix} 2 \\ 2+\epsilon \end{bmatrix} - \begin{bmatrix} 1 & 1 \\ 1+\epsilon & 1 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 3+\epsilon \end{bmatrix} \right\|_\infty \\ &= \left\| \begin{bmatrix} 2 \\ 2+\epsilon \end{bmatrix} - \begin{bmatrix} 2+\epsilon \\ 2 \end{bmatrix} \right\|_\infty \\ &= \left\| \begin{bmatrix} -\epsilon \\ \epsilon \end{bmatrix} \right\|_\infty = \epsilon \quad \square \end{aligned}$$

$$\underbrace{\|\mathbf{x} - \mathbf{x}_a\|_\infty}_{\text{Forward-Error}} = \left\| \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 1+\epsilon \end{bmatrix} \right\|_\infty = \left\| \begin{bmatrix} 2 \\ -2-\epsilon \end{bmatrix} \right\|_\infty = 2 + \epsilon, \quad \epsilon > 0$$

¿Qué es lo que queremos?

Queremos un Forward-Error pequeño.

¿Qué es lo que podemos medir?

Podemos medir en general solamente el Backward-Error.

¿Cómo están relacionados?

4.4.3. Número de condición

El número de condición $\kappa(A) = \frac{\max_i(|\lambda_i|)}{\min_i(|\lambda_i|)}$ donde λ_i son los valores propios de A .

El número de condición de una matriz cuadrada A de $n \times n$ es

$$\text{cond}(A) = \kappa(A) = \|A\| \cdot \|A^{-1}\| \geq 1.$$

¿Y cómo están relacionados los F & B errors?

$$\frac{1}{\kappa(A)} \cdot \frac{\|b - A \cdot x_a\|}{\|b\|} \leq \frac{\|x_a - x\|}{\|x\|} \leq \kappa(A) \cdot \frac{\|b - A \cdot x_a\|}{\|b\|}$$

¿Qué pasa en nuestro ejemplo previo?

$$A = \begin{bmatrix} 1 & 1 \\ 1+\epsilon & 1 \end{bmatrix}, \quad A^{-1} = \begin{bmatrix} -1 & 1 \\ 1+\epsilon & -1 \end{bmatrix} \cdot \frac{1}{\epsilon}$$

$$\begin{aligned} \|A\|_{\infty} &= 2 + \epsilon, \quad \|A^{-1}\|_{\infty} = \frac{2}{\epsilon} + 1 \\ \Rightarrow \kappa_{\infty}(A) &= \|A\|_{\infty} \cdot \|A^{-1}\|_{\infty} = \frac{4}{\epsilon} + 4 + \epsilon \end{aligned}$$

$$\begin{aligned} \|b - A \cdot x_a\|_{\infty} &= \epsilon, \quad \|b\|_{\infty} = 2 + \epsilon \\ \|x - x_a\|_{\infty} &= 2 + \epsilon, \quad \|x\|_{\infty} = 1 \end{aligned}$$

$$\begin{aligned} \frac{\|x - x_a\|_{\infty}}{\|x\|_{\infty}} &\leq \kappa(A) \cdot \frac{\|b - A \cdot x_a\|_{\infty}}{\|b\|_{\infty}} \\ &\leq \left(\frac{4}{\epsilon} + 4 + \epsilon \right) \cdot \frac{\epsilon}{2 + \epsilon} \\ 2 + \epsilon &= \frac{\|x - x_a\|_{\infty}}{\|x\|_{\infty}} \leq \underbrace{\frac{2 + \epsilon}{\epsilon}}_{\text{estimado}} \end{aligned}$$

∴ No importa qué tan pequeño hagamos $\|b - A \cdot x_a\|$, no obtendremos x si $\kappa(A) \rightarrow \infty$.

En general $\log_{10} \kappa(A)$ me dice cuántos dígitos se perderán en la computación de x .

$$\text{Recuerde: } \frac{\|x - x_a\|}{\|x\|} \leq \kappa(A) \cdot \frac{\|b - A \cdot x_a\|}{\|b\|}$$

- Ver ejemplo con la matriz de Hilbert en Jupyter notebook.

4.5. Factorización $PA = LU$

No todas las matrices tienen una factorización LU, por ejemplo $\begin{bmatrix} 0 & 2 \\ 3 & 4 \end{bmatrix}$. Además, en ciertos casos, el error numérico puede hacer que la factorización LU nos lleve a soluciones erróneas. Un ejemplo de esto último se presenta a continuación.

Ejemplo: $\begin{bmatrix} \delta & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \quad \delta = 1e-20,$

$$A = LU = \begin{bmatrix} 1 & 0 \\ 10^{20} & 1 \end{bmatrix} \begin{bmatrix} 10^{-20} & 1 \\ 0 & -10^{20} \end{bmatrix}$$

Solución obtenida con $A = LU$:

$$\Rightarrow \mathbf{x}_a = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$PA\mathbf{x} = P\mathbf{b}, \quad P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 \\ \delta & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$$

$$\hookrightarrow \begin{bmatrix} & L \\ & \end{bmatrix} \begin{bmatrix} U \\ \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 10^{-20} & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$$

Solución obtenida con $PA = LU$:

$$\Rightarrow \mathbf{x}_a = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \quad \text{mucho mejor.}$$

Como vemos, la matriz P , al multiplicarla a ambos lados, lo único que hizo es un cambio de filas. Sin embargo, este cambio de filas resolvió el error numérico. Esta es la idea de la factorización $PA = LU$.

4.5.1. Matriz de Permutación

Una **matriz de permutación** es una matriz P de tamaño $n \times n$ donde todos sus valores son ceros, excepto por un uno en cada fila y cada columna. De forma equivalente, una matriz de permutación se obtiene al efectuar cambios de fila (o de columna) a la matriz identidad I .

Al aplicar una serie de operaciones fila/columna a la matriz identidad y obtener P , la multiplicación PA es equivalente a efectuar estas mismas operaciones fila/columna a la matriz A .

Idea: Dejar en la diagonal el $\max_{i>j} |a_{ij}|$. Esto es llamado **partial pivoting**. Antes de efectuar las operaciones fila, debemos revisar que el elemento "pivote" a usar es mayor que los elementos hacia abajo en la columna. Si no es así, hacemos una permutación de filas.

De ésta forma, los "multiplicadores" que usaremos en cada operación fila nunca serán mayores a 1 en valor absoluto. Guardaremos estos multiplicadores en la matriz L .

4.5.2. Ejemplo

$$\begin{array}{ccc}
\begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} & \underbrace{\begin{bmatrix} 2 & 1 & 5 \\ 4 & 4 & -1 \\ 1 & 3 & 1 \end{bmatrix}}_A & = & \begin{bmatrix} 2 & 1 & 5 \\ 4 & 4 & -1 \\ 1 & 3 & 1 \end{bmatrix} & \begin{array}{c} P \\ \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} \end{array} & \begin{array}{c} L \\ \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} \end{array} \\
\begin{bmatrix} & 1 & \\ 1 & & \\ & & 1 \end{bmatrix} & \underbrace{\begin{bmatrix} 2 & 1 & 5 \\ 4 & 4 & -4 \\ 1 & 3 & 1 \end{bmatrix}}_A & = & \begin{bmatrix} 4 & 4 & -4 \\ 2 & 1 & 5 \\ 1 & 3 & 1 \end{bmatrix} & \begin{array}{c} \\ \begin{bmatrix} & 1 & \\ 1 & & \\ & & 1 \end{bmatrix} \end{array} & \begin{array}{c} \\ \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} \end{array} \\
\begin{bmatrix} 1 & & \\ \frac{-1}{2} & 1 & \\ & & 1 \end{bmatrix} & \begin{bmatrix} 4 & 4 & -4 \\ 2 & 1 & 5 \\ 1 & 3 & 1 \end{bmatrix} & = & \begin{bmatrix} 4 & 4 & -4 \\ 0 & -1 & 7 \\ 1 & 3 & 1 \end{bmatrix} & \begin{array}{c} \\ \\ \begin{bmatrix} & 1 & \\ 1 & & \\ & & 1 \end{bmatrix} \end{array} & \begin{array}{c} \\ \\ \begin{bmatrix} 1 & & \\ \frac{1}{2} & 1 & \\ & & 1 \end{bmatrix} \end{array} \\
\begin{bmatrix} & 1 & \\ \frac{-1}{4} & & \\ & & 1 \end{bmatrix} & \begin{bmatrix} 4 & 4 & -4 \\ 0 & -1 & 7 \\ 1 & 3 & 1 \end{bmatrix} & = & \begin{bmatrix} 4 & 4 & -4 \\ 0 & -1 & 7 \\ 0 & 2 & 2 \end{bmatrix} & \begin{array}{c} \\ \\ \\ \begin{bmatrix} & 1 & \\ 1 & & \\ & & 1 \end{bmatrix} \end{array} & \begin{array}{c} \\ \\ \\ \begin{bmatrix} 1 & & \\ \frac{1}{2} & 1 & \\ \frac{1}{4} & & \\ & & 1 \end{bmatrix} \end{array} \\
\begin{bmatrix} 1 & & \\ & 0 & 1 \\ & +1 & 0 \end{bmatrix} & \begin{bmatrix} 4 & 4 & -4 \\ 0 & -1 & 7 \\ 0 & 2 & 2 \end{bmatrix} & = & \begin{bmatrix} 4 & 4 & -4 \\ 0 & 2 & 2 \\ 0 & -1 & 7 \end{bmatrix} & P_2 \cdot P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, & \begin{bmatrix} 1 & & \\ \frac{1}{4} & 1 & \\ \frac{1}{2} & & \\ & & 1 \end{bmatrix} \\
\begin{bmatrix} 1 & & \\ & 1 & \\ & & \frac{1}{2} \end{bmatrix} & \begin{bmatrix} 4 & 4 & -4 \\ 0 & 2 & 2 \\ 0 & -1 & 7 \end{bmatrix} & = & \begin{bmatrix} 4 & 4 & -4 \\ 0 & 2 & 2 \\ 0 & 0 & 8 \end{bmatrix} & \begin{array}{c} \\ \\ \\ \begin{bmatrix} & 1 & \\ & & 1 \\ 1 & & \end{bmatrix} \end{array} & \begin{array}{c} \\ \\ \\ \begin{bmatrix} 1 & & \\ \frac{1}{4} & 1 & \\ \frac{1}{2} & \frac{-1}{2} & \\ & & 1 \end{bmatrix} \end{array} \\
\therefore \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & 1 & 5 \\ 4 & 4 & -4 \\ 1 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{4} & 1 & 0 \\ \frac{1}{2} & \frac{-1}{2} & 1 \end{bmatrix} \begin{bmatrix} 4 & 4 & -4 \\ 0 & 2 & 2 \\ 0 & 0 & 8 \end{bmatrix}
\end{array}$$

En general P no es conocido a priori y se calcula durante la ejecución:

$$\begin{array}{c}
A \\
P_1 A \\
L'_1 P_1 A \\
P_2 L'_1 P_1 A \\
L'_2 P_2 L'_1 P_1 A \\
\vdots \\
PA = LU
\end{array}$$

$$\begin{bmatrix} \ddots & & & & \\ & \ddots & & & \\ 0 & & \ddots & & \\ \vdots & 0 & (a_{jj} & \cdots & a_{jn}) \\ \vdots & \vdots & \vdots & & \vdots \\ & & (a_{kj} & \cdots & a_{kn}) \\ 0 & 0 & \vdots & & \vdots \end{bmatrix} \quad \leftarrow$$

$$|a_{k,j}| \geq |a_{l,j}|, \quad \forall l \geq j$$

Permutar filas

$$\begin{bmatrix} \ddots & & & \cdots \\ & \ddots & & \cdots \\ 0 & & \ddots & \cdots \\ \vdots & 0 & a_{kj} & \cdots & a_{kn} \\ & & a_{j+1,j} & \cdots & a_{j+1,n} \\ & & \vdots & & \vdots \\ \vdots & \vdots & a_{j,j} & \cdots & a_{j,n} \\ 0 & 0 & \vdots & & \vdots \end{bmatrix}, \text{ luego continuamos la Eliminación Gaussiana.}$$

Una vez que obtenemos las matrices de la igualdad $PA=LU$, para resolver el sistema $A\mathbf{x} = \mathbf{b}$ procedemos a:

- Multiplicar por P por la izquierda a ambos lados para obtener:
 $PA\mathbf{x} = P\mathbf{b}$
 $LU\mathbf{x} = P\mathbf{b}$
- Resolver $L\mathbf{c} = P\mathbf{b}$ para \mathbf{c} , usando Forward Substitution.
- Resolver $U\mathbf{x} = \mathbf{c}$ para \mathbf{x} , usando Backward Substitution.

4.6. La Factorización de Cholesky

Importante: Este algoritmo sólo funciona para matrices definidas positivas y simétricas.

Def 4. La matriz A de $n \times n$ es definida positiva si $\mathbf{x}^T A \mathbf{x} > 0$ para todo vector no nulo \mathbf{x} .

Recordar: A es simétrica si $A = A^T$

4.6.1. Submatriz

Una submatriz principal de A es una submatriz cuadrada de A donde los elementos de su diagonal son también elementos de la diagonal de A .

4.6.2. Propiedades de una matriz definida positiva y simétrica

1. Si una matriz A de $n \times n$ es simétrica, entonces A es definida positiva si y sólo si todos sus valores propios son positivos.
2. Si una matriz A de $n \times n$ es simétrica y definida positiva, y X es una matriz de $m \times n$ de "full rank" con $n \geq m$, entonces $X^T A X$ es simétrica y definida positiva.
3. Cualquier submatriz principal de una matriz A simétrica y definida positiva son también simétricas y definida positiva.

4.6.2.1. Caso base

Para matriz de 2×2 .

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad \det(A) = \underline{\hspace{2cm}} > \underline{\hspace{2cm}}$$

Ahora queremos encontrar la siguiente factorización:

$$A = R^T R = \begin{bmatrix} \diagup & & 0 \\ & \diagup & \\ & & \diagup \end{bmatrix} \begin{bmatrix} \diagdown & & 0 \\ & \diagdown & \\ & & \diagdown \end{bmatrix}$$

donde R es una matriz triangular superior... ¿Cómo se obtiene?

4.6.3. Descomposición de matrices simétricas y definida positiva

Thm 13. Si una matriz A de $n \times n$ es simétrica y definida positiva, se puede descomponer en $A = R^T R$.

4.6.3.1. Caso Base

Caso base: 2×2

$$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix} = \begin{bmatrix} r_{11} & 0 \\ r_{21} & r_{22} \end{bmatrix} \begin{bmatrix} r_{11} & r_{21} \\ 0 & r_{22} \end{bmatrix}$$

Multiplicando $R^T \cdot R$:

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix} = \begin{bmatrix} r_{11}^2 & r_{11} r_{21} \\ r_{11} r_{21} & r_{21}^2 + r_{22}^2 \end{bmatrix}$$

Comparando término a término:

$$a = r_{11}^2 \quad \Rightarrow \quad r_{11} = \pm \sqrt{a} = \sqrt{a}$$

$$b = r_{11} r_{21} \quad \Rightarrow \quad r_{21} = \frac{b}{\sqrt{a}}$$

$$c = r_{21}^2 + r_{22}^2 \quad \Rightarrow \quad c = \frac{b^2}{a} + r_{22}^2 \quad \Rightarrow \quad r_{22} = \frac{\pm \sqrt{ac - b^2}}{\sqrt{a}}$$

$$r_{22} = \frac{\sqrt{ac - b^2}}{\sqrt{a}} \text{ (Por convención se elige la raíz positiva)}$$

¿Es $a \cdot c - b^2 > 0$?

Sí/No ¿Por qué?

$$\begin{aligned} \Rightarrow \begin{bmatrix} a & b \\ b & c \end{bmatrix} &= \begin{bmatrix} \sqrt{a} & 0 \\ \frac{b}{\sqrt{a}} & \sqrt{\frac{ac-b^2}{a}} \end{bmatrix} \begin{bmatrix} \sqrt{a} & \frac{b}{\sqrt{a}} \\ 0 & \sqrt{\frac{ac-b^2}{a}} \end{bmatrix} \\ &= R^T R \end{aligned}$$

¿Cómo se utiliza para resolver un sistema de ecuaciones lineales?

$$\begin{array}{ccc} A & \mathbf{x} & = \mathbf{b} \quad (\text{"A" es simétrica y positiva definida}) \\ \downarrow & & \\ R^T R & \mathbf{x} & = \mathbf{b} \end{array}$$

$$\left. \begin{array}{l} R^T \mathbf{c} = \mathbf{b}, \text{ Forward substitution} \\ R \mathbf{x} = \mathbf{c}, \text{ Backward substitution} \end{array} \right\} \text{ Similar a } PA = LU \text{ o } A = LU.$$

4.6.3.2. Caso General

En general:

$$A = R^T R$$

$$A = \left[\begin{array}{c|c} a & \mathbf{b}^T \\ \hline \mathbf{b} & C \end{array} \right] = \left[\begin{array}{c|ccc} R_{11} & 0 & \cdots & 0 \\ R_{21} & R_{22} & & \\ \vdots & & \ddots & \\ R_{n1} & & & R_{nn} \end{array} \right] \left[\begin{array}{c|ccc} R_{11} & R_{21} & \cdots & R_{n1} \\ 0 & & & \\ \vdots & & \ddots & \\ 0 & & & R_{nn} \end{array} \right]$$

$$\Rightarrow \left[\begin{array}{c} a \\ \mathbf{b} \end{array} \right] = R_{11} \left[\begin{array}{c} R_{11} \\ R_{21} \\ \vdots \\ R_{n1} \end{array} \right] \Rightarrow \begin{aligned} R_{11} &= \sqrt{a} \\ R_{2:n,1} &= \frac{\mathbf{b}}{\sqrt{a}} \end{aligned}$$

$$\begin{aligned} \Rightarrow A &= \left[\begin{array}{c|c} a & \mathbf{b}^T \\ \hline \mathbf{b} & C \end{array} \right] = \left[\begin{array}{c|c} \sqrt{a} & 0 \\ \hline \frac{\mathbf{b}}{\sqrt{a}} & R_1^T \end{array} \right] \cdot \left[\begin{array}{c|c} \sqrt{a} & \mathbf{b}^T/\sqrt{a} \\ \hline 0 & R_1 \end{array} \right] \\ &= \left[\begin{array}{c|c} a & \mathbf{b}^T \\ \hline \mathbf{b} & \frac{\mathbf{b}}{\sqrt{a}} \cdot \frac{\mathbf{b}^T}{\sqrt{a}} + R_1^T R_1 \end{array} \right] \end{aligned}$$

$$\therefore C = \frac{\mathbf{b} \cdot \mathbf{b}^T}{a} + R_1^T R_1$$

¿Cómo encontramos $R_1^T R_1$? (¡Tenemos que hacer Cholesky otra vez!)
$$\Rightarrow C - \frac{1}{a} \mathbf{b} \cdot \mathbf{b}^T = R_1^T R_1, \quad \text{Pero para una matriz más pequeña. Hasta que ... ¡lleguemos a una matriz de } 2 \times 2!$$

Recuerde: $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix} = \begin{bmatrix} \sqrt{a} & 0 \\ \frac{b}{\sqrt{a}} & \sqrt{c - \frac{b^2}{a}} \end{bmatrix} \begin{bmatrix} \sqrt{a} & \frac{b}{\sqrt{a}} \\ 0 & \sqrt{c - \frac{b^2}{a}} \end{bmatrix}$

Ejemplo: Resolver $\begin{bmatrix} 4 & 0 & -2 \\ 0 & 1 & 1 \\ -2 & 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix}$

$$\begin{array}{ccc} \begin{bmatrix} 4 & 0 & -2 \\ 0 & 1 & 1 \\ -2 & 1 & 3 \end{bmatrix} & \rightarrow & \begin{bmatrix} 4 & 0 & -1 \\ 0 & & R_1 \\ 0 & & R_0 \end{bmatrix} \end{array} \quad \Rightarrow C = \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix}, \frac{\mathbf{b}}{\sqrt{a}} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

$$\frac{\mathbf{b} \cdot \mathbf{b}^T}{a} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\Rightarrow R_1^T R_1 = C - \frac{1}{a} \mathbf{b} \cdot \mathbf{b}^T = \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

$$\begin{array}{ccc} \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} & c=2 & \\ \begin{array}{ccc} A_1 & R_1^T & R_1 \end{array} & c - \frac{\mathbf{b} \cdot \mathbf{b}^T}{a} = 2 - \frac{1 \cdot 1}{1} = 1 & \end{array}$$

$$\begin{bmatrix} 4 & 0 & -2 \\ 0 & 1 & 1 \\ -2 & 1 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{Forward Substitution}$$

$$\Rightarrow \begin{array}{ccc} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix} & \Rightarrow & \begin{array}{l} c_1 = 2 \\ c_2 = 2 \\ -1 \cdot 2 + 1 \cdot 2 + c_3 = 0 \\ c_3 = 0 \end{array} \\ \begin{array}{ccc} R^T & \mathbf{c} & = \mathbf{b} \end{array} & & \end{array}$$

$$\mathbf{R} \cdot \mathbf{x} = \mathbf{c}$$

Backward Substitution

$$\begin{bmatrix} 2 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix} \quad \Rightarrow \quad \begin{array}{l} \downarrow \\ x_3 = 0 \\ x_2 = 2 \\ x_1 = 1 \end{array}$$

Recuerde:

$$\begin{bmatrix} 4 & 0 & -2 \\ 0 & 1 & 1 \\ -2 & 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix}$$

o

$$1 \cdot \begin{bmatrix} 4 \\ 0 \\ -2 \end{bmatrix} + 2 \cdot \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} + 0 \cdot \begin{bmatrix} -2 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix}$$

4.7. Métodos Iterativos

La eliminación Gaussiana, $PA=LU$ y el método de Cholesky se conocen como Métodos Directos, los cuales se ejecutan una cantidad finita de operaciones elementales.

El problema de los métodos directos es que son muy costosos computacionalmente (la eliminación Gaussiana y $PA=LU$ tienen complejidad $\frac{2}{3}n^3$, y Cholesky $\frac{1}{3}n^3$).

Los Métodos Iterativos comienzan desde una aproximación de la solución y van mejorando la aproximación en cada iteración.

4.7.1. Método de Jacobi

Este método toma una matriz A , y la separa en tres matrices L , U , D (ADVERTENCIA: ¡ L y U son diferentes de $PA=LU$ o $A=LU$!), donde:

- L : Elementos bajo de la diagonal de A
- U : Elementos sobre de la diagonal de A
- D : Diagonal de A

En otras palabras, $A = L + U + D$. Matricialmente la podemos ver como:

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} 0 & \cdots & 0 \\ a_{21} & 0 & \cdots \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & 0 \end{bmatrix} + \begin{bmatrix} a_{11} & \cdots & 0 \\ 0 & a_{22} & \cdots \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_{nn} \end{bmatrix} + \begin{bmatrix} 0 & a_{12} & \cdots & a_{1n} \\ 0 & 0 & a_{23} & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & a_{n-1,n} \\ 0 & \cdots & 0 & 0 \end{bmatrix}$$

Descomposición "simple" de A

¿Recuerdan la iteración de punto fijo? ¿Qué se hace?

$$f(x) = ax - b$$

↓

$$x_{i+1} = g(x_i)$$

$$x_0 = \text{"dato"}$$

Ahora la derivación:

$$\begin{aligned} A\mathbf{x} &= \mathbf{b} & \iff & \mathbf{F}(\mathbf{x}) = A\mathbf{x} - \mathbf{b} \\ (L + D + U)\mathbf{x} &= \mathbf{b} & & \tilde{\mathbf{F}}(\mathbf{r}) = 0 \\ L\mathbf{x} + D\mathbf{x} + U\mathbf{x} &= \mathbf{b} \end{aligned}$$

Aquí restamos $(L\mathbf{x} + U\mathbf{x})$ en ambos lados de la ecuación, quedando:

$$D\mathbf{x} = \mathbf{b} - L\mathbf{x} - U\mathbf{x}$$

multiplicamos por la izquierda por D^{-1} , donde D es no singular:

$$\mathbf{x} = D^{-1}(\mathbf{b} - L\mathbf{x} - U\mathbf{x})$$

∴ El Método de Jacobi se puede entender como la iteración de punto fijo de lo anterior:

$$\begin{aligned}\mathbf{x}_0 &= \text{"dato inicial"} \\ \mathbf{x}_{n+1} &= D^{-1}(\mathbf{b} - (L + U)\mathbf{x}_n)\end{aligned}$$

la cual es una iteración de punto fijo vectorial de la forma:

$$\mathbf{x}_{n+1} = G(\mathbf{x}_n)$$

Nota:

$$\begin{aligned}\mathbf{x}_{n+1} &= D^{-1}(\mathbf{b} - (L + U)\mathbf{x}_n) = D^{-1}\mathbf{b} - D^{-1} \cdot (L + U)\mathbf{x}_n \\ &= D^{-1}(\mathbf{b} - (L + U + D - D)\mathbf{x}_n) \\ &= D^{-1}(\mathbf{b} - A\mathbf{x}_n + D\mathbf{x}_n) \\ &= D^{-1}(\mathbf{b} - A\mathbf{x}_n) + D^{-1}D\mathbf{x}_n \\ \mathbf{x}_{n+1} &= \mathbf{x}_n + D^{-1}(\mathbf{b} - A\mathbf{x}_n) \\ \mathbf{x}_{n+1} &= \mathbf{x}_n + D^{-1}\mathbf{r}_n\end{aligned}$$

donde $\mathbf{r}_n = \mathbf{b} - A\mathbf{x}_n$ es el vector residual.

4.7.2. Método de Gauss-Seidel

El método de Gauss-Seidel nace de un despeje diferente de la ecuación $(L + U + D)\mathbf{x} = \mathbf{b}$, de la forma: Derivación:

$$\begin{aligned}A\mathbf{x} &= \mathbf{b} \\ (L + D + U)\mathbf{x} &= \mathbf{b} \\ (L + D)\mathbf{x} + U\mathbf{x} &= \mathbf{b} \\ (L + D)\mathbf{x} &= \mathbf{b} - U\mathbf{x} \\ \Rightarrow \mathbf{x}_0 &= \text{"dato inicial"} \\ \mathbf{x}_{n+1} &= (L + D)^{-1}(\mathbf{b} - U\mathbf{x}_n) = (L + D)^{-1}\mathbf{b} - (L + D)^{-1}U\mathbf{x}_n\end{aligned}$$

Donde $(L + D)$ es una matriz triangular inferior, por lo que se utiliza Forward Substitution.

Otra forma en que podemos ver el método de Gauss-Seidel es en base al residuo del sistema, esto nace de:

$$\begin{aligned}\Rightarrow \mathbf{x}_{n+1} &= (L + D)^{-1}(\mathbf{b} - (L + U + D)\mathbf{x}_n + (L + D)\mathbf{x}_n) \\ \mathbf{x}_{n+1} &= \mathbf{x}_n + (L + D)^{-1}\mathbf{r}_n\end{aligned}$$

4.7.3. Successive Over-Relaxation (SOR(ω))

Se define $\omega \in \mathbb{R}$, al cual llamamos el Parámetro de Relajación. Si $\omega > 1$, se refiere como sobre- relajación. Luego, multiplicamos por ω la igualdad $(L + U + D)\mathbf{x} = \mathbf{b}$, y despejamos como:

$$\begin{aligned}
(L + D + U) \mathbf{x} &= \mathbf{b} & / \cdot \omega, \omega \neq 0 \\
\omega(L + D + U) \mathbf{x} &= \omega \mathbf{b} & / + D \mathbf{x} \\
\omega \underline{L} \mathbf{x} + \omega D \mathbf{x} + \omega U \mathbf{x} + \underline{D} \mathbf{x} &= \omega \mathbf{b} + D \mathbf{x} \\
(\omega L + D) \mathbf{x} &= \omega \mathbf{b} + (1 - \omega) D \mathbf{x} - \omega U \mathbf{x}
\end{aligned}$$

Finalmente, el método iterativo nos quedaría como:

$$\begin{aligned}
\Rightarrow \quad \mathbf{x}_0 &= \text{"dato inicial"} \\
\mathbf{x}_{n+1} &= (\omega L + D)^{-1} (\omega \mathbf{b} + [(1 - \omega) D - \omega U] \mathbf{x}_n) \\
&= (\omega L + D)^{-1} \omega \mathbf{b} + (\omega L + D)^{-1} [(1 - \omega) D - \omega U] \mathbf{x}_n
\end{aligned}$$

Y, nuevamente, también podemos ver este método en términos de su residuo como:

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \left(L + \frac{D}{\omega} \right)^{-1} \cdot (\mathbf{b} - A \cdot \mathbf{x}_n) = \mathbf{x}_n + \left(L + \frac{D}{\omega} \right)^{-1} \mathbf{r}_n$$

4.7.4. Tabla Resumen

	$\mathbf{x}_{n+1} = G(\mathbf{x}_n)$	$\mathbf{x}_{n+1} = M \cdot \mathbf{x}_n + \mathbf{b}$	$\mathbf{x}_{n+1} = \mathbf{x}_n + N \cdot \mathbf{r}_n$
Jacobi	$\mathbf{x}_{n+1} = D^{-1} (\mathbf{b} - (L + U) \mathbf{x}_n)$	$\mathbf{x}_{n+1} = -D^{-1} (L + U) \mathbf{x}_n + D^{-1} \mathbf{b}$	$\mathbf{x}_{n+1} = \mathbf{x}_n + D^{-1} \mathbf{r}_n$
G-S	$\mathbf{x}_{n+1} = (L + D)^{-1} (\mathbf{b} - U \mathbf{x}_n)$	$\mathbf{x}_{n+1} = -(L + D)^{-1} U \mathbf{x}_n + (L + D)^{-1} \mathbf{b}$	$\mathbf{x}_{n+1} = \mathbf{x}_n + (L + D)^{-1} \mathbf{r}_n$
SOR(ω)	$\mathbf{x}_{n+1} = (\omega L + D)^{-1} (\omega \mathbf{b} + [(1 - \omega) D - \omega U] \mathbf{x}_n)$	$\mathbf{x}_{n+1} = (\omega L + D)^{-1} [(1 - \omega) D - \omega U] \mathbf{x}_n + (\omega L + D)^{-1} \omega \mathbf{b}$	$\mathbf{x}_{n+1} = \mathbf{x}_n + \left(L + \frac{D}{\omega} \right)^{-1} \mathbf{r}_n$

4.7.5. Convergencia

Para todo método iterativo de la forma $\mathbf{x}_{n+1} = M \cdot \mathbf{x}_n + \mathbf{b}$ tendremos lo siguiente:

$$\begin{aligned}
\mathbf{x}_{n+1} &= M \cdot \mathbf{x}_n + \mathbf{b} \\
\mathbf{x}_{n+2} &= M \cdot \mathbf{x}_{n+1} + \mathbf{b}
\end{aligned}$$

Restando las 2 ecuaciones:

$$\mathbf{x}_{n+2} - \mathbf{x}_{n+1} = M \cdot \mathbf{x}_{n+1} + \mathbf{b} - M \cdot \mathbf{x}_n - \mathbf{b}$$

Aplicando la norma:

$$\|\mathbf{x}_{n+2} - \mathbf{x}_{n+1}\| = \|M \cdot (\mathbf{x}_{n+1} - \mathbf{x}_n)\|$$

Luego, podemos observar que el error en este paso se puede medir como:

$$e_{n+1} = \|\mathbf{x}_{n+2} - \mathbf{x}_{n+1}\|$$

Ahora, podemos observar que se cumple lo siguiente:

$$\begin{aligned}
\Rightarrow \|\mathbf{x}_{n+2} - \mathbf{x}_{n+1}\| &\leq \|M\| \cdot \|\mathbf{x}_{n+1} - \mathbf{x}_n\| \\
e_{n+1} &\leq \|M\| \cdot e_n
\end{aligned}$$

En donde se tiene que si $\|M\| < 1$, el método reducirá el error, por lo tanto convergerá. Una posible norma a utilizar es $\|M\|_\infty$, pero se puede utilizar cualquier norma compatible.

Def5. La matriz $A = (a_{ij})$ cuadrada de dimensión $n \times n$ es estrictamente diagonal dominante si para cada $1 \leq i \leq n$, $|a_{ii}| > \sum_{j \neq i} |a_{ij}|$. En otras palabras, es cuando en cada una de sus filas el valor absoluto de su diagonal es mayor estricto que la suma de los valores absolutos de los demás componentes de dicha fila.

Thm 14. Si la matriz A de dimensión $n \times n$ es estrictamente diagonal dominante se cumple que:

(1) A es una matriz no singular.

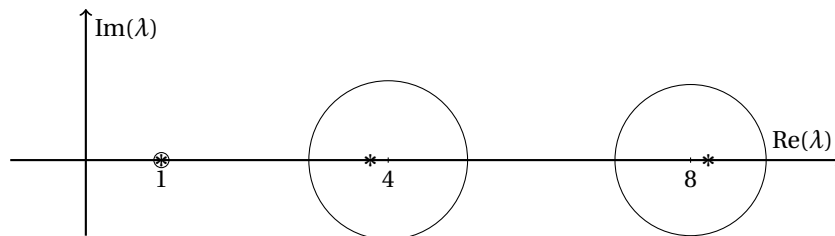
(2) Los métodos de Jacobi y Gauss-Seidel convergen a una solución única de $A\mathbf{x} = \mathbf{b}$, $\forall \mathbf{b}$ y $\forall \mathbf{x}$.

4.7.5.1. Gershgorin circle theorem:

Thm 15. Sea A una matriz de $n \times n$. Cada valor propio λ de A pertenece por lo menos a uno de los discos $|\lambda - a_{ii}| \leq \sum_{j \neq i} |a_{ij}|$.

$$\text{Ej: } \begin{pmatrix} 8 & 1 & 0 \\ 1 & 4 & \epsilon \\ 0 & \epsilon & 1 \end{pmatrix}, |\epsilon| < 1 \Rightarrow \begin{array}{ll} |\lambda - 8| \leq 1 & \lambda_1 = 8,23607 \\ |\lambda - 4| \leq 1 + |\epsilon| & \lambda_2 = 3,76397 \\ |\lambda - 1| \leq |\epsilon| & \lambda_3 = 0,99965 \end{array}$$

$\epsilon = 0,1$



4.7.6. Comentarios varios

- Ahora nos surge la duda ¿Por qué preferir utilizar métodos iterativos a métodos directos? Como ya lo dejamos entrever anteriormente, la razón está en el costo computacional.

Para métodos directos tendremos los siguientes costos:

- Eliminación Gaussiana: $\frac{2}{3}n^3$
- PA=LU: $\frac{2}{3}n^3$
- Cholesky: $\frac{1}{3}n^3$

Mientras tanto, los métodos iterativos tendrán un costo computacional de: $I \cdot n^2$.

Siendo I la cantidad de iteraciones que se necesitaron para llegar al resultado requerido.

- Luego, para los casos en que $I \ll n$, podemos observar que los métodos iterativos son mucho más rápidos que los directos.
- Además, cuando tenemos problemas en donde los valores de A y/o b van cambiando constantemente, también resulta útil hacer uso de métodos iterativos, puesto que cada vez que uno de estos valores cambie, podemos utilizar las raíces obtenidas con el valor anterior como punto inicial de la iteración, como nuestro \mathbf{x}_0 .
- Se conoce como matriz dispersa (*sparse*) a aquellas matrices en donde muchos de sus componentes valen cero. Por otra parte, una matriz densa es aquella en la que muy pocos (o ninguno) de sus componentes son ceros.
- Al utilizar PA=LU en una matriz dispersa, el método tiende a llenar LU con muchos coeficientes distintos de cero, pasando a ser una matriz densa, lo cual requiere más memoria para ser almacenada.
- Esto no se da al utilizar métodos iterativos, por lo que se prefieren en estos casos también.

Ejemplo:
$$\begin{aligned} 1 \cdot u + 3 \cdot v &= -1 \\ 5 \cdot u + 4 \cdot v &= 6 \end{aligned} \Rightarrow \begin{bmatrix} 5 & 4 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 6 \\ -1 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix} + \begin{bmatrix} 0 & 4 \\ 0 & 0 \end{bmatrix} = L + D + U, \quad \mathbf{x}_0 = \begin{bmatrix} 0 & 0 \end{bmatrix}^T$$

Jacobi:

$$\begin{aligned} \mathbf{x}_{n+1} &= D^{-1} \cdot (b - (L + U) \cdot \mathbf{x}_n) \\ \mathbf{x}_{n+1} &= \begin{bmatrix} 1/5 & 0 \\ 0 & 1/3 \end{bmatrix} \cdot \left(\begin{bmatrix} 6 \\ -1 \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ 4 & 0 \end{bmatrix} \mathbf{x}_n \right) \\ \mathbf{x}_1 &= \begin{bmatrix} 1/5 & 0 \\ 0 & 1/3 \end{bmatrix} \cdot \left(\begin{bmatrix} 6 \\ -1 \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ 4 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right) \\ \mathbf{x}_1 &= \begin{bmatrix} 6/5 \\ -1/3 \end{bmatrix} \\ \mathbf{x}_2 &= \begin{bmatrix} 1/5 & 0 \\ 0 & 1/3 \end{bmatrix} \cdot \left(\begin{bmatrix} 6 \\ -1 \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ 4 & 0 \end{bmatrix} \begin{bmatrix} 6/5 \\ -1/3 \end{bmatrix} \right) \\ \mathbf{x}_2 &= \begin{bmatrix} 1,46 \\ -0,73 \end{bmatrix} \\ \vdots \\ \mathbf{x}_{55} &= \begin{bmatrix} 2 \\ -1 \end{bmatrix} \end{aligned}$$

Gauss-Seidel:

$$\begin{aligned} \mathbf{x}_{n+1} &= (L + D)^{-1} \cdot (b - U \cdot \mathbf{x}_n) \\ \mathbf{x}_{n+1} &= \begin{bmatrix} 5 & 0 \\ 1 & 3 \end{bmatrix}^{-1} \cdot \left(\begin{bmatrix} 6 \\ -1 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 4 & 0 \end{bmatrix} \mathbf{x}_n \right) \\ \mathbf{x}_1 &= \begin{bmatrix} 5 & 0 \\ 1 & 3 \end{bmatrix}^{-1} \cdot \left(\begin{bmatrix} 6 \\ -1 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 4 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right) \\ \mathbf{x}_1 &= \begin{bmatrix} 5 & 0 \\ 1 & 3 \end{bmatrix}^{-1} \begin{bmatrix} 6 \\ -1 \end{bmatrix} \\ &\Rightarrow \begin{bmatrix} 5 & 0 \\ 1 & 3 \end{bmatrix} \mathbf{x}_1 = \begin{bmatrix} 6 \\ -1 \end{bmatrix} \text{ iES!} \\ \mathbf{x}_1 &= \begin{bmatrix} 6/5 \\ -0,73 \end{bmatrix} \\ \mathbf{x}_2 &= \begin{bmatrix} 5 & 0 \\ 1 & 3 \end{bmatrix}^{-1} \cdot \left(\begin{bmatrix} 6 \\ -1 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 4 & 0 \end{bmatrix} \begin{bmatrix} 6/5 \\ -0,73 \end{bmatrix} \right) \\ \mathbf{x}_2 &= \begin{bmatrix} 5 & 0 \\ 1 & 3 \end{bmatrix}^{-1} \begin{bmatrix} 8,93 \\ -1 \end{bmatrix} \\ \mathbf{x}_1 &= \begin{bmatrix} 1,78 \\ -0,92 \end{bmatrix} \\ \vdots \\ \mathbf{x}_{28} &= \begin{bmatrix} 2 \\ -1 \end{bmatrix} \end{aligned}$$

dado
↓
SOR(ω), $\omega = 1,09$

$$\begin{aligned} \mathbf{x}_{n+1} &= (\omega \cdot L + D)^{-1} (\omega \cdot b + [(1 - \omega) \cdot D - \omega \cdot U] \mathbf{x}_n) \\ \mathbf{x}_{n+1} &= \begin{bmatrix} 5 & 0 \\ \omega & 3 \end{bmatrix}^{-1} \cdot \left(\begin{bmatrix} 6 \cdot \omega \\ -\omega \end{bmatrix} + \begin{bmatrix} (1 - \omega) \cdot 5 & 4 \cdot \omega \\ 0 & (1 - \omega) \cdot 3 \end{bmatrix} \mathbf{x}_n \right) \\ \mathbf{x}_{n+1} &= \begin{bmatrix} 5 & 0 \\ \omega & 3 \end{bmatrix}^{-1} \cdot \left(\begin{bmatrix} 6 \cdot \omega \\ -\omega \end{bmatrix} + \begin{bmatrix} (1 - \omega) \cdot 5 & 4 \cdot \omega \\ 0 & (1 - \omega) \cdot 3 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right) \\ \mathbf{x}_1 &= \begin{bmatrix} 1,30 \\ -0,83 \end{bmatrix} \\ \mathbf{x}_{n+1} &= \begin{bmatrix} 5 & 0 \\ \omega & 3 \end{bmatrix}^{-1} \cdot \left(\begin{bmatrix} 6 \cdot \omega \\ -\omega \end{bmatrix} + \begin{bmatrix} (1 - \omega) \cdot 5 & 4 \cdot \omega \\ 0 & (1 - \omega) \cdot 3 \end{bmatrix} \begin{bmatrix} 1,30 \\ -0,83 \end{bmatrix} \right) \\ \mathbf{x}_2 &= \begin{bmatrix} 1,78 \\ -0,92 \end{bmatrix} \\ \vdots \\ \mathbf{x}_{17} &= \begin{bmatrix} 1,99 \\ -1,00 \end{bmatrix} \end{aligned}$$

4.8. Método del Gradiente Descendente

La idea principal del gradiente descendente como un problema de optimización, es encontrar el mínimo de una función cuadrática convexa al moverse en la dirección de máximo decrecimiento, es decir es un proceso iterativo. El gradiente ∇f es la dirección de máximo crecimiento, por lo tanto la dirección de máximo decrecimiento es $-\nabla f$.

Después de localizar el mínimo a lo largo de esta dirección de decrecimiento se debe repetir el proceso a partir de este punto y hacer una nueva minimización unidimensional en la nueva dirección.

```

1  for i in range(n) :
2      v = ∇f
3      Minimizar f(xi - sv) para escalar s = s*
4      xi+1 = xi - s* v

```

4.8.1. Minimizar una función cuadrática convexa

$$\Phi(\mathbf{y}) = \frac{1}{2} \mathbf{y}^T A \mathbf{y} - \mathbf{y}^T \cdot \mathbf{b}$$

$$\Rightarrow \nabla \Phi(\mathbf{y}) = \frac{1}{2} (A^T + A) \mathbf{y} - \mathbf{b} = 0$$

$$\begin{array}{l} \text{A es simétrica} \\ \text{y positiva definida} \end{array} \Rightarrow \underbrace{A \mathbf{y} - \mathbf{b} = 0}_{A \mathbf{y} = \mathbf{b}}$$

Recuerde:

$$\text{Jacobi} \Rightarrow \mathbf{x}_{k+1} = \mathbf{x}_k + D^{-1} \cdot \mathbf{r}_k,$$

$$\text{G-S} \Rightarrow \mathbf{x}_{k+1} = \mathbf{x}_k + (L + D)^{-1} \cdot \mathbf{r}_k,$$

$$\text{SOR}(\omega) \Rightarrow \mathbf{x}_{k+1} = \mathbf{x}_k + (L + \frac{D}{\omega})^{-1} \cdot \mathbf{r}_k,$$

$$\Rightarrow \mathbf{x}_{k+1} = \mathbf{x}_k + \text{"Vector"}$$

$$= \mathbf{x}_k - \alpha_k \cdot \mathbf{d}_k$$

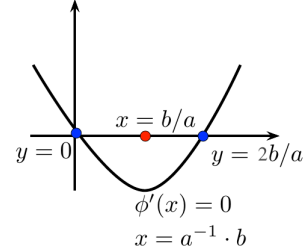
$$\mathbf{r}_k = \mathbf{b} - A \mathbf{x}_k$$

1D-sketch

$$\phi(x) = \frac{1}{2} a x^2 - x b$$

$$\phi'(y) = a y - b$$

$$\phi'(x) = 0 = a x - b \Rightarrow x = \frac{b}{a}$$



Primera opción : Definir \mathbf{d}_k en la dirección de máximo decrecimiento.

$$\Rightarrow \nabla \Phi(\mathbf{x}_k) = A \mathbf{x}_k - \mathbf{b} = -\mathbf{r}_k$$

$$\Rightarrow \mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{r}_k$$

4.8.2. ¿Qué es y cómo se obtiene α_k ?

Nota: Por simplicidad se utilizará α en vez de α_k en el siguiente desarrollo.

$$\Phi(\mathbf{x}_{k+1}) = \frac{1}{2} (\mathbf{x}_k + \alpha \mathbf{r}_k)^T A (\mathbf{x}_k + \alpha \mathbf{r}_k) - (\mathbf{x}_k + \alpha \mathbf{r}_k)^T \mathbf{b}$$

donde $\Phi(\mathbf{x}_{k+1})$ es una función de α cuando \mathbf{x}_k y \mathbf{r}_k son conocidos, $\Rightarrow f(\alpha) = \Phi(\mathbf{x}_k + \alpha \cdot \mathbf{r}_k)$.

4.8.2.1. ¿Cómo se obtiene el "óptimo" de " $f(\alpha)$ "?

$$f'(\alpha) = 0$$

$$\Rightarrow \alpha = \frac{\mathbf{r}_k^T \mathbf{r}_k}{\mathbf{r}_k^T A \mathbf{r}_k}$$

(4.1)

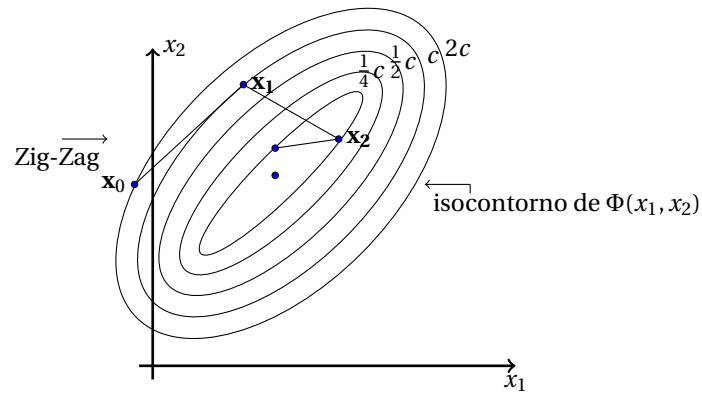
4.8.2.2. Algoritmo

```

1   $\mathbf{x}_0 = \text{"dato"}$ 
2  for  $k$  in  $\text{range}(n)$  :
3       $\mathbf{r}_k = b - A\mathbf{x}_k$ 
4       $\alpha_k = \frac{\mathbf{r}_k^T \mathbf{r}_k}{\mathbf{r}_k^T A \mathbf{r}_k}$ 
5       $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{r}_k$ 

```

4.8.2.3. Gráficamente



$$\Phi(x_1, x_2) = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} x_1 & x_2 \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

4.8.3. Convergencia en matrices simétricas y definida positiva

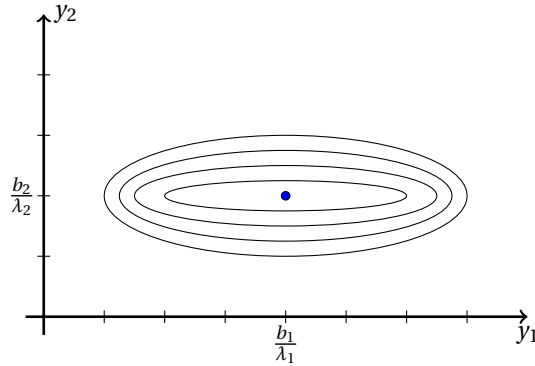
Sea A una matriz simétrica y positiva definida, entonces el método del gradiente descendente converge a la solución de $A\mathbf{x} = \mathbf{b}$, para cualquier \mathbf{x}_0 y \mathbf{b} , y

$$\|\mathbf{e}^{k+1}\|_A \leq \frac{\kappa_2(A) - 1}{\kappa_2(A) + 1} \|\mathbf{e}^k\|_A \quad \left| \quad \kappa(A) = \|A\| \|A^{-1}\| \right.$$

donde $\|\cdot\|_A$ es A -norma, i.e. $\|\mathbf{x}\|_A = \sqrt{\mathbf{x}^T A \mathbf{x}}$ y $\mathbf{e}^k = \mathbf{x}_k - \mathbf{x}$

Ejemplo:

$$\begin{aligned} A &= \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}, \lambda_i > 0, 0 < \lambda_2 \leq \lambda_1, \mathbf{b} = \langle b_1, b_2 \rangle^T \\ \Rightarrow \Phi(\mathbf{y}) &= \frac{1}{2} \langle y_1, y_2 \rangle \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} - \langle y_1, y_2 \rangle \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \\ &= \frac{1}{2} (\lambda_1 y_1^2 + \lambda_2 y_2^2) - b_1 y_1 - b_2 y_2 = C \\ \Rightarrow &\left(\frac{y_1 - b_1/\lambda_1}{\sqrt{2/\lambda_1}} \right)^2 + \left(\frac{y_2 - b_2/\lambda_2}{\sqrt{2/\lambda_2}} \right)^2 = C + \frac{b_1^2}{2\lambda_1} + \frac{b_2^2}{2\lambda_2} \\ &\quad \uparrow \\ &\quad \text{"ellipse"} \end{aligned}$$



4.9. Método del Gradiente Conjugado

Recuerde del capítulo 4.8 que construimos la solución del sistema de ecuaciones lineales iterativamente de la siguiente forma:

$$\begin{aligned}\mathbf{x}_1 &= \mathbf{x}_0 + \alpha_0 \mathbf{r}_0 \\ \mathbf{x}_2 &= \mathbf{x}_1 + \alpha_1 \mathbf{r}^1 \\ &\vdots \\ \mathbf{x}_{k+1} &= \mathbf{x}_k + \alpha_k \mathbf{r}_k\end{aligned}$$

$$\Rightarrow \mathbf{x}_2 = \mathbf{x}_0 + \alpha_0 \mathbf{r}_0 + \alpha_1 \mathbf{r}_1$$

$$\mathbf{x}_2 = \mathbf{x}_0 + \underbrace{\begin{bmatrix} \mathbf{r}_0 & | & \mathbf{r}_1 \end{bmatrix}}_{\text{Base}} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}$$

$$\vdots$$

$$\mathbf{x}_{k+1} = \mathbf{x}_0 + \underbrace{\begin{bmatrix} \mathbf{r}_0 & | & \dots & | & \mathbf{r}_k \end{bmatrix}}_{\text{Base}} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{bmatrix}$$

“k” puede ser mayor que “n”,
i.e. la dimensión del espacio

¿Cuántos vectores “linealmente independientes” necesitamos para construir una base de \mathbb{R}^n ?

¿Necesitamos más de “n” iteraciones entonces?

¿Qué tal si entonces elegimos la dirección convenientemente? i.e. si usamos \mathbf{d}_k en vez de \mathbf{r}_k .

4.9.1. Derivación del Método del Gradiente Conjugado

El método del Gradiente Conjugado (CG) se utiliza para resolver sistemas de ecuaciones lineales, es decir: $A\mathbf{x} = \mathbf{b}$, donde la matriz $A \in \mathbb{R}^{n \times n}$ es simétrica ($A = A^T$) y positiva definida ($\mathbf{x}^T A \mathbf{x} > 0$ para $\mathbf{x} \neq \mathbf{0}$). En este caso se busca construir la solución \mathbf{x} como una combinación lineal de vectores en \mathbb{R}^n linealmente independiente, es decir:

$$\mathbf{x} = \mathbf{x}_0 + \sum_{i=0}^{n-1} \alpha_i \mathbf{d}_i. \quad (4.2)$$

Donde definiremos la siguiente notación:

$$\mathbf{x} = \underbrace{\mathbf{x}_0 + \alpha_0 \mathbf{d}_0}_{\mathbf{x}_1} + \underbrace{\alpha_1 \mathbf{d}_1}_{\mathbf{x}_2} + \underbrace{\alpha_2 \mathbf{d}_2 + \dots + \alpha_{n-1} \mathbf{d}_{n-1}}_{\mathbf{x}_3} = \underbrace{\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3}_{\mathbf{x}_n},$$

donde obtenemos,

$$\begin{aligned} \mathbf{x}_1 &= \mathbf{x}_0 + \alpha_0 \mathbf{d}_0 \\ \mathbf{x}_2 &= \mathbf{x}_1 + \alpha_1 \mathbf{d}_1 \\ &\vdots \\ \mathbf{x}_n &= \mathbf{x}_{n-1} + \alpha_{n-1} \mathbf{d}_{n-1}, \end{aligned}$$

las cuales se pueden expresar de la siguiente forma:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k, \quad k \in \{0, 1, \dots, n-1\}. \quad (4.3)$$

Multiplicando (4.3) por $-A$ por la izquierda,

$$-A\mathbf{x}_{k+1} = -A\mathbf{x}_k - \alpha_k A\mathbf{d}_k$$

y sumando \mathbf{b} en ambos lados obtenemos,

$$\begin{aligned} \underbrace{\mathbf{b} - A\mathbf{x}_{k+1}}_{\mathbf{r}_{k+1}} &= \underbrace{\mathbf{b} - A\mathbf{x}_k}_{\mathbf{r}_k} - \alpha_k A\mathbf{d}_k \\ \mathbf{r}_{k+1} &= \mathbf{r}_k - \alpha_k A\mathbf{d}_k, \end{aligned} \quad (4.4)$$

por lo tanto hemos encontrado una relación entre los vectores residuales en función de las direcciones \mathbf{d}_k . Sin embargo, aún nos falta por definir como encontraremos los α_k y \mathbf{d}_k que se necesitan en las ecuaciones (4.3) y (4.4).

Una primera alternativa a elegir α_k es similar a la utilizada en la ecuación (4.1), pero teniendo la salvedad que ahora la dirección de búsqueda es \mathbf{d}_k y no \mathbf{r}_k como se utilizó en el caso del Gradiente Descendente. En este caso utilizaremos otro camino, el cual consiste en reemplazar (4.2) en $A\mathbf{x} = \mathbf{b}$ obteniendo,

$$\begin{aligned} A \left(\mathbf{x}_0 + \sum_{i=0}^{n-1} \alpha_i \mathbf{d}_i \right) &= \mathbf{b} \\ A\mathbf{x}_0 + \sum_{i=0}^{n-1} \alpha_i A\mathbf{d}_i &= \mathbf{b}. \end{aligned}$$

Re-escribiendo obtenemos,

$$\alpha_0 A\mathbf{d}_0 + \alpha_1 A\mathbf{d}_1 + \dots + \alpha_{n-1} A\mathbf{d}_{n-1} = \mathbf{b} - A\mathbf{x}_0 = \mathbf{r}_0. \quad (4.5)$$

Ahora es conveniente incluir la definición de A -ortogonalidad:

Def 6. *A-ortogonalidad:* Sea $\mathbf{u} \in \mathbb{R}^n$ no nulo, $\mathbf{v} \in \mathbb{R}^n$ no nulo y, $A \in \mathbb{R}^{n \times n}$ simétrica y positiva definida, entonces decimos que \mathbf{u} y \mathbf{v} son *A-ortogonales* (o *conjugados*) si $\langle \mathbf{u}, \mathbf{v} \rangle_A = 0$, donde $\langle \mathbf{u}, \mathbf{v} \rangle_A = \mathbf{u}^T A \mathbf{v}$.

Recuerde que la noción tradicional de ortogonalidad implica $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T \mathbf{v} = \mathbf{u}^T I \mathbf{v} = 0$, i.e. donde la matriz utilizada es la identidad $I \in \mathbb{R}^{n \times n}$.

Multiplicando la ecuación (4.5) por \mathbf{d}_0^T por la izquierda obtenemos,

$$\alpha_0 \mathbf{d}_0^T A \mathbf{d}_0 + \alpha_1 \mathbf{d}_0^T A \mathbf{d}_1 + \cdots + \alpha_{n-1} \mathbf{d}_0^T A \mathbf{d}_{n-1} = \mathbf{d}_0^T \mathbf{r}_0.$$

Entonces si consideramos ahora la definición 6 de A-ortogonalidad para el conjunto de vectores $\{\mathbf{d}_0, \mathbf{d}_1, \dots, \mathbf{d}_{n-1}\}$ obtenemos,

$$\begin{aligned} \alpha_0 \mathbf{d}_0^T A \mathbf{d}_0 + \alpha_1 \underbrace{\mathbf{d}_0^T A \mathbf{d}_1}_0 + \cdots + \alpha_{n-1} \underbrace{\mathbf{d}_0^T A \mathbf{d}_{n-1}}_0 &= \mathbf{d}_0^T \mathbf{r}_0, \\ \alpha_0 \mathbf{d}_0^T A \mathbf{d}_0 &= \mathbf{d}_0^T \mathbf{r}_0, \end{aligned}$$

por lo cual obtenemos $\alpha_0 = \frac{\mathbf{d}_0^T \mathbf{r}_0}{\mathbf{d}_0^T A \mathbf{d}_0}$, lo que es equivalente al valor obtenido por (4.1) considerando la dirección \mathbf{d}_0 en este caso. Antes de pasar al caso general, debemos hacer la siguiente simplificación. Considerando que uno ya obtiene el valor de α_0 y conoce el vector \mathbf{d}_0 uno podría mover al lado derecho de la ecuación lo conocido, es decir obtiene lo siguiente,

$$\alpha_1 A \mathbf{d}_1 + \cdots + \alpha_{n-1} A \mathbf{d}_{n-1} = \underbrace{\mathbf{r}_0 - \alpha_0 A \mathbf{d}_0}_{\mathbf{r}_1}.$$

es decir, va obteniendo vector residual de la siguiente iteración en el lado derecho de la ecuación, ver ecuación (4.4).

En este caso, utilizando la A-ortogonalidad, obtendríamos $\alpha_1 = \frac{\mathbf{d}_1^T \mathbf{r}_1}{\mathbf{d}_1^T A \mathbf{d}_1}$, entonces en el caso general se obtiene,

$$\alpha_k = \frac{\mathbf{d}_k^T \mathbf{r}_k}{\mathbf{d}_k^T A \mathbf{d}_k}. \quad (4.6)$$

Nótese que también se podría usar $\alpha_k = \frac{\mathbf{d}_k^T \mathbf{r}_0}{\mathbf{d}_k^T A \mathbf{d}_k}$, sin embargo por ahora se prefiere utilizar (4.6) dado que necesitaremos \mathbf{r}_k posteriormente. ¿Hay alguna diferencia Matemática en usar α_k como $\frac{\mathbf{d}_k^T \mathbf{r}_0}{\mathbf{d}_k^T A \mathbf{d}_k}$ o $\frac{\mathbf{d}_k^T \mathbf{r}_k}{\mathbf{d}_k^T A \mathbf{d}_k}$? ¿Hay alguna diferencia Computacional¹ en usar α_k como $\frac{\mathbf{d}_k^T \mathbf{r}_0}{\mathbf{d}_k^T A \mathbf{d}_k}$ o $\frac{\mathbf{d}_k^T \mathbf{r}_k}{\mathbf{d}_k^T A \mathbf{d}_k}$?

Hasta este punto solo hemos obtenido α_k , sin embargo también necesitamos definir como obtendremos \mathbf{d}_k , es decir, los vectores linealmente independientes $\{\mathbf{d}_0, \mathbf{d}_1, \dots, \mathbf{d}_{n-1}\}$ que además ¡deben ser A-ortogonales!

Para obtener los vectores \mathbf{d}_k se propone la siguiente ecuación:

$$\mathbf{d}_{k+1} = \mathbf{r}_{k+1} - \beta_k \mathbf{d}_k, \quad (4.7)$$

donde el coeficiente β_k se obtiene nuevamente utilizando la A-ortogonalidad, es decir multiplicando la ecuación (4.7) por $\mathbf{d}_k^T A$ por la izquierda,

$$\underbrace{\mathbf{d}_k^T A \mathbf{d}_{k+1}}_0 = \mathbf{d}_k^T A \mathbf{r}_{k+1} - \beta_k \mathbf{d}_k^T A \mathbf{d}_k,$$

¹¡Le sugiero implementarlo y ver qué ocurre! En realidad ya está implementado en los jupyter notebooks del curso, solo debe modificarlo un poco.

por lo tanto,

$$\beta_k = \frac{\mathbf{d}_k^T A \mathbf{r}_{k+1}}{\mathbf{d}_k^T A \mathbf{d}_k}. \quad (4.8)$$

Por lo cual ahora podemos construir nuestra primera versión del algoritmo,

```

1  $\mathbf{x}_0 = \text{"dato"}$ 
2  $\mathbf{r}_0 = \mathbf{b} - A\mathbf{x}_0$ 
3  $\mathbf{d}_0 = \mathbf{r}_0$ 
4 for  $k$  in  $\text{range}(0, n)$  :
5      $\alpha_k = \mathbf{d}_k^T \mathbf{r}_k / \mathbf{d}_k^T A \mathbf{d}_k \rightarrow A\mathbf{y} = \text{afun}(\mathbf{y})$ 
6      $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$ 
7      $\mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_k A \mathbf{d}_k = \mathbf{b} - A\mathbf{x}_{k+1}$ 
8      $\beta_k = \mathbf{d}_k^T A \mathbf{r}_{k+1} / \mathbf{d}_k^T A \mathbf{d}_k$ 
9      $\mathbf{d}_{k+1} = \mathbf{r}_{k+1} - \beta_k \mathbf{d}_k$ 

```

Notar algo muy importante que hasta ahora no se había definido, esto es, ¿como se define o inicializa \mathbf{d}_0 ? Answer: Look back at the algorithm!

4.9.2. Análisis de la A-ortogonalidad

Para que todo el desarrollo de la sección 4.9.1 se sustente, necesitamos demostrar que:

$$\mathbf{d}_j^T A \mathbf{d}_{k+1} = 0, \quad \text{para } j \in \{0, 1, \dots, k\}, \text{ y } k \in \{0, 1, \dots, n-2\}, \quad (4.9)$$

donde $\mathbf{d}_{k+1} = \mathbf{r}_{k+1} - \beta_k \mathbf{d}_k$, ver ecuación (4.7), y $\beta_k = \frac{\mathbf{d}_k^T A \mathbf{r}_{k+1}}{\mathbf{d}_k^T A \mathbf{d}_k}$, ver ecuación (4.8).

Para demostrar (4.9) procederemos por inducción en k . El caso base es para $k=0$, es decir necesitamos demostrar que $\mathbf{d}_0^T A \mathbf{d}_1 = 0$. Multiplicando por la izquierda la ecuación (4.7) por $\mathbf{d}_0^T A$ y considerando $k=0$ obtenemos,

$$\mathbf{d}_0^T A \mathbf{d}_1 = \mathbf{d}_0^T A \mathbf{r}_1 - \beta_0 \mathbf{d}_0^T A \mathbf{d}_0,$$

reemplazando ahora β_0 en la ecuación anterior. Es decir, utilizar la definición de β_k en ecuación (4.8) con $k=0$, i.e.

$\beta_0 = \frac{\mathbf{d}_0^T A \mathbf{r}_1}{\mathbf{d}_0^T A \mathbf{d}_0}$ obtenemos,

$$\begin{aligned} \mathbf{d}_0^T A \mathbf{d}_1 &= \mathbf{d}_0^T A \mathbf{r}_1 - \underbrace{\beta_0}_{\frac{\mathbf{d}_0^T A \mathbf{r}_1}{\mathbf{d}_0^T A \mathbf{d}_0}} \mathbf{d}_0^T A \mathbf{d}_0 \\ \mathbf{d}_0^T A \mathbf{d}_1 &= \mathbf{d}_0^T A \mathbf{r}_1 - \frac{\mathbf{d}_0^T A \mathbf{r}_1}{\mathbf{d}_0^T A \mathbf{d}_0} \mathbf{d}_0^T A \mathbf{d}_0 \\ \mathbf{d}_0^T A \mathbf{d}_1 &= \mathbf{d}_0^T A \mathbf{r}_1 - \mathbf{d}_0^T A \mathbf{r}_1 \\ \mathbf{d}_0^T A \mathbf{d}_1 &= 0. \end{aligned}$$

Por lo tanto se cumple el caso base $\mathbf{d}_0^T A \mathbf{d}_1 = 0$.

Ahora asumamos que el conjunto de vectores $\{\mathbf{d}_0, \mathbf{d}_1, \dots, \mathbf{d}_{k-1}\}$ son A-ortogonales. Considere ahora la ecuación (4.4) donde la multiplicamos por la izquierda por \mathbf{d}_j^T ,

$$\mathbf{d}_j^T \mathbf{r}_k = \mathbf{d}_j^T \mathbf{r}_{k-1} - \alpha_{k-1} \mathbf{d}_j^T A \mathbf{d}_{k-1}, \quad (4.10)$$

donde tenemos 2 casos: (i) $j = k - 1$ y (ii) $0 \leq j < k - 1$. Para el primer caso, (i), obtenemos,

$$\begin{aligned} \mathbf{d}_{k-1}^T \mathbf{r}_k &= \mathbf{d}_{k-1}^T \mathbf{r}_{k-1} - \alpha_{k-1} \mathbf{d}_{k-1}^T A \mathbf{d}_{k-1} \\ &= \mathbf{d}_{k-1}^T \mathbf{r}_{k-1} - \frac{\mathbf{d}_{k-1}^T \mathbf{r}_{k-1}}{\mathbf{d}_{k-1}^T A \mathbf{d}_{k-1}} \mathbf{d}_{k-1}^T A \mathbf{d}_{k-1} \\ &= \mathbf{d}_{k-1}^T \mathbf{r}_{k-1} - \mathbf{d}_{k-1}^T \mathbf{r}_{k-1} \\ &= 0. \end{aligned}$$

Para el segundo caso, (ii), obtenemos,

$$\begin{aligned} \mathbf{d}_j^T \mathbf{r}_k &= \mathbf{d}_j^T \mathbf{r}_{k-1} - \alpha_{k-1} \underbrace{\mathbf{d}_j^T A \mathbf{d}_{k-1}}_{0 \text{ por } A\text{-ortogonalidad}} \\ &= \mathbf{d}_j^T \mathbf{r}_{k-1}, \end{aligned}$$

reemplazando por la definición de \mathbf{r}_{k-1} obtenemos,

$$\begin{aligned} \mathbf{d}_j^T \mathbf{r}_k &= \mathbf{d}_j^T \mathbf{r}_{k-1} \\ &= \mathbf{d}_j^T (\mathbf{r}_{k-2} - \alpha_{k-2} A \mathbf{d}_{k-2}) \\ &= \mathbf{d}_j^T \mathbf{r}_{k-2} - \alpha_{k-2} \underbrace{\mathbf{d}_j^T A \mathbf{d}_{k-2}}_0 \\ &= \mathbf{d}_j^T \mathbf{r}_{k-2}, \end{aligned}$$

siguiendo el mismo desarrollo obtenemos la misma recurrencia hasta que lleguemos al $j + 1$ -ésimo residuo,

$$\begin{aligned} \mathbf{d}_j^T \mathbf{r}_k &= \mathbf{d}_j^T \mathbf{r}_{k-1} \\ &= \mathbf{d}_j^T \mathbf{r}_{k-2} \\ &\vdots \\ &= \mathbf{d}_j^T \mathbf{r}_{j+1} \\ &= \mathbf{d}_j^T (\mathbf{r}_j - \alpha_j A \mathbf{d}_j), \end{aligned}$$

al reemplazar por la definición de α_j obtenemos,

$$\begin{aligned} \mathbf{d}_j^T \mathbf{r}_k &= \mathbf{d}_j^T (\mathbf{r}_j - \alpha_j A \mathbf{d}_j) \\ &= \mathbf{d}_j^T \mathbf{r}_j - \alpha_j \mathbf{d}_j^T A \mathbf{d}_j \\ &= \mathbf{d}_j^T \mathbf{r}_j - \frac{\mathbf{d}_j^T \mathbf{r}_j}{\mathbf{d}_j^T A \mathbf{d}_j} \mathbf{d}_j^T A \mathbf{d}_j \\ &= \mathbf{d}_j^T \mathbf{r}_j - \mathbf{d}_j^T \mathbf{r}_j \\ &= 0. \end{aligned}$$

Por lo tanto,

$$\mathbf{d}_j^T \mathbf{r}_k = 0, \quad j \in \{0, 1, \dots, k-1\}. \quad (4.11)$$

Note que la ecuación (4.11) es válida para \mathbf{r}_k , ahora consideremos \mathbf{r}_{k+1} , i.e. obtengamos el producto interno entre \mathbf{d}_j y \mathbf{r}_{k+1} ,

$$\mathbf{d}_j^T \mathbf{r}_{k+1} = \underbrace{\mathbf{d}_j^T \mathbf{r}_k}_{0 \text{ por (4.11)}} - \alpha_k \underbrace{\mathbf{d}_j^T A \mathbf{d}_k}_0, \quad \text{para } j \in \{0, 1, \dots, k-1\} \quad (4.12)$$

Por lo tanto podemos concluir que el conjunto $\{\mathbf{d}_0, \mathbf{d}_1, \dots, \mathbf{d}_{k-1}\}$ es ortogonal a \mathbf{r}_{k+1} .

Otra relación interesante es entre los vectores residuales. Considere el producto interno entre \mathbf{r}_k y j -ésimo vector \mathbf{d}_j , ver ecuación (4.7),

$$\underbrace{\mathbf{r}_k^T \mathbf{d}_j}_{0 \text{ por (4.11)}} = \mathbf{r}_k^T \mathbf{r}_j - \beta_{j-1} \underbrace{\mathbf{r}_k^T \mathbf{d}_{j-1}}_{0 \text{ por (4.11)}} \\ 0 = \mathbf{r}_k^T \mathbf{r}_j. \quad (4.13)$$

Por lo tanto $\mathbf{r}_k^T \mathbf{r}_j = 0$ para $j \in \{0, 1, \dots, k-1\}$, i.e. los vectores residuales son ortogonales.

Nos falta el último paso, determinar si efectivamente \mathbf{d}_{k+1} es A -ortogonal a \mathbf{d}_j para $j \in \{0, 1, \dots, k\}$. Multipliquemos entonces \mathbf{d}_{k+1} (ver ecuación (4.7)) por la izquierda por $\mathbf{d}_j^T A$, lo que nos da:

$$\mathbf{d}_j^T A \mathbf{d}_{k+1} = \mathbf{d}_j^T A \mathbf{r}_{k+1} - \beta_k \mathbf{d}_j^T A \mathbf{d}_k.$$

Nuevamente tenemos 2 casos: (i) $j = k$ y (ii) $j < k$. Para el primer caso, (i), obtenemos:

$$\begin{aligned} \mathbf{d}_k^T A \mathbf{d}_{k+1} &= \mathbf{d}_k^T A \mathbf{r}_{k+1} - \beta_k \mathbf{d}_k^T A \mathbf{d}_k \\ &= \mathbf{d}_k^T A \mathbf{r}_{k+1} - \frac{\mathbf{d}_k^T A \mathbf{r}_{k+1}}{\mathbf{d}_k^T A \mathbf{d}_k} \mathbf{d}_k^T A \mathbf{d}_k \\ &= \mathbf{d}_k^T A \mathbf{r}_{k+1} - \mathbf{d}_k^T A \mathbf{r}_{k+1} \\ &= 0. \end{aligned}$$

Ahora, para el segundo caso, (ii), obtenemos,

$$\begin{aligned} \mathbf{d}_j^T A \mathbf{d}_{k+1} &= \mathbf{d}_j^T A \mathbf{r}_{k+1} - \beta_k \underbrace{\mathbf{d}_j^T A \mathbf{d}_k}_{=0} \\ &= \mathbf{d}_j^T A \mathbf{r}_{k+1}. \end{aligned} \quad (4.14)$$

Donde notamos que $\mathbf{d}_j^T A = (A \mathbf{d}_j)^T$, y considerando la ecuación (4.4) para el caso del $j+1$ -ésimo vector residual,

$$\mathbf{r}_{j+1} = \mathbf{r}_j - \alpha_j A \mathbf{d}_j,$$

podemos despejar $A \mathbf{d}_j$ en función de los vectores residuales \mathbf{r}_j y \mathbf{r}_{j+1} con $\alpha_j \neq 0$ ²,

$$A \mathbf{d}_j = \frac{\mathbf{r}_j}{\alpha_j} - \frac{\mathbf{r}_{j+1}}{\alpha_j}. \quad (4.15)$$

Reemplazando la ecuación (4.15) en la ecuación (4.14) obtenemos,

$$\mathbf{d}_j^T A \mathbf{d}_{k+1} = \mathbf{d}_j^T A \mathbf{r}_{k+1} \quad (4.16)$$

$$= (A \mathbf{d}_j)^T \mathbf{r}_{k+1} \quad (4.17)$$

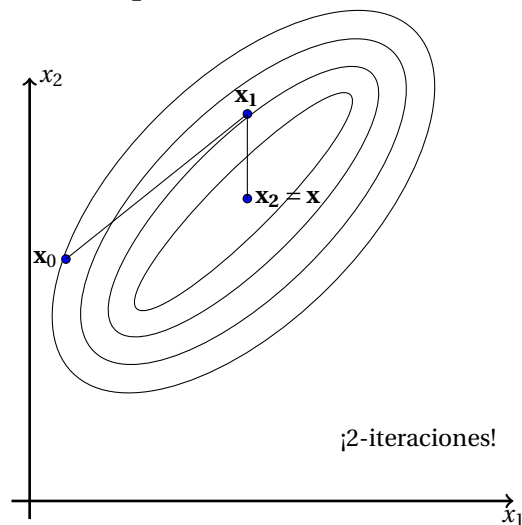
$$= \left(\frac{\mathbf{r}_j}{\alpha_j} - \frac{\mathbf{r}_{j+1}}{\alpha_j} \right)^T \mathbf{r}_{k+1} \quad (4.18)$$

$$= \frac{\mathbf{r}_j^T \mathbf{r}_{k+1}}{\alpha_j} - \frac{\mathbf{r}_{j+1}^T \mathbf{r}_{k+1}}{\alpha_j}, \quad (4.19)$$

por la ecuación (4.13) sabemos que los vectores residuales son ortogonales, por lo tanto $\mathbf{d}_j^T A \mathbf{d}_{k+1} = 0$ para $j \in \{0, 1, \dots, k\}$, y esto completa la demostración.

²¿Qué pasa en el caso de $\alpha_j = 0$?

4.9.3. Gradiente Conjugado - Interpretación Gráfica



4.9.4. Convergencia del Algoritmo del Gradiente Conjugado para una matriz simétrica y definida positiva

Thm 16. Sea A una matriz simétrica y positiva definida. El método del gradiente conjugado para resolver $Ax = b$ converge a lo más en “ n ” pasos usando aritmética exacta.

4.9.5. Gradiente Conjugado - Versión 2

¿Podría usted demostrar que la siguiente implementación es equivalente a la anterior? Warning: Watch out for the sign used for β_k .

```

1  $\mathbf{x}_0 = \text{"dato"}$ 
2  $\mathbf{d}_0 = \mathbf{r}_0 = \mathbf{b} - A \mathbf{x}_0$ 
3 for  $k$  in  $\text{range}(n)$  :
4     if ( $\|\mathbf{r}_k\| == 0$ ) :
5         break
6      $\alpha_k = \frac{\mathbf{r}_k^T \cdot \mathbf{r}_k}{\mathbf{d}_k^T A \mathbf{d}_k}$ 
7      $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$ 
8      $\mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_k A \mathbf{d}_k$ 
9      $\beta_k = \frac{\mathbf{r}_{k+1}^T \cdot \mathbf{r}_{k+1}}{\mathbf{r}_k^T \cdot \mathbf{r}_k}$ 
10     $\mathbf{d}_{k+1} = \mathbf{r}_{k+1} + \beta_k \mathbf{d}_k$ 

```

- ¿Cómo se puede mejorar esta propuesta?
- ¿Qué convendría hacer al momento de la implementación?

4.9.5.1. Ejemplo

p.129, 14.(a)

$$\mathbf{x}_0 = \begin{bmatrix} 0 & 0 \end{bmatrix}^T$$

$$\mathbf{r}_0 = \mathbf{b} - A \mathbf{x}_0 = \begin{bmatrix} 0 & 1 \end{bmatrix}^T$$

$$\mathbf{d}_0 = \mathbf{r}_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\alpha_0 = \mathbf{d}_0^T \cdot \mathbf{r}_0 / \mathbf{d}_0^T A \mathbf{d}_0$$

$$= \frac{\begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}}{\begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}}$$

$$= \frac{\boxed{1}}{\begin{bmatrix} 0 & 1 \end{bmatrix} \underbrace{\begin{bmatrix} -1 \\ 2 \end{bmatrix}}_{A \mathbf{d}_0}} \rightarrow \mathbf{r}_0^T \mathbf{r}_0$$

$$\boxed{\alpha_0 = \frac{1}{2}}$$

$$\mathbf{x}_1 = \mathbf{x}_0 + \alpha_0 \mathbf{d}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 1/2 \end{bmatrix}$$

$$\mathbf{r}_1 = \mathbf{r}_0 - \alpha_0 A \mathbf{d}_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

$$\mathbf{r}_1 = \begin{bmatrix} 1/2 \\ 0 \end{bmatrix}$$

$$\beta_0 = \frac{\mathbf{r}_1^T \mathbf{r}_1}{\mathbf{r}_0^T \mathbf{r}_0} = \frac{\begin{bmatrix} 1/2 & 0 \end{bmatrix} \begin{bmatrix} 1/2 \\ 0 \end{bmatrix}}{1}$$

$$\boxed{\beta_0 = 1/4}$$

$$\mathbf{d}_1 = \mathbf{r}_1 + \beta_0 \mathbf{d}_0$$

$$= \begin{bmatrix} 1/2 \\ 0 \end{bmatrix} + \frac{1}{4} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1/4 \end{bmatrix}$$

$$\underbrace{\begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}}_A \underbrace{\begin{bmatrix} u \\ v \end{bmatrix}}_x = \underbrace{\begin{bmatrix} 0 \\ 1 \end{bmatrix}}_b$$

$$k = 1$$

$$\alpha_1 = \frac{\mathbf{r}_1^T \mathbf{r}_1}{\mathbf{d}_1^T A \mathbf{d}_1} = \frac{1/4}{\begin{bmatrix} 1/2 & 1/4 \end{bmatrix} \underbrace{\begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 1/2 \\ 1/4 \end{bmatrix}}_{\begin{bmatrix} 1/4 \\ 0 \end{bmatrix}}}$$

$$\boxed{\alpha_1 = \frac{1/4}{1/8} = 2}$$

$$\mathbf{x}_2 = \mathbf{x}_1 + \alpha_1 \mathbf{d}_1 = \begin{bmatrix} 0 \\ 1/2 \end{bmatrix} + 2 \begin{bmatrix} 1/2 \\ 1/4 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\mathbf{r}_2 = \mathbf{r}_1 - \alpha_1 A \mathbf{d}_1$$

$$= \begin{bmatrix} 1/2 \\ 0 \end{bmatrix} - 2 \begin{bmatrix} 1/4 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

↑
¡residual nulo!

$$\beta_1 = 0$$

$$\mathbf{d}_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + 0 \begin{bmatrix} 1/2 \\ 1/4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\therefore \mathbf{x} = \mathbf{x}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \& \mathbf{r}_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

¡CG terminó en sólo 2 iteraciones!

4.10. Método de Newton en \mathbb{R}^n

Anteriormente resolvimos sistemas de ecuaciones lineales, ahora estamos interesados en resolver sistemas de ecuaciones no-lineales.

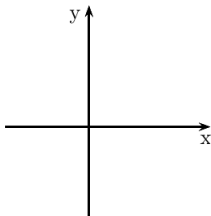
Recuerde: 1D: $f(x) \Rightarrow f(r) = 0$

$$\begin{aligned} \text{FPI: } x_{i+1} &= g(x_i) \\ &\longrightarrow x_i + f(x_i) \\ &\longrightarrow x_i + \frac{1}{a} f(x_i) \\ \text{NM: } x_{i+1} &= x_i - \frac{f(x_i)}{f'(x_i)} \end{aligned}$$

Tanto este método como el método de Newton multivariado se derivan de la aproximación lineal de la expansión de Taylor.

Ej:

$$\begin{aligned} \begin{matrix} x^2 + y^2 = 1 \\ y = x^2 \end{matrix} &\Rightarrow y + y^2 = 1 \Rightarrow y = \frac{-1 + \sqrt{5}}{2} \\ &y_1 \approx 0,61, y_2 \approx -1,61 \end{aligned}$$



$$\begin{aligned} \Rightarrow x_1 &= \underline{\hspace{2cm}} \\ x_2 &= \underline{\hspace{2cm}} \end{aligned}$$

$$\Rightarrow \mathbf{F}(\mathbf{x}) = \mathbf{F}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x^2 + y^2 - 1 \\ y - x^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\text{En 1D: } \underbrace{f(x_{i+1})}_0 = f(x_i) + f'(x_i) (x_{i+1} - x_i)$$

$$\Rightarrow x_{i+1} = x_i - (f'(x_i))^{-1} f(x_i)$$

En n -D:

$$\begin{aligned}\underbrace{\mathbf{F}(\mathbf{x}_{i+1})}_0 &= \mathbf{F}(\mathbf{x}_i) + J(\mathbf{x}_i) (\mathbf{x}_{i+1} - \mathbf{x}_i) + O(\|\mathbf{x}_{i+1} - \mathbf{x}_i\|^2) \\ \Rightarrow -\mathbf{F}(\mathbf{x}_i) &= J(\mathbf{x}_i) (\mathbf{x}_{i+1} - \mathbf{x}_i) \\ -J^{-1}(\mathbf{x}_i) \mathbf{F}(\mathbf{x}_i) &= \mathbf{x}_{i+1} - \mathbf{x}_i \\ \Rightarrow \mathbf{x}_{i+1} &= \mathbf{x}_i - \underbrace{J^{-1}(\mathbf{x}_i) \mathbf{F}(\mathbf{x}_i)}\end{aligned}$$

Matriz Jacobiana de F en \mathbf{x}_i

Donde

$$\mathbf{F}(\mathbf{x}) = \mathbf{F} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_n(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} f_1(x_1, x_2, \dots, x_n) \\ \vdots \\ f_n(x_1, x_2, \dots, x_n) \end{bmatrix}$$

Y la Matriz Jacobiana puede ser descrita de las siguientes formas: $\Rightarrow J(F) \bigg|_{\mathbf{x} = \mathbf{x}_0} = \left[\frac{\partial f_i}{\partial x_j} \right] \bigg|_{\mathbf{x} = \mathbf{x}_0} = \begin{bmatrix} \nabla f_1 \\ \nabla f_2 \\ \vdots \\ \nabla f_n \end{bmatrix} \bigg|_{\mathbf{x} = \mathbf{x}_0}$

Ej: 2D $\Rightarrow \left[\begin{array}{cc} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{array} \right] \bigg|_{\mathbf{x} = \mathbf{x}_0}$

Volviendo al ejemplo:

$$\begin{aligned}\mathbf{F}(\mathbf{x}) &= \begin{bmatrix} x^2 + y^2 - 1 \\ y - x^2 \end{bmatrix} = \begin{bmatrix} f_1(x, y) \\ f_2(x, y) \end{bmatrix} \\ \nabla f_1 &= \left\langle \frac{\partial f_1}{\partial x}, \frac{\partial f_1}{\partial y} \right\rangle = \langle 2x, 2y \rangle \\ \nabla f_2 &= \langle -2x, 1 \rangle \\ \therefore J(\mathbf{F}) &= \begin{bmatrix} 2x & 2y \\ -2x & 1 \end{bmatrix} \\ \therefore \begin{bmatrix} x_{i+1} \\ y_{i+1} \end{bmatrix} &= \begin{bmatrix} x_i \\ y_i \end{bmatrix} - \left(J(\mathbf{F}) \bigg|_{\mathbf{x} = \langle x_i, y_i \rangle} \right)^{-1} \begin{bmatrix} f_1(x_i, y_i) \\ f_2(x_i, y_i) \end{bmatrix} \\ \begin{bmatrix} x_{i+1} \\ y_{i+1} \end{bmatrix} &= \begin{bmatrix} x_i \\ y_i \end{bmatrix} - \begin{bmatrix} 2x_i & 2y_i \\ -2x_i & 1 \end{bmatrix}^{-1} \begin{bmatrix} x_i^2 + y_i^2 - 1 \\ y_i - x_i^2 \end{bmatrix}\end{aligned}$$

4.10.1. Algoritmo

```
1  $\mathbf{x}_0$  = "Initial guess" = "data"
2 for  $i$  in range( $n$ ):
3      $\mathbf{x}_{i+1} = \mathbf{x}_i - (J(\mathbf{x}_i))^{-1} \mathbf{F}(\mathbf{x}_i)$ 
```

Pero obtener la inversa computacionalmente es muy costoso por lo tanto en vez de hacer lo anterior, establecemos un $\mathbf{x}_{i+1} = \mathbf{x}_i + \mathbf{w}$, donde \mathbf{w} es la solución de $J(\mathbf{x}_i) \mathbf{w} = -\mathbf{F}(\mathbf{x}_i)$. Ahora sólo utilizamos la eliminación gaussiana:

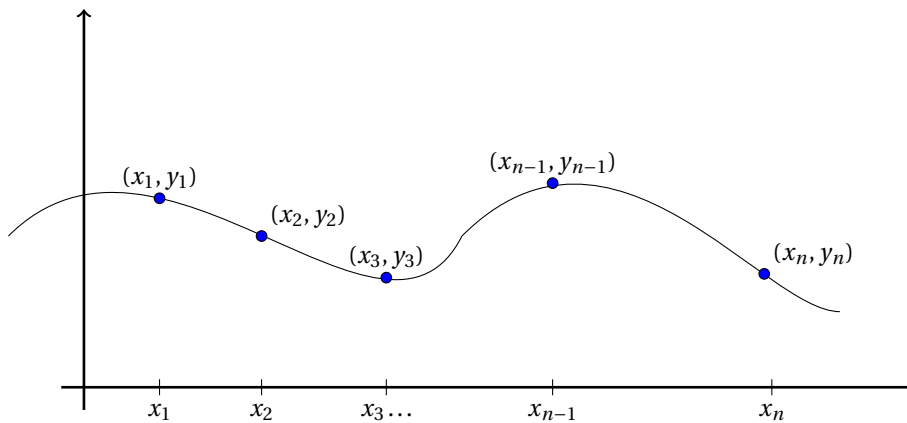
```
1  $\mathbf{x}_0$  = "data"
2 for  $i$  in range( $n$ ):
3     solve  $J(\mathbf{x}_i) \mathbf{w} = -\mathbf{F}(\mathbf{x}_i)$  for  $\mathbf{w}$ 
4      $\mathbf{x}_{i+1} = \mathbf{x}_i + \mathbf{w}$ 
```

La solución encontrada por este método depende de la estimación inicial.

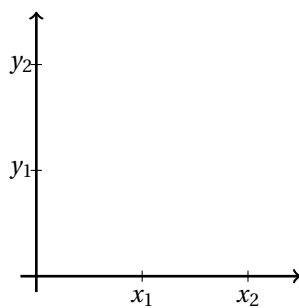
Capítulo 5

Interpolación (1D)

Def 7. La función $y = P(x)$ “interpola” los datos $(x_1, y_1), \dots, (x_n, y_n)$. Si $P(x_i) = y_i$ para cada $i = 1, \dots, n$.



Ejemplo: ¿Cuál es el polinomio de grado mínimo que interpola (x_1, y_1) y (x_2, y_2) ?



¿Lineal? ¿Cuadrático?

5.1. Matriz de Vandermonde

$$y = a_0 + a_1 \cdot x$$

$$\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

$$\begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

Cada fila de la **Matriz de Vandermonde** se compone por: $\{x_i^0, x_i^1, x_i^2, \dots, x_i^{n-1}\}$, donde el super-índice indica exponenciación. Siendo n la cantidad de puntos de interpolación.

Recuerde:

$$\frac{\|\mathbf{x} - \tilde{\mathbf{x}}\|}{\|\mathbf{x}\|} \leq \kappa(A) \cdot \frac{\|\mathbf{b} - A \cdot \tilde{\mathbf{x}}\|}{\|\mathbf{b}\|}$$

Si $\kappa(A) \gg$, la matriz es mal condicionada.

5.2. Interpolación de Lagrange

Para una cantidad de n puntos, podemos generar un polinomio de interpolación de grado $d = n - 1$ de la siguiente forma:

$$P_{n-1}(x) = y_1 \cdot L_1(x) + \dots + y_n \cdot L_n(x)$$

$$= \sum_{i=1}^n y_i \cdot L_i(x)$$

donde $L_i(x_i) = 1, L_i(x_j) = 0$ para $i \neq j$

$$L_k(x) = \frac{(x - x_1)(x - x_2) \dots (x - x_{k-1})(x - x_{k+1}) \dots (x - x_n)}{(x_k - x_1)(x_k - x_2) \dots (x_k - x_{k-1})(x_k - x_{k+1}) \dots (x_k - x_n)}$$

o

$$l_k(x) = \prod_{i=1, i \neq k}^n (x - x_i) = (x - x_1)(x - x_2) \dots (x - x_{k-1})(x - x_{k+1}) \dots (x - x_n)$$

y

$$L_k(x) = \frac{l_k(x)}{l_k(x_k)}$$

5.2.1. Ejemplo

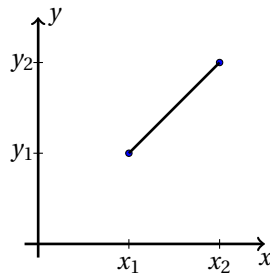
$$\begin{aligned} P_{n-1}(x_j) &= y_1 \cdot L_1(x_j) + \dots + y_j \cdot L_j(x_j) + \dots + y_n \cdot L_n(x_j) \\ &= y_1 \cdot 0 + \dots + \underline{y_j \cdot 1} + \dots + y_n \cdot 0 \end{aligned}$$

$$P_{n-1}(x_j) = y_j$$

Por lo tanto, $P_{n-1}(x)$ interpola la data (x_i, y_i) .

En un ejemplo en específico:

$$P_1(x) = y_1 \cdot L_1(x) + y_2 \cdot L_2(x)$$



Donde

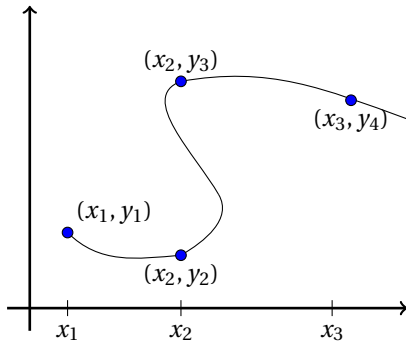
$$L_1(x) = \frac{(x - x_2)}{(x_1 - x_2)}$$

$$L_2(x) = \frac{(x - x_1)}{(x_2 - x_1)}$$

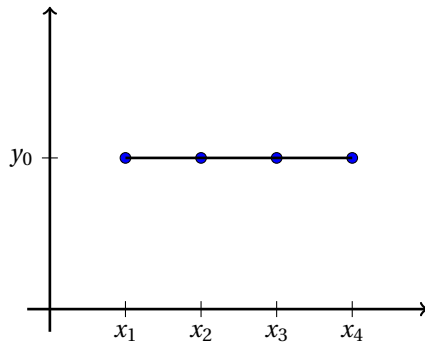
$$\begin{aligned} P_1(x) &= y_1 \cdot \frac{(x - x_2)}{(x_1 - x_2)} + y_2 \cdot \frac{(x - x_1)}{(x_2 - x_1)} \\ &= a_0 + a_1 x \quad (\text{similar a la "Matriz de vandermonde"}) \\ &= \underbrace{\left[\frac{y_1 \cdot (-x_2)}{(x_1 - x_2)} + \frac{y_2 \cdot (-x_1)}{(x_2 - x_1)} \right]}_{a_0} + \underbrace{\left[\frac{y_1}{(x_1 - x_2)} + \frac{y_2}{(x_2 - x_1)} \right]}_{a_1} \cdot x \end{aligned}$$

5.2.2. Teorema de la interpolación polinomial (Unicidad)

Thm 17. Sea $(x_1, y_1), \dots, (x_n, y_n)$, “ n ” puntos en el plano con distinto “ x_i ”, entonces existe uno y sólo un polinomio $P(x)$ de grado $(n-1)$ o “menor” que satisface la siguiente ecuación: $P(x_i) = y_i$ para $i = 1 : n$.



¿Es posible realizar interpolación??



$$P_{n-1}(x) = \underline{\hspace{10em}}$$

Considerando sólo 2 puntos:

$$\begin{aligned} P_1(x) &= y_0 \cdot \frac{(x-x_2)}{(x_1-x_2)} + y_0 \cdot \frac{(x-x_1)}{(x_2-x_1)} \\ &= y_0 \cdot \left(\frac{x-x_2}{x_1-x_2} - \frac{x-x_1}{x_1-x_2} \right) \\ &= y_0 \cdot \left(\frac{x_1-x_2}{x_1-x_2} \right) \\ &= \underline{y_0} \end{aligned}$$

5.3. Diferencias divididas de Newton

Recuerde:

$$P(x) = 1 + 2 \cdot x + 3 \cdot x^2 + 4 \cdot x^3 \sim O(n^2)$$

$$= 1 + x \cdot (2 + x \cdot (3 + 4 \cdot x)) \sim O(n)$$

↑

Mín número de operaciones

$$P_{n-1}(x) = \sum_{i=0}^{n-1} a_i \cdot x^i \iff \text{Matriz de Vandermonde}$$

$$P_{n-1}(x) = \sum_{i=1}^n y_i \cdot L_i(x)$$

Def 8. Denotemos que $\underbrace{f[x_1, x_2, \dots, x_{n-1}, x_n]}_{Y_i}$ es el coeficiente del término x^{n-1} en el único polinomio que interpola $(x_i, f(x_i))$.

$$P_{n-1}(x) = \sum_{i=0}^{n-1} a_i x^i = a_0 + a_1 \cdot x + \dots + \underline{a_{n-1}} \cdot x^{n-1}.$$

En el caso de diferencias divididas de Newton, el polinomio interpolador se define como:

$$\begin{aligned} \Rightarrow P_{n-1}(x) &= f[x_1] + f[x_1 \ x_2] \cdot (x - x_1) \\ &\quad + f[x_1 \ x_2 \ x_3] \cdot (x - x_1) \cdot (x - x_2) \\ &\quad + f[x_1 \ x_2 \ x_3 \ x_4] \cdot (x - x_1) \cdot (x - x_2) \cdot (x - x_3) \\ &\quad \vdots \\ &\quad + f[x_1 \dots x_n] \cdot (x - x_1) \dots (x - x_{n-1}). \end{aligned}$$

5.3.1. Ejemplo

Use las diferencias divididas para encontrar el polinomio de interpolación que pasa a través de los puntos: $(x_1, y_1), (x_2, y_2), (x_3, y_3)$.

Recuerde: $f[x_i] = y_i$

x_1	y_1	$= f[x_1]$	$f[x_1 \ x_2] = \frac{f[x_2] - f[x_1]}{x_2 - x_1}$	$f[x_1 \ x_2 \ x_3] = \frac{f[x_2 \ x_3] - f[x_1 \ x_2]}{x_3 - x_1}$
x_2	y_2	$= f[x_2]$		
x_3	y_3	$= f[x_3]$	$f[x_2 \ x_3] = \frac{f[x_3] - f[x_2]}{x_3 - x_2}$	

$$\Rightarrow P_2(x) = f[x_1] + f[x_1 \ x_2] \cdot (x - x_1) + f[x_1 \ x_2 \ x_3] \cdot (x - x_1) \cdot (x - x_2)$$

5.3.2. Algorítmicamente

```

1 for j in (n):
2     f[xj] = yj
3 for i in (2, n):
4     for j in (n+1-i):
5         f[xj...x(j+1-1)] =  $\frac{f[x_{(j+1)}...x_{(j+1-1)}] - f[x_j...x_{(j+1-2)}]}{x_{(j+1-1)} - x_j}$ 

```

El polinomio de Newton es de la forma:

$$P_{n-1}(x) = \sum_{i=1}^n f[x_1 \dots x_i] \cdot \prod_{j=1}^{i-1} (x - x_j)$$

5.3.3. Ejemplo

Use **Matriz de Vandermonde**, **Interpolación de Lagrange** y **Diferencias Divididas de Newton** para encontrar el polinomio de interpolación que pasa a través de los puntos: (0, 1), (2, 3), (3, 0).

Lagrange:

$$P_2(x) = y_1 \cdot L_1(x) + y_2 \cdot L_2(x) + y_3 \cdot L_3(x)$$

$$L_1(x) = \frac{(x-2)(x-3)}{(0-2)(0-3)}$$

$$L_2(x) = \frac{(x-0)(x-3)}{(2-0)(2-3)}$$

$$L_3(x) = \frac{(x-0)(x-2)}{(3-0)(3-2)}$$

$$\begin{aligned}
 P_2(x) &= 1 \cdot \frac{(x-2)(x-3)}{6} + 3 \cdot \frac{(x) \cdot (x-3)}{-2} + 0 \cdot L_3(x) \\
 &= 1 + \frac{11}{3} \cdot x - \frac{4}{3} \cdot x^2 = a_0 + a_1 \cdot x + a_2 \cdot x^2
 \end{aligned}$$

D.D. Newton:

Vandermonde:

$$P_2(x) = a_0 + a_1 \cdot x + a_2 \cdot x^2$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 0 \end{pmatrix}$$

x_i	y_i
0	1
2	3
3	0

$$\begin{array}{|c|c|c|c|} \hline 1 & 1 & -\frac{4}{3} & \\ \hline \end{array} \Rightarrow P_2(x) = 1 + 1 \cdot (x-0) + \left(-\frac{4}{3}\right) \cdot (x-0)(x-2)$$

¿Qué hago si quiero agregar 1 punto nuevo?

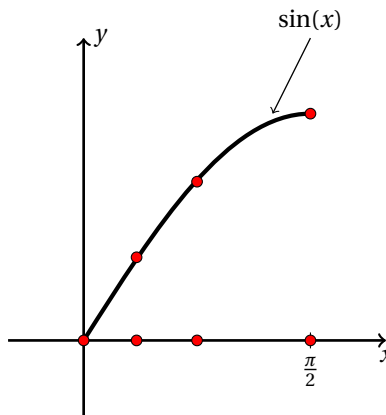
Utilizando el concepto de las Diferencias Divididas de Newton, los nuevos puntos que se agreguen después de calcular el polinomio de interpolación original pueden incorporarse con facilidad.

x_i	y_i			
0	1	1	$\frac{-4}{3}$	$\frac{-5}{3}$
2	3	-3	$\frac{3}{3}$	
3	0	0	-1	
1	0			

$$\begin{aligned}
 P_3(x) &= P_2(x) + \left(\frac{-5}{3}\right) \cdot (x-0)(x-2) \cdot (x-3) \\
 &= 1 + 1 \cdot (x-0) - \frac{4}{3}(x-0)(x-2) \\
 &\quad - \frac{5}{3} \cdot (x-0)(x-2)(x-3)
 \end{aligned}$$

Ejemplo:

Se desea interpolar la función: $f(x) = \sin(x)$, con 4 puntos equiespaciados en el intervalo: $\left[0, \frac{\pi}{2}\right]$.



x_i	y_i			
0	0	0.9549	-0.2443	-0.1139
$\pi/6$	1/2	0.6990	-0.4232	
$2\pi/6$	0.8660	0.2259		
$\pi/2$	1			

\uparrow \uparrow
 Data

$$\begin{aligned}
 \Rightarrow P_3(x) &= 0 + 0.9549 \cdot (x-0) - 0.2443 \cdot (x) \cdot \left(x - \frac{\pi}{6}\right) \\
 &\quad - 0.1139 \cdot (x-0) \cdot \left(x - \frac{\pi}{6}\right) \cdot \left(x - \frac{2\pi}{6}\right)
 \end{aligned}$$

5.4. Error de Interpolación

Thm 18. Asuma que $P(x)$ es el polinomio interpolador (de grado $n-1$ o menor) que ajusta n puntos $(x_1, y_1), \dots, (x_n, y_n)$. El error de interpolación es

$$f(x) - P(x) = \frac{(x-x_1)(x-x_2)\dots(x-x_n)}{n!} \cdot f^{(n)}(c)$$

donde c está entre el menor y el mayor de los números x, x_1, \dots, x_n .

Recuerde: $f(x) = f(x_0) + f'(x_0) \cdot (x - x_0) + \dots + \frac{f^{(n)}(c)}{n!} \cdot (x - x_0)^n$

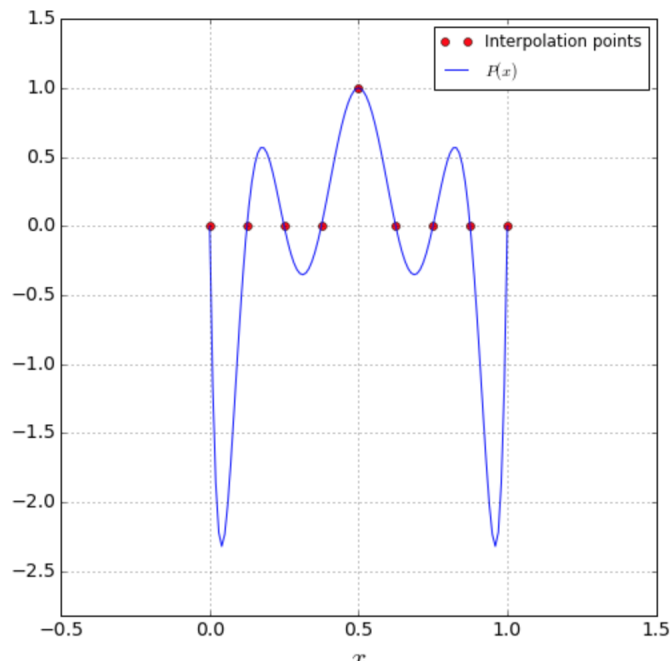
↪ ¿Qué tan buena es la interpolación de $f(x)$ obtenida por $P(x)$?

↪ ¿Cómo decae el error al aproximar una función por medio de un polinomio?

$$\begin{aligned} \text{Error}(x) &= |f(x) - P(x)| \Rightarrow \text{Peor caso} = \max_x |f(x) - P(x)| \\ &\Rightarrow \max_x \text{Error}(x) = \max_x |(x - x_1) \dots (x - x_n)| \cdot \frac{|f^{(n)}(c)|}{n!} \end{aligned}$$

5.4.1. Fenómeno de Runge

A medida que **aumenta** la cantidad de puntos, **aumenta** las oscilaciones en los bordes del intervalo.



5.5. Interpolación de Chebyshev

La interpolación de Chebyshev se refiere a una forma particular de definición de los puntos de interpolación tal que el error de interpolación es minimizado.

5.5.1. Motivación

¿Qué podemos mejorar de la fórmula de error de interpolación?

Recuerde: $\frac{(x-x_1) \cdot (x-x_2) \dots (x-x_n)}{n!} \cdot f^{(n)}(c)$

Consideremos $x \in [-1, 1]$

¿Podemos encontrar x_1, \dots, x_n de tal forma que $(x-x_1) \dots (x-x_n)$ es minimizado?

5.5.1.1. Ejemplo

Consideremos $n=2$,

$$\Rightarrow (x - \hat{x}_1) \cdot (x - \hat{x}_2)$$

$$\Rightarrow w(\hat{x}_1, \hat{x}_2) = \max_{x \in [-1, 1]} |(x - \hat{x}_1) \cdot (x - \hat{x}_2)|$$

$$[x_1, x_2] = \operatorname{argmin}_{x_1, x_2 \in [-1, 1]} w(\hat{x}_1, \hat{x}_2)$$

↪ ¿Cómo obtenemos x_1, x_2 ?

↪ ¿Y en general x_1, \dots, x_n ?

5.5.2. Teorema de Chebyshev

Thm 19. La elección de los números reales $-1 \leq x_1, \dots, x_n \leq 1$ que hace el valor de

$$\max_{-1 \leq x \leq 1} |(x - x_1) \dots (x - x_n)|$$

lo más pequeño posible es:

$$x_i = \cos\left(\frac{(2i-1) \cdot \pi}{2 \cdot n}\right), \quad i = 1, \dots, n$$

y el valor mínimo es $\frac{1}{2^{n-1}}$. De hecho, el mínimo es alcanzado por:

$$(x - x_1) \dots (x - x_n) = \frac{1}{2^{n-1}} \cdot T_n(x)$$

Donde $T_n(x) = \cos(n \cdot \arccos(x))$ es el n -ésimo polinomio de Chebyshev.

A partir del teorema, se llega a la conclusión de que el error de interpolación puede minimizarse si los n puntos de interpolación en $[-1, 1]$ se eligen como las raíces del polinomio de interpolación de Chebyshev $T_n(x)$ de grado n .

■ Observación 1:

$$\begin{aligned} T_0(x) &= 1 \\ T_1(x) &= x \\ T_2(x) &= 2 \cdot x^2 - 1 \\ &\vdots \\ T_{n+1}(x) &= 2 \cdot x \cdot T_n(x) - T_{n-1}(x) \end{aligned}$$

- Observación 2: El valor máximo absoluto de $T_n(x)$ para $-1 \leq x \leq 1$ es 1. Esto se deriva inmediatamente del hecho de que $T_n(x) = \cos(n \cdot \arccos(x))$.
- Observación 3: Todos los ceros de $T_n(x)$ se encuentran entre -1 y 1 . De hecho, los ceros son la solución de: $\cos(n \cdot \arccos(x)) = 0$. Como $\cos(y) = 0$ si y sólo si $y = (2i-1) \cdot (\pi/2)$, notamos que:

$$n \cdot \arccos(x_i) = (2i-1) \cdot (\pi/2)$$

$$x_i = \cos\left(\frac{(2i-1)\pi}{2n}\right)$$

5.5.3. Interpretación Gráfica

$$x_i = \cos(\theta_i), \quad \theta_i = \frac{(2i-1) \cdot \pi}{2n}, \quad i = 1, \dots, n$$

$$i = 1 \Rightarrow \theta_1 = \frac{\pi}{2n}$$

$$n = 5 \Rightarrow \theta_1 = \frac{1}{5} \cdot \frac{\pi}{2}$$

$$\theta_2 = \frac{3\pi}{2n} = \frac{3\pi}{10}, \quad \theta_3 = \frac{5\pi}{2 \cdot 5} = \frac{\pi}{2}$$

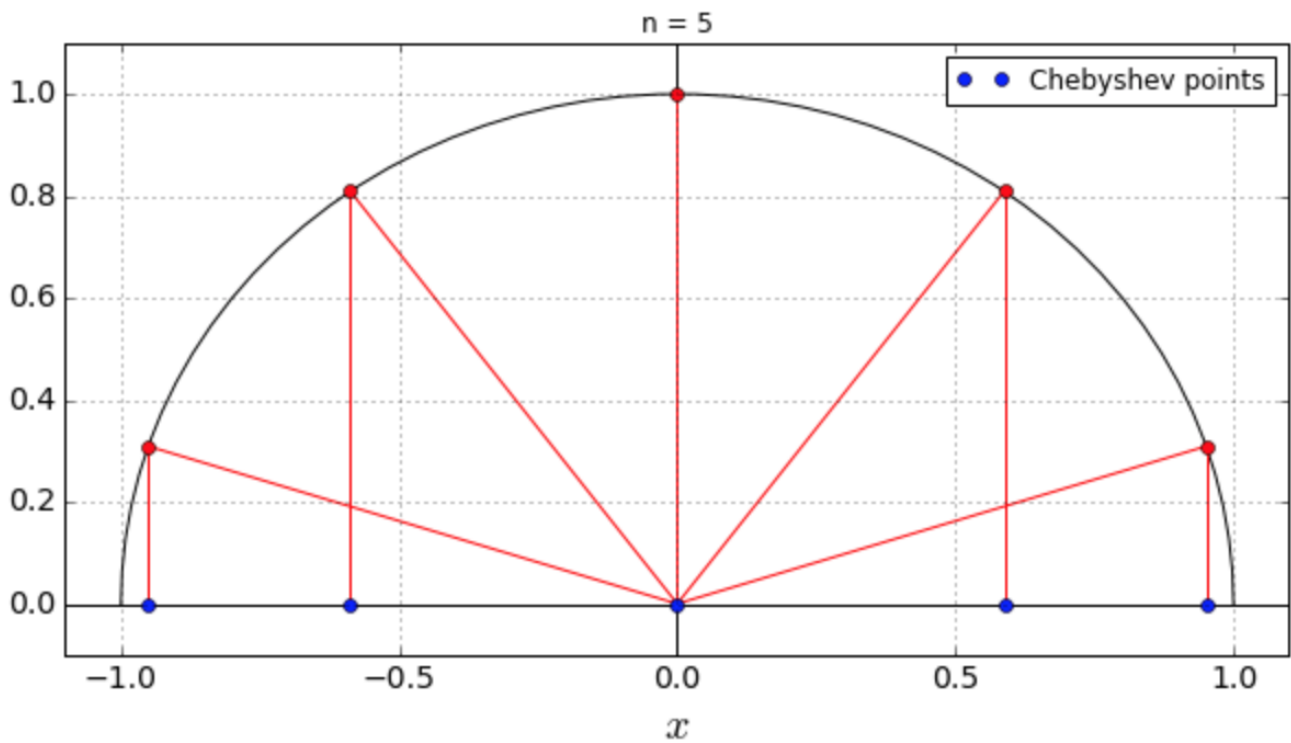
$$\theta_4 = \frac{7\pi}{2 \cdot 5} = \frac{7\pi}{10}, \quad \theta_5 = \frac{9\pi}{10}$$

$$\theta_1 - \theta = \frac{1}{n} \cdot \frac{\pi}{2} = \frac{1}{5} \cdot \frac{\pi}{2}$$

$$\theta_{i+1} - \theta_i = \frac{2}{n} \cdot \frac{\pi}{2} = \frac{\pi}{5}$$

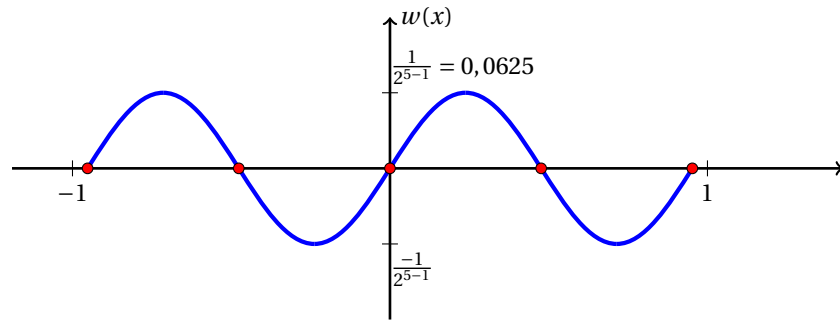
$$\Rightarrow x_i = \cos(\theta_i)$$

$$x_1 = 0,9511, x_2 = 0,5878, x_3 = 0,0000, x_4 = -0,5878, x_5 = -0,9511$$



Recuerde:

$$w(x) = \prod_{i=1}^n (x - x_i)$$



Ejemplo(3.10): Encuentre el peor caso del error para la diferencia de e^x y el polinomio de Chebyshev de grado 4 en el intervalo $x = [-1, 1]$.

$$|e^x - P_4(x)| = \frac{|(x - x_1) \dots (x - x_5)|}{n!} \cdot |f^{(5)}(c)|$$

$$\leq \frac{1}{2^{5-1}} \cdot \frac{1}{n!} |f^{(5)}(c)|$$

$$f(x) = e^x \Rightarrow f^{(5)}(c) = e^x, \max \text{ en } [-1, 1] \Rightarrow e^1$$

$$\Rightarrow |e^x - P_4(x)| \leq \frac{e}{2^4 \cdot 5!} \approx 0,00142$$

5.5.4. Cambio de Intervalo

El análisis de la interpolación de Chebyshev se ha restringido al intervalo $[-1, 1]$, por lo que cambiaremos considerando un intervalo general $[a, b]$.

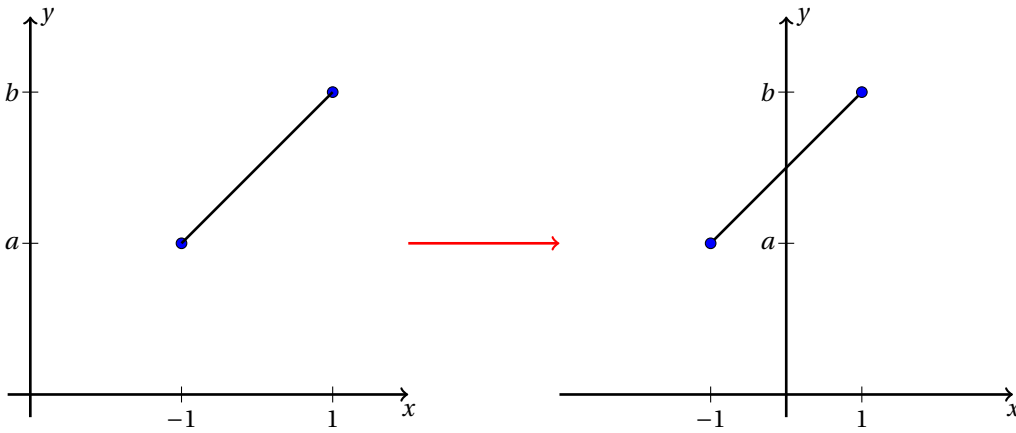
Desplazaremos los puntos originales:

$$x_i = \cos\left(\frac{(2i-1)\pi}{2n}\right)$$

Por:

$$\tilde{x}_i = \frac{b+a}{2} + \frac{b-a}{2} \cos\left(\frac{(2i-1)\pi}{2n}\right)$$

\uparrow Chebyshev en $a \leq x \leq b$ \uparrow Chebyshev points en $-1 \leq x_i \leq 1$



5.5.4.1. Nodos de interpolación de Chebyshev

En el intervalo $[a, b]$ para $i = 1, \dots, n$. La siguiente desigualdad se cumple en $x \in [a, b]$.

$$\Rightarrow |(x - x_1) \dots (x - x_n)| \leq \frac{\left(\frac{b-a}{2}\right)^n}{2^{n-1}}$$

Debido a que la cota superior del término en la fórmula de error de interpolación cambia por el desplazamiento, el valor mínimo queda como:

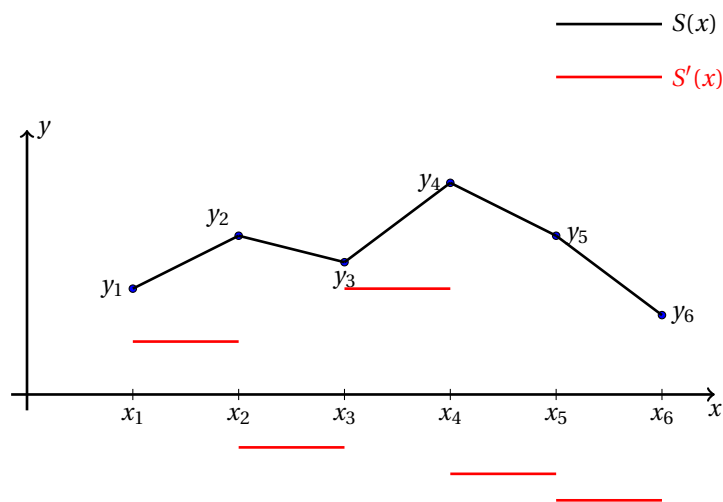
$$\frac{\left(\frac{b-a}{2}\right)^n}{2^{n-1}}$$

Observación: También se podría hacer un cambio de variable de $[a, b]$ a $[-1, 1]$ y usar la misma teoría.

Capítulo 6

Splines

Splines es un enfoque alternativo a la interpolación de datos por medio de polinomios, la idea es definir una función por intervalos. El ejemplo más simple de Spline es la Spline lineal, aquí se conectan los puntos a través de rectas. Es decir para n puntos: $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ con $x_1 < x_2 < \dots < x_n$ la spline lineal se compone de $n - 1$ segmentos. Veamos un ejemplo para $n = 6$:



Algebraicamente:

$$S_1(x) = y_1 + b_1(x - x_1), \quad x \in [x_1, x_2]$$

$$S_2(x) = y_2 + b_2(x - x_2), \quad x \in [x_2, x_3]$$

$$S_3(x) = y_3 + b_3(x - x_3), \quad x \in [x_3, x_4]$$

$$S_4(x) = y_4 + b_4(x - x_4), \quad x \in [x_4, x_5]$$

$$S_5(x) = y_5 + b_5(x - x_5), \quad x \in [x_5, x_6]$$

¿Cómo encontramos los b_i 's?

$$\begin{aligned}
S_1(x_2) &= S_2(x_2) \\
y_1 + b_1(x_2 - x_1) &= y_2 \\
b_1 \cdot (x_2 - x_1) &= y_2 - y_1
\end{aligned}$$

Repitiendo el mismo análisis para x_3 , x_4 , x_5 y x_6 y re-escribiendo las ecuaciones como un sistema de ecuaciones lineales, obtenemos:

$$\begin{pmatrix}
(x_2 - x_1) & 0 & 0 & 0 & 0 \\
0 & (x_3 - x_2) & 0 & 0 & 0 \\
0 & 0 & (x_4 - x_3) & 0 & 0 \\
0 & 0 & 0 & (x_5 - x_4) & 0 \\
0 & 0 & 0 & 0 & (x_6 - x_5)
\end{pmatrix}
\begin{pmatrix}
b_1 \\
b_2 \\
b_3 \\
b_4 \\
b_5
\end{pmatrix}
=
\begin{pmatrix}
y_2 - y_1 \\
y_3 - y_2 \\
y_4 - y_3 \\
y_5 - y_4 \\
y_6 - y_5
\end{pmatrix}$$

Resolviendo el sistema de ecuaciones lineales obtenemos:

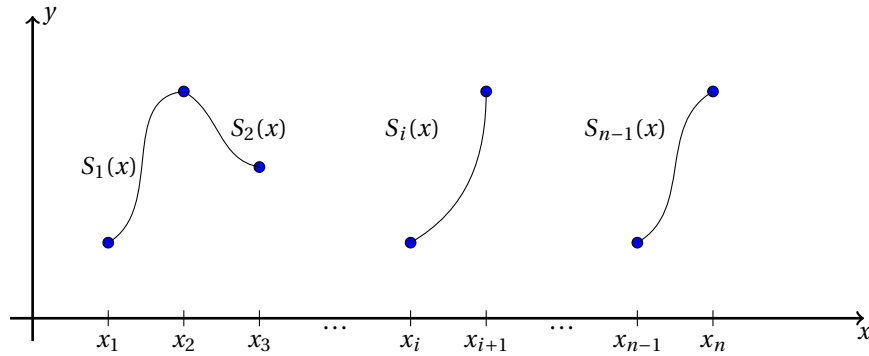
$$b_i = \frac{y_{i+1} - y_i}{x_{i+1} - x_i}, \quad \forall i \in \{1, 2, 3, 4, 5\}$$

6.1. Propiedades de Spline Cúbica

Una Spline cúbica con n puntos $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ donde los x_i son distintos y en orden creciente, se define de la siguiente forma:

$$S(x) = \begin{cases} S_1(x), & x \in [x_1, x_2] \\ S_2(x), & x \in]x_2, x_3] \\ \vdots \\ S_{n-1}(x), & x \in]x_{n-1}, x_n] \end{cases}$$

donde $S_i(x) = y_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3$, para $x \in [x_i, x_{i+1}]$. Y gráficamente



Además se requiere que se cumplan las siguientes propiedades para ser considerada una spline cúbica.

6.1.1. Propiedad 1 (Continuidad)

$$\begin{aligned}
S_i(x_i) &= y_i \\
S_i(x_{i+1}) &= y_{i+1}
\end{aligned}
\quad \text{para } i = 1, 2, \dots, n-1$$

6.1.2. Propiedad 2 (Diferenciabilidad)

$$S'_{i-1}(x_i) = S'_i(x_i) \text{ para } i = 2, \dots, n-1$$

6.1.3. Propiedad 3 (Continuidad en la segunda derivada)

$$S''_{i-1}(x_i) = S''_i(x_i) \text{ para } i = 2, \dots, n-1$$

6.2. Interpretación de las 3 propiedades

- La propiedad 1 garantiza que la spline interpole los puntos de datos.
- La 2da propiedad obliga a las pendientes de las partes adyacentes de la spline a ser iguales.
- La 3ra propiedad hace lo mismo pero con la segunda derivada.

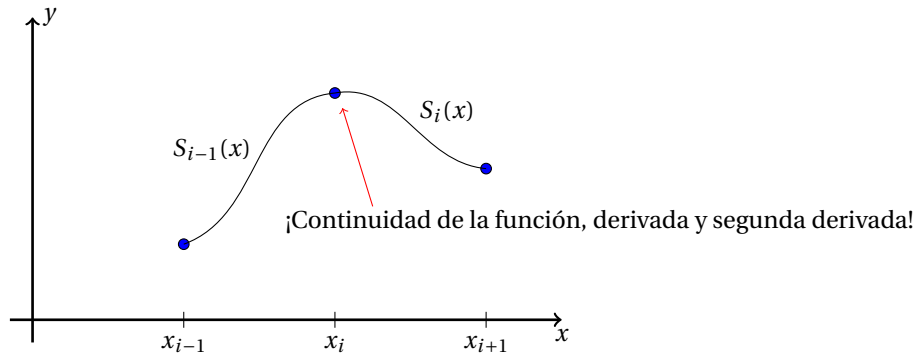
6.3. Cantidad de ecuaciones a satisfacer

La construcción de un Spline a partir de un conjunto de puntos significa que debemos encontrar los coeficientes: b_i, c_i, d_i , que hacen que se cumplan las 3 propiedades.

¿Cuántas ecuaciones se deben satisfacer?

La propiedad 1 entrega $(n-1)$ ecuaciones independientes, por otra parte las propiedades 2 y 3 nos dan $(n-2)$ cada una. En total tenemos $n-1 + 2(n-2) = 3n-5$ ecuaciones.

Pero tenemos 3 coeficientes por cada ecuación, es decir: $3(n-1) = 3n-3$. Finalmente la determinación de los coeficientes consiste en resolver un sistema de $3n-5$ ecuaciones lineales con $3n-3$ incógnitas. Por lo tanto aún necesitamos dos ecuaciones más para la unicidad de la solución.



Resumen:

- Cantidad de variables: $3n - 3$
- Cantidad de ecuaciones: $3n - 5$
- ¿Es cuadrada la matriz?¹
- ¿Qué falta o qué sobra?²
- ¿Qué opciones tenemos?³

Las spline también tiene otras condiciones, dependiendo de estas la spline adquiere otro nombre, veremos las más usadas.

6.4. Tipos de condiciones de borde para splines cúbicas

6.4.1. Spline Natural

Se llama **spline natural** a la spline cúbica que cumple las condiciones de borde:

$$S_1''(x_1) = S_{n-1}''(x_n) = 0$$

6.4.2. Spline con curvatura ajustada

Ahora en vez de definir las condiciones de borde en 0, podemos elegir sus valores arbitrariamente, es decir establecer las curvaturas deseadas en los extremos, es decir:

$$\begin{aligned} S_1''(x_1) &= k_1 \\ S_{n-1}''(x_n) &= k_2 \end{aligned}$$

A estas se les llama **spline con curvatura ajustada**.

6.4.2.1. Spline cúbica fijada (Clamped cubic spline)

De manera similar, definimos los valores a las pendientes en el borde de la siguiente forma:

$$\begin{aligned} S_1'(x_1) &= k_1 \\ S_{n-1}'(x_n) &= k_2 \end{aligned}$$

6.4.3. Spline Terminada parabólicamente

Si obligamos a las primera parte de la spline y a la última a ser un polinomio de grado como máximo 2, al hacer $d_1 = d_n = 0$, de esta forma requerimos que $c_1 = c_2$ y $c_{n-1} = c_n$.

¹No

²Faltan dos ecuaciones

³¡Aún falta definir las condiciones de borde!

6.4.4. Spline cúbica sin nodo (Not-a-knot cubic spline)

Como S_1 y S_2 son polinomios de grado 3 o menor, se requiere que sus terceras derivadas concuerden en x_2 , como ya concuerda en su primera y segunda derivada. Para obtener la otra condición hacemos lo mismo para S_{n-1} y S_{n-2} , es decir debe cumplir

$$\begin{aligned} S_1'''(x_2) &= S_2'''(x_2) \\ S_{n-2}'''(x_{n-1}) &= S_{n-1}'''(x_{n-1}) \end{aligned}$$

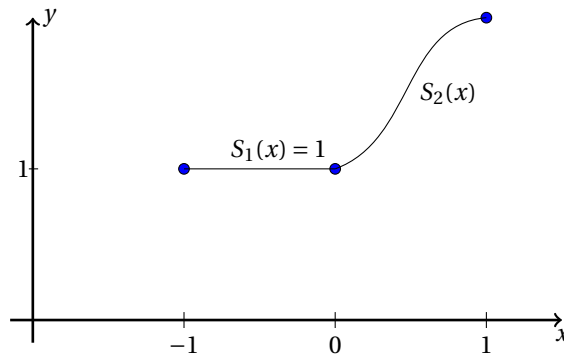
6.5. Unicidad de la Spline Cúbica

Thm 20. Asuma $n \geq 2$. Entonces, para el conjunto de datos $(x_1, y_1), \dots, (x_n, y_n)$ y para cualquiera de las condiciones de borde (6.4.1, 6.4.2, 6.4.2.1), existe una única spline cúbica que satisface las condiciones de borde y se ajusta a los puntos dados.

Lo mismo es válido para 6.4.3 con $n \geq 3$ y para 6.4.4 con $n \geq 4$.

6.5.1. Ejemplo

Asuma que la parte izquierda de una spline cúbica natural es $S_1(x) = 1$ en el intervalo $[-1, 0]$. Encuentre 3 diferentes posibilidades para la definición de $S_2(x)$ en $[0, 1]$.



$$S_2(x) = 1 + b(x-0) + c(x-0)^2 + d(x-0)^3$$

$$S_1(0) = S_2(0) \quad \checkmark \checkmark$$

$$S_1'(0) = 0, S_2'(x) = b + 2cx + 3dx^2$$

$$S_2'(0) = b$$

$$\therefore b = 0$$

$$S_1''(0) = 0, S_2'(x) = 2c + 6dx$$

$$S_2''(0) = 2c$$

$$\therefore c = 0$$

$$\Rightarrow S_2(x) = 1 + dx^3$$

- Sub-sección 6.4.1 $\Rightarrow S_2''(1) = 6d = 0 \Rightarrow d = 0 \Rightarrow S_2(x) = 1$
- Sub-sección 6.4.2 $\Rightarrow S_2''(1) = 6d = K_2 \Rightarrow d = K_2/6 \Rightarrow S_2(x) = 1 + x^3 \cdot K_2/6$
- Sub-sección 6.4.2.1 $\Rightarrow S'(1) = 3d \cdot 1^2 = P_2 \Rightarrow d = P_2/3 \Rightarrow S_2(x) = 1 + x^3 \cdot P_2/3$
- Sub-sección 6.4.3 $\Rightarrow d = 0 \Rightarrow S_2(x) = 1$
- Sub-sección 6.4.4 $\Rightarrow S_1'''(0) = 0, S_2'''(0) = 6d \Rightarrow d = 0 \Rightarrow S_2(x) = 1$

6.6. Ejercicios Propuestos

1. (0,0), (1,1), (2,4). Determine las ecuaciones para construir la spline cúbica considerando todas las posibles condiciones de borde.
2. Encuentre "c" para las siguientes splines cúbicas.

a)

$$S(x) = \begin{cases} 4 - \frac{11}{4}x + \frac{3}{4}x^3, & x \in [0, 1] \\ 2 - \frac{1}{2}(x-1) + c(x-1)^2 - \frac{3}{4}(x-1)^3, & x \in [1, 2] \end{cases}$$

b)

$$S(x) = \begin{cases} 3 - 9x + 4x^2, & x \in [0, 1] \\ -2 - (x-1) + c(x-1)^2, & x \in [1, 2] \end{cases}$$

3. Decida si las siguientes funciones son splines cúbicas.

a)

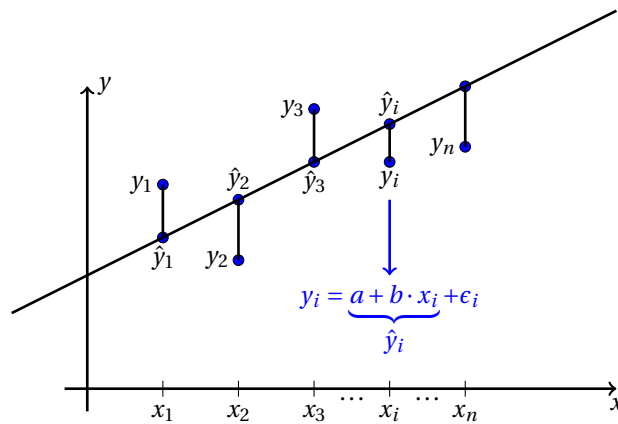
$$S(x) = \begin{cases} x^3 + x - 1, & x \in [0, 1] \\ -(x-1)^3 + 3(x-1)^2 + 3(x-1) + 1, & x \in [1, 2] \end{cases}$$

b)

$$S(x) = \begin{cases} 2x^3 + x^2 + 4x + 5, & x \in [0, 1] \\ (x-1)^3 + 7(x-1)^2 + 12(x-1) + 12, & x \in [1, 2] \end{cases}$$

Capítulo 7

Mínimos Cuadrados



¿Cómo encontramos los coeficientes “ a ” y “ b ” tal que se minimice el error total? En otras palabras, ¿Cómo minimizamos el error cuadrático, entre y_i y $\hat{y}_i = a + b x_i$? Además, ¿Qué es ϵ_i ?

$$\text{Error total: } \sum_{i=1}^n (\underbrace{y_i}_{\text{data}} - \underbrace{\hat{y}_i}_{\text{estimado}})^2 = F(a, b)$$

¿Cómo encontramos el mín de $F(a, b)$?¹

$$\nabla F = \langle 0, 0 \rangle$$

$$(1) \Rightarrow \frac{\partial F}{\partial a} = F_a = \frac{\partial}{\partial a} \sum_{i=1}^n (y_i - a - b x_i)^2 = \sum_{i=1}^n 2(y_i - a - b x_i) (-1) = 0$$

$$(2) \Rightarrow \frac{\partial F}{\partial b} = F_b = \frac{\partial}{\partial b} \sum_{i=1}^n (y_i - a - b x_i)^2 = \sum_{i=1}^n 2(y_i - a - b x_i) (-x_i) = 0$$

¹¡Igualando su gradiente al vector nulo!

(1) \Rightarrow

$$\begin{aligned}\sum_{i=1}^n (y_i - a - b x_i) &= 0 \\ \left(\sum_{i=1}^n y_i \right) - a n - b \left(\sum_{i=1}^n x_i \right) &= 0 \\ \therefore n a + \left(\sum_{i=1}^n x_i \right) b &= \left(\sum_{i=1}^n y_i \right)\end{aligned}$$

(2) \Rightarrow

$$\begin{aligned}\sum_{i=1}^n (y_i x_i - a x_i - b x_i^2) &= 0 \\ \left(\sum_{i=1}^n y_i x_i \right) - a \left(\sum_{i=1}^n x_i \right) - b \left(\sum_{i=1}^n x_i^2 \right) &= 0 \\ \therefore \left(\sum_{i=1}^n x_i \right) a + \left(\sum_{i=1}^n x_i^2 \right) b &= \left(\sum_{i=1}^n y_i x_i \right)\end{aligned}$$

 \therefore

$$\begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n y_i x_i \end{pmatrix} \quad (7.1)$$

$$\begin{aligned}a &= \frac{(\sum_{i=1}^n x_i^2) (\sum_{i=1}^n y_i) - (\sum_{i=1}^n x_i) (\sum_{i=1}^n y_i x_i)}{n (\sum_{i=1}^n x_i)^2 - (\sum_{i=1}^n x_i^2)} \\ b &= \frac{n (\sum_{i=1}^n y_i x_i) - (\sum_{i=1}^n x_i) (\sum_{i=1}^n y_i)}{n (\sum_{i=1}^n x_i)^2 - (\sum_{i=1}^n x_i^2)}\end{aligned}$$

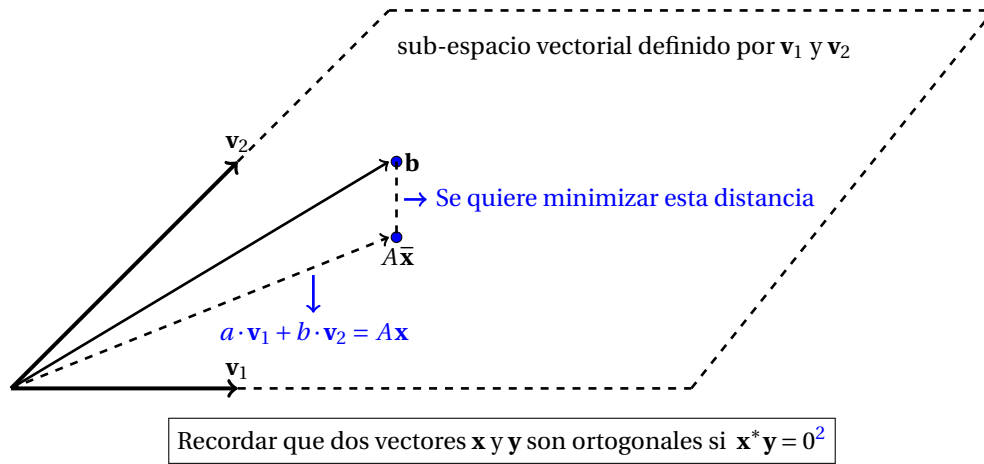
Otra manera de ver esto desde el álgebra lineal:Para cada $x_i \Rightarrow a + b x_i = y_i$, re-escribiéndolo matricialmente:

$$A \cdot \begin{pmatrix} a \\ b \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}}_{(\mathbf{v}_1 \quad \mathbf{v}_2)} \underbrace{\begin{pmatrix} a \\ b \end{pmatrix}}_{\mathbf{x}} = \underbrace{\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix}}_{\mathbf{b}}$$

 \Rightarrow

$$A = (\mathbf{v}_1 \quad \mathbf{v}_2)$$

7.1. Generalización



Ahora, para $\bar{\mathbf{x}}$ (el minimizador) sabemos que $\mathbf{A}\bar{\mathbf{x}}$ es ortogonal al vector $\mathbf{r} = \mathbf{b} - \mathbf{A}\bar{\mathbf{x}}$, por lo tanto su producto interno debe ser 0, es decir, son ortogonales:

$$(\mathbf{A}\bar{\mathbf{x}})^* (\mathbf{b} - \mathbf{A}\bar{\mathbf{x}}) = 0$$

$$\bar{\mathbf{x}}^* \mathbf{A}^* (\mathbf{b} - \mathbf{A}\bar{\mathbf{x}}) = 0$$

Considerando $\mathbf{b} \neq \mathbf{0}$, ¿Cómo hacemos 0 la ecuación anterior?

$$\mathbf{A}^* (\mathbf{b} - \mathbf{A}\bar{\mathbf{x}}) = 0$$

$$\mathbf{A}^* \mathbf{b} - \mathbf{A}^* \mathbf{A} \bar{\mathbf{x}} = 0$$

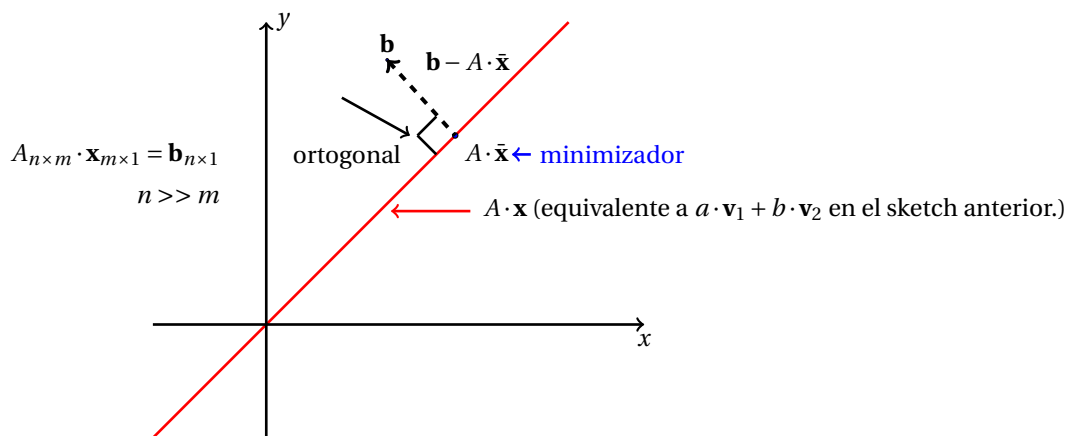
\Rightarrow

$$\mathbf{A}^* \mathbf{A} \bar{\mathbf{x}} = \mathbf{A}^* \mathbf{b}$$

esto es conocido como las ecuaciones normales.

$$\therefore \bar{\mathbf{x}} = (\mathbf{A}^* \mathbf{A})^{-1} \mathbf{A}^* \mathbf{b}$$

donde $(\mathbf{A}^* \mathbf{A})^{-1} \mathbf{A}^*$ es la pseudo-inversa de More-Penrose.



²i.e. su producto interno es 0.

7.1.1. Ejemplo

$$\underbrace{\begin{pmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}}_{\mathbf{x}} = \underbrace{\begin{pmatrix} -3 \\ 15 \\ 9 \end{pmatrix}}_{\mathbf{b}}$$

$$\Rightarrow A^* A \bar{\mathbf{x}} = A^* \mathbf{b} \Rightarrow \underbrace{\begin{pmatrix} 1 & 2 & 2 \\ -4 & 3 & 2 \end{pmatrix} \begin{pmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{pmatrix}}_{\begin{pmatrix} 9 & 6 \\ 6 & 29 \end{pmatrix}} \bar{\mathbf{x}} = \begin{pmatrix} 1 & 2 & 2 \\ -4 & 3 & 2 \end{pmatrix} \begin{pmatrix} -3 \\ 15 \\ 9 \end{pmatrix}$$

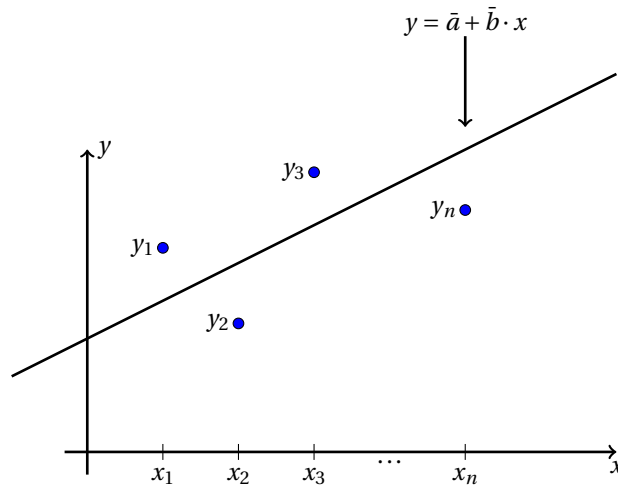
$$\begin{pmatrix} 9 & 6 \\ 6 & 29 \end{pmatrix} \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix} = \begin{pmatrix} 45 \\ 75 \end{pmatrix}$$

$$\therefore \bar{x}_1 = 3,8 \text{ \& } \bar{x}_2 = 1,8$$

$$\mathbf{r} = \mathbf{b} - A \bar{\mathbf{x}} = \begin{pmatrix} -3 \\ 15 \\ 9 \end{pmatrix} - \begin{pmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} 3,8 \\ 1,8 \end{pmatrix} = \begin{pmatrix} 0,4 \\ 2 \\ -2,2 \end{pmatrix}$$

$$\|\mathbf{x}\|_2 = \sqrt{(0,4)^2 + (2)^2 + (-2,2)^2} = 3 \rightarrow \text{Min.}$$

Recordar:



$$\underbrace{\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}}_A \underbrace{\begin{pmatrix} a \\ b \end{pmatrix}}_{\mathbf{x}} = \underbrace{\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix}}_{\mathbf{b}}$$

Ecuaciones normales: $A^* A \bar{\mathbf{x}} = A^* \mathbf{b}$

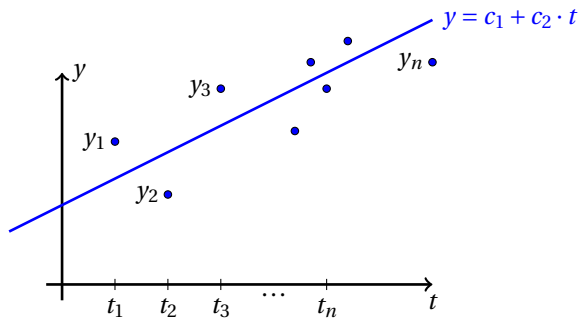
$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \bar{a} \\ \bar{b} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ x_1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$\begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix} \begin{pmatrix} \bar{a} \\ \bar{b} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{pmatrix}$$

Lo cual es lo mismo que obtuvimos antes en la ecuación (7.1), pero en ese caso se vio el problema como una minimización del error cuadrático.

7.2. Tipos de modelos

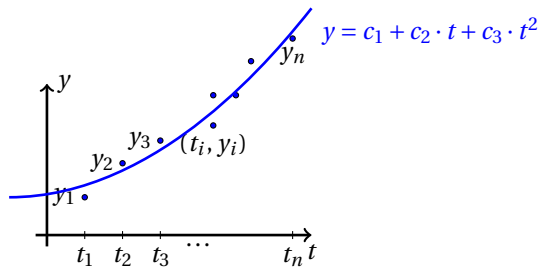
7.2.1. Modelo lineal



$$\Rightarrow \underbrace{\begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ 1 & t_3 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}}_{\mathbf{x}} = \underbrace{\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix}}_{\mathbf{b}}$$

$$\Rightarrow A^* A \bar{\mathbf{x}} = A^* \mathbf{b}$$

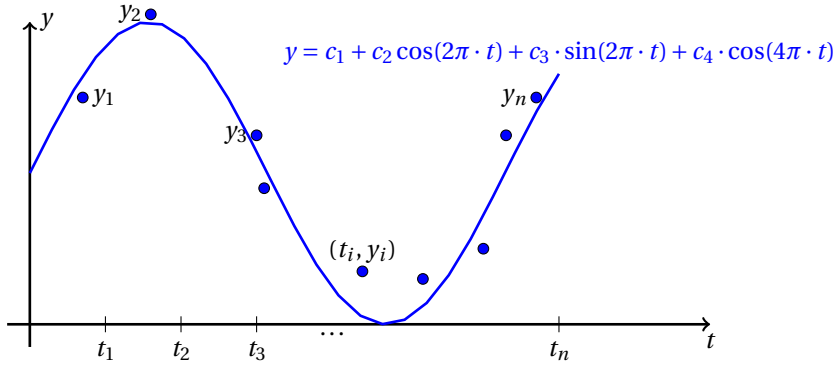
7.2.2. Modelo cuadrático



$$\Rightarrow \underbrace{\begin{pmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_n & t_n^2 \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}}_{\mathbf{x}} = \underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}}_{\mathbf{b}}$$

$$\Rightarrow A^* A \bar{\mathbf{x}} = A^* \mathbf{b}$$

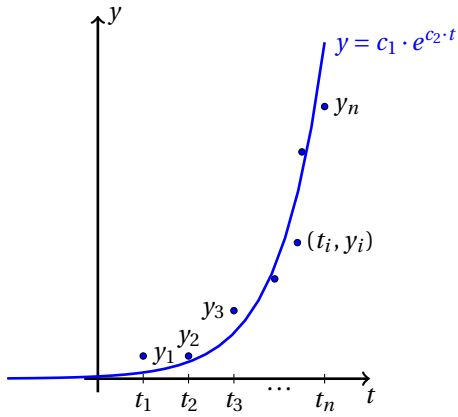
7.2.3. Modelo periódico



$$\Rightarrow \begin{pmatrix} 1 & \cos(2\pi t_1) & \sin(2\pi t_1) & \cos(4\pi t_1) \\ 1 & \cos(2\pi t_2) & \sin(2\pi t_2) & \cos(4\pi t_2) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \cos(2\pi t_n) & \sin(2\pi t_n) & \cos(4\pi t_n) \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$\Rightarrow A^* A \bar{\mathbf{x}} = A^* \mathbf{b}$$

7.2.4. Modelo exponencial

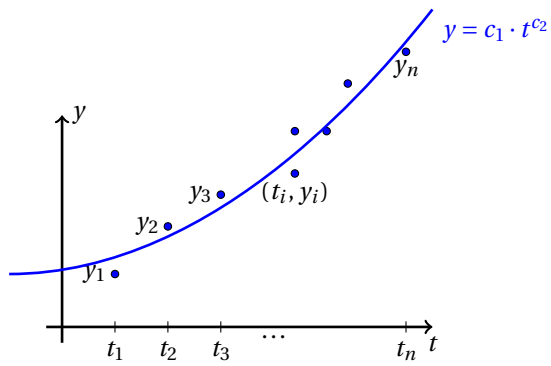


$$\Rightarrow \log y_i = \underbrace{\log(c_1)}_k + c_2 \cdot t$$

$$\begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix} \begin{pmatrix} K \\ c_2 \end{pmatrix} = \begin{pmatrix} \log y_1 \\ \log y_2 \\ \vdots \\ \log y_n \end{pmatrix}, \text{ entonces } \overline{C_1} = e^{\overline{K}}$$

$$\Rightarrow A^* A \bar{\mathbf{x}} = A^* \mathbf{b}$$

7.2.5. Ley de potencia



$$\Rightarrow \log y = \underbrace{\log(c_1)}_k + c_2 \cdot \log t$$

$$\begin{pmatrix} 1 & \log t_1 \\ 1 & \log t_2 \\ \vdots & \vdots \\ 1 & \log t_n \end{pmatrix} \begin{pmatrix} k \\ c_2 \end{pmatrix} = \begin{pmatrix} \log y_1 \\ \log y_2 \\ \vdots \\ \log y_n \end{pmatrix}$$

$$\Rightarrow A^* A \bar{\mathbf{x}} = A^* \mathbf{b}$$

Capítulo 8

Descomposición QR

Sea $A = (\mathbf{a}_1 \mid \mathbf{a}_2 \mid \cdots \mid \mathbf{a}_n)$ una matriz de dimensiones $m \times n$ con $m > n$, donde \mathbf{a}_j es la j -ésima columna de A .

Idea principal: Esta descomposición es similar a la descomposición $A = LU$ o $PA = LU$ pero usando ortogonalidad. Notar que es algo distinto a la descomposición SVD, que también es muy útil.

8.1. Ortogonalización de Gram-Schmidt y mínimos cuadrados

$$\begin{array}{ccccc}
 A_{m \times n} & = & \hat{Q}_{m \times n} & \cdot & \hat{R}_{n \times n} \\
 \uparrow & & \uparrow & & \uparrow \\
 \text{Matriz original} & & \text{matriz ortonormal} & & \text{matriz triangular superior}
 \end{array}$$

$$A = (\mathbf{a}_1 \mid \mathbf{a}_2 \mid \cdots \mid \mathbf{a}_n) = (\mathbf{q}_1 \mid \mathbf{q}_2 \mid \cdots \mid \mathbf{q}_n) \cdot \begin{pmatrix} r_{11} & r_{12} & r_{13} & \cdots & \cdots & r_{1n} \\ 0 & r_{22} & & & & \vdots \\ 0 & 0 & r_{33} & & & \vdots \\ \vdots & & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & r_{nn} \end{pmatrix} = QR$$

\Downarrow
 $\|\mathbf{q}_i\|_2^2 = \mathbf{q}_i^* \cdot \mathbf{q}_i = 1$
 $\&$
 $\mathbf{q}_i^* \cdot \mathbf{q}_j = 0, \text{ if } i \neq j$

8.1.1. Matriz ortonormal

Una matriz cuadrada Q es ortonormal si $Q^{-1} = Q^*$.

↑
Esto significa que es muy fácil obtener su inversa
⇒ es muy simple resolver un sistema de
ecuaciones lineales cuando la matriz es ortonormal.

Thm 21. Si Q es una matriz ortonormal de dimensiones $m \times m$ y \mathbf{x} es un vector m -dimensional, entonces $\|Q\mathbf{x}\|_2 = \|\mathbf{x}\|_2$. — This is only valid for norm 2

Demostración. $\|Q\mathbf{x}\|_2^2 = (Q\mathbf{x})^* (Q\mathbf{x}) = \mathbf{x}^* \underbrace{Q^* Q}_I \mathbf{x} = \mathbf{x}^* \mathbf{x} = \|\mathbf{x}\|_2^2$

□

8.1.2. Descomposición QR reducida y full

A continuación se observa la relación entre la descomposición QR reducida y full para matrices $A \in \mathbb{R}^{m \times n}$ rectangulares con $m > n$,

$$A_{m \times n} = \underbrace{\hat{Q}_{m \times n} \cdot \hat{R}_{n \times n}}_{\text{QR reducida}} = \underbrace{\left(\begin{array}{c|c|c|c} \hat{Q}_{m \times n} & \mathbf{q}_{n+1} & \cdots & \mathbf{q}_m \end{array} \right)}_{\text{QR full}} \cdot \overbrace{\begin{pmatrix} \hat{R} \\ 0 \cdots 0 \\ \vdots \ddots \vdots \\ 0 \cdots 0 \end{pmatrix}}^{R_{m \times n}}$$

¿Cómo obtenemos \hat{Q} y \hat{R} ?

$$A = \hat{Q} \hat{R} = \left(\begin{array}{c|c|c|c} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{array} \right) = \left(\begin{array}{c|c|c|c} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{array} \right) \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_{nn} \end{pmatrix}$$

Multiplicamos \hat{Q} por \hat{R} y obtenemos:

$$\left(\begin{array}{c|c|c|c} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{array} \right) = \left(\begin{array}{c|c|c|c} r_{11} \cdot \mathbf{q}_1 & r_{12} \cdot \mathbf{q}_1 + r_{22} \cdot \mathbf{q}_2 & \cdots & \sum_{i=1}^n r_{in} \cdot \mathbf{q}_i \end{array} \right)$$

↓

el k -ésimo termino es $\sum_{i=1}^k r_{ik} \cdot \mathbf{q}_i$

Idea principal: Como tenemos una igualdad, en este caso matricial, podemos utilizar el argumento que cada columna de la matriz del lado izquierdo (LHS, del inglés left hand side) debe ser igual a la respectiva columna de la matriz del lado derecho (RHS, del inglés right hand side), es decir la k -ésima columna de la matriz del lado izquierdo es igual a la k -ésima columna de la matriz del lado derecho.

	LHS	=	RHS
1ra columna:	\mathbf{a}_1	=	$r_{11} \cdot \mathbf{q}_1$
2da columna:	\mathbf{a}_2	=	$r_{12} \cdot \mathbf{q}_1 + r_{22} \cdot \mathbf{q}_2$
\vdots	\vdots		\vdots
k -ésima columna:	\mathbf{a}_k	=	$\sum_{i=1}^k r_{ik} \cdot \mathbf{q}_i$
\vdots	\vdots		\vdots
n -ésima columna	\mathbf{a}_n	=	$\sum_{i=1}^n r_{in} \cdot \mathbf{q}_i$

1ra ecuación vectorial:

$$\underbrace{\mathbf{a}_1}_{\text{Conocido}} = \underbrace{r_{11}}_{\text{Desconocido}} \cdot \underbrace{\mathbf{q}_1}_{\text{Desconocido}}$$

¿Qué otra información conocemos? $\mathbf{q}_i^* \cdot \mathbf{q}_i = 1 \iff \|\mathbf{q}_i\|_2 = 1$, y $\mathbf{q}_i^* \cdot \mathbf{q}_j = 0$, para $i \neq j$. ¿Cómo usamos esa información para encontrar r_{11} y \mathbf{q}_1 ?

Tenemos 2 opciones:

$$\begin{array}{ll} \mathbf{q}_1^* \cdot \mathbf{a}_1 = r_{11} \cdot \underbrace{\mathbf{q}_1^* \cdot \mathbf{q}_1}_1 & \searrow \\ \mathbf{q}_1^* \cdot \mathbf{a}_1 = r_{11} & \|\mathbf{a}_1\|_2 = |r_{11}| \cdot \underbrace{\|\mathbf{q}_1\|}_1 \\ & \|\mathbf{a}_1\| = \|r_{11}\| \end{array}$$

¿Qué opción usamos y por qué?

Sí, la segunda opción dado que podemos utilizar la identidad “ $\|\mathbf{q}_1\|=1$ ” y deshacernos de ese término. Sin embargo, la identidad $\|\mathbf{a}_1\| = |r_{11}|$ nos da 2 opciones para el signo de r_{11} , por lo cual elegimos $r_{11} > 0 \Rightarrow \boxed{r_{11} = \|\mathbf{a}_1\|}$. Por lo tanto ahora la ecuación vectorial tiene otro término conocido:

$$\underbrace{\mathbf{a}_1}_{\text{Conocido}} = \underbrace{r_{11}}_{\text{Conocido}} \cdot \underbrace{\mathbf{q}_1}_{\text{Desconocido}}$$

¿Cómo encontramos \mathbf{q}_1 ?

$$\mathbf{q}_1 = \boxed{\frac{\mathbf{a}_1}{r_{11}}} \quad :)$$

2da ecuación vectorial:

$$\underbrace{\mathbf{a}_2}_{\text{Conocido}} = \underbrace{r_{12}}_{\text{Desconocido}} \cdot \underbrace{\mathbf{q}_1}_{\text{Conocido}} + \underbrace{r_{22}}_{\text{Desconocido}} \cdot \underbrace{\mathbf{q}_2}_{\text{Desconocido}}$$

¿Qué podemos hacer acá?

¿Podemos obtener la norma de \mathbf{a}_2 para encontrar r_{22} al igual que antes?

$$\begin{aligned}
\|\mathbf{a}_2\|_2^2 &= \mathbf{a}_2^T \cdot \mathbf{a}_2 = (r_{12} \cdot \mathbf{q}_1 + r_{22} \cdot \mathbf{q}_2)^* \cdot (r_{12} \cdot \mathbf{q}_1 + r_{22} \cdot \mathbf{q}_2) \\
&= (r_{12} \cdot \mathbf{q}_1^T + r_{22} \cdot \mathbf{q}_2^T) \cdot (r_{12} \cdot \mathbf{q}_1 + r_{22} \cdot \mathbf{q}_2) \\
&= r_{12}^2 + r_{22}^2
\end{aligned}$$

$\therefore \|\mathbf{a}_2\|_2 = \sqrt{r_{12}^2 + r_{22}^2} \Rightarrow$ ¡Una ecuación y 2 incógnitas!, entonces no nos sirve este camino.

¿Qué otra propiedad/información podemos usar acá?

Solo tenemos estas propiedades: $\mathbf{q}_i^* \cdot \mathbf{q}_i = 1$ & $\mathbf{q}_i^* \cdot \mathbf{q}_j = 0, \quad i \neq j$

Entonces tenemos 2 opciones nuevamente

$$\mathbf{q}_2^* \cdot \mathbf{a}_2 = r_{12} \cdot \underbrace{\mathbf{q}_2^* \cdot \mathbf{q}_1}_0 + r_{22} \cdot \underbrace{\mathbf{q}_2^* \cdot \mathbf{q}_2}_1$$

Desconocido Desconocido

$$\underbrace{\mathbf{q}_2^*}_{\text{Desconocido}} \cdot \underbrace{\mathbf{a}_2}_{\text{Conocido}} = \underbrace{r_{22}}_{\text{Desconocido}}$$

Una ecuación y 2 incógnitas, no sirve...

$$\mathbf{q}_1^* \cdot \mathbf{a}_2 = r_{12} \cdot \underbrace{\mathbf{q}_1^* \cdot \mathbf{q}_1}_1 + r_{22} \cdot \underbrace{\mathbf{q}_1^* \cdot \mathbf{q}_2}_0$$

Desconocido

$$\underbrace{\mathbf{q}_1^*}_{\text{Conocido}} \cdot \underbrace{\mathbf{a}_2}_{\text{Conocido}} = \underbrace{r_{12}}_{\text{Desconocido}}$$

¡Tenemos un ganador!

¿Qué es lo que tenemos ahora entonces?

$$\underbrace{\mathbf{a}_2}_{\text{Conocido}} = \underbrace{r_{12} \cdot \mathbf{q}_1}_{\text{Conocido}} + \underbrace{r_{22} \cdot \mathbf{q}_2}_{\text{Desconocido}} \quad (8.1)$$

¿Qué podemos hacer ahora para conocer los Desconocido? ¿Recuerda los pasos que hicimos antes para obtener r_{11} y \mathbf{q}_1 de la ecuación $\mathbf{a}_1 = r_{11} \cdot \mathbf{q}_1$? Estos fueron los pasos.

1er paso: $r_{11} = \|\mathbf{a}_1\|$, considerando $r_{ii} > 0$.

2do paso: $\mathbf{q}_1 = \frac{\mathbf{a}_1}{r_{11}}$.

Entonces, moviendo lo conocido al lado izquierdo de la ecuación (8.1) obtenemos:

$$\underbrace{\mathbf{a}_2 - r_{12} \cdot \mathbf{q}_1}_{\text{¡Esto es un vector conocido ahora!}} = \underbrace{r_{22} \cdot \mathbf{q}_2}_{\text{Esto es aún desconocido}}$$

¿Qué hacemos entonces?

$$r_{22} = \|\mathbf{a}_2 - r_{12} \cdot \mathbf{q}_1\|_2$$

y

$$\mathbf{q}_2 = \frac{\mathbf{a}_2 - r_{12} \cdot \mathbf{q}_1}{r_{22}}$$

En general:

$$\underbrace{\mathbf{a}_k}_{\text{Conocido}} = \underbrace{r_{1k}}_{\text{Desconocido}} \cdot \underbrace{\mathbf{q}_1}_{\text{Conocido}} + \underbrace{r_{2k}}_{\text{Desconocido}} \cdot \underbrace{\mathbf{q}_2}_{\text{Conocido}} + \underbrace{r_{3k}}_{\text{Desconocido}} \cdot \underbrace{\mathbf{q}_3}_{\text{Conocido}} + \dots + \underbrace{r_{k-1,k}}_{\text{Desconocido}} \cdot \underbrace{\mathbf{q}_{k-1}}_{\text{Conocido}} + \underbrace{r_{k,k}}_{\text{Desconocido}} \cdot \underbrace{\mathbf{q}_k}_{\text{Desconocido}}$$

\therefore Hemos construido la Ortonormalización Clásica de Gram-Schmidt.

8.1.3. Algoritmo Clásico de Gram-Schmidt para obtener la descomposición QR

Sea \mathbf{a}_j , $j \in \{1, \dots, n\}$ un conjunto de vectores linealmente independientes¹, el algoritmo es el siguiente:

```

1  for j in range(1, n+1):
2      y = a_j
3      for i in range(1, j):
4          r_ij = q_i^T · a_j
5          y = y - r_ij · q_i
6          r_jj = ||y||_2
7          q_j = y / r_jj

```

Si cambiamos la línea 4 del algoritmo por la siguiente: $r_{ij} = \mathbf{q}_i^* \cdot \mathbf{y}$ obtenemos la ortonormalización modificada de Gram-Schmidt, la cual es ¡MUCHO MEJOR! ¿Por qué?

8.1.4. ¿Cómo usamos la descomposición QR en problemas de mínimos cuadrados?

Dado

$$A_{m \times n} \cdot \mathbf{x}_{n \times 1} = \mathbf{b}_{m \times 1}. \quad (8.2)$$

Obtenga la descomposición QR full, aunque en realidad al final solo utilizará la descomposición QR reducida, pero es necesario utilizar la descomposición QR full en la derivación del algoritmo final,

$$A_{m \times n} = Q_{m \times m} \cdot \begin{bmatrix} \hat{R}_{n \times n} \\ 0 \end{bmatrix}.$$

Reemplace la matriz A por la descomposición QR en (8.2):

$$Q_{m \times m} \cdot \begin{bmatrix} \hat{R}_{n \times n} \\ 0 \end{bmatrix} \cdot \mathbf{x} = \mathbf{b}$$

Multiplicando por la matriz ortonormal $Q_{m \times m}^*$ por la izquierda,

$$\underbrace{Q_{m \times m}^* \cdot Q_{m \times m}}_I \cdot \begin{bmatrix} \hat{R}_{n \times n} \\ 0 \end{bmatrix} \cdot \mathbf{x} = Q_{m \times m}^* \cdot \mathbf{b} \quad (8.3)$$

Pero recuerde $Q_{m \times m} = \left(\underbrace{\hat{Q}_{m \times n}}_{Q \text{ reducida}} \mid \mathbf{q}_{n+1} \mid \dots \mid \mathbf{q}_m \right)$, lo que implica que la ecuación (8.3) se reduce a,

$$\begin{pmatrix} \hat{R}_{n \times n} \\ 0 \end{pmatrix} \cdot \mathbf{x} = \begin{pmatrix} \hat{Q}_{m \times n}^* \\ \mathbf{q}_{n+1} \\ \vdots \\ \mathbf{q}_m \end{pmatrix} \cdot \mathbf{b}$$

Dado que la parte inferior de $\begin{pmatrix} \hat{R}_{n \times n} \\ 0 \end{pmatrix}$ es la matriz nula, se obtiene la siguiente simplificación:

$$\underbrace{\hat{R}_{n \times n} \cdot \bar{\mathbf{x}}}_{\text{Esto se puede resolver con Backward Substitution}} = \underbrace{\hat{Q}_{m \times n}^* \cdot \mathbf{b}}_{\text{Esto es solo un producto matriz vector}} \quad (8.4)$$

¹¿Qué pasaría hubieran vectores linealmente dependientes?

- El resultado obtenido en la ecuación (8.4) también puede obtenerse desde las ecuaciones normales.
- Notar que en la ecuación (8.4) se obtuvo la solución de mínimos cuadrados de (8.2), por eso se usó la notación \bar{x} . Si uno minimizara un funcional distinto, se podría obtener otra solución.

8.1.5. Ejemplo (4.14) del libro guía

$$\underbrace{\begin{pmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{pmatrix}}_A \cdot \underbrace{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}}_b = \underbrace{\begin{pmatrix} -3 \\ 15 \\ 9 \end{pmatrix}}_b$$

↙

$$A = \frac{1}{15} \cdot \begin{pmatrix} 5 & -14 & 2 \\ 10 & 5 & 10 \\ 10 & 2 & -11 \end{pmatrix} \cdot \begin{pmatrix} 3 & 2 \\ 0 & 5 \\ 0 & 0 \end{pmatrix} \Rightarrow \begin{pmatrix} 3 & 2 \\ 0 & 5 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{15} \cdot \begin{pmatrix} 5 & 10 & 10 \\ -14 & 5 & 2 \\ 2 & 10 & -11 \end{pmatrix} \cdot \begin{pmatrix} -3 \\ 15 \\ 9 \end{pmatrix} = \begin{pmatrix} 15 \\ 9 \\ 3 \end{pmatrix}$$

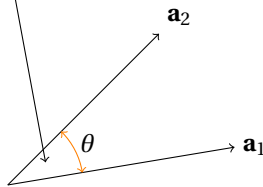
$\uparrow \quad \uparrow \quad \uparrow$
 $\mathbf{q}_1 \quad \mathbf{q}_2 \quad \mathbf{q}_3$

$$\Rightarrow \underbrace{\begin{pmatrix} 3 & 2 \\ 0 & 5 \end{pmatrix}}_{\hat{R}_{n \times n}} \cdot \underbrace{\begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix}}_{\hat{Q}_{m \times n}^*} = \frac{1}{15} \cdot \begin{pmatrix} 5 & 10 & 10 \\ -14 & 5 & 2 \end{pmatrix} \cdot \underbrace{\begin{pmatrix} -3 \\ 15 \\ 9 \end{pmatrix}}_b = \begin{pmatrix} 15 \\ 9 \end{pmatrix}$$

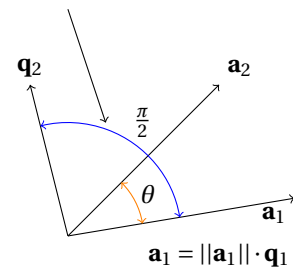
$$\therefore \bar{x}_1 = 3,8 \quad , \quad \bar{x}_2 = 1,8$$

8.1.6. Interpretación gráfica del proceso de ortonormalización

Si θ es muy pequeño se generan bases mal condicionadas lo cual es malo para la computación.



\Rightarrow



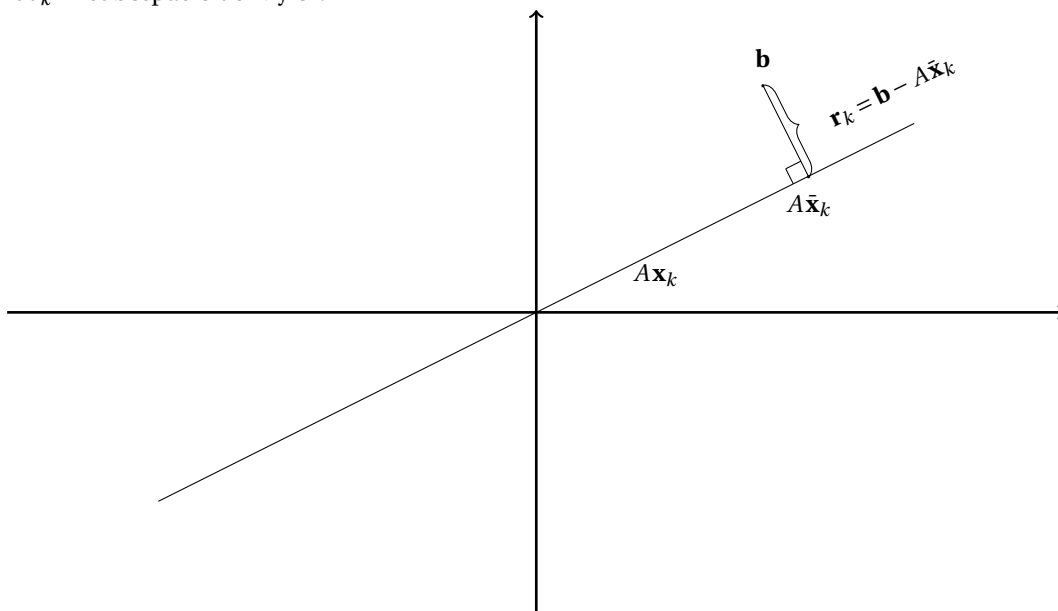
Los 2 vectores nuevos generan el mismo espacio vectorial que los vectores originales

Capítulo 9

Método del residuo mínimo generalizado (GMRes)

9.1. Derivación de GMRes

Idea: Encontrar una solución a $A\mathbf{x} = \mathbf{b}$, donde $A \in \mathbb{R}^{m \times m}$ y $\mathbf{x}, \mathbf{b} \in \mathbb{R}^m$, por medio de una aproximación de \mathbf{x} con $\mathbf{x}_k \in \mathcal{K}_k \rightarrow$ subespacio de Krylov.



¿El subespacio generado por $A\mathbf{x}_k$ es un subconjunto de \mathbb{R}^m ?

¿Qué es \mathcal{K}_k ? \rightarrow Es el subespacio de Krylov definido por

$$\mathcal{K}_k = \text{span}\{\mathbf{r}_0, A\mathbf{r}_0, A^2\mathbf{r}_0, A^3\mathbf{r}_0, \dots, A^{k-1}\mathbf{r}_0\}$$

donde

$$\mathbf{r}_0 = \mathbf{b} - A\mathbf{x}_0$$

¿Qué significa $\mathbf{x}_k \in \mathcal{K}_k$?

$$\begin{aligned}\mathbf{x}_k &= \tilde{c}_1 \mathbf{r}_0 + \tilde{c}_2 A \mathbf{r}_0 + \cdots + \tilde{c}_k A^{k-1} \mathbf{r}_0 \\ &= \underbrace{\begin{pmatrix} \mathbf{r}_0 & A \mathbf{r}_0 & A^2 \mathbf{r}_0 & \cdots & A^{k-1} \mathbf{r}_0 \end{pmatrix}}_{K_{m \times k} \text{ es la base de Krylov}} \begin{pmatrix} \tilde{c}_1 \\ \tilde{c}_2 \\ \vdots \\ \tilde{c}_k \end{pmatrix}\end{aligned}$$

¿Qué tan útil es que $\mathbf{x}_k \in \mathcal{K}_k$? Entonces, $\mathbf{x} \approx \mathbf{x}_k + \mathbf{x}_0$

$$\begin{aligned}\Rightarrow \mathbf{x} &\approx \mathbf{x}_0 + \begin{pmatrix} \mathbf{r}_0 & A \mathbf{r}_0 & A^2 \mathbf{r}_0 & \cdots & A^{k-1} \mathbf{r}_0 \end{pmatrix} \begin{pmatrix} \tilde{c}_1 \\ \tilde{c}_2 \\ \vdots \\ \tilde{c}_k \end{pmatrix} \\ \Rightarrow A \mathbf{x} &= \mathbf{b} \\ A(\mathbf{x}_k + \mathbf{x}_0) &= \mathbf{b} \\ \Rightarrow A \mathbf{x}_k &= \mathbf{b} - A \mathbf{x}_0 \\ A \mathbf{x}_k &= \mathbf{r}_0 \\ \Rightarrow \underbrace{A}_{A_{m \times m}} \underbrace{\begin{pmatrix} \mathbf{r}_0 & A \mathbf{r}_0 & A^2 \mathbf{r}_0 & \cdots & A^{k-1} \mathbf{r}_0 \end{pmatrix}}_{K_{m \times k}} \underbrace{\begin{pmatrix} \tilde{c}_1 \\ \tilde{c}_2 \\ \vdots \\ \tilde{c}_k \end{pmatrix}}_{\tilde{\mathbf{c}}_{k \times 1}} &= \underbrace{\mathbf{r}_0}_{m \times 1} \\ \Rightarrow A_{m \times m} K_{m \times k} \tilde{\mathbf{c}}_{k \times 1} &= \tilde{\mathbf{r}}_{0 m \times 1} \quad \circledast\end{aligned}$$

¿Que es \circledast ?

¡Sí, es un problema de mínimos cuadrados!

\Rightarrow queremos minimizar

$$\| \underbrace{A}_{m \times k} \underbrace{\tilde{\mathbf{c}}}_{k \times 1} - \underbrace{\mathbf{r}_0}_{m \times 1} \|_2$$

↑

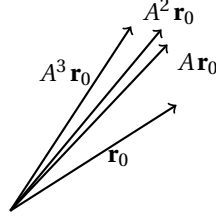
¡esto es sólo una matriz rectangular!

¡Bien! ¿Hay algún problema?

Sí... pero también hay una solución

¿Cuál es el problema?

$\mathcal{K}_k = \text{span}\{\mathbf{r}_0, A\mathbf{r}_0, A^2\mathbf{r}_0, \dots, A^{k-1}\mathbf{r}_0\}$ es una base mal condicionada.
¡son casi paralelos!



¿Qué podemos hacer?

¿Cómo?

Sí, ¡necesitamos hablar con Gram-Schmidt!

Esto significa encontrar una base ortonormal a partir de $\mathcal{K}_k = \text{span}\{\mathbf{r}_0, A\mathbf{r}_0, A^2\mathbf{r}_0, \dots, A^{k-1}\mathbf{r}_0\} = \text{span}\{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_k\}$

∴ Ajustando $\mathbf{q}_1 = \mathbf{r}_0 / \|\mathbf{r}_0\|_2$ y usando el algoritmo modificado de Gram-Schmidt (mejor conocido en este contexto como la iteración de Arnoldi) podemos encontrar \mathbf{q}_i , $i = 2 \dots k$.

Entonces, el problema de minimización original

$$\|A K_k \tilde{\mathbf{c}} - \mathbf{r}_0\|_2$$

considerando, $\mathbf{x}_0 + K_{m \times k} \tilde{\mathbf{c}} = \mathbf{x}_0 + Q_k \mathbf{c}$

$$\downarrow$$

$$(\mathbf{q}_1 | \mathbf{q}_2 | \dots | \mathbf{q}_k)$$

$$\Rightarrow \|A(\mathbf{x}_0 + Q_k \mathbf{c}) - \mathbf{b}\|_2 = \|A Q_k \mathbf{c} - (\mathbf{b} - A \mathbf{x}_0)\|_2$$

$$= \|\underbrace{A Q_k}_{m \times k} \mathbf{c} - \mathbf{r}_0\|_2$$

donde las columnas de Q_k son una mejor base para \mathcal{K}_k .

Sin embargo, $\|\underbrace{A Q_k}_{m \times k} \mathbf{c} - \mathbf{r}_0\|_2$ aún es un problema de mínimos cuadrados. ¿Podemos hacerlo aún mejor?

Desde antes: $\langle \mathbf{r}_0, A\mathbf{r}_0, A^2\mathbf{r}_0, \dots, A^{k-1}\mathbf{r}_0 \rangle = \langle \mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_k \rangle$

entonces, desde que dijimos $\mathbf{q}_1 = \mathbf{r}_0 / \|\mathbf{r}_0\|_2$

$$\Rightarrow \quad A\mathbf{r}_0 = \hat{h}_{11}\mathbf{q}_1 + \hat{h}_{12}\mathbf{q}_2 \text{ (descomponiendo } A\mathbf{r}_0 \text{ en la} \\ \text{dirección conocida } \mathbf{q}_1 \text{ y} \\ \text{la dirección desconocida } \mathbf{q}_2)$$

¿Cómo encontramos \mathbf{q}_2 ? ¡Al igual como lo hicimos para QR!

$$\left. \begin{aligned} \hat{h}_{11} &= \mathbf{q}_1^T A\mathbf{r}_0 \\ \hat{h}_{21} &= \|A\mathbf{r}_0 - \hat{h}_{11}\mathbf{q}_1\|_2 \\ \mathbf{q}_2 &= \frac{A\mathbf{r}_0 - \hat{h}_{11}\mathbf{q}_1}{\hat{h}_{21}} \end{aligned} \right\} \text{Esta es una iteración de Arnoldi (o} \\ \text{solo QR modificado)}$$

$$\Rightarrow \quad A\mathbf{r}_0 = \begin{pmatrix} \mathbf{q}_1 & \mathbf{q}_2 \end{pmatrix} \begin{pmatrix} \hat{h}_{11} \\ \hat{h}_{21} \end{pmatrix}$$

Ahora, en vez de usar $A\mathbf{r}_0$ en la izquierda, usamos $A\mathbf{q}_1$, esto debido a que:

$$\langle \mathbf{r}_0, A\mathbf{r}_0, A^2\mathbf{r}_0, \dots, A^{k-1}\mathbf{r}_0 \rangle = \langle \mathbf{q}_1, A\mathbf{q}_1, A\mathbf{q}_2, \dots, A\mathbf{q}_{k-1} \rangle = \langle \mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_k \rangle$$

$$\begin{aligned} A\mathbf{q}_1 &= h_{11}\mathbf{q}_1 + h_{21}\mathbf{q}_2 \\ \Rightarrow \quad h_{11} &= \mathbf{q}_1^T A\mathbf{q}_1 \\ h_{21} &= \|A\mathbf{q}_1 - h_{11}\mathbf{q}_1\|_2 \\ \mathbf{q}_2 &= (A\mathbf{q}_1 - h_{11}\mathbf{q}_1) / h_{21} \end{aligned}$$

En general, esto nos da la descomposición parcial de Hessemberg

$$AQ_k = Q_{k+1} \tilde{H}_k$$

¡nótese la diferencia en los sub-índices de Q !

$$A_{m \times m} \underbrace{\begin{pmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_k \end{pmatrix}}_{Q_k \quad m \times k} = \underbrace{\begin{pmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_k & \mathbf{q}_{k+1} \end{pmatrix}}_{Q_{k+1} \quad m \times (k+1)} \underbrace{\begin{pmatrix} h_{11} & h_{12} & & & h_{1,k} \\ h_{21} & h_{22} & & & h_{2,k} \\ 0 & h_{32} & \ddots & & \\ \vdots & 0 & & \ddots & \\ \vdots & \vdots & & & h_{k,k} \\ 0 & 0 & & & h_{k+1,k} \end{pmatrix}}_{\tilde{H}_k \quad (k+1) \times k}$$

¿Es esto correcto? ¹

¹¡Sí!

$$\Rightarrow \left\| A Q_k \mathbf{c} - \mathbf{r}_0 \right\|_2 = \left\| Q_{k+1} \tilde{H}_k \mathbf{c} - \mathbf{r}_0 \right\|_2$$

$$\text{recordar } \mathbf{q}_1 = \frac{\mathbf{r}_0}{\|\mathbf{r}_0\|_2} \Rightarrow \mathbf{r}_0 = \|\mathbf{r}_0\|_2 \mathbf{q}_1 = \|\mathbf{r}_0\|_2 Q_{k+1} \mathbf{e}_1$$

↓

$$\mathbf{e}_1^T = \underbrace{\langle 1, 0, \dots, 0 \rangle^T}_{k+1 \text{ componentes}}$$

Reemplazando \mathbf{r}_0 :

$$\left\| Q_{k+1} \tilde{H}_k \mathbf{c} - \mathbf{r}_0 \right\|_2 = \left\| Q_{k+1} \tilde{H}_k \mathbf{c} - Q_{k+1} \|\mathbf{r}_0\|_2 \mathbf{e}_1 \right\|_2$$

$$= \left\| Q_{k+1} (\tilde{H}_k \mathbf{c} - \|\mathbf{r}_0\|_2 \mathbf{e}_1) \right\|_2$$

$$\text{recordar: } \|\mathbf{v}\|_2 = \sqrt{\mathbf{v}^* \mathbf{v}}$$

$$= \sqrt{(\tilde{H}_k \mathbf{c} - \|\mathbf{r}_0\|_2 \mathbf{e}_1)^* \underbrace{Q_{k+1}^* Q_{k+1}}_{I_{k+1}} (\tilde{H}_k \mathbf{c} - \|\mathbf{r}_0\|_2 \mathbf{e}_1)}$$

$$= \left\| \underbrace{\tilde{H}_k}_{(k+1) \times k} \cdot \underbrace{\mathbf{c}}_{k \times 1} - \underbrace{\|\mathbf{r}_0\|_2 \mathbf{e}_1}_{(k+1) \times 1} \right\|_2$$

¿Qué es eso?

$$\begin{pmatrix} h_{11} & h_{12} & \cdots & \cdots & h_{1,k} \\ h_{21} & h_{22} & & & h_{2,k} \\ 0 & h_{32} & \ddots & & \vdots \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & h_{k,k} \\ 0 & \cdots & & 0 & h_{k+1,k} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ \vdots \\ c_k \end{pmatrix} = \begin{pmatrix} \|\mathbf{r}_0\| \\ 0 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix}$$

“¡diminuto!”
↓
es un problema de mínimos cuadrados
de tamaño $(k+1) \times k$ y antes teníamos uno
de tamaño $m \times k$.
↑
“¡GIGANTE!”

$$\therefore \left\| A K_k \tilde{\mathbf{c}} - \mathbf{b} \right\|_2 = \left\| \tilde{H}_k \mathbf{c} - \|\mathbf{r}_0\|_2 \mathbf{e}_1 \right\|_2$$

Es decir, reducimos un problema de mínimos cuadrados de $m \times n$ a un problema de mínimos cuadrados de $(k+1) \times k$

9.2. Algorítmicamente

```

1  $\mathbf{x}_0 = \text{"initial guess"}$ 
2  $\mathbf{r}_0 = \mathbf{b} - A\mathbf{x}_0$ 
3  $\mathbf{q}_1 = \frac{\mathbf{r}_0}{\|\mathbf{r}_0\|_2}$ 
4 for  $k$  in  $\text{range}(1, m+1)$  :
5      $\mathbf{y} = A \cdot \mathbf{q}_k$ 
6     for  $j$  in  $\text{range}(1, k+1)$  :
7          $h_{jk} = \mathbf{q}_j^* \cdot \mathbf{y}$ 
8          $\mathbf{y} = \mathbf{y} - h_{jk} \mathbf{q}_j$ 
9      $h_{k+1,k} = \|\mathbf{y}\|_2$ 
10    if  $h_{k+1,k} > 0$  :
11         $\mathbf{q}_{k+1} = \frac{\mathbf{y}}{h_{k+1,k}}$ 
12     $\text{argmin}_{\mathbf{c}_k} \|\tilde{H}\mathbf{c}_k - \|\mathbf{r}_0\|\mathbf{e}_1\|_2$ 
13     $\mathbf{x}_k = Q_k \mathbf{c}_k + \mathbf{x}_0$ 

```

donde $\tilde{H} = \begin{pmatrix} h_{11} & & & h_{1,k} \\ h_{21} & \ddots & & \\ & \ddots & \ddots & \\ & 0 & \ddots & h_{k,k} \\ & & & h_{k+1,k} \end{pmatrix}_{(k+1) \times k}$

Capítulo 10

Agradecimientos

Los siguientes apuntes están basados en el texto guía **NUMERICAL ANALYSIS** Second Edition - Timothy Sauer y en los apuntes del Profesor Claudio Torres. Estos aún pueden presentar errores, por lo que deben ser utilizados como ayuda para el estudio.

Enviar cualquier sugerencia o error al profesor Claudio Torres a ctorres@inf.utfsm.cl, quien las considerará para su inclusión.

En la corrección y generación de los apuntes se han recibido muchas sugerencias y correcciones, aquí se listan los nombres de algunos voluntarios pero existen varios que posiblemente e involuntariamente, no han sido incluidos/as.

Se agradece de todos modos a quienes han querido mejorar estos apuntes a lo largo de estos años.

Las secciones fueron traspasadas a Latex por:

Sección/es	Autor	e-mail
1	Patricio Ramirez	luis.ramirez.12@sansano.usm.cl
2	Ian Zamorano	ian.zamorano.12@sansano.usm.cl
3	Felipe Osses	fosses@alumnos.inf.utfsm.cl
4	Francisco Salazar	francisco.salazar.12@sansano.usm.cl
5	Laura Bermeo	laura.bermeo.12@sansano.usm.cl
6	Yerson Escobar	yerson.escobar.12@sansano.usm.cl
7	Yerson Escobar	yerson.escobar.12@sansano.usm.cl
8	Javier Ortiz	javier.ortizg@alumnos.usm.cl
9	Germán Marcelo Treimun	german.treimun@alumnos.usm.cl
10	Marco Rojas	marco.rojaso@alumnos.usm.cl
11	Diego Reyes	diego.reyes@alumnos.usm.cl
12	Alfredo Gallardo	agallard@alumnos.inf.utfsm.cl
13	Alfredo Gallardo	agallard@alumnos.inf.utfsm.cl
14	Renata Mella	renata.mella.12@sansano.usm.cl

Cuadro 10.1: Autores Apuntes CC1 - 2015.

Las secciones fueron corregidas por:

Sección/es	Autor	e-mail
2 y 3	María José Astillo	mastudil@alumnos.inf.utfsm.cl
4 y 5	Daniel Quinteros	daniel.quinteros.12@sansano.usm.cl
1,7,8,9 y 10	Fernando Iturbe	fernando.iturbe.12@sansano.usm.cl

Cuadro 10.2: Correctores Apuntes CC1 - 2016.

Las gráficos fueron corregidos por:

Autor	e-mail
Roberto Fuentes	roberto.fuentes@alumnos.usm.cl

Cuadro 10.3: Corrector gráficos Apuntes CC1 - 2016.

Las secciones fueron mejoradas por:

Sección	Autor	e-mail
1 y 2	Daniel San Martín	dsanmartinreyes@gmail.com
3	Pablo Ibarra	pablo.ibarras@alumnos.usm.cl
4	Patricio Horth Medina	patricio.horth@alumnos.usm.cl
5 y 6	Victor Zuñiga Millán	victor.zuniga@alumnos.usm.cl
7	Diego Villegas Aravena	diego.villegas.12@sansano.usm.cl
8 y 10	Fabián Da Silva	fabian.dasilva.12@sansano.usm.cl
9	Iván Lazo R	ivan.lazo@alumnos.usm.cl
11, 12 y 14	Germán Treimun	german.treimun@alumnos.usm.cl

Cuadro 10.4: CC1-2016

Los anexos fueron creados por:

Sección	Autor	e-mail
Review de calculo	Katherine Barros	katherine.barros@alumnos.usm.cl
Tutorial de instalación iPython	Diego Reyes	diego.reyes@alumnos.usm.cl

Cuadro 10.5: CC1-2016

Apunte corregido y re-estructurado por:

Autor	e-mail
Fernando Iturbe	fernando.iturbe.12@sansano.usm.cl

Cuadro 10.6: Correctores Apuntes CC1 - 2017.

Apunte corregido:

Autor
Rodrigo Hernández

Cuadro 10.7: Correctores Apuntes CC1 - 2018.

Sugerencias:

Autor
Prof. Cristopher Arenas
Prof. Ariel Sanhueza

Cuadro 10.8: Apuntes CC1 - 2019.