

Un Curso de Álgebra Lineal

con notación asociativa y un módulo para Python

Edición curso 2022/2023

Versión: 18 de marzo de 2023

Marcos Bujosa



Licencia: Creative Commons Reconocimiento-CompartirIgual 4.0 Internacional

Copyright © 2008–2023 Marcos Bujosa

Puede encontrar la última versión de este libro en: <https://github.com/mbujosab/CursoDeAlgebraLineal>

Este libro ha sido escrito con L^AT_EX.

Copyright © 2008–2023 Marcos Bujosa

Licencia: Creative Commons Reconocimiento-CompartirIgual 4.0 Internacional



Algunos derechos reservados. Esta obra está bajo una licencia Creative Commons Reconocimiento-CompartirIgual 4.0 Internacional. Para ver una copia de esta licencia, visite <http://creativecommons.org/licenses/by-sa/4.0/> o envíe una carta a Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.

Nunca hubiera podido escribir este libro sin la ayuda y enseñanzas de mi hermano Andrés.
Si este libro tiene algo de valor, se debe en gran medida a todo lo que él me ha enseñado.

Prefacio

Aunque muchos perciben las matemáticas como una batería de procedimientos mecánicos para resolver problemas tipo, en realidad las matemáticas son un lenguaje. La principal intención de este curso es que el estudiante lo perciba y juegue con las reglas de dicho lenguaje.

También es frecuente escuchar que “*las matemáticas están en todas las cosas*”. Yo discrepo. Creo que las matemáticas son algo al margen de la realidad material del mundo, y comparto la frase de Eduardo Sáenz de Cabezón de que “*las matemáticas son el lenguaje en el que nosotros leemos el mundo*” (en su charla “*Las matemáticas nos hacen más libres y menos manipulables*”; minuto 6:40). Es decir, son el lenguaje que usamos para describir la realidad. A las descripciones matemáticas del mundo las llamamos *modelos*... y cuando nos “ponemos estupendos” incluso las denominamos *modelos científicos*. Las matemáticas nos sirven para expresar ideas y describir relaciones, pero no necesariamente relacionadas con el mundo real. Por eso considero que las matemáticas son algo muy muy distinto al mundo material que denominamos “realidad”.

Las matemáticas son un *lenguaje formal*. Consisten en un conjunto de símbolos y unas reglas para su manipulación. Así es como entiendo la cita de Charles F. Van Loan “Notation is everything”. Las matemáticas son notación. No obstante, solemos dar una interpretación subjetiva a dicha notación. Las interpretaciones nos dotan de un esquema mental que nos ayuda a pensar, y por eso son importantes. Pero las interpretaciones corresponden a un nivel distinto. De hecho, una misma notación puede tener interpretaciones alternativas. Por ejemplo, utilizamos el Teorema de Pitágoras cuando trabajamos con distancias euclídeas, con números complejos, al deducir algunas identidades trigonométricas, al analizar señales o series temporales, para tener una interpretación geométrica de la media y la desviación típica, etc... y en todos estos ejemplos el formalismo del Teorema de Pitágoras es el mismo. El formalismo es el pilar sobre el que se apoyan las distintas interpretaciones. Así pues, las matemáticas son un conjunto de símbolos y unas reglas de uso para dichos símbolos. En matemáticas LA NOTACIÓN LO ES TODO.

Fiel a este espíritu, este libro presta especial importancia a la notación y al juego de manipulación de símbolos. También muestra algunas interpretaciones geométricas, como la representación gráfica de un vector o la proyección de un vector sobre un subespacio, pero el peso recae en la manipulación de los símbolos, y no tanto en la interpretación. Así pues, al escribirlo, he tenido un especial cuidado en elegir una notación que evite la ambigüedad¹ y que sea operativa. Ello me ha empujado a seguir algunas reglas de notación que no son estándar (véase el capítulo de introducción al libro). Espero que la claridad y consistencia de la notación propuesta compense el coste de haberme alejado de la notación habitual en algunos aspectos.

La potencia de esta notación como lenguaje se demuestra en su implementación como módulo de Python.

¡Qué mejor muestra de que las matemáticas son un lenguaje, que usar dicho lenguaje con un PC para “enseñar” al PC a resolver problemas matemáticos y que todo funcione como se espera!

El módulo **NACAL** es una implementación literal de la notación empleada en el libro. Así se puede verificar que siguiendo las reglas de reescritura podemos lograr que el ordenador sea capaz de operar con vectores, matrices, resolver sistemas de ecuaciones, diagonalizar matrices, calcular determinantes, etc.

¹Por ejemplo, hay una nítida distinción entre vectores y matrices... aquí no encontrara ¡la “transpuesta” de ningún vector!

El módulo se instala ejecutando² `pip3 install nacal` desde un terminal. También puede descargar el código fuente y la documentación desde

<https://github.com/mbujosab/nacallib>

Puede usar el módulo, incluso sin instalar nada, si abre los Notebooks de **Jupyter** en su navegador desde:

<https://mybinder.org/v2/gh/mbujosab/nacal-Jupyter-Notebooks/master>

Otras características del libro

Todos los resultados están demostrados³. Muchas de las demostraciones son tan sencillas que están propuestas como ejercicios (frecuentemente se ofrecen pistas que indican los pasos necesarios en la demostración); consecuentemente dichas demostraciones aparecen en la sección de soluciones a los ejercicios⁴. El propósito es doble: por una parte se aligera el texto, y por otra permite al estudiante demostrar los resultados por su cuenta (sin tener la solución a la vista).

El curso gira en torno al *método de eliminación* y las *transformaciones elementales*. Hay numerosos libros de texto y manuales que siguen este modelo; por ejemplo Cullen (1972); Larson et al. (2004); Lay (2007); Poole (2004); Strang (2007, 2003), o el libro de problemas Arvesú Carballo et al. (2005). Pero este texto presenta una diferencia “operativa” respecto a los anteriores manuales. Por ejemplo, el Profesor Strang indica que “*el sistema de ecuaciones $\mathbf{Ax} = \mathbf{b}$ tiene solución si el vector \mathbf{b} es una combinación lineal de las columnas de \mathbf{A}* ”, pero aplica el método de eliminación gaussiana operando con las *filas*, por lo que realmente acaba resolviendo un sistema distinto aunque equivalente a $\mathbf{Ax} = \mathbf{b}$; ese es el modo de aplicar la eliminación en prácticamente todos los manuales de Álgebra Lineal. No obstante, aquí emplearemos el método de eliminación por columnas, operando en el espacio columna de \mathbf{A} para encontrar las combinaciones de las columnas que son iguales a \mathbf{b} . Es más, con un único algoritmo (la eliminación por columnas y sin necesidad de la sustitución hacia atrás) calcularemos casi todo (la inversa de una matriz, el conjunto de soluciones a un sistema de ecuaciones lineales, realizaremos operaciones con subespacios de \mathbb{R}^n , calcularemos determinantes, etc.). Tan solo al diagonalizar necesitaremos tanto la eliminación por filas como por columnas... ¡ambas transformaciones a la vez!.

Mi experiencia es que los estudiantes de economía no suelen tener una especial afinidad por las matemáticas. Y en aquellos casos en que se sienten a gusto con ellas, es porque conocen una serie de algoritmos que les dan seguridad, pues es reconfortante saber que si se siguen los pasos correctamente se llega a la solución de ciertos problemas. Apoyándome en esto, la exposición de este libro es fundamentalmente algorítmica.⁵ Tirando del hilo de los algoritmos irán surgiendo distintas estructuras conceptuales, cuya generalización permitirá reconsiderar lo ya visto, pero con un mayor grado de abstracción. No obstante, si uno analiza las demostraciones con perspectiva, se evidencia que de nuevo todo acaba reduciéndose a unas reglas precisas en la manipulación de unos símbolos (“*La notación lo es todo*”).

Pero limitarse a la parte algorítmica dejaría fuera cuestiones más abstractas que son importantes. Por ello he incluido al final de ciertas lecciones una sección de “Generalizaciones o visiones alternativas (★)”. Estas secciones se pueden omitir sin perder el hilo del libro (aunque espero que puedan servir como referencia cuando el alumno se encuentre con otros escenarios distintos a \mathbb{R}^n en futuras asignaturas).

Algunas lecciones disponen de un apéndice cuya lectura no se debe pasar por alto; aunque en algunos casos se puede posponer. Por ejemplo, los apéndices de las lecciones 3 y 4 contienen material prescindible para dichas lecciones, pero no para lecciones posteriores. Por el contrario, el apéndice de la Lección 6 contiene material imprescindible para esa misma la lección.

²Previamente debe tener instalado **Python 3** en su PC.

³Salvo algunas propiedades de los números reales y el *Teorema Fundamental del Álgebra* cuya demostración está fuera del alcance de este curso.

⁴Si está usando la versión en pdf, sepa que la numeración de los ejercicios es también un enlace que lleva a la solución; y que al final de cada solución hay un cuadrado que es un enlace al correspondiente enunciado.

⁵El desarrollo en paralelo del módulo NAcAL para Python me ha ayudado en este sentido.

Instrucciones para los alumnos

Antes y durante las clases en el aula. Este libro está dividido en lecciones. Para poder aprovechar adecuadamente las clases recibidas en el aula es IMPRESCINDIBLE leer con antelación la lección correspondiente. El profesor puede dar por supuesto que usted así lo ha hecho.

En clase se dará una exposición general de lo que aparece en este libro; los detalles de cada lección deben ser preparados por el alumno estudiando las secciones correspondientes.

Recuerde que usted debe haber leído la lección correspondiente antes de cada clase.

Para ayudar a tener una idea general de cada lección, puede encontrar un resumen por lección en el Apéndice A. (**está por hacer**). También hay unas transparencias que sirven de guión y de apoyo para el profesor; pero a usted le puede venir bien llevarlas a clase para centrarse principalmente en entender qué es lo que le explica su profesor, en lugar de centrarse en copiar la pizarra y tomar apuntes. Por otra parte, el tiempo dedicado a resolver problemas en clase será un buen momento para resolver dudas con su profesor.

Al final del libro encontrará las soluciones a los problemas, pero no debe mirarlas hasta que haya dado con “su” solución. Consultar la solución de otro sin haber resuelto el ejercicio por cuenta propia sirve de muy poco cuando se estudia. **Recuerde que el aprendizaje es una tarea activa, es decir, usted debe encontrar la solución activamente, y no consultar la solución de otro** (la mía en particular).

También me gustaría acompañar las lecciones con vídeos (**por hacer**)... aunque no creo que sea capaz de igualar las excelentes clases de Grant Sanderson de su canal [3blue1brown](#); clases que recomiendo que el alumno vea en paralelo al curso.

Al estudiar recuerde que el estudio individual (o en tándem) es imprescindible; dicho de otra forma, limitarse a atender las explicaciones en clase no es en modo alguno suficiente, ni limitarse a mirar las transparencias de clase tampoco. Es importante que estudie el contenido del libro. Si hay alguna parte que no se entiende, dígamelo e intentaré dicha parte para que se entienda mejor.

Resolución de ejercicios con el ordenador Además de resolver los ejercicios con lápiz y papel, también podrá usar el ordenador empleando el módulo **NACAL** de Python. Está pensado para ser usado con los Notebooks de Jupyter que puede encontrar en <https://github.com/mbujosab/nacal-Jupyter-Notebooks>.

Aprenderá mucho si experimenta con el ordenador. No se desanime si le cuesta un poco al principio.

Índice general

Prefacio	v
Índice general	ix
Introducción al libro	xvii
0.1. Motivación para escribir un libro <i>ligeramente diferente</i>	xvii
0.2. Algunos procedimientos no son los habituales	xviii
Usaremos tres métodos de eliminación (principalmente por columnas)	xix
La eliminación por columnas de izquierda a derecha es más directa	xix
Eliminación Gaussiana solo si es necesario	xx
Otra ventaja indirecta de la eliminación por columnas	xx
Transformaciones elementales como base de los algoritmos	xx
0.3. La notación propuesta es quizá la aportación más interesante	xxi
Evita ambigüedades habituales	xxi
Una notación operativa frente a notaciones descriptivas	xxiv
Notación clara, operativa y con semántica	xxiv
Esta notación respeta las definiciones y resuelve los citados problemas	xxviii
Implementación de la notación en Python (el módulo NAcAL)	xxix
 I Álgebra matricial	 1
1. Vectores y matrices (notación y operaciones)	3
1.1. Vectores	3
1.1.1. Definiciones de algunos vectores especiales	5
1.2. Sumas de vectores y productos de vectores por escalares	6
1.2.1. Propiedades	8
1.3. Vectores como representación de datos	10
1.4. Matrices	11
1.4.1. Filas de una matriz	13
1.4.2. Columnas de una matriz	13
1.4.3. Resumen de la notación para seleccionar componentes, filas y columnas	14
1.4.4. Extensión de la notación vectorial y reglas de reescritura.	15
1.4.5. La transposición (o trasposición)	16
1.4.6. Definiciones de algunas matrices especiales	17
1.5. Suma de matrices y producto de matrices por escalares	19
1.5.1. Propiedades	19
1.5.2. Operaciones componente a componente	20
1.5.3. La transposición es un operador lineal	20
1.5.4. Operaciones fila a fila	21
1.6. Extensión de la notación matricial y las reglas de reescritura	21
<i>Apendices a la lección</i>	22

1.A. Subvectores mediante selección de una lista de índices	22
1.B. Submatrices mediante selección de una lista de índices	22
2. Combinaciones lineales	25
2.1. Producto punto	25
2.2. Matriz por vector	26
2.2.1. Combinación lineal de vectores	26
2.2.2. Producto de una matriz por un vector	26
2.2.3. Propiedades del producto matriz por vector	28
2.3. Vector por matriz	29
2.3.1. Propiedades del producto de un vector por una matriz	30
3. Multiplicación matricial	31
3.1. Producto de matrices	31
3.1.1. Otras dos formas de calcular el producto de matrices	33
Cálculo del producto de matrices componente a componente (filas por columnas) . . .	33
Cálculo del producto de matrices operando con las filas	33
3.2. Nuevas reglas de reescritura.	34
Transpuesta de un producto	34
<i>Apendices a la lección</i>	35
3.A. Productos de matrices triangulares	35
3.B. Producto matricial como suma de productos de submatrices	35
3.B.1. Producto matricial expresado como una suma de matrices	38
3.B.2. Submatriz de un producto como suma de productos de submatrices	39
Matrices partidas en bloques	39
3.B.3. Matriz por vector como suma de submatrices por subvectores	40
3.C. Propiedades de la traza	40
II Transformaciones elementales, métodos de eliminación y matriz inversa	41
4. Transformaciones elementales y métodos de eliminación	43
4.1. Transformaciones y matrices elementales	43
4.1.1. Transformaciones elementales (y matrices elementales) de Tipo I	43
4.1.2. Transformaciones elementales (y matrices elementales) de Tipo II	45
4.1.3. Transformación de las columnas de una matriz	46
4.1.4. Sobre las matrices elementales y su notación	46
4.1.5. Secuencias de transformaciones elementales. Parte I	48
4.1.6. Intercambios	49
4.1.7. Permutaciones	50
4.2. Eliminación por columnas	51
4.2.1. Método de eliminación (o eliminación de “Izquierda a derecha”)	51
4.2.2. Método de eliminación Gaussiano	52
4.2.3. Método de eliminación Gauss-Jordan	53
<i>Apendices a la lección</i>	54
4.A. Transformaciones elementales de las filas	54
4.A.1. Matriz escalonada por filas	57
4.B. Transformación elemental “ <i>espejo</i> ” de otra transformación	57
5. Inversa de una matriz. Rango de una matriz	61
5.1. Matrices invertibles	61
5.1.1. Inversa de las matrices (y las transformaciones) elementales.	62
5.1.2. Matrices pre-escalonadas con inversa.	64

5.1.3.	Matrices invertibles	65
5.1.4.	Secuencias de transformaciones elementales. Parte II	68
5.1.5.	Inversa de una matriz triangular	69
5.2.	Rango de una matriz	71
5.2.1.	Algunas propiedades del rango de una matriz	72
<i>Apendices a la lección</i>		73
5.A.	“Espejo” de la inversa de una secuencia de transformaciones elementales	73
 III Subespacios y resolución de sistemas de ecuaciones lineales		75
 6. Espacios vectoriales y funciones lineales		77
6.1.	Espacios vectoriales	77
6.1.1.	Algunas propiedades	79
6.2.	Funciones lineales	80
6.2.1.	Ejemplos de funciones lineales que ya hemos usado	80
6.2.2.	Algunas propiedades de las funciones lineales	81
6.3.	Creación de nuevos espacios vectoriales	82
6.3.1.	Subespacios	82
	Intersección de subespacios	83
6.3.2.	Espacios vectoriales de productos cartesianos	83
6.3.3.	Espacios vectoriales de funciones	84
	Ejemplos de este tipo de espacios que probablemente encontrará en futuras asignaturas	85
6.3.4.	Subespacio de funciones lineales	85
6.4.	Generalizaciones o visiones alternativas (★)	85
6.4.1.	Subespacios y funciones lineales	85
<i>Apendices a la lección</i>		87
6.A.	Funciones	87
6.A.1.	Notación	88
6.A.2.	Invertibilidad	89
6.A.3.	Composición de funciones	89
6.A.4.	Los sistemas (listas ordenadas) son funciones	90
 7. Resolviendo $Ax = 0$		91
7.1.	Sistemas de ecuaciones lineales	91
7.1.1.	Notación matricial de los sistemas de ecuaciones lineales	92
7.2.	Sistemas de ecuaciones lineales homogéneos	92
7.3.	Espacio nulo de una matriz $\mathcal{N}(A)$	93
7.4.	Resolución de un Sistema Homogéneo por eliminación	94
 8. Resolviendo $Ax = b$		99
8.1.	Eliminación sobre la matriz ampliada	99
8.1.1.	El Teorema de de Rouché-Frobenius	103
8.2.	Espacio columna de una matriz	104
	El espacio columna y la eliminación	104
8.3.	Generalizaciones o visiones alternativas (★)	105
 9. Independencia, base y dimensión		107
9.1.	Combinaciones lineales de vectores de \mathcal{V}	107
9.1.1.	Extendiendo la notación matricial a los espacios vectoriales	107
	Producto de un sistema de n vectores de \mathcal{V} por un vector de \mathbb{R}^n	107
	Sistema de n vectores de \mathcal{V} por una matriz de $\mathbb{R}^{n \times p}$	108
9.2.	Sistemas generadores	109

9.3.	Transformaciones elementales sobre sistemas de vectores	111
	Sistemas equivalentes	111
9.4.	Sistemas linealmente dependientes y sistemas linealmente independientes	112
9.5.	Bases y dimensión	112
9.5.1.	Eliminación “de izquierda a derecha” y sistemas “acoplados” de vectores	114
	Encontrando dos bases de $\mathcal{C}(\mathbf{A})$ y una base de $\mathcal{N}(\mathbf{A})$	116
9.6.	Generalizaciones o visiones alternativas (★)	118
	Conjunto generador de un subespacio	118
	Sistemas de vectores y funciones lineales	118
	Sistemas linealmente independientes y coordenadas	120
10.	Los cuatro subespacios fundamentales de una matriz \mathbf{A}	121
10.1.	Suma de subespacios	121
10.1.1.	Suma directa y subespacios suplementarios	123
10.2.	Los cuatro subespacios fundamentales de una matriz \mathbf{A}	124
10.2.1.	El espacio fila	124
10.2.2.	El espacio nulo por la izquierda	125
10.2.3.	El gran esquema	125
10.3.	Encontrando bases para los espacios $\mathcal{C}(\mathbf{A}^\top)$ y $\mathcal{N}(\mathbf{A}^\top)$	127
	Encontrando una base para el espacio fila de \mathbf{A}	128
	Encontrando una base para el espacio nulo por la izquierda.	128
IV	Ortogonalidad	131
11.	Vectores ortogonales. Subespacios ortogonales	133
11.1.	Geometría	133
11.1.1.	Longitud de un vector en \mathbb{R}^2 y en \mathbb{R}^3	133
11.1.2.	Ángulo formado por dos vectores tanto en \mathbb{R}^2 como en \mathbb{R}^3	134
11.1.3.	Generalizando al espacio euclídeo \mathbb{R}^n	135
11.2.	Subespacios ortogonales en \mathbb{R}^n	138
11.2.1.	Ortogonalidad de los 4 subespacios fundamentales de \mathbf{A}	139
	El método de eliminación como generador de bases del complemento ortogonal	140
11.2.2.	De las ecuaciones paramétricas a las ecuaciones cartesianas	140
	El camino recorrido hasta ahora: de las ecuaciones cartesianas a las paramétricas	141
	Recorriendo el camino inverso: de las ecuaciones paramétricas a las cartesianas	141
11.2.3.	Puntos, rectas, planos e hiper-planos de \mathbb{R}^n . Espacios afines	142
11.3.	Generalizaciones o visiones alternativas (★)	143
11.3.1.	Productos escalares en \mathcal{V}	144
	Ejemplos de productos escalares en \mathbb{R}^n	144
	Ejemplos de productos escalares en otros subespacios	145
11.3.2.	Generalizando los conceptos geométricos a otros espacios vectoriales	145
11.3.3.	Subespacios ortogonales en espacios vectoriales abstractos	147
11.3.4.	Espacios pre-Hilbert	147
	Complementos ortogonales	147
	Un subespacio \mathcal{H} de dimensión infinita y el conjunto \mathcal{H}^\perp no siempre son suplementarios	149
12.	Proyecciones sobre subespacios	151
12.1.	Proyección ortogonal y mínima distancia	151
12.1.1.	Sistema de ecuaciones normales	152
12.1.2.	Alternativa a un sistema de ecuaciones lineales sin solución	152
12.2.	Expresión matricial de la proyección ortogonal	153

12.2.1. Matrices proyección	153
12.3. Generalizaciones o visiones alternativas (★)	154
Funciones idempotentes	154
Funciones autoadjuntas	154
12.3.1. Proyección ortogonal	155
Sobre algunos subespacios de dimensión infinita no existe la proyección ortogonal . . .	157
V Determinantes	159
13. Propiedades de los determinantes	161
13.1. Función determinante y función volumen	161
13.1.1. Tres propiedades de la función volumen de un paralelogramo	161
13.1.2. Las tres primeras propiedades (P-1 to P-3)	162
La relación entre la función <i>Volumen</i> y la función <i>Determinante</i>	162
13.2. Seis propiedades del determinante	163
13.2.1. Determinante de una matriz con una columna de ceros	163
13.2.2. Determinantes de matrices elementales	163
13.2.3. Sucesión de transformaciones elementales por columnas	164
Intercambio de columnas. Propiedad antisimétrica	164
13.2.4. Matriz singular, matriz inversa y producto de matrices	164
Determinante de matrices singulares y determinante de matrices inversas	164
Determinante del producto.	165
13.2.5. Determinante de la matriz transpuesta.	165
14. Fórmulas para el cálculo del determinante y aplicaciones	167
14.1. Determinante de matrices triangulares	167
14.2. Cálculo del determinante por eliminación Gaussiana	168
14.3. Determinante de matrices diagonales por bloques	168
14.4. Fórmulas sencillas para matrices de orden menor que 4	169
14.5. No hay fórmulas sencillas para matrices de orden mayor a 3	170
14.6. Expansión de Laplace (desarrollo por cofactores).	170
14.6.1. Propiedad multilineal	171
14.6.2. Menores y cofactores	171
Nueva notación y definición de menores y cofactores	172
14.6.3. Desarrollo del determinante por cofactores (Expansión de Laplace).	172
14.7. Aplicación de los determinantes	173
14.7.1. Regla de Cramer para la resolución de un sistema de ecuaciones	173
14.7.2. Cálculo de la inversa de una matriz	174
VI Autovalores y autovectores. Diagonalización y formas cuadráticas	177
15. Autovalores y autovectores	179
15.1. Cálculo de los autovalores y los autovectores	179
16. Diagonalización en bloques triangulares por semejanza	183
16.1. Matrices semejantes	183
16.2. Diagonalización por bloques triangulares	185
16.2.1. Autovalores, determinante y traza de matrices semejantes	191
16.2.2. Autovectores	191
16.3. Matrices diagonalizables	192

17. Diagonalización ortogonal de matrices simétricas	193
17.1. Método de Gram-Schmidt	193
17.2. Matrices ortogonales	194
17.3. Nota sobre la conjugación de números complejos	195
17.4. Diagonalización de matrices simétricas	195
18. Diagonalización por congruencia de matrices simétricas. Formas cuadráticas	199
18.1. Formas cuadráticas y matrices definidas positivas	199
18.1.1. Formas cuadráticas reales	199
18.1.2. Matrices definidas positivas	199
18.1.3. Algunos tipos de formas cuadráticas	201
18.2. Diagonalización de matrices simétricas por congruencia	201
18.2.1. Ley de inercia	204
18.3. Completar el cuadrado para clasificar las formas cuadráticas	205
<i>Apendices a la lección</i>	206
18.A. Factorización de Cholesky de una matriz definida positiva	206
18.B. Productos escalares en \mathbb{R}^n	208
 VII Estadística y econometría	
<i>“por hacer” (en estado muy preliminar)</i>	209
19. Visión geométrica de la Estadística	211
19.1. El producto escalar más habitual en estadística descriptiva	211
19.2. La media aritmética	212
La media aritmética, $\mu_{\mathbf{y}}$, es el producto escalar de \mathbf{y} con el vector constante $\mathbf{1}$	212
La media aritmética, $\mu_{\mathbf{y}}$, es el valor por el que hay que multiplicar el vector $\mathbf{1}$ para lograr <i>el vector constante más próximo a \mathbf{y}</i>	212
19.3. La desviación típica	213
La desviación típica $\sigma_{\mathbf{y}}$ es la longitud de la componente de \mathbf{y} ortogonal a $\mu_{\mathbf{y}}$	213
19.4. La varianza $\sigma_{\mathbf{y}}^2$ y el Teorema de Pitágoras	214
19.5. Covarianza y correlación	215
19.6. Regresión lineal o ajuste MCO	215
19.6.1. Tres casos sencillos	216
El vector constante $\mathbf{1}$ como único regresor	216
Regresión lineal simple y la recta de regresión	216
Regresión lineal con tres regresores (de ellos uno constante). El plano de regresión . .	218
19.6.2. Caso general	220
19.7. Interpretación estadística	221
19.7.1. El vector de medias como predictor	221
19.7.2. La desviación típica como medida del error de predicción de la media.	222
19.7.3. Regresión lineal como predictor	222
19.8. Medida de ajuste en la regresión lineal	222
<i>Apendices a la lección</i>	222
19.A. Otros productos escalares usados en estadística descriptiva	222
 A. Resumen de los temas por lecciones	225
A.1. Resumen del Tema 1	225
 Soluciones a los ejercicios	231
Soluciones a los Ejercicios	231
 Bibliografía	267

Glosario	269
Símbolos	271
Índice alfabético	273
Índice de figuras	277
Índice de cuadros	279

0.1. Motivación para escribir un libro *ligeramente diferente*

(QWERTY): La tradición no siempre elige el camino que más conviene La tradición tiene un enorme peso en las actividades académicas (tanto docentes como de investigación). Quizá se deba al origen eclesiástico y medieval de la universidad. Sea como fuere, el hecho es que habitualmente reproducimos lo recibido por nuestros maestros sin cuestionarlo. Esta conducta colectiva pone en pié de igualdad el “cómo” con el “qué”. Sin embargo, “qué son las cosas” y “cómo se exponen” no son lo mismo.

Si un autor de prestigio acuña una determinada terminología, notación o forma de exposición, lo habitual es que quienes venimos detrás la reproduzcamos fielmente... sin cuestionarnos seriamente si es la más correcta o la que más nos conviene. Quienes usamos las matemáticas (sin ser matemáticos) frecuentemente olvidamos que son un lenguaje en el que se han establecido ciertos convenios, y que dichos convenios son solo eso: “convenios”. De ahí que olvidemos que está dentro de “las reglas del juego” el poder introducir aquellos cambios que nos resulten convenientes (de manera similar a cuando programamos en Matlab algunas funciones a nuestra conveniencia, aunque ya existan funciones parecidas en las librerías originales de Matlab).

Hay algunos hábitos en los manuales de Álgebra Lineal que, o bien no considero completamente correctos, o bien no ayudan a la exposición. Darse cuenta de ello no es inmediato. Si he llegado a esa conclusión ha sido gracias a mi experiencia como docente, a discutirlo con otros profesores y al esfuerzo realizado programando el módulo **NAcAL**. Al intentar implementar tanto la notación como los objetos y los algoritmos del curso, me he encontrado tanto con “atajos” como con dificultades inesperadas que me han permitido tomar conciencia de qué cambios conviene introducir respecto a los “usos de la tradición”.

Empecemos por aquello que no considero correcto. En Álgebra Lineal números, vectores y matrices son objetos distintos, pero es habitual emplear una notación que no los distingue claramente (véase la Sección 0.3). Quizá algunos consideren que introducir explícitamente esa distinción incorpora una dificultad añadida. Quien comparta dicha visión debería leer la siguiente cita de Godement (1968) donde se sugiere justo lo contrario:

Many people— in particular the majority of those whose attitude to mathematics is purely utilitarian— are of the opinion that, in textbooks designed to be read by beginners, it is useless or even dangerous to put great emphasis on rigour, to prove every statement, to introduce notions of great generality or to use carefully and strictly defined terminology. If this view were correct it would imply that, contrary to all professional mathematicians and common sense, the worst-written textbooks would be those most easily understood by the beginner (Godement, 1968).

Coincido completamente con la cita del Profesor Godement y considero que la notación debe ser clara y precisa. Solo de ese modo el alumno puede adquirir confianza en lo que hace. La notación que propongo es consecuente con el hecho de que escalares, vectores y matrices son objetos diferentes... lo opuesto a **Matlab** (MATrix LABoratory)⁶ donde todo son matrices.

⁶O su versión de código abierto **GNU Octave**.

Y continuemos con otros aspectos de la “tradición” que considero que no ayudan.

1. Resolver sistemas de ecuaciones mediante eliminación *por filas* requiere un paso adicional: “la sustitución hacia atrás” (innecesario con la eliminación *por columnas*). No solo eso; además resulta que el razonamiento tras el algoritmo de resolución de sistemas de ecuaciones (homogéneos o no-homogéneos) y tras el cálculo de la inversa de una matriz, es el mismo cuando empleamos la eliminación *por columnas*; pero no si empleamos la eliminación *por filas*.
2. Usar eliminación Gaussiana (que escalona una matriz) o la eliminación Gauss-Jordan (que además reduce la matriz) entraña más pasos de los necesarios en la mayoría de las aplicaciones, y oculta relaciones de dependencia lineal entre las columnas (o las filas) de una matriz y las subsiguientes matrices transformadas. En muchos casos basta aplicar la eliminación de izquierda a derecha (si es por columnas) o de arriba a abajo (si es por filas). En resumen, intercambiar la posición de vectores, o generar más ceros de los necesarios, no añade nada en la mayoría de aplicaciones, pero oscurece relaciones importantes.
3. Definir el intercambio de vectores como “una” transformación elemental no ayuda. Es mejor describirlo como una sucesión de transformaciones elementales.
4. Usar la misma notación para una transformación (o una matriz) intercambio y una (matriz) permutación no ayuda.
5. Describir las transformaciones elementales con productos de matrices elementales oscurece la notación innecesariamente, pues requiere tener en cuenta el orden de las matrices.
6. No usar la congruencia para clasificar formas cuadráticas introduce dificultades; y hacerlo únicamente por relaciones de semejanza despista al alumno, que puede pensar que ese es un método general cuando no lo es (pues requiere conocer los autovalores previamente). Clasificar formas cuadráticas siguiendo el Criterio de Sylvester es complicado, poco operativo y nada intuitivo.

Creo que estos ejemplos bastan para ilustrar que la tradición no siempre elige el camino más conveniente. Pero sin duda los dos primeros ejemplos son los más importantes (pues están *absolutamente extendidos*).

0.2. Procedimientos del libro que se alejan de lo habitual

El algoritmo más usado en la mayoría de manuales de álgebra lineal es el método de Gauss. Por tradición todos los manuales aplican la eliminación por filas y tan solo emplean dos variantes: la eliminación Gaussiana (que escalona una matriz) y la eliminación Gauss-Jordan (que escalona y reduce, es decir, convierte los pivotes en “unos” con ceros por encima y por debajo, como en el caso de la matriz identidad).

Con frecuencia aplicar estos algoritmos supone realizar más pasos de los necesarios. Pensemos en el siguiente sistema de ecuaciones lineales:

$$\begin{cases} 2y + 2z + w = 0 \\ 2x + 2y - 2w = 2 \end{cases}$$

El método tradicional de resolución trabaja en el espacio dual del espacio columna de la matriz de coeficientes y requiere emplear dos pasos:

a) Escalonar y reducir la matriz ampliada mediante eliminación por filas hasta obtener un sistema de ecuaciones triangular pero “equivalente”:⁷

$$\left[\begin{array}{cccc|c} 0 & 2 & 2 & 6 & 0 \\ 2 & 2 & 0 & -2 & 2 \end{array} \right] \xrightarrow{[(-1)\tau_1 + 2]} \left[\begin{array}{cccc|c} 0 & 2 & 2 & 6 & 0 \\ 2 & 0 & -2 & -8 & 2 \end{array} \right] \xrightarrow{[\tau_1 \div 2]} \left[\begin{array}{cccc|c} 2 & 0 & -2 & -8 & 2 \\ 0 & 2 & 2 & 6 & 0 \end{array} \right] \xrightarrow{\left[\begin{smallmatrix} \tau_1 \\ (\frac{1}{2})2 \\ ((\frac{1}{2})1 \end{smallmatrix} \right]} \left[\begin{array}{cccc|c} 1 & 0 & -1 & -4 & 1 \\ 0 & 1 & 1 & 3 & 0 \end{array} \right];$$

⁷Es decir, otro sistema de ecuaciones que es distinto pero con el mismo conjunto solución.

identificar las variables pivote (o endógenas) y las variables libres (o exógenas), y despejar las variables endógenas para obtener el siguiente sistema mucho más simple:

$$\begin{cases} x &= 1 + z + 4w \\ y &= -z - 3w \end{cases}.$$

b) Paso de “sustitución hacia atrás”, es decir, dar valores a las variables exógenas para obtener una solución particular y una base para las soluciones del sistema homogéneo Primero se asigna el valor cero a las variables exógenas ($z = 0, w = 0$) y se deduce el valor de las endógenas para obtener una solución particular $\mathbf{x}_p = (1, 0, 0, 0)$.

Después, iterativamente se asigna el valor 0 a todas las variables exógenas excepto una, a la que se le asigna el valor 1; y se deduce el valor de las endógenas para obtener una solución del sistema homogéneo. En el ejemplo, primero se asignan los valores ($z = 1, w = 0$) para obtener el vector $\mathbf{x}_a = (1, -1, 1, 0)$ y luego se asignan los valores ($z = 0, w = 1$) para obtener el vector $\mathbf{x}_b = (4, -3, 0, 1)$. Consecuentemente el conjunto solución es

$$\left\{ \mathbf{v} \in \mathbb{R}^4 \mid \exists a, b \in \mathbb{R}, \mathbf{v} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + a \begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix} + b \begin{pmatrix} 4 \\ -3 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

Usaremos tres métodos de eliminación (principalmente por columnas)

La eliminación por columnas de izquierda a derecha es más directa

El paso de “sustitución hacia atrás” es innecesario con la eliminación por columnas. Basta “hacer ceros” a la derecha del primer elemento no nulo de cada fila empezando por la primera (pero apilando primero la matriz ampliada⁸ sobre la matriz identidad):

$$\left[\begin{array}{cccc|c} 0 & 2 & 2 & 6 & 0 \\ 2 & 2 & 0 & -2 & -2 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array} \right] \xrightarrow{\begin{matrix} \tau \\ [(-1)2+3] \\ [(-3)2+4] \end{matrix}} \left[\begin{array}{cccc|c} 0 & 2 & 0 & 0 & 0 \\ 2 & 2 & -2 & -8 & -2 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & -3 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array} \right] \xrightarrow{\begin{matrix} [(-1)1+2] \\ [(1)1+3] \\ [(4)1+4] \\ [(1)1+5] \end{matrix}} \left[\begin{array}{cccc|c} 0 & 2 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 \\ 1 & -1 & 1 & 4 & 1 \\ 0 & 1 & -1 & -3 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array} \right]$$

La solución particular aparece en la última columna nula (vector vertical entre las líneas horizontales), y una base del subespacio de soluciones del sistema homogéneo aparecen bajo las columnas nulas de la parte superior (vectores verticales entre las líneas horizontales de las columnas 3 y 4). Aplicar la eliminación de izquierda a derecha para hacer ceros a la derecha de los pivotes es el único paso necesario (a esto lo llamo “pre-escalonar” la matriz).⁹ No es necesario escalonar la matriz. Basta con hacer ceros a la derecha de los pivotes de la submatriz superior.

Lo ventajoso desde un punto de vista didáctico no es haber evitado el paso de sustitución hacia atrás. Lo más ventajoso es que la explicación de por qué funciona es EXACTAMENTE LA MISMA que la que se emplea para invertir una matriz o para resolver un sistema de ecuaciones homogéneo. Es decir, una vez que el alumno entiende el procedimiento para invertir una matriz, ya puede entender todo lo demás.

No solo eso, el hecho de que la eliminación solo emplee pasos de izquierda a derecha implica que la matriz que aparece por debajo es triangular superior y que los subespacios generados por las j primeras

⁸Al operar por columnas, la última columna de la matriz ampliada es el opuesto del vector del lado derecho.

⁹Una matriz está *pre-escalonada* si las componentes a la derecha de cada pivote son nulas; y está *escalonada* si además los pivotes de sus columnas describen una escalera descendente, dejando todas las columnas nulas (si hay) a la derecha de las columnas no nulas.

columnas de la matriz de coeficientes generan el mismo subespacio que las j primera columnas de la *matriz pre-escalada*, para $j = 1 : k$ donde k es el número de columnas. Estos dos hechos permiten explicar otras cuestiones teóricas al margen del mero objetivo de resolver un sistema de ecuaciones.

Eliminación Gaussiana solo cuando es necesario

La eliminación *Gaussiana* y la eliminación *Gauss-Jordan* son variantes que añaden pasos adicionales a la eliminación *de izquierda a derecha*, pero el paso fundamental es pre-escalonar (precisamente el que no se destaca en los manuales de Álgebra Lineal). La eliminación Gaussiana o la eliminación Gauss-Jordan solo son necesarias para resolver algunas cuestiones muy puntuales:

1. Para calcular un determinante por eliminación es necesario escalar; por tanto es necesario aplicar la eliminación Gaussiana.
2. Para invertir una matriz es necesario escalar y reducir; es decir, es necesario aplicar la eliminación Gauss-Jordan.

Además, al diagonalizar es necesario combinar la eliminación por filas y por columnas:

- La diagonalización por semejanza emplea la eliminación Gaussiana por filas y columnas.
- La diagonalización de una matriz simétrica por congruencia requiere escalar por filas y columnas con un especial cuidado si aparecen ceros en la diagonal.

Esta disección permite al alumno tener una idea clara de los procedimientos usados en el curso y por qué son necesarios ciertos pasos en algunos problemas y no en otros.

Otra ventaja indirecta de la eliminación por columnas

Los alumnos llegan a la universidad con las (mal aprendidas) recetas “del cole”. Generalmente no entienden su funcionamiento y con frecuencia tampoco las saben aplicar correctamente (salvo en problemas triviales).

Desde un punto de vista didáctico, encuentro muy ventajoso usar la eliminación por columnas toda vez que el razonamiento empleado es nuevo para el alumno... ¡y completamente algebraico!¹⁰

$$\text{por una parte } \begin{bmatrix} \mathbf{A} \\ \mathbf{I} \end{bmatrix}_{\tau_1 \dots \tau_k} = \begin{bmatrix} \mathbf{A}_{\tau_1 \dots \tau_k} \\ \mathbf{I}_{\tau_1 \dots \tau_k} \end{bmatrix} \quad \text{y por otra} \quad (\mathbf{A}_{\tau_1 \dots \tau_k})_{|j} = \mathbf{A}(\mathbf{I}_{\tau_1 \dots \tau_k})_{|j}.$$

De este modo el alumno no puede acudir al uso de las “mal aprendidas recetas del cole”. La nueva vía les exige (y por tanto les permite) entender por qué funciona lo que van viendo a lo largo del curso (léase la advertencia al comienzo de la Lección 7).

Por añadidura, la simplicidad del algoritmo de eliminación por columnas permite deducir resultados teóricos más abstractos de manera sencilla (por ejemplo en la Sección 9.5.1).

Transformaciones elementales como base de los algoritmos

En este curso doy un peso importante a las transformaciones elementales. Ello permite diseccionar el método de eliminación (como ya he indicado más arriba). Pero no solo eso. Al estudiar la relación entre las transformaciones elementales de las filas y las transformaciones elementales de las columnas, doy cabida a la exposición de dos algoritmos que no aparecen en los manuales habituales para cursos de Álgebra Lineal; y esto pese a que el segundo resulta de enorme “utilidad práctica” para clasificar matrices simétricas de cualquier orden (pues no requiere conocer los autovalores de la matriz).

La demostración de que *toda matriz cuadrada es diagonalizable por bloques triangulares* describe un algoritmo basado en transformaciones elementales que logra obtener dicho resultado. Al igual que con otros métodos de diagonalización por semejanza, es necesario conocer los autovalores de la matriz. Puesto que cuando la matriz es diagonalizable el algoritmo la diagonaliza, es una generalización del habitual resultado de diagonalización presentado en los cursos de Álgebra Lineal que permite trabajar con las potencias de una

¹⁰Para una explicación de la notación véase la Sección 0.3.

matriz sin necesidad de calcular la forma canónica de Jordan (que está fuera del programa). Del algoritmo se deducen resultados teóricos muy conocidos. Por ejemplo, que la suma de los autovalores es igual a la traza y el producto igual al determinante, o que la matriz es diagonalizable si y solo si la multiplicidad geométrica y algebraica de los autovalores coincide. Es decir, expongo un algoritmo más general y del que se deducen múltiples resultados teóricos. Como con el resto de algoritmos del curso, también éste está implementado en el módulo de Python **NAcAL**; algo muy práctico, pues el cálculo con lápiz y papel resulta pesado.

El segundo algoritmo no habitual y basado en transformaciones elementales es *la diagonalización por congruencia* (que expongo junto con la Ley de Inercia de Sylvester). Es sorprendente que los manuales insistan en clasificar formas cuadráticas mediante el cálculo de los autovalores (método que en general no es aplicable para matrices de orden mayor a 4)¹¹ o aplicando el criterio de Sylvester (que solo es aplicable para matrices definidas positivas y cuya extensión a matrices semidefinidas requiere un cálculo excesivo).

Sorprende que la mayoría de los manuales no mencionen la diagonalización por congruencia que, mediante la eliminación gausiana, muestra los signos de todos los autovalores. ¡Todo un contrasentido para aquellos manuales que presentan el Álgebra Lineal desde un punto de vista “instrumental”! (supongo que de nuevo se debe al peso de la tradición).

0.3. La notación propuesta es quizá la aportación más interesante

Notation is everything.

CHARLES F. VAN LOAN, FFTs and the Sparse Factorization Idea (1992)

A good notation should be unambiguous, pregnant, easy to remember: it should avoid harmful second meanings, and take advantage of useful second meanings; the order and connection of signs should suggest the order and connection of things.

GEORGE POLYA, How to Solve It (1957)

Estas dos citas señalan la importancia de la notación en matemáticas y cómo debe ser para ser buena. Al inicio de este preámbulo he dedicado algunos párrafos a comentar la lúcida y provocativa afirmación de Charles F. Van Loan. En esta sección me detendré en la segunda cita.

Lo primero que señalaré es que los manuales de álgebra lineal (y econometría) han extendido una notación que no cumple ni la primera de las indicaciones de George Polya, ni tampoco la última (pues con “*connection of things*” entiendo que George Polya se refiere a la conexión entre *objetos matemáticos*).

También incidiré en algo no señalado por George Polya pero que creo importante para una buena notación: *su uso debe ser similar al empleo de un lenguaje de programación*; es decir, debe ser “operativa”. Con “operativa” no me refiero a que deba describir el modo de calcular, sino que su uso permita operar fácilmente con símbolos para deducir resultados.

Se debe evitar las ambigüedades e incoherencias habituales

En Álgebra Lineal se emplean habitualmente operaciones entre tres objetos distintos: los *escalares*, los *vectores* de \mathbb{R}^n y las *matrices*. Por ejemplo: el producto de un *vector* por un *escalar* (que arroja como resultado un *vector*), el producto punto¹² entre dos *vectores* (que arroja como resultado un *escalar*), el producto de una *matriz* por un *vector*, (que es un *vector*), o el producto entre *matrices* (que arroja como resultado otra *matriz*). Es decir, el lenguaje parece indicar que hay productos entre tres objetos matemáticos distintos. Si no fuera así, ¿a qué viene tanto nombre para una misma operación producto?

Por otra parte (y como veremos a continuación) muchas operaciones del Álgebra Lineal parecen lograrse mediante productos matriciales. Esto da pie a simplificar la notación, pues el producto matricial es asociativo.¹³ Pero la simplificación debe hacerse con cuidado para evitar ambigüedades, ya que el *producto de dos*

¹¹Pues no existe un modo general de encontrar las raíces de polinomios de grado 5 o más.

¹²Es decir, el producto escalar usual en \mathbb{R}^n .

¹³Es decir, como $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$, los paréntesis son innecesarios y podemos escribir sin ambigüedad \mathbf{ABC} .

matrices siempre es una matriz. Para ilustrarlo indicaré algunas operaciones que se pueden lograr mediante un producto de matrices.

Ejemplo 1: Selección de componentes. Si tenemos una matriz \mathbf{A} y la multiplicamos por su derecha por otra matriz cuya única columna es como la columna j -ésima de la matriz identidad, logramos una *matriz* cuya única columna es la j -ésima de \mathbf{A} .

Si además multiplicamos por la izquierda por una matriz cuya única fila es la i -ésima fila de la matriz identidad, seleccionamos la i -ésima componente de la j -ésima columna de \mathbf{A} , es decir, obtenemos la *matriz* cuya única componente es a_{ij} :

$$\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}; \quad \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = [3]. \quad (1)$$

Ejemplo 2: Eliminación Gaussiana. Tratemos de “escalonar” la matriz \mathbf{A} del ejemplo anterior... Lo usual es operar con las filas para generar ceros por debajo de la primera componente no nula de la primera fila. Por tanto es necesario restar el doble de la primera fila a la segunda. Esta operación se puede realizar aplicando la misma operación sobre la matriz identidad para obtener lo que se llama una matriz elemental; y luego multiplicar la matriz elemental por la matriz \mathbf{A} :

$$\begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 0 & -2 \end{bmatrix}. \quad (2)$$

Ejemplo 3: Producto punto de vectores de \mathbb{R}^n . Si tenemos dos vectores de \mathbb{R}^n y multiplicamos la matriz cuya única fila es el primero de los vectores con otra matriz cuya única columna es el segundo de los vectores, logramos una *matriz* cuya única componente es el producto punto de ambos *vectores*.

$$\begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = [3]. \quad (3)$$

Estos tres ejemplos son suficientes para entender la filosofía de MatLab (Matrix Laboratory): si consideramos que todo son matrices, gran cantidad de operaciones se pueden realizar usando únicamente *el producto matricial* que, por ser asociativo, no requiere prestar atención al orden en que se realizan las operaciones.

Pero los nombres de las operaciones en Álgebra Lineal parecen sugerir que hay distintos objetos matemáticos. ¿Es porque el lenguaje en Álgebra Lineal es artificialmente recargado y poco preciso? No. La realidad es que hay objetos distintos y que la anterior impresión es fruto de una descuidada (y muy extendida) simplificación. Para mayor confusión, los ejemplos anteriores dan lugar a nombres como *vector fila* y *vector columna*... Pero... si hay dos tipos de vectores ¿los vectores de \mathbb{R}^n son filas o son columnas? ¿Por qué no se indica algo como \mathbb{R}_{filas}^n ?... Sencillamente porque no hay tal distinción.

Un vector de \mathbb{R}^n es una lista ordenada de números; escribir dicha lista en horizontal o vertical no cambia la lista. Así que cuando se escribe $\mathbf{y}^T \mathbf{A}$ es que en realidad el símbolo “ \mathbf{y} ” representa una *matriz*, aunque el libro lo llame “*vector fila*”.

El resultado numérico de la Ecuación 1 es una matriz; pero como Matlab no muestra los corchetes alrededor de las matrices, visualmente parece que obtenemos un número. Consecuentemente, muchos manuales no distinguen entre número, vector de \mathbb{R}^1 o matriz de orden 1.

Gilbert Strang indica en la página 1 de su manual que *un par de números es un vector de \mathbb{R}^2 que se escribe en forma de columna*. En la página 4 nombra los vectores fila e indica que un vector fila es completamente diferente de un vector columna incluso cuando tiene las mismas componentes, ya que es la

transpuesta de un vector columna. Pero... ¿escribir el par de números en vertical cambia el par?... no se entiende. En la página 10 muestra un primer ejemplo de producto punto o producto interior

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

y en la página 16 se indica que es más frecuente escribir dicho producto como

$$\begin{bmatrix} 4 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix};$$

cuyo resultado debe ser la matriz $[0]$ (pues *matriz por matriz es otra matriz*). Por tanto, no distingue entre números y matrices de orden 1. ¿Qué sentido tiene hablar de números, vectores y matrices si todo son matrices?

Voy a poner otro ejemplo de un manual distinto que comparte la misma notación y, por tanto, los mismos problemas. Tras indicar esencialmente lo mismo que G. Strang, en la página 133 de Poole (2004) se define una matriz como *un arreglo rectangular de números...* pero si los números son lo mismo que las matrices de orden 1, esta definición es equivalente a que *una matriz es un arreglo rectangular de matrices 1 por 1*. De nuevo no se entiende (¿el objeto definido es parte de su propia definición?). Que las secciones siguientes se dediquen respectivamente a la multiplicación por un escalar y a la multiplicación de matrices no ayuda. ¿No es suficiente hablar de producto de matrices? ¿A qué viene tanto nombre si lo que se está multiplicado en cualquier caso son matrices?

En Poole (2004) se denotan las matrices con las letras $A, B, C \dots$. En algunas páginas A_1, A_2, A_3 denotan tres matrices distintas (el subíndice es parte del nombre), pero I_n es la matriz identidad de orden n (aquí el subíndice indica el orden de la matriz). En algunas páginas, con $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3 \dots$ se denota a tres vectores distintos, pero en otras con $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ se denota a las columnas 1, 2 y 3 de la matriz B . Las filas de A se denotan con $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$. Este batiburrillo de fuentes aparece en la demostración de $A(B+C) = AB + AC$:

$$\begin{aligned} [A(B+C)]_{ij} &= \mathbf{A}_i \cdot (\mathbf{b}_j + \mathbf{c}_j) \\ &= \mathbf{A}_i \cdot \mathbf{b}_j + \mathbf{A}_i \cdot \mathbf{c}_j \\ &= (AB)_{ij} + (AC)_{ij} \\ &= (AB + AC)_{ij} \end{aligned} \tag{4}$$

La demostración se apoya en *el modo de calcular* la componente ij del AB (que es el producto escalar entre la fila i -ésima de A y la columna j -ésima de B) pero acaba usando demasiadas fuentes para un resultado que atañe exclusivamente a las matrices A, B y C .

Casi idéntica notación es usada en Strang (2003), pero en ocasiones este manual también denota la fila i -ésima de B con \mathbf{b}_i . Así, en la página 50 se indica una forma particular para el cálculo del producto matricial:

$$AB = \left[\begin{array}{ccc|c} | & & & | \\ \mathbf{a}_1 & \cdots & & \mathbf{a}_n \\ | & & & | \end{array} \right] \left[\begin{array}{cc} - & \mathbf{b}_1 & - \\ & \vdots & \\ - & \mathbf{b}_n & - \end{array} \right] = \left[\mathbf{a}_1 \mathbf{b}_1 + \cdots + \mathbf{a}_n \mathbf{b}_n \right]; \tag{5}$$

donde \mathbf{a}_j es una columna pero \mathbf{b}_j es una fila. Para lidiar con esta ambigüedad se emplea un recurso visual (los guiones verticales y horizontales) que indican al lector el *modo usado para calcular* (“columnas por filas”). Este recurso es incómodo para explicar la factorización LU, así que el Prof. Strang presenta la siguiente ecuación en la página 85:

$$(\text{Row 3 of } A) = l_{31}(\text{Row 3 of } U) + l_{32}(\text{Row 2 of } U) + l_{33}(\text{Row 3 of } U). \tag{6}$$

Una notación operativa frente a notaciones descriptivas

Los ejemplos anteriores muestran que hay una tendencia a emplear notaciones que son visuales o que describen un *modo de calcular*. Éstas no forman parte de las características indicadas por George Polya¹⁴.

La notación descriptiva dificulta la operatividad. Por ejemplo, en la página 49 de Larson et al. (2004) encontramos la descripción más frecuente del producto matricial. Es una descripción que se centra en una forma de calcular el producto pero que no ayuda a interpretar el resultado:

Si $A = [a_{ij}]$ es una matriz $m \times n$ y $B = [b_{ij}]$ es una matriz $n \times p$, el producto AB es la matriz $m \times p$

$$AB = [c_{ij}]$$

donde

$$c_{ij} = \sum_{k=1}^n a_{ik}b_{kj} = a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j} + \cdots + a_{in}b_{nj}. \quad (7)$$

Esta notación no destaca que se está calculando el producto punto entre una fila de A y una columna de B . Larson et al. dejan para el lector la demostración del conocido resultado $(AB)^T = B^T A^T$; pero tan escasos recursos de notación obligan a realizar la siguiente demostración:

Por una parte,

$$(AB)_{ij}^T = (AB)_{ji} = \sum_{k=1}^n a_{jk}b_{ki}, \quad (8)$$

y por otra

$$(B^T A^T)_{ij} = b_{1i}a_{j1} + b_{2i}a_{j2} + \cdots + b_{ni}a_{jn} = \sum_{k=1}^n b_{ki}a_{jk} = \sum_{k=1}^n a_{jk}b_{ki}. \quad (9)$$

(la segunda parte requiere pensar qué elementos de A y de B es necesario emplear para calcular el elemento ij del producto cuando las matrices están transpuestas; algo que no ayuda ni a la escritura ni a la lectura de la demostración).

Considero que estos ejemplos son suficientes para ilustrar que la notación habitual es confusa y no ayuda a operar. Tras 15 años impartiendo Matemáticas II sé que esto supone una barrera importante para el estudiante.

Usar una notación alternativa no sale gratis Quizá la contribución más interesante del libro es la notación. Creo haber fijado una notación precisa, clara y fácil de manejar. Cumple las indicaciones de George Polya y además es *operativa*; tan operativa que la he implementado en el módulo **NAcAL**. Un módulo para Python que constituye un cuasi-lenguaje de programación.

La adopción de una notación distinta (junto a la decisión de priorizar la eliminación por columnas) viene acompañada de otro coste de naturaleza muy distinta: no poder apoyar mis clases en los manuales de texto habituales. La firme convicción de que los alumnos deben disponer de un texto de referencia me empujó a escribir este libro. Por otro lado, ignorar la notación dominante es un salto audaz que me ha forzado a implementar la notación como módulo de Python para que no pueda caber duda alguna de que la propuesta funciona y que es útil y operativa (el funcionamiento del módulo es la mejor demostración). En suma, cambiar la notación ha traído aparejada una cantidad de trabajo muy considerable.

Notación clara, operativa y con semántica

Veamos algunas de las características de la notación que propongo y que permite evitar las dificultades indicadas en los ejemplos de más arriba.

¹⁴Este tipo de notación es muy frecuente en estadística y probabilidad. Por ejemplo, suele expresarse el valor esperado o la varianza con un sumatorio o una integral, subrayando así un modo de calcular la esperanza o la varianza, en lugar de sugerir su significado.

La notación es consecuente con la definición de los objetos matemáticos. Los números, los vectores (que son listas de números) y las matrices (listas de vectores) son objetos distintos y, por tanto, deben ser diferenciados por la notación. A tal efecto, los vectores \mathbf{a} , \mathbf{b} , $\mathbf{c} \dots$ se escriben encerrando la lista de números, a , b , $c \dots$ entre paréntesis:

$$\mathbf{a} = (1, 2, 3,) = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

Cuando el vector está escrito en horizontal, detrás de cada elemento de la lista se escribe una coma. De este modo podemos distinguir el número dos entre paréntesis: (2) del vector de \mathbb{R}^1 cuya lista solo contiene el número dos: $(2,)$.

Las matrices \mathbf{A} , \mathbf{B} , $\mathbf{C} \dots$ se encierran entre corchetes:

$$\mathbf{A} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \neq \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \mathbf{a}.$$

Una lista de vectores de \mathbb{R}^m entre corchetes denota la matriz cuyas columnas son dichos vectores. No existen vectores fila o vectores columna. Los vectores son listas ordenadas de números (da igual que los escribamos vertical u horizontalmente puesto que la lista no cambia). Consecuentemente no tiene sentido transponer un vector. Por tanto, el producto punto entre dos vectores se escribe así:

$$\mathbf{a} \cdot \mathbf{b},$$

y el producto del vector \mathbf{a} con la matriz \mathbf{B} así:

$$\mathbf{aB}.$$

Cuando nos referimos a una fila o una columna, siempre es fila o columna *de una matriz*.

$$\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \rightarrow \text{fila}_2(\mathbf{A}) = (2, 4,).$$

Podemos hablar de *matrices fila* (matrices con una única fila) o *matrices columna* (matrices con una columna); pero las matrices no son vectores (son listas de vectores).

La notación usa dos símbolos especiales como operadores. Estos símbolos son:

- el operador “ $|$ ” que selecciona componentes,
- el operador “ τ ” que realiza transformaciones elementales.

Junto al operador selector “ $|$ ” *es necesario indicar* el índice (o índices) del elemento (o elementos) seleccionado(s). Bajo el operador “ τ ”, y entre corchetes, *se puede* describir la transformación aplicada del modo siguiente:

$\tau_{[(\lambda)i+j]}$ Suma λ veces el elemento i -ésimo al j -ésimo (solo cambia el elemento j -ésimo)

$\tau_{[(\lambda)j]}$ Multiplica por λ el j -ésimo elemento (solo cambia el elemento j -ésimo)

$\tau_{[i \rightleftharpoons j]}$ intercambia de posición los elementos i -ésimo y j -ésimo

Con los vectores es irrelevante si los operadores actúan por la derecha o por la izquierda de un vector; el resultado es idéntico:

- Selección de la componente i -ésima: ${}_i|\mathbf{a} = \mathbf{a}|_i = a_i$ (es un número)

- Transformación elemental de componentes: $\boxed{\tau \mathbf{a} = \mathbf{a}_\tau}$ (es un vector)

Por ejemplo, para $\mathbf{a} = (2, 5, 0)$ tenemos que: ${}_1|\mathbf{a} = 2$ y $\mathbf{a}_{|3} = 0$.

Así pues, usando “ $|$ ” evitamos una de las ambigüedades anteriores, ya que con

$$\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3$$

se denota una serie de tres *vectores* y con

$$\mathbf{a}_{|1} \quad \mathbf{a}_{|2} \quad \mathbf{a}_{|3}$$

se denota a los componentes a_1 , a_2 y a_3 del vector \mathbf{a} . De esta manera podemos denotar sin ambigüedad la componente k -ésima del vector \mathbf{a}_h con $\mathbf{a}_{h|k}$.

Con matrices es diferente operar por un lado u otro. Por la izquierda de una matriz actuamos sobre la filas y por la derecha sobre las columnas:

- $\boxed{\mathbf{A}_{|j}}$ Columna j -ésima de una matriz (es un vector).
- $\boxed{{}_i|\mathbf{A}}$ Fila i -ésima de una matriz (es un vector).
- $\boxed{{}_i|\mathbf{A}_{|j} = a_{ij}}$ Elemento (i, j) -ésimo (es un número); puesto que ${}_i|(\mathbf{A}_{|j}) = ({}_i|\mathbf{A})_{|j}$.
- $\boxed{\mathbf{A}_{\tau_1 \cdots \tau_k}}$ Secuencia de transformaciones sobre las columnas de \mathbf{A} (es una matriz).
- $\boxed{\tau_1 \cdots \tau_k \mathbf{A}}$ Secuencia de transformaciones sobre las filas de \mathbf{A} (es una matriz).
- $\boxed{\tau_1 \cdots \tau_k \mathbf{A}_{\tau_{(k+1)} \cdots \tau_p}}$ Secuencia de transformaciones sobre las filas y columnas de \mathbf{A} (es una matriz); puesto que $(\tau_1 \cdots \tau_k \mathbf{A})_{\tau_{(k+1)} \cdots \tau_p} = \tau_1 \cdots \tau_k (\mathbf{A}_{\tau_{(k+1)} \cdots \tau_p})$.

Incluir el **producto de matrices por bloques** me empujó a **extender la notación del operador selector** de manera que se puedan seleccionar submatrices. La extensión consiste en que cuando el argumento de “ $|$ ” es una *lista* de índices (i, j, k, \dots) creamos la submatriz con las filas o columnas de la lista:

- $\boxed{(1,3,)|\mathbf{A}}$ es la submatriz formada por las filas 1 y 3 de \mathbf{A} .
- $\boxed{\mathbf{A}_{|(2,4,)}}$ es la submatriz formada por las columnas 2 y 4 de \mathbf{A} .
- $\boxed{(1,3,)|\mathbf{A}_{|(2,4,)}}$ submatriz formada por las filas 1 y 3 y columnas 2 y 4.

Esta extensión de la notación no solo simplifica algunas demostraciones del libro, *su implementación en NAcAL también ha simplificado el código de la librería*.

Para finalizar esta breve descripción de los dos operadores... de lo anterior se deriva que:

la trasposición cambia de lado los operadores “ $|$ ” y “ τ ”. Es decir

$${}_i|(\mathbf{A}^\top) = \mathbf{A}_{|i} \quad \text{y} \quad (\tau \mathbf{A})^\top = \mathbf{A}_\tau.$$

(Fíjese que en este libro la trasposición actúa sobre matrices ¹⁵, pero jamás sobre vectores).

¹⁵y también sobre secuencias de transformaciones elementales (véase la Sección 4.A).

Al ser asociativa, la notación es operativa y adquiere semántica. Como ambos operadores (“|” y “ τ ”) son lineales, la notación hereda la asociatividad del producto matricial; por ejemplo $(\mathbf{AB})_\tau = \mathbf{A}(\mathbf{B}_\tau)$. Esto dota de operatividad y semántica a la notación...

Para explicarlo volveré sobre el manual de Poole (2004). En dicho texto no se hace una distinción clara entre vectores y matrices, de manera que el producto *matriz por vector* se presenta como un caso particular del producto de matrices. Y por otra parte, el producto de matrices es definido como en la Ecuación 7. Así, el *modo de calcular* \mathbf{Ab} adquiere un mayor protagonismo que la propia naturaleza del objeto \mathbf{Ab} .

Sin embargo, aquí ¹⁶ se define el producto *matriz por vector*, \mathbf{Ab} , como *el vector combinación lineal de las columnas de \mathbf{A} cuyos parámetros son los elementos de \mathbf{b}* . Esta definición no se centra en el *modo de calcular*; se centra en la naturaleza del objeto: es una combinación lineal de las columnas de \mathbf{A} .¹⁷

$$\mathbf{Ab} = (\mathbf{A}_{|1})b_{|1} + \cdots + (\mathbf{A}_{|n})b_{|n}.$$

El modo de calcular queda al margen de la definición. Lo sorprendente es que la notación indica cómo calcular \mathbf{Ab} : si nos fijamos en su componente i -ésima, vemos que es el producto de la *fila* i -ésima de \mathbf{A} por \mathbf{b} :

$${}_i|\mathbf{Ab}.$$

A su vez, el producto *vector por matriz*, \mathbf{aB} , se define como *el vector combinación lineal de las filas de \mathbf{B} donde los parámetros de la combinación son los elementos de \mathbf{a}* .¹⁸

$$\mathbf{aB} = {}_1|\mathbf{a}({}_1|\mathbf{B}) + \cdots + {}_n|\mathbf{a}({}_n|\mathbf{B}).$$

Consecuentemente tenemos que *¡la transposición también cambia de lado al vector!* pues $\mathbf{aB} = (\mathbf{B}^\tau)\mathbf{a}$.

Empleando la definición de *matriz por vector*, se definirá el *producto de matrices*, \mathbf{AB} , como *aquella matriz cuya columna j -ésima es*

$$(\mathbf{AB})_{|j} = \mathbf{A}(\mathbf{B}_{|j})$$

es decir, \mathbf{AB} es una matriz cuyas columnas son combinaciones de las columnas de \mathbf{A} , pues son el producto de la matriz \mathbf{A} por cada uno de los vectores $\mathbf{B}_{|j}$.

Compárese la expresión $(\mathbf{AB})_{|j} = \mathbf{A}(\mathbf{B}_{|j})$ con la Ecuación 7 en la página xxiv. No solo es mucho más breve, también indica qué relación hay entre las columnas de \mathbf{A} y las de \mathbf{AB} . ¡La definición no se basa en la descripción de cómo calcular el producto!... Sin embargo, como la notación es asociativa (y por tanto los paréntesis son innecesarios), si denotamos el elemento de la fila i y columna j , estamos indicando simultáneamente que dicho elemento *es el producto de la fila i -ésima de \mathbf{A} con la columna j -ésima de \mathbf{B}* :

$${}_i|\mathbf{AB}_{|j}.$$

También se deduce que la fila i -ésima de \mathbf{AB} es combinación lineal de las filas de \mathbf{B} , pues ${}_i|\mathbf{AB} = ({}_i|\mathbf{A})\mathbf{B}$ (por ser un vector del tipo “*vector por matriz*”). Es decir,

de la propia notación se deducen propiedades. ¡La notación cobra semántica!

Es un buen momento para releer la cita de George Polya de la página xxi.

Aunque las reglas de notación se extienden a más operaciones y objetos, las que he enunciado aquí son suficientes para ilustrar cómo se resuelven los problemas de notación señalados más arriba.

¹⁶Al igual que en Strang (2003).

¹⁷Que también podemos escribir del siguiente modo: $\mathbf{Ab} = (\mathbf{A}_{|1})b_1 + \cdots + (\mathbf{A}_{|n})b_n$.

¹⁸Que también podemos escribir del siguiente modo: $\mathbf{aB} = a_1({}_1|\mathbf{B}) + \cdots + a_n({}_n|\mathbf{B})$.

Esta notación respeta las definiciones y resuelve los citados problemas

- Con esta notación la Ecuación 1 en la página xxii quedaría así

$$\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}_{|_2} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}; \quad {}_1| \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}_{|_2} = 3.$$

Nótese la clara distinción entre matrices, vectores y escalares.

- Para la Ecuación 2 en la página xxii tenemos varias formas de escritura

$$\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}_{[(-2)\tau\mathbf{1}+\mathbf{2}]} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}_{[(-2)\tau\mathbf{1}+\mathbf{2}]} \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 0 & -2 \end{bmatrix}.$$

Pero es más fácil de leer $\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}_{[(-2)\tau\mathbf{1}+\mathbf{2}]}$, que indica que hay que restar dos veces la primera fila de la segunda, que interpretar qué operación realiza la matriz $\begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}$. Esta ventaja es importante cuando se aplica una secuencia de transformaciones o cuando la matriz es de un orden mayor.

- Hay distintas formas de escribir el producto punto entre dos vectores, por ejemplo

$$\begin{pmatrix} 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix} = 3.$$

Nótese la clara distinción entre escalares y vectores (aquí no hay matrices).

- La demostración de la Ecuación 4 en la página xxiii queda así

$$\mathbf{A}(\mathbf{B} + \mathbf{C})_{|_j} = \mathbf{A}(\mathbf{B}_{|_j} + \mathbf{C}_{|_j}) = \mathbf{A}\mathbf{B}_{|_j} + \mathbf{A}\mathbf{C}_{|_j} = (\mathbf{A}\mathbf{B} + \mathbf{A}\mathbf{C})_{|_j}.$$

(ahorro considerable de tipos de fuentes y razonamiento sencillo y directo.)

- La descripción del cálculo del producto matricial de la Ecuación 5 en la página xxiii queda así

$$\mathbf{A}\mathbf{B} = \sum_{k=1}^n (\mathbf{A}_{|(k,)})_{((k,)|} \mathbf{B}).$$

Como el argumento de los operadores selector es una lista (de un único índice), dentro del sumatorio aparecen productos de matrices columna (selector actúa por la derecha) con matrices fila (selector actúa por la izquierda). Es decir, $\mathbf{A}\mathbf{B}$ es una suma de matrices resultantes de multiplicar matrices columna $(\mathbf{A}_{|(k,)})$ por matrices fila $((k,)|\mathbf{B})$.

- La Ecuación 6 en la página xxiii se podría escribir así

$${}_3|\mathbf{A} = l_{31}({}_1|\mathbf{U}) + l_{32}({}_2|\mathbf{U}) + l_{33}({}_3|\mathbf{U}),$$

pero es muchísimo mejor escribir simplemente:

$${}_3|\mathbf{A} = {}_3|\mathbf{L}\mathbf{U}.$$

Es decir, la tercera fila de \mathbf{A} es una combinación lineal de las filas de \mathbf{U} (por ser el producto del vector ${}_3|\mathbf{L}$ por la matriz \mathbf{U}). La notación indica simultáneamente el significado (o relación entre los objetos¹⁹) y cómo calcular, ya que el elemento i -ésimo del vector ${}_3|\mathbf{A}$ es un producto punto: ${}_3|\mathbf{A}_{|_i} = {}_3|\mathbf{L}\mathbf{U}_{|_i} = ({}_3|\mathbf{L}) \cdot (\mathbf{U}_{|_i})$.

¹⁹Como sugiere la cita de George Polya.

- La demostración de que $(\mathbf{AB})^\top = (\mathbf{B}^\top)(\mathbf{A}^\top)$ se reduce a:²⁰

$$(\mathbf{AB})^\top_{|j} = {}_j|\mathbf{AB} = (\mathbf{B}^\top)({}_j|\mathbf{A}) = (\mathbf{B}^\top)(\mathbf{A}^\top)_{|j}.$$

Éste es un buen ejemplo de lo que quiero decir con “notación operativa”. Aquí solo es necesario fijarse que ${}_j|\mathbf{A}$ es un vector y aplicar dos reglas de manipulación de símbolos:

1. la transposición cambia de lado el operador: $(\mathbf{A}^\top)_{|j} = {}_j|\mathbf{A}$,
2. la transposición cambia de lado el vector: $\mathbf{aB} = (\mathbf{B}^\top)\mathbf{a}$;

Esta demostración no se basa en cómo calcular el producto \mathbf{AB} .²¹ ¡Se basa en el uso de las reglas de notación! Es decir, en el uso del lenguaje.

Para usar esta notación con L^AT_EX, puede descargarse un paquete con las macros desde mi repositorio de GitHub: <https://github.com/mbujosab/nacal-latex-package>

Implementación de la notación en Python (el módulo NAcAL)

El código del módulo **NAcAL** es una implementación *literal* de la notación empleada en el curso (véase la **documentación**). Un par de ejemplos serán suficientes para ilustrarlo.

Ejemplo 1: La suma de vectores se define en la Lección 1 como una operación componente a componente:

$$(\mathbf{a} + \mathbf{b})_{|i} = \mathbf{a}_{|i} + \mathbf{b}_{|i} \quad \text{para } i = 1 : n,$$

y su implementación en NAcAL es literalmente:

```
Vector([ (self|i) + (other|i) for i in range(1, len(self)+1) ])
```

donde **self** es **a** y **other** es **b**; y donde `len(self)+1` es el número de componentes de **a**.

Ejemplo 2: El producto de Matrices se define en la Lección 3 como una operación columna a columna:

$$(\mathbf{AX})_{|j} = \mathbf{A}(\mathbf{X}_{|j}) \quad \text{para } j = 1 : n,$$

y su implementación en NAcAL es literalmente:

```
Matrix([ self*(x|j) for j in range(1, x.n+1) ])
```

donde **self** es **A** y **x** es **X**; y donde `x.n+1` es el número de columnas de **X**.

Como se puede apreciar, debido a la literalidad de la implementación (incluido el uso de la barra vertical), el código de Python es tan conciso como las propias definiciones.

Más llamativo aún es que las demostraciones por inducción referidas a la eliminación y a la diagonalización de matrices son descripciones *literales* del código para la implementación del correspondiente algoritmo programado en **NAcAL**.

Las demostraciones por inducción describen los algoritmos correspondientes.

²⁰Creo que es imposible encontrar una demostración más sucinta de este conocido resultado.

²¹En contraste con las ecuaciones (8) y (9) en la Página xxiv.

NacAL en acción. El módulo **NacAL** trabaja con vectores de \mathbb{R}^n , con matrices, con transformaciones elementales, con subespacios en \mathbb{R}^n y con subespacios afines en \mathbb{R}^n . Implementa el álgebra matricial, la eliminación de izquierda a derecha (por columnas) y de arriba a abajo (por filas), la eliminación Gaussiana tanto por filas como por columnas y la eliminación Gauss-Jordan tanto por filas como por columnas. Con estos procedimientos **NacAL** es capaz de invertir matrices, resolver sistemas de ecuaciones homogéneos y no homogéneos, calcular determinantes,²² diagonalizar por bloques triangulares por semejanza, diagonalizar ortogonalmente matrices simétricas y también diagonalizar matrices simétricas por congruencia (lo que permite clasificar formas cuadráticas). También realiza operaciones con subespacios de \mathbb{R}^n y subespacios afines de \mathbb{R}^n (incluyendo la representación paramétrica y cartesiana, la suma, la intersección, el complemento ortogonal, la inclusión, la pertenencia de un vector a un subespacio o subespacio afín, etc.). Emplea \LaTeX tanto para representar los objetos como para mostrar el desarrollo de muchos cálculos. Además es capaz de trabajar simbólicamente. Veamos algunos ejemplos.

Ejemplo 1. Cálculo de la inversa de $\mathbf{B} = \begin{bmatrix} 2 & \sqrt{3} \\ 0 & \frac{1}{\pi} \end{bmatrix}$ por eliminación Gauss-Jordan:

Librería NAcAL para Python

```
B = Matrix([[2,sympy.sqrt(3)],[0,fracc(1,sympy.pi)]])
InvMat(B,1)
```

$$\left[\begin{array}{cc|c} 2 & \sqrt{3} & \\ 0 & \frac{1}{\pi} & \end{array} \right] \xrightarrow{\left[\left(-\frac{\sqrt{3}}{2} \right) 1+2 \right]} \left[\begin{array}{cc|c} 2 & 0 & \\ 0 & \frac{1}{\pi} & \end{array} \right] \xrightarrow{\left[\left(\frac{1}{2} \right) 1 \right], \left[\left(\pi \right) 2 \right]} \left[\begin{array}{cc|c} 1 & 0 & \\ 0 & 1 & \end{array} \right]$$

(Por tanto $\mathbf{B}^{-1} = \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}\pi}{2} \\ 0 & \pi \end{bmatrix}$)

Ejemplo 2. Resolución de un sistema de ecuaciones lineales:

Librería NAcAL para Python

```
A = Matrix( [ [2,-2,0], [-1,3,4*sympy.sqrt(7)] ] );
b = Vector( [2*sympy.pi, sympy.exp(1)] )
SEL( A, b, 1 ) # resuelve el Sistema de Ecuaciones Lineales
```

$$\left[\begin{array}{ccc|c} 2 & -2 & 0 & -2\pi \\ -1 & 3 & 4\sqrt{7} & -e \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right] \xrightarrow{\left[\left(1 \right) 1+2 \right], \left[\left(\pi \right) 1+4 \right]} \left[\begin{array}{ccc|c} 2 & 0 & 0 & 0 \\ -1 & 2 & 4\sqrt{7} & -\pi - e \\ 1 & 1 & 0 & \pi \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right] \xrightarrow{\left[\left(-2\sqrt{7} \right) 2+3 \right], \left[\left(\frac{e}{2} + \frac{\pi}{2} \right) 2+4 \right]} \left[\begin{array}{ccc|c} 2 & 0 & 0 & 0 \\ -1 & 2 & 0 & 0 \\ 1 & 1 & -2\sqrt{7} & \frac{e}{2} + \frac{3\pi}{2} \\ 0 & 1 & -2\sqrt{7} & \frac{e}{2} + \frac{\pi}{2} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right]$$

$$\left\{ \mathbf{v} \in \mathbb{R}^3 \mid \exists \mathbf{p} \in \mathbb{R}^1, \mathbf{v} = \begin{pmatrix} \frac{e}{2} + \frac{3\pi}{2} \\ \frac{e}{2} + \frac{\pi}{2} \\ 0 \end{pmatrix} + \begin{bmatrix} -2\sqrt{7} \\ -2\sqrt{7} \\ 1 \end{bmatrix} \mathbf{p} \right\} = \left\{ \mathbf{v} \in \mathbb{R}^3 \mid \begin{bmatrix} -1 & 1 & 0 \\ \frac{\sqrt{7}}{14} & 0 & 1 \end{bmatrix} \mathbf{v} = \begin{pmatrix} -\pi \\ \frac{\sqrt{7}(e+3\pi)}{28} \end{pmatrix} \right\}$$

(El conjunto solución, que es un espacio afín, se representa con unas ecuaciones paramétricas a la izquierda y con unas cartesianas a la derecha).

²²Aunque también está implementado el cálculo recursivo mediante la expansión de Laplace.

Ejemplo 3. Diagonalización de una matriz simétrica por congruencia:

Librería NAcAL para Python

```
f = sympy.symbols('f(\lambda)')
A = Matrix([[0,f],[f, sympy.exp(-2)]])
DiagonalizaC(A,1)
```

$$\begin{bmatrix} 0 & f(\lambda) \\ f(\lambda) & e^{-2} \end{bmatrix} \xrightarrow[\begin{smallmatrix} \tau \\ \tau \end{smallmatrix}]{\begin{smallmatrix} \tau \\ 1 \Rightarrow 2 \end{smallmatrix}} \begin{bmatrix} e^{-2} & f(\lambda) \\ f(\lambda) & 0 \end{bmatrix} \xrightarrow[\begin{smallmatrix} \tau \\ (-f(\lambda)e^2)1+2 \end{smallmatrix}]{\begin{smallmatrix} \tau \\ (-f(\lambda)e^2)1+2 \end{smallmatrix}} \begin{bmatrix} e^{-2} & 0 \\ 0 & -f(\lambda)^2 e^2 \end{bmatrix}$$

(las transformaciones representadas por encima de las flechas son transformaciones de las columnas, las que aparecen por debajo son transformaciones de las filas).

Estos ejemplos muestran una de las características ÚNICAS de NAcAL... No solo realiza los cálculos, además los muestra PASO A PASO. ¡Es una especie de profesor particular virtual que trabaja indistintamente con números o símbolos!

Para poder visualizar la representación en L^AT_EX se debe trabajar con NAcAL en un Notebook de Jupyter, o usar el paquete PythonT_EX y llamar a NAcAL desde el propio código L^AT_EX del documento.

Parte I

Álgebra Matricial

Vectores y matrices (notación y operaciones)

En matemáticas cada objeto tiene su definición. Comencemos con uno que usaremos mucho:

Definición 1.1. Un sistema¹ es una lista ordenada de objetos².

Dos sistemas son distintos si las listas son diferentes; así, los siguientes sistemas de números son diferentes:

$$[1; 2; 1;] \neq [1; 2;] \neq [2; 1;] \neq [2; 1; 2;] \neq [2; 1; 1;] \neq [2; 1; 1; 1; 1; 2;].$$

Un sistema no es un conjunto (o colección) de objetos. En un conjunto no hay orden en la disposición de sus elementos y, de hecho, si todo elemento del conjunto A está en el conjunto B , y todo elemento de B está en el conjunto A , se dice que los conjuntos A y B son iguales; por tanto los siguientes conjuntos son iguales:

$$\{1, 2, 1\} = \{1, 2\} = \{2, 1\} = \{2, 1, 2\} = \{2, 1, 1\} = \{2, 1, 1, 1, 1, 2\}.$$

Distinguiremos conjuntos y sistemas del siguiente modo: por una parte encerramos los elementos de un conjunto entre llaves y separados por comas. Por otra, encerramos los elementos de un *sistema* genérico entre corchetes y pondremos un “punto y coma” después de cada elemento.

En este curso trataremos con dos importantes tipos de sistemas: los vectores de \mathbb{R}^n (listas de números) y las matrices (listas de vectores de \mathbb{R}^n).

1.1. Vectores de \mathbb{R}^n

Denotamos con \mathbb{R} al *conjunto de números reales*. Ejemplos de números reales son: 7 , $\frac{-3}{11}$, $\sqrt{2}$, ó $-\pi$. Con frecuencia, a los números reales también los llamaremos *escalares*.

Definición 1.2. Llamamos *vector* de \mathbb{R}^n a un sistema (o lista ordenada) de n números reales.

El símbolo “ \mathbb{R} ” nos indica que el sistema está compuesto por números *reales*, y el superíndice “ n ” nos indica la cantidad de números que tiene la lista.

Recuerde que los sistemas están ordenados y, consecuentemente, podemos identificar los elementos de la lista por la posición que ocupan: hay un primer elemento, un segundo elemento, etc. La forma de escribir la lista no es importante siempre que quede claro quien es el primer elemento, quien el segundo, etc. Por eso podemos escribir un *mismo* vector tanto en horizontal como en vertical:³

$$(\pi, -1, 0, 1,) \quad \text{ó} \quad \begin{pmatrix} \pi \\ -1 \\ 0 \\ 1 \end{pmatrix}.$$

Ambas son representaciones del vector de \mathbb{R}^4 cuya primera componente es π , la segunda -1 , la tercera 0 y la cuarta 1 .

¹Una definición más formal aparecerá en la página 90

²Con “objeto” siempre nos referiremos a *objetos matemáticos*: números, funciones, ecuaciones, variables, etc.

³Otros libros distinguen entre vectores *fila* y vectores *columna*; pero aquí, dada la Definición 1.2, no cabe tal distinción.

Notación. Aunque los vectores de \mathbb{R}^n son *sistemas*, para ellos usaremos una notación específica:⁴ escribiremos cada vector con su correspondiente lista de números entre *paréntesis* (y cuando los escribamos en horizontal, pondremos una coma detrás de cada elemento).⁵

Denotaremos los vectores de \mathbb{R}^n con letras minúsculas en negrita cursiva: ***a***, ***b***, ***c***, etc. Y nos referiremos genéricamente a los componentes de un vector con la misma letra con la que denotamos al vector completo, pero en cursiva y con un subíndice que indica la posición del componente dentro del vector. Por ejemplo:

$$\mathbf{a} = (a_1, a_2, a_3, a_4, a_5, a_6);$$

es un vector de \mathbb{R}^6 cuyo tercer componente es a_3 . Para indicar que el segundo componente de $\mathbf{b} = (0, 4, 6)$ es un cuatro, escribiremos $b_2 = 4$.

Decimos que dos vectores \mathbf{x} e \mathbf{y} de \mathbb{R}^n son iguales si lo son sus correspondientes listas ordenadas de componentes, es decir, si y solo si: $x_i = y_i$ para $i = 1 : n$ (i.e. si son iguales componente a componente).⁶ Por ejemplo, con los números 1, 3 y 5, podemos formar dos vectores *distintos* del siguiente modo:

$$\mathbf{x} = (1, 3, 5) \quad \text{y} \quad \mathbf{z} = (3, 5, 1).$$

Recuerde que dos vectores son iguales solo si lo son sus respectivas listas. Por eso \mathbf{x} y \mathbf{z} son distintos ($\mathbf{x} \neq \mathbf{z}$).

El operador selector de componentes “|”.

Para homogeneizar la notación con vectores y matrices, también denotaremos la i -ésima componente de \mathbf{a} añadiendo un subíndice en el que aparece el símbolo “|” junto al índice de la componente: $\mathbf{a}_{|i}$. De esta manera indicamos que el operador selector “|” *actúa* sobre el vector para seleccionar la componente i -ésima. Así, las siguientes formas de denotar la tercera componente de \mathbf{b} son equivalentes:

$$\mathbf{b}_{|3} \equiv b_3.$$

Uso de NAcAL



Durante el curso usaremos el módulo **NAcAL** para Python. Así podremos experimentar en el ordenador con los objetos y procedimientos que vayamos viendo a lo largo de este curso de Álgebra Lineal. El módulo es una implementación literal de los objetos y procedimientos que veremos en el libro.

El primer paso para usar el módulo es importar el módulo en memoria

```
from nacal import * # importamos el módulo en memoria
```

Librería NAcAL para Python

Una vez importado el módulo, tecleando `Vector([0, 4, 6])` en un terminal obtenemos:

terminal de Python

```
>>> from nacal import *
>>> b = Vector( [0, 4, 6] ) # Vector b cuya lista de componentes es 0, 4 y 6
>>> b
Vector([0, 4, 6])
```

⁴Si usáramos la notación genérica de *sistema*, escribiríamos $[\pi; -1; 0; 1;]$, es decir, una lista entre corchetes con sus elementos seguidos por puntos y comas (como en los ejemplos dados tras la Definición 1.1)

⁵Así $(2,)$ es un *vector* de \mathbb{R}^1 (es decir, una lista con un único elemento); y (2) es un *número* entre paréntesis.

⁶Con “ $n : m$ ” denotamos la secuencia $n, n+1, \dots, m$, es decir, la lista ordenada de números naturales que van de n a m . Así, con $i = 1 : n$ indicamos que el índice i recorre la sucesión de enteros que va de 1 a n , es decir $1, 2, \dots, n$. Cuando el primer índice, n , sea mayor que el segundo, m , entenderemos que “ $n : m$ ” es una lista vacía: $[\]$. Empleando una notación “pythonesca”, con $n : m$ denotamos la lista $[k \in \mathbb{N} \mid n \leq k \leq m]$.

donde el resultado aparece en gris en el anterior recuadro de **terminal** de **Python**; y si ejecutamos el código en un Notebook de Jupyter obtenemos:

$$\begin{pmatrix} 0 \\ 4 \\ 6 \end{pmatrix}$$

Para obtener el tercer componente de **b** tecleamos:

```
b[3]
```

Librería NAcAL para Python

y tanto en un terminal de Python como en un Notebook de Jupyter obtendremos el número 6:



Hemos definido un vector de \mathbb{R}^n como un sistema de n números reales. Al encerrar una lista de números entre paréntesis denotamos un vector (cuyas componentes son los números encerrados en el orden en el que aparecen en la lista dentro del paréntesis). Por ejemplo

$$\mathbf{a} = (1, 4, 9, 2).$$

Para seleccionar una componente de la lista usamos el operador “[” y el índice de la componente que queremos seleccionar. Por ejemplo, la componente a_3 (la tercera componente de \mathbf{a}) es:

$$\mathbf{a}_{|3} = 9.$$

1.1.1. Definiciones de algunos vectores especiales

Definición 1.3. Llamamos **vector nulo** (o **vector cero**) de \mathbb{R}^n al vector cuyas n componentes son cero, y lo denotamos con **0**. Por ejemplo, los vectores nulos de \mathbb{R}^1 , \mathbb{R}^2 , de \mathbb{R}^3 y de \mathbb{R}^8 son respectivamente:

$$(0), \quad \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{y} \quad (0, 0, 0, 0, 0, 0, 0, 0).$$

¡Nótese que estos cuatro vectores nulos **no son iguales**, pues no tienen el mismo número de componentes (y por tanto sus correspondientes listas ordenadas no son iguales)!⁷

```
V0( 3 )
```

```
# vector nulo de tres componentes
```

Librería NAcAL para Python

$$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Definición 1.4. Sea el vector \mathbf{a} . Llamamos **vector opuesto** de \mathbf{a} al vector $-\mathbf{a}$, que es el vector cuyos componentes son los opuestos de los de \mathbf{a} ; es decir $(-\mathbf{a})_{|i} = -(\mathbf{a}_{|i})$. Por ejemplo:

$$\text{si } \mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}, \text{ entonces } -\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}; \text{ y si } \mathbf{c} = (1, -3\pi, \sqrt{2}), \text{ entonces } -\mathbf{c} = (-1, 3\pi, -\sqrt{2}).$$

⁷Nótese la coma en $(0,)$ que es el vector nulo de \mathbb{R}^1 , en contraste con (0) que es el número cero entre paréntesis.

`-b` `# vector opuesto de b`

$$\begin{pmatrix} 0 \\ -4 \\ -6 \end{pmatrix}$$

1.2. Sumas de vectores y productos de vectores por escalares

Las dos operaciones fundamentales con vectores son la suma y el producto de un vector por un número.

Definición 1.5 (suma de vectores). La suma, $\mathbf{a} + \mathbf{b}$, de dos vectores \mathbf{a} y \mathbf{b} de \mathbb{R}^n es el vector de \mathbb{R}^n cuya componente i -ésima es la suma de las componentes $\mathbf{a}_{|i}$ y $\mathbf{b}_{|i}$. Es decir, sumamos componente a componente:

$$(\mathbf{a} + \mathbf{b})_{|i} = \mathbf{a}_{|i} + \mathbf{b}_{|i} \quad \text{para } i = 1 : n.$$

Por tanto:

$$\mathbf{a} + \mathbf{b} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} + \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ \vdots \\ a_n + b_n \end{pmatrix}.$$

Observación. Nótese que el operador “ $|i$ ” es distributivo⁸ para la suma: $(\mathbf{a} + \mathbf{b})_{|i} = \mathbf{a}_{|i} + \mathbf{b}_{|i}$.

Definición 1.6 (producto de un vector por un escalar). El producto, $\lambda \mathbf{b}$, de un vector \mathbf{b} de \mathbb{R}^n por un escalar λ es el vector de \mathbb{R}^n cuya componente i -ésima es el producto de λ por $\mathbf{b}_{|i}$. Es decir, multiplicamos cada componente de \mathbf{b} por λ :

$$(\lambda \mathbf{b})_{|i} = \lambda(\mathbf{b}_{|i}) \quad \text{para } i = 1 : n.$$

Por tanto

$$\lambda \mathbf{b} = \lambda \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} \lambda b_1 \\ \vdots \\ \lambda b_n \end{pmatrix}.$$

Al vector $\lambda \mathbf{b}$ lo denominamos *múltiplo* (o múltiplo escalar) de \mathbf{b} .

La operación de multiplicar un vector por un número también se denomina “escalar el vector” (es decir, cambiarlo de escala: piense que si tiene un vector con datos de salarios en miles de euros, expresar esos datos en euros es escalar los datos; y se logra multiplicando por 1000).

Observación. El operador “ $|i$ ” es asociativo⁹ para el producto por escalares: $(\lambda \mathbf{b})_{|i} = \lambda(\mathbf{b}_{|i})$. Fíjese cómo la región encerrada por el paréntesis se desplaza en la expresión anterior... en un lado del la igualdad “ λ ” está dentro del paréntesis e “ $|i$ ” fuera; y lo contrario ocurre al otro lado. Cuando es posible “desplazar” los paréntesis manteniendo la igualdad, significa que el orden en que se realizan las operaciones no es importante y que podemos omitir los paréntesis.¹⁰ Por tanto podemos escribir sencillamente $\lambda \mathbf{b}_{|i}$.

Definición 1.7 (Operador lineal). Un operador distributivo respecto de la suma y asociativo respecto al producto por escalares se dice que es *lineal*.

Así pues, el operador “ $|i$ ” (que selecciona el componente i -ésimo de un vector) es un operador lineal.

⁸De manera similar a la propiedad distributiva en la suma $(a + b)x = ax + bx$.

⁹De manera similar a la propiedad asociativa del producto entre escalares: $(ab)c = a(bc)$.

¹⁰De manera similar a cuando escribimos el producto de tres escalares: abc .



La definición de las operaciones de suma de vectores y producto de un vector por un escalar convierten al operador “ $|_i$ ” en un **operador lineal**; es decir, *distributivo* respecto a la suma:

$$(\mathbf{a} + \mathbf{b})|_i = \mathbf{a}|_i + \mathbf{b}|_i$$

y *asociativo* respecto al producto por *escalares*:

$$(\lambda \mathbf{b})|_i = \lambda(\mathbf{b}|_i).$$

La omisión de los paréntesis y del símbolo del producto en Python En un texto la expresión $2\mathbf{b}$ significa “multiplicar por 2 el vector \mathbf{b} ”, es decir

$$2\mathbf{b} \equiv 2 \cdot \mathbf{b}$$

donde el punto denota el producto. Nosotros estamos acostumbrados a *omitir* el símbolo de la operación producto (el punto) sin que nos ocasione ningún problema (así logramos expresiones más compactas).

Pero tenga en cuenta que Python no funciona como un humano. El intérprete de Python requiere la presencia *explícita* del símbolo de la operación producto (el asterisco “ $*$ ”):

```
2*b          # multiplica el Vector b por 2 (también podemos escribir b*2 )
```

Librería NAcAL para Python

$$\begin{pmatrix} 0 \\ 8 \\ 12 \end{pmatrix}$$

El intérprete de Python tiene establecidas una serie de precedencias entre operadores. En el siguiente cuadro de operadores, los de arriba preceden a (*se ejecutan antes que*) los de abajo:

Precedencia en las operaciones	
Símbolo en Python	Operación asociada en NAcAL
$()$, $[]$, $\{ \}$	
$**$	exponenciación
$\sim A$	transposición
$*$, $/$	multiplicación, división
$+$, $-$	suma, resta
$\&$	transformación elemental
$ $	selector

Cuadro 1.1: Precedencia de operadores (detalles [aquí](#): 6.16 *Orden de evaluación* y 6.17 *Prioridad de operador*)

Fíjese que Python ejecuta antes la operación *producto* ($*$) que la selección de un elemento ($|$). Por tanto, en Python podemos producir errores si no escribimos con suficiente cuidado la expresiones (hay que recordar que las expresiones se ejecutan de izquierda a derecha pero con unas reglas de precedencia estrictas).

Como la precedencia no es fácil de recordar, por claridad en el código (*y para asegurar que realizamos las operaciones en el orden deseado*) es recomendable usar los paréntesis: lo que haya dentro de los paréntesis será lo primero en ser evaluado.

EJERCICIO 1. ¿Cómo evalúa el módulo NAcAL las siguientes expresiones: $(3*\mathbf{b})|_2$; $3*(\mathbf{b}|_2)$ y $3*\mathbf{b}|_2$?

EJERCICIO 2. ¿Cuál es la primera operación realizada en cada una de estas operaciones? $\mathbf{b}+\mathbf{b}*2$ y $\mathbf{b}+\mathbf{b}|2$.

EJERCICIO 3. ¿Qué problema puede tener la expresión $\mathbf{b}+\mathbf{b}|2$ para nosotros?

Moraleja: lo más conveniente es poner paréntesis tanto al escribir con lapiz y papel como al escribir en Python. Con ello evitamos ambigüedades (en el primer caso) y sorpresas (en el segundo).

1.2.1. Propiedades de la suma y del producto por un escalar



Explotando las propiedades de la notación (junto con algunas propiedades de los *números reales*) vamos a demostrar ocho propiedades que verifican la *suma de vectores* y el *producto de un vector por un escalar* (Proposición 1.2.1). (ocho propiedades que nos permitirán definir el *espacio vectorial* \mathbb{R}^n en la Lección 6).

Para empezar recordemos algunas propiedades (¡no todas!) de los números reales que seguro que conoce y que usted ha manejado desde la escuela. . .

Recordatorio de algunas propiedades del conjunto de números reales

Sean a , b y c números reales. Entonces:

- | | |
|----------------------------------|---------------------------|
| 1. $a + b = b + a$. | 5. $ab = ba$. |
| 2. $a + (b + c) = (a + b) + c$. | 6. $a(b + c) = ab + ac$. |
| 3. $a + 0 = a$. | 7. $a(bc) = (ab)c$. |
| 4. $a + (-a) = 0$. | 8. $1a = a$. |

La cuestión que queremos abordar ahora es: ¿también se verificarán estas propiedades cuando operamos con vectores? En concreto, hagámonos las siguientes ocho preguntas (puede investigar con el módulo **NAcAL**):

1. ¿Es verdad que

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}?$$

¿Es cierto solo para los vectores del ejemplo? o ¿es una regla general para cualquier par de vectores?

2. ¿Es verdad que

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix} + \left(\begin{pmatrix} 4 \\ 5 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right) = \left(\begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 4 \\ 5 \end{pmatrix} \right) + \begin{pmatrix} 1 \\ 0 \end{pmatrix}?$$

El paréntesis significa que primero hay que hacer la suma dentro del paréntesis, y al vector resultante sumarle el vector de fuera.

¿Es cierto solo para los vectores del ejemplo? o ¿es una regla general para cualquier trío de vectores?

3. ¿Es verdad que

$$\begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}?$$

¿Es cierto solo para los vectores del ejemplo? o ¿es una regla general para cualquier par de vectores?

4. ¿Es verdad que

$$\begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} + \begin{pmatrix} -4 \\ -5 \\ -6 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}?$$

¿Es cierto solo para los vectores del ejemplo? o ¿es una regla general para cualquier par de vectores?

5. ¿Es verdad que

$$2 \left(\begin{pmatrix} 1 \\ -1 \end{pmatrix} + \begin{pmatrix} 4 \\ 5 \end{pmatrix} \right) = 2 \begin{pmatrix} 1 \\ -1 \end{pmatrix} + 2 \begin{pmatrix} 4 \\ 5 \end{pmatrix}?$$

(recuerde que primero se realizan las operaciones de dentro del paréntesis) ¿Es una regla general para cualquier par de vectores? ¿Y si cambiamos el escalar 2 por otro número?

6. ¿Es verdad que

$$(2 + 3) \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + 3 \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}?$$

(recuerde que primero se realizan las operaciones de dentro del paréntesis) ¿Es una regla general para cualquier par de vectores? ¿y si cambiamos los escalares 2 y 3 por otros números?

7. ¿Es verdad que

$$2 \left(3 \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right) = (2 \cdot 3) \begin{pmatrix} 1 \\ 2 \end{pmatrix}?$$

(recuerde, primero las operaciones dentro de los paréntesis). Hágase las mismas preguntas de arriba.

8. ¿Es verdad que si $a = 1$

$$a \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}?$$

¿Es cierto solo para el escalar $a = 1$?

En el EJERCICIO 4 se le pide que demuestre las ocho propiedades que acabamos de revisar¹¹. Su solución aparece al final del libro ¡pero usted debe resolverlo por su cuenta y sin mirar! Si “cotillea” previamente la solución, el ejercicio le servirá de muy poco.

Para resolver el ejercicio únicamente necesita aplicar las reglas de notación vistas más arriba (definiciones de *suma* y *producto por escalares* de la Página 6); y cuando tenga que operar con los componentes de los vectores (que son números reales) aplique la propiedad de las operaciones entre números reales que necesite (de entre las del *recordatorio* de la Página 8).

EJERCICIO 4. Demuestre las propiedades de la siguiente proposición:

Proposición 1.2.1 (Propiedades de las operaciones entre vectores). *Para cualesquiera vectores \mathbf{a} , \mathbf{b} y \mathbf{c} con el mismo número de componentes, y para cualesquiera escalares (números) λ y φ , se verifican las siguientes ocho propiedades:*

1. $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$.

2. $\mathbf{a} + (\mathbf{b} + \mathbf{c}) = (\mathbf{a} + \mathbf{b}) + \mathbf{c}$.

3. $\mathbf{0} + \mathbf{a} = \mathbf{a}$.

4. $\mathbf{a} + (-\mathbf{a}) = \mathbf{0}$.

5. $\lambda(\mathbf{a} + \mathbf{b}) = \lambda\mathbf{a} + \lambda\mathbf{b}$.

6. $(\lambda + \varphi)\mathbf{a} = \lambda\mathbf{a} + \varphi\mathbf{a}$.

7. $\lambda(\varphi\mathbf{a}) = (\lambda\varphi)\mathbf{a}$.

8. $1\mathbf{a} = \mathbf{a}$.

¹¹Se le propone como ejercicio una ¡demostración! Las demostraciones son imprescindibles en matemáticas. Es IMPOSIBLE que adquiera una buena formación matemática si no realiza las demostraciones. Solo se adquiere una verdadera comprensión al realizar las demostraciones. Las que se le propondrán en el curso o son muy sencillas (como las de este primer ejercicio) o son “casi” una repetición de otras ya vistas, o las pistas que se le dan describen los pasos necesarios para dar con la demostración.

Aunque las demostraciones son sencillas es posible que le cuesten (quizá sea la primera vez que intenta demostrar algo... y la primera vez suele resultar difícil). No se desanime; *es importante lograr demostrar proposiciones sencillas* (si trabaja el curso adecuadamente cuando lo termine habrá aprendido mucho).

Pista. Dispone de dos estrategias. La primera se asemeja a la forma de operar mostrada en las ocho preguntas de más arriba (por ello es posible que le parezca más sencilla). La segunda estrategia explota las propiedades de *linealidad* del operador “ $|$ ”. Es importante que se familiarice con este segundo modo de trabajar: es más abstracto pero mucho más rápido y versátil (pues la demo solo se centra en aplicar las propiedades a la componente i -ésima).

Estrategia 1 Como estas propiedades provienen de los escalares (los números reales), la estrategia es operar con las componentes dentro de los vectores. Puede escribir sus vectores o como filas o como columnas (en ambos casos la demostración es idéntica).

Estrategia 2 Recuerde que dos vectores \mathbf{v} y \mathbf{w} son iguales si son idénticos componente a componente: $\mathbf{v}_{|i} = \mathbf{w}_{|i}$. Use las propiedades del operador “ $|$ ” siempre que pueda; es decir, tenga en cuenta que son intercambiables las expresiones:

$$\blacksquare (\mathbf{a} + \mathbf{b})_{|i} \quad \text{y} \quad \mathbf{a}_{|i} + \mathbf{b}_{|i} \quad \text{(propiedad distributiva de “ } |i \text{”).}$$

$$\blacksquare (\lambda \mathbf{a})_{|i} \quad \text{y} \quad \lambda(\mathbf{a}_{|i}) \quad \text{(propiedad asociativa de “ } |i \text{”).}$$

Además recuerde que $\mathbf{a}_{|i}$, $\mathbf{b}_{|i}$ y $\mathbf{c}_{|i}$ son números reales; por tanto, puede usar las **propiedades de las operaciones entre números reales** del mismo modo que en la primera estrategia.

Nota 1. Estas ocho propiedades nos permitirán definir el espacio vectorial \mathbb{R}^n en la Lección 6.

1.3. Vectores como representación de datos

Como los vectores son un *un sistema* de números, podemos interpretar cada componente en función de la posición que ocupa. Por ejemplo, imagine que tiene datos del año 2003 sobre la tasa de crecimiento interanual del PIB de cinco países: España (1), Francia (2), Alemania (3), R. Unido (4) y EEUU (5); entonces podemos interpretar los distintos componentes del siguiente vector en función de su posición en la lista:

$$\mathbf{datos}(2003) = \begin{pmatrix} 2.9 \\ 0.9 \\ -0.2 \\ 2.5 \\ 1.8 \end{pmatrix}.$$

Puesto que el número 1.8 está en la quinta posición, sabemos que corresponde al quinto país, y por tanto sabemos que EEUU creció a una tasa del 1.8 % en el año 2003¹². ¿Qué país no creció en el año 2003?

En algunos manuales verá una interpretación de los vectores como segmentos con norma y dirección. Ésta es una interpretación frecuente en aplicaciones físicas, pero poco natural para los economistas. Para un economista la interpretación más natural es que los componentes de los vectores de \mathbb{R}^n son datos (precios, cantidades, número de parados, tasa de inflación, etc.) y la posición de cada componente indica a quien corresponde el dato (como en el ejemplo anterior). Ésta será la interpretación que use en la asignatura de Econometría dentro de un par de cursos.



Ahora pasamos a definir las matrices como un sistema de n vectores de \mathbb{R}^m (es decir, una lista ordenada de n vectores de \mathbb{R}^m). La siguiente introducción así lo ilustra.

¹²fuentes: Informe ICAE sobre la economía española (Octubre 2006). **Instituto Complutense de Análisis Económico.**

1.4. Matrices de $\mathbb{R}^{m \times n}$

Imagine que además tiene datos anuales de los cinco países de arriba durante el periodo: 2003 a 2006.

$$\text{datos}(\mathbf{2003}) = \begin{pmatrix} 2.9 \\ 0.9 \\ -0.2 \\ 2.5 \\ 1.8 \end{pmatrix}, \quad \text{datos}(\mathbf{2004}) = \begin{pmatrix} 3.1 \\ 2.0 \\ 1.1 \\ 3.2 \\ 3.1 \end{pmatrix}, \quad \text{datos}(\mathbf{2005}) = \begin{pmatrix} 3.3 \\ 1.6 \\ 0.7 \\ 1.9 \\ 3.7 \end{pmatrix}, \quad \text{datos}(\mathbf{2006}) = \begin{pmatrix} 3.3 \\ 2.0 \\ 0.9 \\ 1.5 \\ 3.5 \end{pmatrix}.$$

Podemos escribir todos los datos juntos y dar al resultado el nombre de **PIB**. Así tendremos recogidos en **PIB** los datos de crecimiento de los cinco países en los años 2003–2006. Esto se logra con una *matriz*:

Llamamos *matriz* a una lista ordenada de vectores *con el mismo número de componentes* (... es decir ¡un vector de vectores!). De esta manera podremos definir la matriz **PIB** como una *sistema* de vectores con los datos de los cinco años disponibles:

$$[\text{datos}(\mathbf{2003}); \text{datos}(\mathbf{2004}); \text{datos}(\mathbf{2005}); \text{datos}(\mathbf{2006})].$$

Ahora bien, como los vectores se pueden escribir tanto en vertical como en horizontal, si lo disponemos todo en horizontal la escritura puede arrojar una lista muy larga y visualmente difícil de leer:

$$\left[(2.9, 0.9, -0.2, 2.5, 1.8); (3.1, 2.0, 1.1, 3.2, 3.1); (3.3, 1.6, 0.7, 1.9, 3.7); (3.3, 2.0, 0.9, 1.5, 3.5); \right].$$

Sabemos que las cuartas componentes de cada uno de los vectores dentro de la matriz corresponden al Reino Unido... pero esta disposición de los datos no es fácil de leer ¹³. Para facilitar la lectura dispondremos los datos según el siguiente convenio: disponemos los datos referidos al año 2003 en una primera columna, los del año 2004 en la segunda, los del año 2005 en la tercera y los del año 2006 en la cuarta. De ese modo se obtiene el siguiente arreglo rectangular de los datos:

$$\mathbf{PIB} = \begin{bmatrix} 2.9 & 3.1 & 3.3 & 3.3 \\ 0.9 & 2.0 & 1.6 & 2.0 \\ -0.2 & 1.1 & 0.7 & 0.9 \\ 2.5 & 3.2 & 1.9 & 1.5 \\ 1.8 & 3.1 & 3.7 & 3.5 \end{bmatrix}.$$

Así cada columna corresponde a un año distinto, y cada fila a un país distinto... ¡Mucho más fácil de leer!

Definición 1.8. Llamamos *matriz* de $\mathbb{R}^{m \times n}$ a un sistema de n vectores de \mathbb{R}^m .

$$[\mathbf{a}; \mathbf{b}; \dots \mathbf{v};] \quad \text{donde } \mathbf{a}, \mathbf{b}, \dots \mathbf{v} \text{ son } n \text{ vectores de } \mathbb{R}^m.$$

Cada uno de los vectores del sistema es una *columna* de la matriz.

Con las matrices usaremos indistintamente tanto la representación genérica de los *sistemas*, como otra que mantiene los corchetes pero omite los puntos y comas detrás de cada columna:

$$[\mathbf{a} \ \mathbf{b} \ \dots \ \mathbf{v}].$$

Consecuentemente tenemos el siguiente convenio de notación: encerrando un vector entre corchetes escribimos una matriz con una única columna (*matriz columna*). Por ejemplo, si $\mathbf{x} = (1, 5, 9)$, entonces

$$[\mathbf{x}] = \begin{bmatrix} 1 \\ 5 \\ 9 \end{bmatrix}.$$

¹³Aunque un ordenador no tiene ningún problema con esta representación, los humanos si lo tenemos... ¡imagine una lista de 8 vectores con 20 componentes cada uno. Algo similar ocurre si escribimos varios vectores en columna, unos debajo de otros.

Encerrando entre corchetes una lista de vectores (todos con el mismo número de componentes), escribimos la matriz cuyas columnas son dichos vectores en el orden en el que aparecen en la lista. Por ejemplo, si:

$$\mathbf{a} = (1, 0, 3, 0), \quad \mathbf{f} = (2, 2, 2, 2), \quad \text{y} \quad \mathbf{0} = (0, 0, 0, 0),$$

encerrando “entre corchetes” algunos de estos vectores, creamos matrices cuyas columnas son los vectores de la lista escritos en vertical:

$$\begin{bmatrix} \mathbf{a} & \mathbf{f} & \mathbf{0} & \mathbf{f} \end{bmatrix} = \left[\begin{pmatrix} 1 \\ 0 \\ 3 \\ 0 \end{pmatrix}; \begin{pmatrix} 2 \\ 2 \\ 2 \\ 2 \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}; \begin{pmatrix} 2 \\ 2 \\ 2 \\ 2 \end{pmatrix} \right] = \begin{bmatrix} 1 & 2 & 0 & 2 \\ 0 & 2 & 0 & 2 \\ 3 & 2 & 0 & 2 \\ 0 & 2 & 0 & 2 \end{bmatrix}.$$

Definición 1.9. El **orden** de la matriz indica el número de filas y columnas que posee. Está compuesto por dos números: el primero corresponde al número de filas (m) y el segundo al número de columnas (n). El orden se suele expresar cómo “ $m \times n$ ” (y se lee “ m por n ”).

En el ejemplo con datos económicos, el orden de la matriz **PIB** es 5 por 4. Nótese que siempre *se indica primero el número de filas y luego el de columnas*. Con frecuencia expresaremos el orden de la matriz bajo su nombre; por ejemplo: **PIB**.

5 × 4

A cada número dentro de la matriz lo denominamos **componente** (o elemento), y al igual que en el caso de los vectores, identificamos los componentes de la matriz por la posición que ocupan. De esta manera podemos decir que el componente de **PIB** ubicado en la tercera fila y primera columna es -0.2 .

Notación. Denotamos las matrices de $\mathbb{R}^{m \times n}$ con letras mayúsculas y en negrita: **A**, **B**, **C**; y a los componentes genéricos de la matriz con la misma letra con la que denotamos a la matriz completa, pero en minúscula cursiva y con dos subíndices que indican la posición que ocupa el componente dentro de la matriz, comenzando siempre por la fila, y luego por la columna (por tanto el componente a_{ij} de **A** se encuentra en la fila i -ésima y columna j -ésima). Otras formas que usaremos para denotar dicho componente son:

$$a_{ij} \equiv \text{elem}_{ij}(\mathbf{A}) \equiv {}_i|\mathbf{A}|_j.$$

Así escribimos la matriz genérica **A** de orden 2 por 3 como:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix};$$

y de la matriz

$$\mathbf{B} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & 7 \end{bmatrix}$$

podemos afirmar que su orden es 2×3 y que su componente b_{21} es cero, y su componente ${}_2|\mathbf{B}|_3$ es 7.

terminal de Python

```
>>> a = Vector( [1, 0] )
>>> b = Vector( [2,-1] )
>>> c = Vector( [3, 7] )
>>> B = Matrix( [a, b, c] ) # Matrix cuya lista de vectores es a, b, c
>>> 2|B|3                    # elemento de B situado en la segunda fila, segunda columna
7
```

Nótese que usamos un solo índice para identificar las componentes de un vector \mathbf{x} , pero que necesitamos *dos* índices para identificar las componentes de una matriz \mathbf{A} . Así, cuando nos queremos referir a la componente i -ésima del vector \mathbf{x} escribimos:

$$x_i \quad \text{ó} \quad \mathbf{x}_{|i};$$

y cuando queremos referirnos al componente situado en la fila i -ésima y columna j -ésima de la matriz \mathbf{A} :

$$\text{elem}_{ij}(\mathbf{A}), \quad \text{ó} \quad a_{ij}, \quad \text{ó} \quad \mathbf{A}_{|i|j}.$$

Pero ¿qué pasa si nos queremos referir a toda la fila i -ésima o toda la columna j -ésima de \mathbf{A} ?

1.4.1. Filas de una matriz

La fila i -ésima de \mathbf{A} (de orden $m \times n$) es el *vector* de \mathbb{R}^n cuyos componentes son a_{ik} , donde el índice k recorre todas las columnas ($k = 1 : n$) y que denotamos como:

$$\text{fila}_i(\mathbf{A}) \equiv \mathbf{A}_{|i} = (a_{i1}, \dots, a_{in}) \in \mathbb{R}^n$$

(que por ser un vector de \mathbb{R}^n se puede escribir tanto en horizontal como en vertical).

El operador selector de filas: “ $|i$ ”. Al escribir “ $|i$ ” como subíndice a la *izquierda* de la matriz seleccionamos la *fila* i -ésima de la matriz.

Por ejemplo, si $\mathbf{B} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & 7 \end{bmatrix}$, entonces $\mathbf{B}_{|1} = (1, 2, 3) = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$.

<code>1 B</code>	<code># primera fila de B</code>	Librería NAcAL para Python
$(1, 2, 3)$		

1.4.2. Columnas de una matriz

La columna j -ésima de \mathbf{A} (de orden $m \times n$) es el *vector* de \mathbb{R}^m con componentes a_{kj} , donde el índice k recorre todas las filas ($k = 1 : m$). Denotaremos dicha columna como $\text{col}_j(\mathbf{A})$ o bien como $\mathbf{A}_{|j}$:

$$\text{col}_j(\mathbf{A}) \equiv \mathbf{A}_{|j} = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix} = (a_{1j}, \dots, a_{mj}) \in \mathbb{R}^m,$$

(que por ser un vector de \mathbb{R}^m se puede escribir tanto en horizontal como en vertical).

El operador selector de columnas: “ $|j$ ”. Al escribir “ $|j$ ” como subíndice a la *derecha* de la matriz, seleccionamos la *columna* j -ésima de la matriz.

Así, si $\mathbf{B} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & 7 \end{bmatrix}$; entonces $\mathbf{B}_{|2} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$.

<code>B 2</code>	<code># segunda columna de B</code>	Librería NAcAL para Python
------------------	-------------------------------------	----------------------------

¿Y cómo escribir la submatriz de \mathbf{B} de orden $m \times 1$ cuya única columna es la j -ésima de \mathbf{B} ? Basta encerrar entre “corchetes” dicha columna:

$$[\mathbf{B}_{|2}] = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

```
Matrix( [ B|2 ] ) # Matriz cuya única columna es la segunda columna de B
```

$$\begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

Nótese que $\begin{pmatrix} 2 \\ -1 \end{pmatrix}$ es un *vector* (cuya lista contiene dos números que se puede escribir en vertical u horizontal), pero que $\begin{bmatrix} 2 \\ -1 \end{bmatrix}$ es una *matriz* cuya lista contiene una única *columna*. Vectores y matrices son objetos matemáticos distintos (lista de números *vs* lista de vectores). Gracias a la notación adoptada serán fácilmente distinguibles: los vectores se encierran entre “paréntesis” y matrices entre “corchetes”.

Así $\mathbf{a} = (1, 2, 3)$ es un vector y $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$ es una matriz (y por tanto **son objetos distintos**).¹⁴

Con la notación que hemos establecido podemos describir una matriz mediante sus columnas:

$$\mathbf{A}_{m \times n} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{A}_{|1;} & \mathbf{A}_{|2;} & \dots & \mathbf{A}_{|n;} \end{bmatrix}}_{\text{lista de columnas}}; \text{ donde } \mathbf{A}_{|j} = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}.$$

Diremos que *dos matrices* del mismo orden *son iguales* si y solo si lo son sus correspondientes sistemas de vectores, es decir: $\mathbf{A} = \mathbf{B}$ si y solo si $\mathbf{A}_{|j} = \mathbf{B}_{|j}$ para $j = 1 : n$ (i.e. si son iguales columna a columna).

1.4.3. Resumen de la notación para seleccionar componentes, filas y columnas

Disponemos de varias alternativas para denotar tanto los componentes, como las filas y columnas de $\mathbf{A}_{n \times m}$.

Notaciones alternativas		
Componente ij (escalar)	Fila i -ésima (vector de \mathbb{R}^n)	Columna j -ésima (vector de \mathbb{R}^m)
a_{ij}	$\mathbf{A}_{ i}$	$\mathbf{A}_{ j}$
$\text{elem}_{ij}(\mathbf{A})$	$\text{fila}_i(\mathbf{A})$	$\text{col}_j(\mathbf{A})$

En el ejemplo con datos económicos, $_{1|}\mathbf{PIB}$ (primera fila) corresponde a los datos de España y $_{5|}\mathbf{PIB}$ (quinta fila) a los de EEUU; y por otra parte, $\mathbf{PIB}_{|3}$ (tercera columna) corresponde a los datos del año 2005.

¹⁴Nota para profesores de matemáticas y/o lectores “puntillosos”. Tradicionalmente se sigue el siguiente convenio

$$\begin{cases} \mathbb{R} = \mathbb{R}^1 \\ \mathbb{R}^{n+1} = \mathbb{R} \times \mathbb{R}^n \end{cases}$$

Sin embargo, aquí sigo un convenio alternativo acorde con el funcionamiento interno de Python:

$$\begin{cases} \mathbb{R}^0 = \{\emptyset\} \\ \mathbb{R}^{n+1} = \mathbb{R} \times \mathbb{R}^n \end{cases}$$

(donde \emptyset denota al conjunto vacío). De tal manera que $\mathbb{R}^1 = \mathbb{R} \times \{\emptyset\} = \{(\alpha, \emptyset) \mid \alpha \in \mathbb{R}\}$. Así pasa en Python, y por ello una tupla con un único elemento se escribe con $(\mathbf{a},)$. Nótese la coma detrás de la \mathbf{a} (que se puede entender que separa el primer elemento, \mathbf{a} , del último, que es vacío). De este modo la selección de los elementos sigue la siguiente función recursiva: si $\mathbf{v} = (v_1, \mathbf{w}) \in \mathbb{R}^n$, donde $\mathbf{w} \in \mathbb{R}^{n-1}$ entonces

$$\begin{cases} (v_1, \mathbf{w})_{|1} = v_1 \\ (v_1, \mathbf{w})_{|k+1} = \mathbf{w}_{|k} \end{cases}$$

Por tanto, $\mathbf{a} = (1, 2, 3)$ es un sistema de tres *números* de \mathbb{R} , y $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$ es un sistema de tres *vectores* de \mathbb{R}^1 , a saber $(1,)$, $(2,)$ y $(3,)$. Es decir, la matriz \mathbf{A} escrita con la notación genérica de sistema es $\mathbf{A} = [(1,); (2,); (3,)]$.



La notación de las matrices tiene paralelismos con la de los vectores; **pero con importantes diferencias**.

Una lista de vectores encerrada entre *corchetes* denota una matriz, cuyas *columnas* son los vectores de la lista en el orden en el que aparecen en la lista dentro del corchete. Por ejemplo, si

$$\mathbf{a} = (1, 2,), \quad \mathbf{b} = (1, 0,), \quad \mathbf{c} = (9, 2,); \quad \text{entonces} \quad \mathbf{A} = [\mathbf{a} \ \mathbf{b} \ \mathbf{c}] = \begin{bmatrix} 1 & 1 & 9 \\ 2 & 0 & 2 \end{bmatrix}.$$

Con las matrices **NO hay la libertad** para escribir horizontal o verticalmente la lista que hemos encerrado (es una importante diferencia con los vectores).

Los subíndices “ i ” (por la izquierda) y “ j ” (por la derecha) operan sobre la matriz, pero su comportamiento es distinto. El subíndice “ i ” **a la izquierda de la matriz selecciona la fila** i -ésima y el subíndice “ j ” **a la derecha selecciona la columna** j -ésima (tanto las filas como las columnas son *vectores*):

$$\mathbf{A}_{|3} = \begin{pmatrix} 9 \\ 2 \end{pmatrix}, \quad {}_2|\mathbf{A} = (2, \quad 0, \quad 2).$$

Así, $\mathbf{A}_{|3}$ es un *vector* formado por las componentes de la tercera columna de \mathbf{A} ; pero, $[\mathbf{A}_{|3}]$ es una *matriz columna* cuya única columna es igual que la tercera columna de \mathbf{A} .

$$\mathbf{A}_{|3} = \mathbf{c} = (9, \quad 2,) = \begin{pmatrix} 9 \\ 2 \end{pmatrix} \neq [\mathbf{A}_{|3}] = [\mathbf{c}] = \begin{bmatrix} 9 \\ 2 \end{bmatrix} \neq \begin{bmatrix} 9 & 2 \end{bmatrix}.$$

Cuando se emplean dos subíndices (uno por cada lado) se selecciona la *componente* situada en la fila i -ésima (indicada por el índice de la izquierda) y columna j -ésima (indicada por el índice de la derecha)

$${}_{1|}\mathbf{A}_{|3} = a_{13} = 9.$$

1.4.4. Extensión de la notación vectorial y reglas de reescritura.

Nótese que ${}_i|\mathbf{A}_{|j}$ es la componente j -ésima de la fila i , es decir $({}_i|\mathbf{A})_{|j}$, pero también es la componente i -ésima de la columna j , por lo que sería conveniente poder escribir ${}_i|(\mathbf{A}_{|j})$, de manera que

$${}_i|\mathbf{A}_{|j} = {}_i|(\mathbf{A}_{|j}) = ({}_i|\mathbf{A})_{|j}.$$

Para ello aceptaremos que el operador “ $|$ ” también pueda operar por la izquierda de un *vector*. Así

$${}_i|\mathbf{a} = \mathbf{a}_{|i} = a_i.$$

Con esta flexibilización de la notación, y puesto que tanto las filas como las columnas de una matriz son vectores, cuando seleccionemos las *componentes* de \mathbf{A} , nos dará igual operar dos veces por el mismo lado o una vez por cada lado (ambas operaciones arrojan necesariamente el mismo resultado). Así

$${}_i|(\mathbf{A}_{|j}) = (\mathbf{A}_{|j})_{|i} \quad \text{y} \quad ({}_i|\mathbf{A})_{|j} = {}_j|({}_i|\mathbf{A}).$$

Para no perder la *asociatividad* de la notación cuando “ $|$ ” opera por la izquierda de los *vectores*, también aceptaremos que los escalares aparezcan multiplicando por la derecha; es decir,¹⁵ $\lambda \mathbf{b} = \mathbf{b} \lambda$. Esto da lugar (jugando con multiplicar por la derecha o seleccionar por la izquierda) a dos nuevas reglas de reescritura:

¹⁵Como en estas notas propongo una notación propia, aquí aceptaremos una expresión como “ $\mathbf{a}2$ ”... ¡expresión que nunca aparece en los textos! **El modo habitual de escribir dicho producto es “ $2\mathbf{a}$ ” con “el coeficiente primero”**. Aquí lo aceptaremos para mantener la asociatividad por la izquierda, pero tenga en cuenta que no es una escritura aceptada generalmente.



Primera. - “*podemos desplazar los paréntesis*” para sacar un símbolo por un lado e introducir otro por el otro lado:

$$(\lambda \mathbf{b})_{|i} = \lambda(\mathbf{b}_{|i}) = \left({}_i \mathbf{b} \right) \lambda = {}_i (\mathbf{b} \lambda).$$

Segunda. - “*podemos intercambiar las posiciones del escalar y el selector*” sacando uno de ellos fuera del paréntesis y metiendo el otro:

$$(\mathbf{b}_{|i}) \lambda = (\mathbf{b} \lambda)_{|i} \quad \text{y} \quad \lambda({}_i \mathbf{b}) = {}_i (\lambda \mathbf{b}).$$

De hecho, puesto que las ocho expresiones de más arriba arrojan el mismo resultado (aunque el orden de ejecución de las operaciones difiera entre ellas) podemos omitir el paréntesis y escribir sencillamente $\lambda \mathbf{b}_{|i}$ o también cualquiera de las siguientes reordenaciones $\lambda \mathbf{b}_{|i} = \lambda {}_i \mathbf{b} = \mathbf{b}_{|i} \lambda = {}_i \mathbf{b} \lambda$.

No obstante a lo anterior, en las demostraciones siempre usaré expresiones con paréntesis, pues ayudan a identificar la propiedad empleada en cada paso de la demostración.

Recuerde que en Python también es conveniente usar paréntesis para evitar errores y sorpresas.

EJERCICIO 5. ¿Cómo evalúa el módulo NAcAL las siguientes expresiones? (Revise la Tabla 1.2 en la página 7)

- | | |
|--------------------------|--------------------------|
| (a) $3 * \mathbf{b} 1$ | (b) $3 * 1 \mathbf{b}$ |
| (c) $1 * 3 \mathbf{b}$ | (d) $1 \mathbf{b} * 3$ |

(Recuerde que poner paréntesis nos aclara el orden en que se realizan las operaciones).

1.4.5. La transposición (o trasposición)

Encerrando \mathbf{a} entre corchetes denotamos una matriz con una única columna; para denotar una matriz cuya única fila es \mathbf{a} necesitamos una operación adicional: la *transposición*, (\top),...

Definición 1.10 (Matriz transpuesta (o traspuesta)). Sea \mathbf{A} de orden m por n , la matriz transpuesta de \mathbf{A} , que denotamos con \mathbf{A}^\top , es la matriz de orden $n \times m$ cuyas m columnas son iguales a las m filas de \mathbf{A} :

$$(\mathbf{A}^\top)_{|i} = {}_i \mathbf{A}; \quad i = 1 : m.$$

Es decir, la transpuesta de \mathbf{A} es la matriz n por m cuyas columnas son las filas de \mathbf{A} escritas en vertical

$m \times n$

$$\mathbf{A}^\top = \left[{}_1 \mathbf{A}; \quad \dots \quad {}_m \mathbf{A}; \right].$$

Ejemplo: $\mathbf{A} = \begin{bmatrix} 1 \\ 0 \\ 5 \\ 1 \end{bmatrix} \Rightarrow \mathbf{A}^\top = \begin{bmatrix} 1 & 0 & 5 & 1 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & 7 \end{bmatrix} \Rightarrow \mathbf{B}^\top = \begin{bmatrix} 1 & 0 \\ 2 & -1 \\ 3 & 7 \end{bmatrix}.$

`~B`

`# Transpuesta de la matriz B`

Librería NAcAL para Python

$$\begin{bmatrix} 1 & 0 \\ 2 & -1 \\ 3 & 7 \end{bmatrix}$$

Escribimos una *matriz columna* (matriz con una única columna) poniendo un vector entre corchetes; ahora podemos escribir una *matriz fila* (matriz con una única fila) trasponiendo una matriz columna.

Así $[i|\mathbf{A}]^\top$ es una matriz cuya única fila es igual a la fila i -ésima de \mathbf{A} ; y si

$$\mathbf{B} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & 7 \end{bmatrix}; \quad \text{entonces} \quad \begin{cases} [1|\mathbf{B}]^\top = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \\ [\mathbf{B}|_3]^\top = \begin{bmatrix} 3 & 7 \end{bmatrix} \end{cases}.$$

Librería NAcAL para Python

```
~Matrix( [1|B] )    # Matriz cuya única fila es la primera fila de B
```

$$\begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$$

Librería NAcAL para Python

```
~Matrix( [B|3] )    # Matriz cuya única fila es la tercera columna de B
```

$$\begin{bmatrix} 3 & 7 \end{bmatrix}$$

EJERCICIO 6. Demuestre las siguientes propiedades elementales de la transposición:

- (a) La transposición intercambia los índices de las componentes de la matriz: $_{k|}(\mathbf{A}^\top)_{|j} = _{j|}\mathbf{A}_{|k}$.
- (b) $(\mathbf{A}^\top)^\top = \mathbf{A}$.
- (c) Si $\mathbf{A}_{m \times n}$ y $1 \leq i \leq n$ entonces $\mathbf{A}_{|i} = _{i|}(\mathbf{A}^\top)$.



Pese a que las expresiones $_{i|}(\mathbf{A}^\top)$, $(\mathbf{A}^\top)_{|j}$ y $_{i|}(\mathbf{A}^\top)_{|j}$ solo tienen una interpretación posible incluso si se quitaran los paréntesis (puesto que *la transposición es una operación sobre matrices* y no sobre vectores) siempre escribiré el paréntesis. Así la notación es más clara.

Fíjese además en las reglas de reescritura: “*puedo transponer (o quitar la transpuesta) si a la vez cambio de lado los subíndices*” (fíjese en su aplicación en las demostraciones de las propiedades anteriores).

1.4.6. Definiciones de algunas matrices especiales



Al igual que hicimos con los vectores, añadiremos algunas definiciones relativas a las matrices. *Preste atención al uso de la notación.*

Definición 1.11. Llamamos *matriz nula* (o *matriz cero*) de orden $m \times n$, $\mathbf{0}$, a la matriz cuyas n columnas son vectores nulos de \mathbb{R}^m : $\mathbf{0}_{|j} = \mathbf{0} \in \mathbb{R}^m$ para $j = 1 : n$.

Librería NAcAL para Python

```
M0(2,4)    # matriz nula de orden 2 por 4
```

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Definición 1.12. Llamamos *matriz opuesta* de \mathbf{A} a la matriz $-\mathbf{A}$, del mismo orden que \mathbf{A} , cuyas columnas son las de \mathbf{A} multiplicadas por -1 , es decir, $(-\mathbf{A})_{|j} = -(\mathbf{A}_{|j})$.

-B

matriz opuesta de B

Librería NAcAL para Python

$$\begin{bmatrix} -1 & -2 & -3 \\ 0 & 1 & -7 \end{bmatrix}$$

También daremos nombre a ciertas matrices en función de la disposición de sus componentes:

Definición 1.13. Decimos que una matriz es *cuadrada* cuando tiene el mismo número de filas que de columnas. Es decir, una matriz cuadrada es de orden $n \times n$.

Cuando tenemos una matriz cuadrada de orden n por n abreviamos y decimos sencillamente que es de *orden n* . Así, si decimos que una matriz es *de orden n* (y nada más) estamos indicando que es cuadrada.

Definición 1.14. A las matrices que no son cuadradas las denominamos *rectangulares*.

Recuerde: al expresar el orden de matrices rectangulares es necesario indicar el número de filas y columnas.

Definición 1.15 (Matriz simétrica). Decimos que la matriz \mathbf{A} es *simétrica* cuando $\mathbf{A} = \mathbf{A}^\top$.

Así, $\mathbf{A}_{|j} = (\mathbf{A}^\top)_{|j} = {}_j|\mathbf{A}$. Por tanto, toda matriz simétrica es cuadrada. Y puesto que la “columna” j -ésima es igual a la “fila” j -ésima, las componentes i -ésimas de dichos vectores son iguales: ${}_i|\mathbf{A}_{|j} = {}_j|\mathbf{A}_{|i}$.

Definición 1.16. Decimos que una matriz es *diagonal* cuando todos los componentes fuera de la diagonal principal son nulos, es decir, $d_{ij} = 0$ cuando $i \neq j$.

Por ejemplo

$$\mathbf{D} = \begin{bmatrix} d_{11} & 0 & \cdots & 0 \\ 0 & d_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_{mm} \end{bmatrix}; \quad \mathbf{F} = \begin{bmatrix} f_{11} & 0 & \cdots & 0 & 0 \\ 0 & f_{22} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & f_{mp} & 0 \end{bmatrix}; \quad \text{y} \quad \mathbf{G} = \begin{bmatrix} g_{11} & 0 & \cdots & 0 \\ 0 & g_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & g_{nn} \\ 0 & 0 & \cdots & 0 \end{bmatrix}.$$

Fíjese que de las tres matrices diagonales, solo \mathbf{D} es cuadrada. Tenga en cuenta además, que los componentes de la diagonal principal pueden tomar cualquier valor. Así pues, ¡toda matriz nula $\mathbf{0}$ es diagonal! Además, *toda matriz diagonal y cuadrada es simétrica*, ya que por una parte

$${}_i|\mathbf{D}_{|j} = \begin{cases} 0 & (i \neq j) \\ d_{ii} & (i = j) \end{cases}, \quad \text{y por otra} \quad {}_i|(\mathbf{D}^\top)_{|j} = {}_j|\mathbf{D}_{|i} = \begin{cases} 0 & (i \neq j) \\ d_{ii} & (i = j) \end{cases}; \quad \text{por tanto} \quad \mathbf{D}^\top = \mathbf{D}.$$

Definición 1.17. Llamamos matriz *identidad* (que denotamos con \mathbf{I}) a la matriz cuadrada y diagonal cuyas componentes en la diagonal principal son unos, es decir

$${}_i|\mathbf{I}_{|j} = \begin{cases} 0 & (i \neq j) \\ 1 & (i = j) \end{cases}.$$

Por ejemplo, las matrices identidad de órdenes 2 y 1 son: $\mathbf{I}_{2 \times 2} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; \quad \mathbf{I}_{1 \times 1} = [1]$.

I(2)

matriz identidad de orden 2

Librería NAcAL para Python

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Nótese que las matrices identidad son simétricas (pues son cuadradas y diagonales).

1.5. Suma de matrices y producto de matrices por escalares

Definición 1.18 (suma de matrices). Definimos la suma de dos matrices \mathbf{A} y \mathbf{B} , de orden $m \times n$, como la matriz del mismo orden cuya columna j -ésima es la suma de la columna j -ésima de \mathbf{A} y la j -ésima de \mathbf{B} :

$$(\mathbf{A} + \mathbf{B})_{|j} = \mathbf{A}_{|j} + \mathbf{B}_{|j} \quad \text{para } j = 1 : n,$$

es decir, $\mathbf{A} + \mathbf{B} = \left[(\mathbf{A}_{|1} + \mathbf{B}_{|1}); \dots (\mathbf{A}_{|n} + \mathbf{B}_{|n}); \right]$.

Definición 1.19 (producto de una matriz por un escalar). Definimos el producto de una matriz \mathbf{A} por un escalar λ como la matriz resultante de multiplicar las columnas de \mathbf{A} por el escalar λ : $m \times n$

$$(\lambda \mathbf{A})_{|j} = \lambda (\mathbf{A}_{|j}) \quad \text{para } j = 1 : n,$$

es decir, $\lambda \mathbf{A} = \left[\lambda (\mathbf{A}_{|1}); \dots \lambda (\mathbf{A}_{|n}); \right]$.

Definición 1.20. Para cualquier λ decimos que $\lambda \mathbf{A}$ es un múltiplo de \mathbf{A} .



De nuevo la definición de las operaciones de suma y producto por un escalar convierten al operador “ $|j$ ” en un **operador lineal**; es decir, *distributivo respecto a la suma de matrices*

$$(\mathbf{A} + \mathbf{B})_{|j} = \mathbf{A}_{|j} + \mathbf{B}_{|j}$$

y *asociativo respecto al producto de una matriz por un escalar*

$$(\lambda \mathbf{A})_{|j} = \lambda (\mathbf{A}_{|j}).$$

1.5.1. Propiedades de la suma de matrices y del producto por un escalar

La suma de matrices y producto de matrices por escalares verifican propiedades análogas a las de la Proposición 1.2.1; y su demostración es idéntica a la del Ejercicio 4.

EJERCICIO 7. Demuestre las propiedades de la siguiente proposición.

Proposición 1.5.1 (Propiedades de las operaciones entre matrices). Para cualesquiera matrices \mathbf{A} , \mathbf{B} y \mathbf{C} de idéntico orden y para cualesquiera escalares λ y φ , se verifica que:

- | | |
|--|---|
| 1. $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$. | 5. $\lambda(\mathbf{A} + \mathbf{B}) = \lambda\mathbf{A} + \lambda\mathbf{B}$. |
| 2. $\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$. | 6. $(\lambda + \varphi)\mathbf{A} = \lambda\mathbf{A} + \varphi\mathbf{A}$. |
| 3. $\mathbf{0} + \mathbf{A} = \mathbf{A}$. | 7. $\lambda(\varphi\mathbf{A}) = (\lambda\varphi)\mathbf{A}$. |
| 4. $\mathbf{A} + (-\mathbf{A}) = \mathbf{0}$. | 8. $1\mathbf{A} = \mathbf{A}$. |

Pista. Emplee la linealidad del operador “ $|j$ ” (es decir, que es distributivo para la suma de matrices y asociativo para el producto por escalares). Use las *propiedades de las operaciones con vectores* de la Proposición 1.2.1 (ya que $\mathbf{A}_{|j}$, $\mathbf{B}_{|j}$ y $\mathbf{C}_{|j}$ son vectores de \mathbb{R}^n).

Nota 2. Estas ocho propiedades (de manera similar a las ocho propiedades de las operaciones con vectores de la Proposición 1.2.1) nos permitirán definir en la Lección 6 el espacio vectorial de matrices $\mathbb{R}^{m \times n}$.

1.5.2. Operaciones componente a componente

La mayoría de manuales definen la suma y el producto por escalares componente a componente:

“la suma de dos matrices \mathbf{A} y \mathbf{B} (del mismo orden) es la matriz que resulta de sumar los componentes de \mathbf{A} a los componentes de \mathbf{B} ”.

y

“el producto de una matriz \mathbf{A} por un escalar λ : es la matriz resultante de multiplicar los componentes de \mathbf{A} por el escalar λ ”.

EJERCICIO 8. Demuestre que las definiciones 1.18 y 1.19 implican que las operaciones de suma y producto por escalares se pueden calcular componente a componente; es decir, que:

$${}_i(\mathbf{A} + \mathbf{B})_{|j} = {}_i\mathbf{A}_{|j} + {}_i\mathbf{B}_{|j}, \quad \text{y} \quad {}_i(\lambda\mathbf{A})_{|j} = \lambda({}_i\mathbf{A}_{|j}); \quad \text{con } i = 1 : m, \quad \text{y} \quad j = 1 : n.$$

Por tanto,

$$\begin{aligned} \mathbf{A} + \mathbf{B} &= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix} \\ &= \begin{bmatrix} (a_{11} + b_{11}) & (a_{12} + b_{12}) & \cdots & (a_{1n} + b_{1n}) \\ \vdots & \vdots & \ddots & \vdots \\ (a_{m1} + b_{m1}) & (a_{m2} + b_{m2}) & \cdots & (a_{mn} + b_{mn}) \end{bmatrix}; \end{aligned}$$

y

$$\lambda\mathbf{A} = \lambda \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} \lambda(a_{11}) & \lambda(a_{12}) & \cdots & \lambda(a_{1n}) \\ \vdots & \vdots & \ddots & \vdots \\ \lambda(a_{m1}) & \lambda(a_{m2}) & \cdots & \lambda(a_{mn}) \end{bmatrix}.$$

1.5.3. La transposición es un operador lineal

Antes de mostrar las operaciones por filas necesitamos demostrar las propiedades de linealidad de la transposición.¹⁶

EJERCICIO 9. Demuestre las siguientes proposiciones.

(a) **Proposición 1.5.2** (Transpuesta de un múltiplo). Sea \mathbf{A} entonces $(\lambda\mathbf{A})^\top = \lambda(\mathbf{A}^\top)$.

(b) **Proposición 1.5.3** (Transpuesta de una suma). Sean \mathbf{A} y \mathbf{B} entonces $(\mathbf{A} + \mathbf{B})^\top = \mathbf{A}^\top + \mathbf{B}^\top$.

☞ Así pues, el operador transposición es un operador lineal.

¹⁶De la transposición ya solo nos falta demostrar que $(\mathbf{AB})^\top = (\mathbf{B}^\top)(\mathbf{A}^\top)$, pero para ello antes tenemos que ver las definiciones de producto de un vector por una matriz \mathbf{aB} y producto de matrices \mathbf{AB} .

1.5.4. Operaciones fila a fila

Y ahora veamos las operaciones por filas. Si de nuevo aceptamos el producto de un escalar por el lado derecho, $\mathbf{A}\lambda = \lambda\mathbf{A}$, entonces el operador selector por la izquierda también es asociativo:

EJERCICIO 10. Demuestre que las definiciones 1.18 y 1.19 implican que se puede operar “por filas”, es decir:

$${}_i|(\mathbf{A} + \mathbf{B}) = {}_i|\mathbf{A} + {}_i|\mathbf{B}, \quad \text{y} \quad {}_i|(\mathbf{A}\lambda) = ({}_i|\mathbf{A})\lambda; \quad \text{donde } \mathbf{A} \text{ y donde } i = 1 : m.$$

$m \times n$



Por tanto el operador selector de *filas*, “ ${}_i|$ ”, es **lineal**; es decir, *distributivo* respecto a la suma de matrices

$${}_i|(\mathbf{A} + \mathbf{B}) = {}_i|\mathbf{A} + {}_i|\mathbf{B}$$

y *asociativo* respecto al producto de una matriz por un *escalar*

$${}_i|(\lambda\mathbf{A}) = \lambda({}_i|\mathbf{A}).$$

1.6. Extensión de la notación matricial y las reglas de reescritura

Así que disponemos de *reglas de reescritura* similares a las que obtuvimos en el caso del los vectores.



Reglas de re-escritura: como el operador selector de filas es lineal, podemos

- “*distribuir el operador*” entre sumandos

$$(\mathbf{A} + \mathbf{B})_{|j} = \mathbf{A}_{|j} + \mathbf{B}_{|j} \quad \text{y} \quad {}_i|(\mathbf{A} + \mathbf{B}) = {}_i|\mathbf{A} + {}_i|\mathbf{B};$$

- “*desplazar los paréntesis*” para sacar un símbolo por un lado e introducir otro por el otro lado:

$$(\lambda\mathbf{A})_{|j} = \lambda(\mathbf{A}_{|j}) \quad \text{y} \quad ({}_i|\mathbf{A})\lambda = {}_i|(\mathbf{A}\lambda).$$

Y además también podemos

- “*intercambiar las posiciones del escalar y el selector*” dentro y fuera del paréntesis:

$$(\mathbf{A}_{|j})\lambda = (\mathbf{A}\lambda)_{|j} \quad \text{y} \quad {}_i|(\lambda\mathbf{A}) = \lambda({}_i|\mathbf{A}).$$

Como las cuatro últimas igualdades arrojan el mismo resultado (pese al distinto orden de ejecución de las operaciones) podemos omitir el paréntesis y escribir tanto $\lambda\mathbf{A}_{|j}$ como ${}_i|\mathbf{A}\lambda$. En aras de la claridad, *en las demostraciones siempre usaré los paréntesis para destacar qué regla o propiedad se aplica en cada paso*, pero en otras ocasiones es práctico omitirlos si no son necesarios (*¡esa es la ventaja de la notación asociativa!*).¹⁷

La Proposición 1.5.2 sobre la “transpuesta de un múltiplo” (que volvemos a copiar como primer punto del recuadro de más abajo) añadida a permitir multiplicar por la derecha, $\mathbf{A}\lambda = \lambda\mathbf{A}$; arroja otra regla de reescritura que permite intercambiar la posición entre el escalar y el símbolo de trasposición.

¹⁷Aunque debe recordar que, dada la precedencia establecida en la ejecución de las operaciones en Python, programando con el ordenador siempre es conveniente poner los paréntesis.



Reglas de re-escritura: podemos

- “*intercambiar las posiciones del escalar y el operador transposición*” dentro y fuera del paréntesis:

$$(\mathbf{A}^\top)\lambda = (\mathbf{A}\lambda)^\top.$$

Además, como el **operador transposición es lineal**, también podemos

- “*desplazar los paréntesis*” para sacar un símbolo por un lado e introducir otro por el otro lado:

$$(\lambda\mathbf{A})^\top = \lambda(\mathbf{A}^\top) \quad (\text{y que podemos escribir sencillamente como } \lambda\mathbf{A}^\top)$$

- “*distribuir el operador*” en una suma

$$(\mathbf{A} + \mathbf{B})^\top = \mathbf{A}^\top + \mathbf{B}^\top$$

Apendices a la lección

1.A. Subvectores mediante selección de una lista de índices

Sea \mathbf{b} un vector de \mathbb{R}^n y sea α una lista de índices menores o iguales a n . Definimos $\mathbf{b}_{|\alpha}$ como el vector (de tantos componentes como *elementos en* α) cuyos elementos son los indicados en la lista α . Es decir,

$$\text{si } \mathbf{b} \in \mathbb{R}^n, \text{ entonces; } \mathbf{b}_{|(i,j,\dots,p)} = (a_i, a_j, \dots, a_p); \quad \text{para } i, j, \dots, p \leq n.$$

Por ejemplo, si $\mathbf{b} = (10, 20, 30, 40,)$ y $\alpha = (3, 3, 1)$, entonces $\mathbf{b}_{|\alpha}$ es el vector de \mathbb{R}^3 cuyas dos primera componentes son como la tercera de \mathbf{b} y su última componente como la primera componente de \mathbf{b} , es decir

$$\mathbf{b}_{|\alpha} = (30, 30, 10,).$$

Evidentemente, para todo vector $\mathbf{b} \in \mathbb{R}^n$ se verifica que $\mathbf{b} = \mathbf{b}_{|(1:n)}$.

1.B. Submatrices mediante selección de una lista de índices

De manera similar, sea \mathbf{I} la matriz identidad de orden n y sea α una lista de índices menores o iguales a n , por ejemplo, $n = 4$ y $\alpha = (2, 3, 4,)$. Definimos $\mathbf{I}_{|\alpha}$ como la matriz (de n filas y tantas columnas como *elementos en* α) cuyas columnas son las columnas de \mathbf{I} indicadas en la lista α . Por ejemplo

$$\mathbf{I}_{|(2,3,4,)} = [\mathbf{I}_{|2} \quad \mathbf{I}_{|3} \quad \mathbf{I}_{|4}] \neq [\mathbf{I}_{|4} \quad \mathbf{I}_{|2} \quad \mathbf{I}_{|3}] = \mathbf{I}_{|(4,2,3,)}.$$

Y definimos ${}_{\alpha}|\mathbf{I}$ como: ${}_{\alpha}|\mathbf{I} = (\mathbf{I}_{|\alpha})^\top$.

Selección de una submatriz por filas o por columnas Ahora con ayuda del producto de matrices podemos definir dos nuevas operaciones. Sea α una lista de índices, entonces podemos seleccionar una lista de columnas (o filas) de una matriz \mathbf{A} para construir una submatriz:

$$\mathbf{A}_{|\alpha} = \mathbf{A}(\mathbf{I}_{|\alpha}), \quad \text{y} \quad {}_{\alpha}|\mathbf{A} = ({}_{\alpha}|\mathbf{I})\mathbf{A}.$$

Consideremos de nuevo la matriz **PIB** de la Página 11. La submatriz con las filas de 2 a 4 de **PIB** (datos de Francia, Alemania y Reino Unido) o la submatriz con sus dos primeras columnas (años 2003 y 2004) son

$${}_{(2,3,4,)}|\mathbf{PIB} = \begin{bmatrix} 0.9 & 2.0 & 1.6 & 2.0 \\ -0.2 & 1.1 & 0.7 & 0.9 \\ 2.5 & 3.2 & 1.9 & 1.5 \end{bmatrix}; \quad \text{y} \quad \mathbf{PIB}_{|(1,2,)} = \begin{bmatrix} 2.9 & 3.1 \\ 0.9 & 2.0 \\ -0.2 & 1.1 \\ 2.5 & 3.2 \\ 1.8 & 3.1 \end{bmatrix}.$$

Y podemos crear submatrices seleccionando simultáneamente filas y columnas; si β es otra lista de índices:

$${}_{\alpha|}(\mathbf{A}_{|\beta}) = ({}_{\alpha|}\mathbf{I})\mathbf{A}(\mathbf{I}_{|\beta}) = ({}_{\alpha|}\mathbf{A})_{|\beta},$$

que, aprovechando la asociatividad, podemos escribir sencillamente como: ${}_{\alpha|}\mathbf{A}_{|\beta}$.

Así, la submatriz con los datos de Francia, Alemania y Reino Unido de los años 2003 y 2004 es

$${}_{(2,3,4,)|}\mathbf{PIB}_{|(1,2,)} = \begin{bmatrix} 0.9 & 2.0 \\ -0.2 & 1.1 \\ 2.5 & 3.2 \end{bmatrix}.$$

Nótese que $({}_{\alpha|}\mathbf{A})(\mathbf{B}_{|\beta}) = ({}_{\alpha|}\mathbf{I})\mathbf{AB}(\mathbf{I}_{|\beta}) = {}_{\alpha|}(\mathbf{AB})_{|\beta}$; en particular $(\mathbf{AB})_{|\beta} = \mathbf{A}(\mathbf{B}_{|\beta})$ y ${}_{\alpha|}(\mathbf{AB}) = ({}_{\alpha|}\mathbf{A})\mathbf{B}$; por lo que podemos escribir sencillamente: ${}_{\alpha|}\mathbf{AB}_{|\beta}$. También se verifica que

$${}_{\alpha|}(\mathbf{A} + \mathbf{B})_{|\beta} = {}_{\alpha|}\mathbf{A}_{|\beta} + {}_{\alpha|}\mathbf{B}_{|\beta} \quad \text{y} \quad \lambda({}_{\alpha|}\mathbf{A}_{|\beta}) = {}_{\alpha|}(\lambda\mathbf{A})_{|\beta}. \quad (1.1)$$

Definición 1.21. Llamamos **submatriz principal** de \mathbf{A} a cualquier submatriz de la forma ${}_{\alpha|}\mathbf{A}_{|\alpha}$; donde α es una lista de k índices tal que $1 \leq \alpha_1 < \alpha_2 < \cdots < \alpha_k \leq n$. Es decir, los índices son distintos, ordenados y comunes en la selección de filas y columnas. El caso particular en el que se seleccionan las k primeras filas y columnas: ${}_{(1:k)|}\mathbf{A}_{|(1:k)}$ se denomina **submatriz principal superior**.

Hiperenlaces a los Notebooks de Jupyter en mybinder.org correspondientes a esta lección.

- **Notebook** Sección 1.1
- **Notebook** Sección 1.2
- **Notebook** Sección 1.4
- **Notebook** Sección 1.5

La url de acceso a todos los Notebooks de Jupyter en mybinder.org es:

- <https://mybinder.org/v2/gh/mbujosab/nacal-jupyter-notebooks/master/>

Combinaciones lineales

2.1. Producto punto (o producto escalar usual en \mathbb{R}^n)

Antes de definir las combinaciones lineales, definamos el producto punto entre dos vectores de \mathbb{R}^n :

Definición 2.1. El *producto punto* $\mathbf{a} \cdot \mathbf{b}$ (o producto escalar usual en \mathbb{R}^n) de dos vectores \mathbf{a} y \mathbf{b} es

$$\mathbf{a} \cdot \mathbf{b} = a_1b_1 + a_2b_2 + a_3b_3 + \cdots + a_nb_n = \sum_{i=1}^n a_ib_i = \sum_{i=1}^n \mathbf{a}_{|i} \mathbf{b}_{|i}.$$

Este producto será muy importante en la segunda parte del curso para definir la ortogonalidad. Pero su introducción en este momento nos permitirá dotar de propiedades muy potentes a la notación del producto de matrices con vectores, además de lograr demostraciones mucho más “limpias” (pues en lugar de sumatorios podremos usar productos punto) y con pasos más sencillos (tan solo habrá que aplicar repetidamente la linealidad tanto del producto punto como del operador selector).

Librería NAcAL para Python

```
a = Vector( (1, 2, 3, 4) )
b = Vector( (1,-1, 1,-1) )
a * b          # producto punto entre los vectores a y b
```

−2

Propiedades del producto punto

El producto punto (o producto escalar usual de \mathbb{R}^n) satisface los siguientes axiomas para cualesquiera vectores \mathbf{x} , \mathbf{y} y \mathbf{z} de \mathbb{R}^n y para cualquier escalar λ .

EJERCICIO 11. Demuestre que el producto punto cumple con los siguientes axiomas:

- (a) **Simetría:** $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$
- (b) **Linealidad respecto al primer argumento:**
 - 1. $(\lambda \mathbf{x}) \cdot \mathbf{y} = \lambda(\mathbf{x} \cdot \mathbf{y})$
 - 2. $(\mathbf{x} + \mathbf{y}) \cdot \mathbf{z} = \mathbf{x} \cdot \mathbf{z} + \mathbf{y} \cdot \mathbf{z}$
- (c) **Positivo:** $\mathbf{x} \cdot \mathbf{x} \geq 0$
- (d) **Definido:** $\mathbf{x} \cdot \mathbf{x} = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$.

Fíjese que como el producto punto es simétrico, también es lineal en el segundo argumento

$$\begin{aligned}\mathbf{x} \cdot (\lambda \mathbf{y}) &= (\lambda \mathbf{y}) \cdot \mathbf{x} = \lambda(\mathbf{y} \cdot \mathbf{x}) = \lambda(\mathbf{x} \cdot \mathbf{y}); \\ \mathbf{x} \cdot (\mathbf{y} + \mathbf{z}) &= (\mathbf{y} + \mathbf{z}) \cdot \mathbf{x} = \mathbf{y} \cdot \mathbf{x} + \mathbf{z} \cdot \mathbf{x} = \mathbf{x} \cdot \mathbf{y} + \mathbf{x} \cdot \mathbf{z}\end{aligned}$$

y además se verifica que $\mathbf{x} \cdot (\lambda \mathbf{y}) = (\lambda \mathbf{x}) \cdot \mathbf{y}$.

Caso especial 1 (Producto punto de un vector por una fila (o columna) de \mathbf{I}). Al multiplicar un vector de \mathbb{R}^n por la fila (o columna) i -ésima de la matriz identidad de orden n , se *selecciona la componente i -ésima del vector*, $(\mathbf{I}_i) \cdot \mathbf{x} = {}_i\mathbf{x} = \mathbf{x}_{|i} = \mathbf{x} \cdot (\mathbf{I}_i)$:

$$({}_i\mathbf{I}) \cdot \mathbf{x} = 0x_1 + 0x_2 + \cdots + 1x_i + \cdots + 0x_n = x_i = {}_i\mathbf{x}.$$

2.2. Producto de una matriz por un vector (a su derecha)

2.2.1. Combinación lineal de vectores

Hay una operación *muy importante* (¡la más importante de todas las que veamos en este curso!): la suma de múltiplos de vectores. Por ejemplo

$$3\mathbf{a} + \mathbf{b} - 7\mathbf{c} + 2\mathbf{d}$$

donde 3, 1, -7 y 2 son los coeficientes. Es una operación tan importante que tiene nombre propio:

Definición 2.2 (Combinación lineal de vectores de \mathbb{R}^n). Sean los vectores $\mathbf{b}_1, \dots, \mathbf{b}_n$. Llamamos *combinación lineal* a cualquier suma de múltiplos de dichos vectores:

$$a_1\mathbf{b}_1 + a_2\mathbf{b}_2 + \cdots + a_n\mathbf{b}_n$$

donde los números “ a_i ” son los *coeficientes* de la combinación lineal.

EJERCICIO 12. Sean los vectores

$$\mathbf{x} = \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}; \quad \mathbf{y} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}; \quad \mathbf{z} = \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}.$$

Compruebe que la combinación lineal $2\mathbf{x} + 2\mathbf{y} + 2\mathbf{z}$ es el vector nulo.

Ejemplo 1. Nótese que todo vector \mathbf{a} de \mathbb{R}^n es combinación lineal de las columnas de $\mathbf{I}_{n \times n}$, pues

$$(\mathbf{I}_{|1})a_1 + \cdots + (\mathbf{I}_{|n})a_n = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} a_1 + \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} a_2 + \cdots + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} a_n = \begin{pmatrix} a_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ a_2 \\ \vdots \\ 0 \end{pmatrix} + \cdots + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \mathbf{a}.$$

2.2.2. Producto de una matriz por un vector

El producto de una matriz por un vector a su derecha, $\mathbf{A}\mathbf{b}$, es una combinación lineal de las columnas de la matriz. Para recordarlo, *escribiremos el vector de la derecha en forma de columna*. Llamaremos a esta operación *combinación lineal de las columnas de \mathbf{A}* o *producto de una matriz por un vector a su derecha*.

Definición 2.3 (matriz por un vector a su derecha). El producto de $\mathbf{A}_{m \times n}$ por un vector $\mathbf{b} \in \mathbb{R}^n$ a su derecha es la *combinación lineal de las n columnas* de \mathbf{A} cuyos coeficientes son los componentes de \mathbf{b} :

$$\mathbf{A}\mathbf{b} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \equiv \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} b_1 + \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} b_2 + \cdots + \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} b_n. \quad (2.1)$$

O expresado de forma alternativa: $\mathbf{A}\mathbf{b} = (\mathbf{A}_{|1})b_1 + (\mathbf{A}_{|2})b_2 + \cdots + (\mathbf{A}_{|n})b_n$.

Fíjese que el número de columnas de la matriz \mathbf{A} *debe ser igual* al número de componentes del vector \mathbf{b} (¡una componente de \mathbf{b} por cada columna de \mathbf{A} para poder expresar la suma de múltiplos de las columnas!)

```
A = Matrix( [ Vector([1,1,1]), Vector([1,0,1]), Vector([-2,2,1]) ] )
b = Vector( [1,2,-3] )
A*b
```

$$\mathbf{A}\mathbf{b} = \begin{bmatrix} 1 & 1 & -2 \\ 1 & 0 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix} = 1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + 2 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} - 3 \begin{pmatrix} -2 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 9 \\ -5 \\ 0 \end{pmatrix}.$$

EJERCICIO 13. Escriba como producto de una matriz por un vector la combinación lineal del ejercicio 12.

Y ahora veamos un par de casos especialmente sencillos:

Caso especial 2 (Producto de una matriz identidad por un vector). *Ahora podemos expresar el producto del Ejemplo 1 en la página anterior de manera mucho más compacta*

$$\mathbf{I}\mathbf{a} = \mathbf{a}.$$

(es decir, todo vector de \mathbb{R}^n es combinación lineal de las columnas de la matriz identidad de orden n)

Caso especial 3 (Producto de una matriz por una columna de \mathbf{I}). *Por otra parte, de la Definición 2.3 resulta evidente que*

$$\mathbf{A}(\mathbf{I}_{|j}) = \mathbf{A}_{|j}$$

es decir, multiplicar \mathbf{A} por la columna j -ésima de la identidad es lo mismo que seleccionar su columna j -ésima.

EJERCICIO 14. Demuestre que $\mathbf{A}(\mathbf{I}_{|j}) = \mathbf{A}_{|j}$.

Pues bien, hay una estrecha relación entre el producto $\mathbf{A}\mathbf{b}$ y los productos punto de \mathbf{b} y las filas de \mathbf{A} :

EJERCICIO 15. Demuestre que la componente i -ésima del vector $\mathbf{A}\mathbf{b}$ es el producto punto $({}_i\mathbf{A}) \cdot \mathbf{b}$:

$${}_i(\mathbf{A}\mathbf{b}) = ({}_i\mathbf{A}) \cdot \mathbf{b} \quad \text{donde } i = 1 : m.$$

```
2|(A*b) == (2|A)*b # ¿son iguales?
```

```
True
```



Hemos visto que la operación **MATRIZ** (de orden $m \times n$) por un **vector** (de n componentes) es un **vector** formado por una suma de múltiplos de las n columnas:

$$\mathbf{A}\mathbf{b} = (\mathbf{A}_{|1})b_1 + (\mathbf{A}_{|2})b_2 + \cdots + (\mathbf{A}_{|n})b_n \in \mathbb{R}^m.$$

A un vector que es suma de múltiplos de vectores se le llama **combinación lineal**. Por tanto $\mathbf{A}\mathbf{b}$ es una **combinación lineal de las columnas** de \mathbf{A} .

También hemos visto que ${}_i(\mathbf{A}\mathbf{b}) = ({}_i\mathbf{A}) \cdot \mathbf{b}$, que $\mathbf{I}\mathbf{a} = \mathbf{a}$ y que $\mathbf{A}(\mathbf{I}_{|j}) = \mathbf{A}_{|j}$.

2.2.3. Propiedades del producto de una matriz por un vector a su derecha

El producto de una matriz por un vector posee dos importantísimas propiedades: las *propiedades de linealidad en el vector*. La primera respecto a la suma: $\mathbf{A}(\mathbf{b} + \mathbf{c}) = \mathbf{A}\mathbf{b} + \mathbf{A}\mathbf{c}$, y la segunda respecto al producto por un escalar: $\mathbf{A}(\lambda\mathbf{b}) = \lambda(\mathbf{A}\mathbf{b})$. Veámoslo.

EJERCICIO 16. Demuestre las siguientes proposiciones (inténtelo primero sin mirar las pistas).

(a) **Proposición 2.2.1.** Sea la matriz \mathbf{A} y los vectores \mathbf{b} y \mathbf{c} de \mathbb{R}^n , entonces: $\mathbf{A}(\mathbf{b} + \mathbf{c}) = \mathbf{A}\mathbf{b} + \mathbf{A}\mathbf{c}$.

Pista. Demuestre que ambos vectores son iguales componente a componente, es decir, que ${}_i(\mathbf{A}(\mathbf{b} + \mathbf{c}))$ es igual a ${}_i(\mathbf{A}\mathbf{b} + \mathbf{A}\mathbf{c})$. Para ello emplee que cada componente es un producto punto y que los productos escalares también son lineales en el segundo argumento. Use también que el operador selector es lineal.

(b) **Proposición 2.2.2.** Sea la matriz \mathbf{A} , el vector \mathbf{b} de \mathbb{R}^n y el escalar λ , entonces: $\mathbf{A}(\lambda\mathbf{b}) = \lambda(\mathbf{A}\mathbf{b})$.

Pista. Siga la misma estrategia que en el apartado anterior.

El producto de una matriz por un vector es un **operador lineal** en el vector; es decir, *distributivo respecto a la suma*

$$\mathbf{A}(\mathbf{b} + \mathbf{c}) = \mathbf{A}\mathbf{b} + \mathbf{A}\mathbf{c}.$$

y *asociativo respecto al producto por un escalar*

$$\mathbf{A}(\lambda\mathbf{b}) = \lambda(\mathbf{A}\mathbf{b}).$$

EJERCICIO 17. Demuestre el siguiente corolario (inténtelo primero sin mirar las pistas).

Corolario 2.2.3. Sea la matriz \mathbf{A} , la matriz \mathbf{B} , y el vector \mathbf{c} de \mathbb{R}^n , entonces:

$$\mathbf{A}(\mathbf{B}\mathbf{c}) = \begin{bmatrix} \mathbf{A}(\mathbf{B}_{|1}); & \dots & \mathbf{A}(\mathbf{B}_{|n}); \end{bmatrix} \mathbf{c},$$

donde $\mathbf{B}_{|i}$ es la i -ésima columna de \mathbf{B} .

Pista. Comience con la expresión de la izquierda. Expresé el producto dentro de paréntesis como una combinación lineal de columnas (Definición 2.3) y aplique las propiedades de linealidad (proposiciones 2.2.1 y 2.2.2) y de nuevo la Definición 2.3 para obtener la expresión de la derecha de la igualdad.

Para terminar la sección, añadamos dos proposiciones más... la primera se parece a la Proposición 2.2.2 pero no es igual (fíjese en la posición de los paréntesis del lado derecho de las igualdades).

EJERCICIO 18. Demuestre las siguientes proposiciones (inténtelo sin mirar las pistas).

(a) **Proposición 2.2.4.** Sea la matriz \mathbf{A} , el vector \mathbf{b} de \mathbb{R}^n y el escalar λ , entonces: $\mathbf{A}(\lambda\mathbf{b}) = (\lambda\mathbf{A})\mathbf{b}$.

Pista. Recuerde que el producto punto es lineal en el primer argumento: $\mathbf{x} \cdot (\lambda\mathbf{y}) = (\lambda\mathbf{x}) \cdot \mathbf{y}$.

(b) **Proposición 2.2.5.** Sean las matrices \mathbf{A} y \mathbf{B} y el vector \mathbf{c} de \mathbb{R}^n entonces: $(\mathbf{A} + \mathbf{B})\mathbf{c} = \mathbf{A}\mathbf{c} + \mathbf{B}\mathbf{c}$.

Más reglas de reescritura.


De las proposiciones 2.2.2 y 2.2.4 se deduce que

$$(\lambda\mathbf{A})\mathbf{b} = \mathbf{A}(\lambda\mathbf{b}) = \lambda(\mathbf{A}\mathbf{b})$$

y por tanto no son necesarios los paréntesis. Así que sin ninguna ambigüedad podemos escribir:

$$\lambda\mathbf{A}\mathbf{b} = \mathbf{A}\lambda\mathbf{b}.$$

Observe además que (como aceptamos que el escalar multiplique a los vectores y matrices por la derecha) el escalar puede aparecer en cualquier lugar, lo único que se mantiene invariante es la posición relativa del vector respecto de la matriz, pues siempre está a su derecha): $\lambda \mathbf{A} \mathbf{b} = \mathbf{A} \lambda \mathbf{b} = \mathbf{A} \mathbf{b} \lambda$.

 Hemos visto varias propiedades que cumple la operación **MATRIZ** por *vector*


Propiedades de linealidad en el vector

- $\mathbf{A}(\mathbf{b} + \mathbf{c}) = \mathbf{A}\mathbf{b} + \mathbf{A}\mathbf{c}$
- $\mathbf{A}(\lambda \mathbf{b}) = \lambda(\mathbf{A}\mathbf{b})$

Otras propiedades

- $\mathbf{A}(\mathbf{B}\mathbf{c}) = \begin{bmatrix} \mathbf{A}(\mathbf{B}_{|1}); & \dots & \mathbf{A}(\mathbf{B}_{|n}); \end{bmatrix} \mathbf{c}$
- Lineal en la matriz $\begin{cases} (\mathbf{A} + \mathbf{B})\mathbf{c} = \mathbf{A}\mathbf{c} + \mathbf{B}\mathbf{c} \\ (\lambda \mathbf{A})\mathbf{b} = \lambda(\mathbf{A}\mathbf{b}) \end{cases}$

2.3. Producto de un vector por una matriz (vector a la izquierda)

 Hasta aquí hemos visto que matriz por vector ($\mathbf{A}\mathbf{b}$) es una combinación lineal de las *columnas* de la matriz. A continuación veremos que vector por matriz ($\mathbf{a}\mathbf{B}$) es una combinación lineal de las *filas* de la matriz (con propiedades análogas).

Definición 2.4 (producto de un vector por una matriz). *El producto de un vector $\mathbf{a} \in \mathbb{R}^m$ que multiplica por la izquierda a una matriz \mathbf{B} de m filas es la combinación lineal de las m filas de \mathbf{B} cuyos coeficientes son los componentes de \mathbf{a} :*

$$\mathbf{a}\mathbf{B} = \mathbf{B}^T \mathbf{a}$$

Puesto que el producto $\mathbf{a}\mathbf{B}$ es una combinación lineal de las filas de \mathbf{B} , dicho producto es un vector de \mathbb{R}^n .

Para recordar que operamos sobre las filas de \mathbf{B} usualmente escribiremos \mathbf{a} en horizontal:

$$\mathbf{a}\mathbf{B} = (a_1, a_2, \dots, a_m) \begin{bmatrix} b_{11} & \dots & b_{1n} \\ b_{21} & \dots & b_{2n} \\ \vdots & \ddots & \vdots \\ b_{m1} & \dots & b_{mn} \end{bmatrix} = a_1(\mathbf{B}_{|1}) + \dots + a_m(\mathbf{B}_{|m}) \in \mathbb{R}^n.$$

```
print( b*A == (~A)*b )
b*A
```

Librería NAcAL para Python

True

(0, -2, -1,)

Cada componente j -ésima del vector $\mathbf{a}\mathbf{B}$ es el producto punto entre \mathbf{a} y la columna j -ésima de \mathbf{B} :

$$(\mathbf{a}\mathbf{B})_{|j} = ((\mathbf{B}^T)\mathbf{a})_{|j} = \mathbf{B}_{|j}^T \mathbf{a} = (\mathbf{B}_{|j}) \cdot \mathbf{a} = \mathbf{a} \cdot (\mathbf{B}_{|j}), \quad \text{donde } j = 1 : n.$$

```
print( (b*A)|3 == b*(A|3) )
b*(A|3)
```

True

-1

2.3.1. Propiedades del producto de un vector por una matriz

EJERCICIO 19. Demuestre las siguientes propiedades del producto \mathbf{aB} .

- (a) Sea el vector \mathbf{a} de m componentes y la matriz $\mathbf{I}_{m \times m}$, entonces: $\mathbf{aI} = \mathbf{a}$.

Pista. Recuerde que \mathbf{I} es simétrica: $\mathbf{I}^T = \mathbf{I}$.

- (b) Sea $({}_i\mathbf{I})$ la fila i -ésima de $\mathbf{I}_{m \times m}$ y la matriz $\mathbf{A}_{m \times n}$, entonces: $({}_i\mathbf{I})\mathbf{A} = {}_i\mathbf{A}$.

Pista. Recuerde el caso especial 3 en la página 27 y que como \mathbf{I} es simétrica, $\mathbf{I}_{|i} = {}_i\mathbf{I}$.

- (c) Sean \mathbf{a} y \mathbf{b} vectores de \mathbb{R}^m y sea la matriz $\mathbf{C}_{m \times n}$, entonces: $(\mathbf{a} + \mathbf{b})\mathbf{C} = \mathbf{aC} + \mathbf{bC}$

Pista. Recuerde la Proposición 2.2.1 y recuerde que $\mathbf{vA} = (\mathbf{A}^T\mathbf{v})$.

- (d) Sea el escalar λ , el vector \mathbf{a} de \mathbb{R}^m y sea la matriz $\mathbf{B}_{m \times n}$, entonces: $(\lambda\mathbf{a})\mathbf{B} = \lambda(\mathbf{aB})$

Pista. Recuerde la Proposición 2.2.2

- (e) Sean \mathbf{a} un vector de \mathbb{R}^m y λ un escalar; y sea la matriz $\mathbf{B}_{m \times n}$, entonces: $\mathbf{a}(\lambda\mathbf{B}) = (\lambda\mathbf{a})\mathbf{B}$.

Pista. Recuerde la Proposición 2.2.4

- (f) Sean \mathbf{a} un vector de \mathbb{R}^m , y las matrices \mathbf{B} y \mathbf{C} de orden m por n , entonces: $\mathbf{a}(\mathbf{B} + \mathbf{C}) = \mathbf{aB} + \mathbf{aC}$.

Pista. Recuerde la Proposición 2.2.5.



El producto de un **vector** (de m componentes) por una **MATRIZ** (de orden $m \times n$) es una **combinación lineal de las m filas** de \mathbf{B} :

$$\mathbf{aB} = a_1({}_1\mathbf{B}) + \cdots + a_m({}_m\mathbf{B}) \in \mathbb{R}^n.$$

Cada componente del vector \mathbf{aB} se puede expresar como un producto punto:

$$(\mathbf{aB})_{|j} = \mathbf{a} \cdot (\mathbf{B}_{|j}), \quad \text{donde } j = 1 : n.$$

El producto \mathbf{aB} tiene propiedades análogas a las del producto \mathbf{Ab} .

- $\mathbf{aI} = \mathbf{a}$
- $({}_i\mathbf{I})\mathbf{A} = {}_i\mathbf{A}$;

Propiedades de linealidad en el vector

- $(\mathbf{a} + \mathbf{b})\mathbf{C} = \mathbf{aC} + \mathbf{bC}$
- $(\lambda\mathbf{a})\mathbf{B} = \lambda(\mathbf{aB})$

Propiedades de linealidad en la matriz

- $\mathbf{a}(\mathbf{B} + \mathbf{C}) = \mathbf{aB} + \mathbf{aC}$
- $\mathbf{a}(\lambda\mathbf{B}) = \lambda(\mathbf{aB})$

Multiplicación matricial

3.1. Producto de matrices

Definición 3.1 (producto de matrices (por columnas)). Sean \mathbf{A} y \mathbf{B} , definimos la matriz producto, \mathbf{AB} , como aquella matriz cuya columna j -ésima es el producto de \mathbf{A} por la columna j -ésima de \mathbf{B} :

$$(\mathbf{AB})_{|j} = \mathbf{A}(\mathbf{B}_{|j}).$$

Es decir, las columnas de \mathbf{AB} son combinaciones lineales de las columnas de \mathbf{A} (y los coeficientes de cada combinación son los componentes de cada una de las columnas de \mathbf{B}).

$$\mathbf{AB} \equiv [\mathbf{A}(\mathbf{B}_{|1}); \quad \dots \quad \mathbf{A}(\mathbf{B}_{|n})].$$

Ejemplo 2. Si $\mathbf{A} = \begin{bmatrix} 1 & 1 & -2 \\ 1 & -2 & 1 \\ -2 & 1 & 1 \end{bmatrix}$ y $\mathbf{B} = \begin{bmatrix} -2 & 1 \\ 1 & 1 \\ 1 & -2 \end{bmatrix}$ entonces

$$\mathbf{AB} = [\mathbf{A}(\mathbf{B}_{|1}); \quad \mathbf{A}(\mathbf{B}_{|2})] = \left[\mathbf{A} \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}; \quad \mathbf{A} \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} \right] = \left[\begin{pmatrix} -3 \\ -3 \\ 6 \end{pmatrix}; \quad \begin{pmatrix} 6 \\ -3 \\ -3 \end{pmatrix} \right] = \begin{bmatrix} -3 & 6 \\ -3 & -3 \\ 6 & -3 \end{bmatrix};$$

nótese que calculamos multiplicando \mathbf{A} por cada columna de \mathbf{B} .¹

Deduzcamos algunas propiedades que verifica el producto de matrices...

EJERCICIO 20. Demuestre las siguientes proposiciones.

(a) **Proposición 3.1.1.** Sean \mathbf{A} y \mathbf{B} , y un vector \mathbf{c} de \mathbb{R}^n , entonces $\mathbf{A}(\mathbf{B}\mathbf{c}) = (\mathbf{AB})\mathbf{c}$.

(b) **Proposición 3.1.2** (El producto de matrices es asociativo). Sean las matrices \mathbf{A} , \mathbf{B} y \mathbf{C} , entonces

$$\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}.$$

Nota importante respecto al operador “ $|j$ ”, el producto de matrices y la asociatividad: La definición de producto matricial mantiene la *asociatividad* del operador “ $|j$ ”, **POR LO QUE NO NECESITAMOS PARÉNTESIS**. Así, $\mathbf{AB}_{|j}$ se refiere indistintamente a la columna j -ésima del producto $(\mathbf{AB})_{|j}$, o al producto de \mathbf{A} por $\mathbf{B}_{|j}$ (i.e., la columna j -ésima), $\mathbf{A}(\mathbf{B}_{|j})$. Y puesto que $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$; es decir, que podemos escribir sencillamente \mathbf{ABC} .

¹Aunque conozca otro método para calcular el producto, no lo emplee por el momento. AHORA ES MUY IMPORTANTE que se acostumbre a considerar las columnas de \mathbf{AB} como combinaciones de las columnas de \mathbf{A} .



Hemos definido producto (\mathbf{AB}) de las matrices \mathbf{A} y \mathbf{B} como aquella matriz cuya columna j -ésima es el producto de \mathbf{A} por la columna j -ésima de \mathbf{B}

$$(\mathbf{AB})_{|j} = \mathbf{A}(\mathbf{B})_{|j}.$$

Por tanto las columnas de \mathbf{AB} son combinaciones lineales de las columnas de \mathbf{A} .

¡De nuevo nótese la asociatividad de la notación!

El producto solo está definido si la primera matriz tiene tantas columnas como filas tiene la segunda.

EJERCICIO 21. Demuestre las siguientes proposiciones.

(a) **Proposición 3.1.3.** Sean las matrices \mathbf{A} , \mathbf{B} y \mathbf{C} , entonces $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$.

(b) **Proposición 3.1.4.** Sean las matrices \mathbf{A} , \mathbf{B} y \mathbf{C} , entonces $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$.

(c) **Proposición 3.1.5.** Sean las matrices \mathbf{A} , \mathbf{B} y el escalar λ , entonces $\mathbf{A}(\lambda\mathbf{B}) = \lambda(\mathbf{AB}) = (\lambda\mathbf{A})\mathbf{B}$.

Pista. Recuerde que $\mathbf{A}(\lambda\mathbf{b}) \begin{cases} = \lambda(\mathbf{Ab}) \\ = (\lambda\mathbf{A})\mathbf{b} \end{cases}$; (Proposiciones 2.2.2 y 2.2.4 de la Página 28).

(d) **Proposición 3.1.6.** Sean la matriz identidad \mathbf{I} de orden m y la matriz \mathbf{A} , entonces $\mathbf{IA} = \mathbf{A}$.

Pista. Recuerde que $\mathbf{Ia} = \mathbf{a}$ (Caso Especial 2 en la página 27).

(e) **Proposición 3.1.7.** Sean la matriz \mathbf{A} y la matriz identidad \mathbf{I} de orden n , entonces $\mathbf{AI} = \mathbf{A}$.

Pista. Recuerde que $\mathbf{A}(\mathbf{I}_{|j}) = \mathbf{A}_{|j}$ (Caso Especial 3 en la página 27).

Muy muy importante: hay que subrayar que no todas las propiedades de las matrices replican las propiedades de los números que usted conoce: por ejemplo, en general \mathbf{AB} es distinto de \mathbf{BA} . Por ejemplo:²

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \neq \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Por otra parte: entre los escalares si $\lambda^2 = \lambda$ entonces λ es necesariamente 0 o 1; pero con las matrices $\mathbf{A}^2 = \mathbf{A}$ no implica que la matriz \mathbf{A} sea cero $\mathbf{0}$ o la identidad \mathbf{I} . Por ejemplo, si $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ entonces $\mathbf{A}^2 = \mathbf{A}$.

En los escalares si $\lambda^2 = 0$ entonces λ es necesariamente 0, pero tampoco esto es cierto con las matrices. Por ejemplo, para $\mathbf{B} = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$

$$\mathbf{B}^2 = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \mathbf{0}.$$

²para que ambos productos estén definidos es necesario que ambas matrices sean cuadradas y del mismo orden.

☞ El producto de matrices verifica las siguientes propiedades:

- $\mathbf{A}(\mathbf{B}c) = (\mathbf{A}\mathbf{B})c$
- $\mathbf{A}(\mathbf{B}\mathbf{C}) = (\mathbf{A}\mathbf{B})\mathbf{C}$
- $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{A}\mathbf{C} + \mathbf{B}\mathbf{C}$
- $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{A}\mathbf{B} + \mathbf{A}\mathbf{C}$
- $\mathbf{A}(\lambda\mathbf{B}) = \lambda(\mathbf{A}\mathbf{B}) = (\lambda\mathbf{A})\mathbf{B}$
- $\mathbf{I}\mathbf{A} = \mathbf{A}$
- $\mathbf{A}\mathbf{I} = \mathbf{A}$

Pese a las dos últimas propiedades, en general $\mathbf{A}\mathbf{B} \neq \mathbf{B}\mathbf{A}$.

Por tanto...

☞ El producto de matrices es un **operador lineal**; pues es *distributivo* respecto a la suma de matrices

$$\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{A}\mathbf{B} + \mathbf{A}\mathbf{C}.$$

y *asociativo* respecto al producto de una matriz por un escalar

$$\mathbf{A}(\lambda\mathbf{B}) = \lambda(\mathbf{A}\mathbf{B}).$$

Y también es lineal por la izquierda, pues

$$(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{A}\mathbf{C} + \mathbf{B}\mathbf{C} \quad \text{y} \quad (\lambda\mathbf{A})\mathbf{B} = \lambda(\mathbf{A}\mathbf{B}).$$

3.1.1. Otras dos formas de calcular el producto de matrices

Aplicando las propiedades que ya conocemos es sencillo deducir otras formas de calcular el producto. Veamos dos nuevas (la primera seguramente le resultará familiar...).

Cálculo del producto de matrices componente a componente (filas por columnas)

EJERCICIO 22. Demuestre la siguiente proposición:

Proposición 3.1.8 (Cálculo del producto componente a componente). Sean las matrices \mathbf{A} , \mathbf{B} , entonces:

$${}_i(\mathbf{A}\mathbf{B})_{|j} = ({}_i\mathbf{A}) \cdot (\mathbf{B}_{|j}).$$

Verifíquelo con el producto del Ejemplo 2 en la página 31 (hágalo a mano y con el módulo NAcAL).

Pero hay más formas de calcular el producto...

Cálculo del producto de matrices operando con las filas

EJERCICIO 23. Demuestre la siguiente proposición:

Proposición 3.1.9. Sean \mathbf{A} y \mathbf{B} , entonces:

$${}_i(\mathbf{A}\mathbf{B}) = ({}_i\mathbf{A})\mathbf{B}.$$

Pista. Recuerde que $\mathbf{a} \cdot (\mathbf{B}_{|j}) = (\mathbf{a}\mathbf{B})_{|j}$, para $j = 1 : n$.

Verifíquelo con el producto del Ejemplo 2 en la página 31 (hágalo a mano y con el módulo NAcAL).

Por tanto, las filas de \mathbf{AB} son combinaciones lineales de las filas de \mathbf{B} .



Explotando el producto punto hemos obtenido otras dos formas de calcular \mathbf{AB} .

La primera alternativa calcula el producto componente a componente usando el *producto punto*.

$${}_i(\mathbf{AB})_{|j} = ({}_i\mathbf{A}) \cdot (\mathbf{B}_{|j}).$$

de manera que los elementos de \mathbf{AB} son el producto punto de una fila de \mathbf{A} y una columna de \mathbf{B} .

La segunda alternativa calcula el producto fila a fila:

$${}_i(\mathbf{AB}) = ({}_i\mathbf{A})\mathbf{B},$$

de manera que las filas de \mathbf{AB} son combinaciones lineales de las filas de \mathbf{B} .

3.2. Nuevas reglas de reescritura.

Realmente no es necesario escribir el “punto” en las siguientes expresiones

$$({}_i\mathbf{A}) \cdot b, \quad a \cdot (\mathbf{B}_{|j}), \quad ({}_i\mathbf{A}) \cdot (\mathbf{B}_{|j})$$

(como tampoco lo es en el producto $2 \cdot x$). Si omitimos el “punto” recuperamos la asociatividad

$${}_i(\mathbf{Ab}) = ({}_i\mathbf{A})b$$

$$(\mathbf{aB})_{|j} = a(\mathbf{B}_{|j})$$

$${}_i(\mathbf{AB})_{|j} = {}_i((\mathbf{AB})_{|j}) = {}_i(\mathbf{A}(\mathbf{B}_{|j})) = ({}_i\mathbf{A})(\mathbf{B}_{|j}).$$

Por tanto podemos escribir sin ambigüedad las siguientes expresiones:

$${}_i\mathbf{Ab} \quad ; \quad a\mathbf{B}_{|j} \quad ; \quad {}_i\mathbf{AB}_{|j}$$

pues da el mismo resultado si se interpretan como selección de un componente o como un producto punto.

Transpuesta de un producto

Para finalizar esta lección demuestre la siguiente propiedad de la transposición que usaremos a menudo:

EJERCICIO 24. Recordando que $(\mathbf{A}^\top)_{|j} = {}_j\mathbf{A}$ y que $\mathbf{aB} = (\mathbf{B}^\top)\mathbf{a}$; demuestre que

Proposición 3.2.1. Sean \mathbf{A} y \mathbf{B} , entonces:

$$m \times p \quad p \times n$$

$$(\mathbf{AB})^\top = (\mathbf{B}^\top)(\mathbf{A}^\top). \quad (3.1)$$

Por tanto, para toda matriz \mathbf{A} se verifica que las matrices $(\mathbf{A}^\top)\mathbf{A}$ y $\mathbf{A}(\mathbf{A}^\top)$ son simétricas:

$$((\mathbf{A}^\top)\mathbf{A})^\top = (\mathbf{A}^\top)((\mathbf{A}^\top)^\top) = (\mathbf{A}^\top)\mathbf{A}; \quad \text{y} \quad (\mathbf{A}(\mathbf{A}^\top))^\top = ((\mathbf{A}^\top)^\top)(\mathbf{A}^\top) = \mathbf{A}(\mathbf{A}^\top).$$

Nótese que la transposición no es asociativa respecto al producto por una *matriz* ya que $(\mathbf{AB})^\top \neq \mathbf{A}(\mathbf{B}^\top)$; por tanto el uso de los paréntesis cobra importancia pues la expresión \mathbf{AB}^\top es ambigua. Lo usual es interpretar que la transposición solo afecta a \mathbf{B} en la anterior expresión, pero es preferible escribir $\mathbf{A}(\mathbf{B}^\top)$.

Apéndices a la lección

3.A. Productos de matrices triangulares

Definición 3.2. Decimos que una matriz \mathbf{L} es *triangular inferior* cuando todos los componentes por encima de la diagonal principal son nulos, es decir, si $l_{ij} = 0$ cuando $i < j$.

Por ejemplo las matrices

$$\mathbf{A} = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{bmatrix}; \quad \mathbf{B} = \begin{bmatrix} b_{11} & 0 & \cdots & 0 & 0 \\ b_{21} & b_{22} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mp} & 0 \end{bmatrix}; \quad \text{y} \quad \mathbf{C} = \begin{bmatrix} c_{11} & 0 & \cdots & 0 \\ c_{21} & c_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \cdots & c_{(m-1)n} \\ c_{m1} & c_{m2} & \cdots & c_{mn} \end{bmatrix}.$$

Las tres son triangulares inferiores, aunque solo \mathbf{A} es cuadrada. Fíjese que las componentes de la diagonal principal y por debajo pueden tomar cualquier valor (por eso el nombre de “inferior”)... ¡Por tanto, una matriz nula $\mathbf{0}$ siempre es triangular inferior... pero también es triangular superior:

Definición 3.3. Decimos que una matriz \mathbf{U} es *triangular superior* cuando todos los componentes por debajo de la diagonal principal son nulos, es decir, si $u_{ij} = 0$ cuando $i \geq j$.

Por tanto, una matriz es *triangular superior* si su transpuesta es *triangular inferior*.

Definición 3.4. Una matriz triangular es *unitaria* si los elementos de la diagonal son iguales a uno.

Las matrices *diagonales* son simultáneamente triangulares superiores e inferiores. Las matrices identidad son *cuadradas*, simultáneamente triangulares superiores e inferiores y además *unitarias*.

EJERCICIO 25. Demuestre las siguientes proposiciones.

- (a) **Proposición 3.A.1.** El producto de dos matrices triangulares superiores es triangular superior.
- (b) **Proposición 3.A.2.** Si \mathbf{A} es triangular superior unitaria y \mathbf{B} es triangular superior, \mathbf{AB} es triangular superior y mantiene al diagonal de \mathbf{B} .
- (c) **Proposición 3.A.3.** Si \mathbf{A} es triangular superior y \mathbf{B} es triangular superior unitaria, \mathbf{AB} es triangular superior y mantiene al diagonal de \mathbf{A} .

Como el producto de matrices triangulares superiores unitarias mantiene la diagonal (formada por unos), tenemos el siguiente

Corolario 3.A.4. El producto de dos matrices triangulares superiores unitarias es también triangular superior unitario.

Resultados análogos son ciertos con matrices triangulares inferiores, basta transponer para demostrarlo.

3.B. Producto matricial como suma de productos de submatrices



En el Ejercicio 83 en la página 168 de la Sección 14.3 se emplea el *producto de matrices por bloques* para demostrar cómo es el determinante de una matriz diagonal por bloques (resultado que se empleará en el Corolario 16.2.4 en la página 187 de la lección sobre diagonalización por semejanza).

Este apéndice muestra cómo calcular el *producto de matrices por bloques* (que es una aplicación de la Ecuación 3.6 en la página 39 cuando las submatrices son el resultado de “cortar” las matrices en bloques).

Considere una lista $\alpha = (\alpha_1, \dots, \alpha_q,)$ de q índices no repetidos, así como la submatriz $\alpha|\mathbf{I}$, de la matriz identidad \mathbf{I} de orden mayor o igual al mayor de los índices en α . Si α_j es la j -ésima componente de la lista α , entonces la columna $(\alpha|\mathbf{I})_{|\alpha_j}$ resulta ser igual a $\mathbf{I}_{|j} \in \mathbb{R}^q$ (donde q es la longitud de la lista α):

$$\alpha|\mathbf{I}_{|k} =_{(\alpha_1, \dots, \alpha_q,)} \mathbf{I}_{|k} =_{(\alpha_1, \dots, \alpha_q,)} (\mathbf{I}_{|k}) = \begin{cases} \mathbf{I}_{|j} \in \mathbb{R}^q & \text{si } k = \alpha_j \\ \mathbf{0} \in \mathbb{R}^q & \text{si } k \notin \alpha \end{cases} \in \mathbb{R}^q, \quad (3.2)$$

puesto que si de $\mathbf{I}_{|k}$ (cuya única componente no nula es la k -ésima) seleccionamos un subvector de q componentes que coloca su k -ésima componente en la posición j , obtenemos $\mathbf{I}_{|j} \in \mathbb{R}^q$, y si seleccionamos un subvector que no toma su única componente no nula, obtenemos un subvector nulo: $\mathbf{0} \in \mathbb{R}^q$.

Ejemplo 3. Si consideremos la lista de 3 índices $\alpha = (4, 1, 2,)$ (cuyo mayor índice es el 4) y una matriz identidad de orden mayor o igual que 4; por ejemplo \mathbf{I} de orden 4. Entonces

$$\alpha|\mathbf{I} =_{(4,1,2,)} \mathbf{I} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

Fijándonos en las columnas de la anterior matriz observamos que:

- como 4 es el *primer* componente de la lista α , entonces la cuarta columna $(_{(4,1,2,)} \mathbf{I})_{|4}$ es igual a la *primera* columna de la identidad de orden 3 (pues 3 es la longitud de la lista):

$$_{(4,1,2,)} \mathbf{I}_{|4} =_{(4,1,2,)} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

- como 1 es la *segunda* componente de la lista α , entonces la primera columna $(_{(4,1,2,)} \mathbf{I})_{|1}$ es igual a la *segunda* columna de la identidad:

$$_{(4,1,2,)} \mathbf{I}_{|1} =_{(4,1,2,)} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

- como 2 es la *tercera* componente de la lista α , entonces la segunda columna $(_{(4,1,2,)} \mathbf{I})_{|2}$ es igual a la *tercera* columna de la identidad:

$$_{(4,1,2,)} \mathbf{I}_{|2} =_{(4,1,2,)} \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}.$$

- y como 3 no está en la lista α , entonces la tercera columna $(_{(4,1,2,)} \mathbf{I})_{|3}$ es nula

$$_{(4,1,2,)} \mathbf{I}_{|3} =_{(4,1,2,)} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

```
alpha = (4,1,2,) # índices
Sistema([ alpha|I(99)|j for j in (4,1,2,3,) ]) # las cuatro columnas del ejemplo
```

$$\left[\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right]$$

Y ahora considere las columnas del producto de matrices $(\mathbf{I}_{|\alpha})(\alpha|\mathbf{I})$. Usando (3.2) se deduce que

$$\left((\mathbf{I}_{|\alpha})(\alpha|\mathbf{I}) \right)_{|k} = (\mathbf{I}_{|\alpha})(\alpha|\mathbf{I}_k) = \begin{cases} (\mathbf{I}_{|\alpha})_{|j} = (\mathbf{I}_{|\alpha})_{|j} = \mathbf{I}_{|k} & \text{si } k = \alpha_j; \\ (\mathbf{I}_{|\alpha})\mathbf{0} = \mathbf{0} & \text{si } k \notin \alpha \end{cases}; \quad (3.3)$$

Así, para $\alpha = (4, 1, 2,)$ e $\mathbf{I}_{4 \times 4}$ tenemos que

$$(\mathbf{I}_{|(4,1,2,)})((4,1,2,)|\mathbf{I}) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

de manera que solo hay unos en las posiciones 1, 2 y 4 de la diagonal.

```
(I(4)|alpha) * (alpha|I(4)) # fíjese en la necesidad de los paréntesis
```

Ahora considere las matrices \mathbf{A} y \mathbf{B} ; y las listas de índices $\gamma_1, \dots, \gamma_h$ tales que cada índice $j = 1 : p$ está en una, y solo una, de las listas γ_i . Por ejemplo, si $p = 5$, $\gamma_1 = (3, 1,)$, $\gamma_2 = (5,)$ y $\gamma_3 = (4, 2,)$; entonces

$$\mathbf{A}_{|\gamma_1 \# \gamma_2 \# \gamma_3} = \mathbf{A}_{|(3,1,5,4,2,)};$$

donde $\#$ es el operador que concatena. Es decir, $\gamma_1 \# \gamma_2 \# \gamma_3 = (3, 1, 5, 4, 2,)$ es la concatenación de las listas γ_1 , γ_2 y γ_3 . Entonces $\mathbf{A}_{|\gamma_1 \# \gamma_2 \# \gamma_3}$ es un reordenamiento (o permutación) de las columnas, pues todos sus índices aparecen una, y solo una, vez en la lista $\gamma_1 \# \gamma_2 \# \gamma_3 = (3, 1, 5, 4, 2,)$.

Si empleamos las listas $\gamma_1, \dots, \gamma_h$ para reordenar tanto las columnas de \mathbf{A} y como las filas de \mathbf{B} , el producto de las matrices resultantes resulta ser igual al producto \mathbf{AB} :

$$(\mathbf{A}_{|\gamma_1 \# \dots \# \gamma_h})(\gamma_1 \# \dots \# \gamma_h|\mathbf{B}) = \mathbf{A}(\mathbf{I}_{|\gamma_1 \# \dots \# \gamma_h})(\gamma_1 \# \dots \# \gamma_h|\mathbf{I})\mathbf{B} = \mathbf{AIB} = \mathbf{AB},$$

pues, como todos los índices están incluidos por (3.3), sabemos que $(\mathbf{I}_{|\gamma_1 \# \dots \# \gamma_h})(\gamma_1 \# \dots \# \gamma_h|\mathbf{I}) = \mathbf{I}$.

Ejemplo 4. Si usamos las listas $\gamma_1 = (3, 1,)$ y $\gamma_2 = (4, 2,)$ con $\mathbf{I}_{4 \times 4}$, entonces

```
g1=(3,1,); g2=(4,2,)
(I(4)|g1)*(g1|I(4)) + (I(4)|g2)*(g2|I(4))
```

$$(\mathbf{I}_{|(3,1,)})((3,1,)|\mathbf{I}) + (\mathbf{I}_{|(4,2,)})((4,2,)|\mathbf{I}) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Consecuentemente podemos calcular \mathbf{AB} mediante la suma de k productos de submatrices:

$$\sum_{k=1}^h (\mathbf{A}_{|\gamma_k|})(\gamma_k|\mathbf{B}) = \sum_{k=1}^h (\mathbf{A}(\mathbf{I}_{|\gamma_k|})(\gamma_k|\mathbf{I})\mathbf{B}) = \mathbf{A} \left(\sum_{k=1}^h (\mathbf{I}_{|\gamma_k|})(\gamma_k|\mathbf{I}) \right) \mathbf{B} = \mathbf{AIB} = \mathbf{AB}, \quad (3.4)$$

donde, por (3.3), cada sumando de $\sum_{k=1}^h (\mathbf{I}_{|\gamma_k|})(\gamma_k|\mathbf{I})$ es una matriz cuadrada cuya j -ésima columna es $\mathbf{I}_{|j|}$, cuando $j \in \gamma_k$, o $\mathbf{0}$, en caso contrario. Y como cada índice está incluido en una y solo una de las listas γ_k , la suma de dichas matrices es igual a la identidad.

Ejemplo 5. Para $\mathbf{A} = \begin{bmatrix} 1 & 2 & -1 \\ 2 & -2 & 0 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \\ -1 & 1 \end{bmatrix}$; y las listas $\gamma_1 = (3, 1,)$ y $\gamma_2 = (2,)$; tenemos

$$\begin{aligned} (\mathbf{A}_{|\gamma_1+\gamma_2|})(\gamma_1+\gamma_2|\mathbf{B}) &= (\mathbf{A}_{|\gamma_1|})(\gamma_1|\mathbf{B}) + (\mathbf{A}_{|\gamma_2|})(\gamma_2|\mathbf{B}) = (\mathbf{A}_{|(3,1,)|})(\gamma_1|\mathbf{B}) + (\mathbf{A}_{|(2,)|})(\gamma_2|\mathbf{B}) \\ &= \begin{bmatrix} -1 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 2 & 1 \end{bmatrix} + \begin{bmatrix} 2 \\ -2 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 4 & 2 \end{bmatrix} + \begin{bmatrix} 2 & 2 \\ -2 & -2 \end{bmatrix} = \begin{bmatrix} 5 & 2 \\ 2 & 0 \end{bmatrix} = \mathbf{AB}. \end{aligned}$$

Librería NAcAL para Python

```
A = Matrix([ [1,2,-1], [2,-2,0] ])
B = Matrix([ Vector([2,1,-1]), Vector([1,1,1]) ])
g1=(3,1,); g2=(2,)
(A|g1)*(g1|B) + (A|g2)*(g2|B) )
```

Fíjese en la coma tras el 2

Nota: las listas de un solo elemento se escriben en Python con una coma después del único elemento. Por ejemplo (2,) es la lista que solo contiene el 2. Así se distingue el *numero* 2 entre paréntesis de la *lista* que contiene al 2. Por claridad, y puesto que la distinción es importante, hemos seguido el mismo criterio de notación:³ así, $\mathbf{A}_{|2|}$ es una columna de \mathbf{A} (un vector) y $\mathbf{A}_{|(2,)|}$ es una submatriz (una matriz columna).

3.B.1. Producto matricial expresado como una suma de matrices

Como caso particular, consideremos que cada lista k -ésima contiene únicamente el índice k , es decir, $\gamma_k = (k,)$. Entonces el producto \mathbf{AB} queda descrito como una suma matrices:

$$\left(\begin{matrix} \mathbf{A} \\ m \times p \end{matrix} \right) \left(\begin{matrix} \mathbf{B} \\ p \times n \end{matrix} \right) = \sum_{k=1}^p (\mathbf{A}_{|(k,)|})(\gamma_k|\mathbf{B}), \quad (3.5)$$

donde cada sumando $(\mathbf{A}_{|(k,)|})(\gamma_k|\mathbf{B})$ es una matriz⁴ de orden m por n . Ejemplo:

$$\begin{aligned} \mathbf{AB} &= \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} -1 & 2 \\ 0 & 0 \\ 1 & 0 \end{bmatrix} = (\mathbf{A}_{|(1,)|})(\gamma_1|\mathbf{B}) + (\mathbf{A}_{|(2,)|})(\gamma_2|\mathbf{B}) + (\mathbf{A}_{|(3,)|})(\gamma_3|\mathbf{B}) \\ &= \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} -1 & 2 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix} \begin{bmatrix} 0 & 0 \end{bmatrix} + \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 2 \\ -1 & 2 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 3 & 0 \\ 1 & 0 \\ 2 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 2 \\ 0 & 2 \\ 2 & 0 \end{bmatrix}. \end{aligned}$$

³véase la nota a pie 5 en la página 4

⁴cundo veamos el concepto de rango entenderá que esta expresión es una descripción del producto matricial como suma de matrices de rango uno como máximo.

3.B.2. Submatriz de un producto como suma de productos de submatrices

Considere una lista de listas $\gamma_1, \dots, \gamma_h$, donde cada índice $j = 1 : p$ está incluido en una y solo una de las listas γ_k ; por (3.4) sabemos que:

$$\mathbf{AB} = \sum_{k=1}^h (\mathbf{A}_{|\gamma_k|})(\gamma_k|\mathbf{B}).$$

Por tanto, seleccionando algunas filas de \mathbf{A} y algunas columnas de \mathbf{B} , la submatriz $\alpha|(\mathbf{AB})|_\beta$ resulta ser

$$\alpha|(\mathbf{AB})|_\beta = \sum_{k=1}^h (\alpha|\mathbf{A}_{|\gamma_k|})(\gamma_k|\mathbf{B}|_\beta). \quad (3.6)$$

Matrices partidas en bloques

Y por fin podemos presentar un modo particular de multiplicar matrices (que era objetivo de este apéndice).

Fíjese que en (3.6) no se requiere que las sub-listas γ_k estén formadas por índices consecutivos; pero éste suele ser el caso más habitual en los textos. En tal caso las matrices quedan “cortadas” en bloques. Fíjese en los dos siguientes ejemplos (en los que hemos marcado los bloques mediante líneas horizontales y verticales).

Ejemplo 6. Sean $\alpha = ((1, 2,), (3, 4,))$, $\gamma = ((1, 2,), (3, 4, 5,))$; y las matrices

$$\mathbf{A} = \left[\begin{array}{cc|ccc} \mathbf{C} & \mathbf{E} \\ \mathbf{D} & \mathbf{F} \end{array} \right] = \left[\begin{array}{cc|ccc} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 2 & -1 & 4 & 2 & 1 \\ 3 & 1 & -1 & 7 & 5 \end{array} \right] \quad \text{y} \quad \mathbf{B} = \left[\begin{array}{c|c} \mathbf{G} \\ \mathbf{H} \end{array} \right] = \left[\begin{array}{c|c} 4 & -2 \\ 5 & 6 \\ 7 & 3 \\ -1 & 0 \\ 1 & 6 \end{array} \right];$$

donde $\mathbf{C} = {}_{(1:2)}|\mathbf{A}|_{(1:2)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\mathbf{E} = {}_{(1:2)}|\mathbf{A}|_{(3:5)} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $\mathbf{D} = {}_{(3:4)}|\mathbf{A}|_{(1:2)} = \begin{bmatrix} 2 & -1 \\ 3 & 1 \end{bmatrix}$ y $\mathbf{F} = {}_{(3:4)}|\mathbf{A}|_{(3:5)} = \begin{bmatrix} 4 & 2 & 1 \\ -1 & 7 & 5 \end{bmatrix}$. Y donde $\mathbf{G} = {}_{(1:2)}|\mathbf{B} = \begin{bmatrix} 4 & -2 \\ 5 & 6 \end{bmatrix}$ y $\mathbf{H} = {}_{(3:5)}|\mathbf{B} = \begin{bmatrix} 7 & 3 \\ -1 & 0 \\ 1 & 6 \end{bmatrix}$.

$$\text{Entonces,} \quad \mathbf{AB} = \left[\begin{array}{cc|cc} \mathbf{C} & \mathbf{E} \\ \mathbf{D} & \mathbf{F} \end{array} \right] \left[\begin{array}{c|c} \mathbf{G} \\ \mathbf{H} \end{array} \right] = \left[\begin{array}{cc|cc} \mathbf{CG} + \mathbf{EH} \\ \mathbf{DG} + \mathbf{FH} \end{array} \right] = \left[\begin{array}{cc|cc} 4 & -2 \\ 5 & 6 \\ 30 & 8 \\ 8 & 27 \end{array} \right].$$

Ejemplo 7. De manera similar, sean $\alpha = ((1, 2,), (3, 4,))$, $\gamma = ((1, 2,), (3, 4, 5,))$, $\beta = ((1, 2,), (3,))$,

$$\mathbf{A} = \left[\begin{array}{cc|cc} \mathbf{C} & \mathbf{E} \\ \mathbf{D} & \mathbf{F} \end{array} \right] = \left[\begin{array}{cc|cc} 2 & -1 & 3 & 1 \\ 1 & 0 & 1 & 2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right] \quad \text{y} \quad \mathbf{B} = \left[\begin{array}{c|c} \mathbf{G} & \mathbf{J} \\ \mathbf{H} & \mathbf{K} \end{array} \right] = \left[\begin{array}{c|c} 1 & 2 & 0 \\ -1 & 0 & 0 \\ 0 & 5 & 1 \\ 1 & -1 & 0 \end{array} \right]; \quad \text{entonces}$$

$$\mathbf{AB} = \left[\begin{array}{cc|cc} \mathbf{CG} + \mathbf{EH} & \mathbf{CJ} + \mathbf{EK} \\ \mathbf{DG} + \mathbf{FH} & \mathbf{DH} + \mathbf{FK} \end{array} \right] = \left[\begin{array}{cc|cc} 4 & 18 & 3 \\ 3 & 5 & 1 \\ 0 & 5 & 1 \\ 1 & -1 & 0 \end{array} \right].$$

3.B.3. Matriz por vector como suma de submatrices por subvectores

Definición 3.5. Sea $_{\gamma|}\mathbf{b}$ el vector formado por las componentes de \mathbf{b} indicadas en γ , es decir, $_{\gamma|}\mathbf{b} = (_{\gamma|}\mathbf{I})\mathbf{b}$.

Si \mathbf{A} tiene n columnas y $\mathbf{b} \in \mathbb{R}^n$; y si $\gamma_1, \dots, \gamma_h$ son listas de índices tales que cada índice $j = 1 : n$ está en una, y solo una, de las listas γ_i , entonces es inmediato ver que la combinación lineal $\mathbf{A}\mathbf{b}$ se puede expresar como suma de varias combinaciones lineales

$$\mathbf{A}\mathbf{b} = \sum_{k=1}^h (\mathbf{A}_{|\gamma_k}) (_{\gamma_k|}\mathbf{b}). \quad (3.7)$$

Por ejemplo, si $n = 4$, $\gamma_1 = (1, 3,)$, y $\gamma_2 = (2, 4,)$. Entonces, reordenando los sumandos tenemos

$$\mathbf{A}\mathbf{b} = \overbrace{(\mathbf{A}_{|1}) (_{1|}\mathbf{b}) + (\mathbf{A}_{|3}) (_{3|}\mathbf{b})}^{\text{combinación de columnas de índice impar}} + \overbrace{(\mathbf{A}_{|2}) (_{2|}\mathbf{b}) + (\mathbf{A}_{|4}) (_{4|}\mathbf{b})}^{\text{combinación de columnas de índice par}} = (\mathbf{A}_{|\gamma_1}) (_{\gamma_1|}\mathbf{b}) + (\mathbf{A}_{|\gamma_2}) (_{\gamma_2|}\mathbf{b}),$$

es decir, la suma de una combinación lineal de las columnas de índice impar, $(\mathbf{A}_{|\gamma_1}) (_{\gamma_1|}\mathbf{b})$, más otra combinación de las columnas de índice par, $(\mathbf{A}_{|\gamma_2}) (_{\gamma_2|}\mathbf{b})$.

3.C. Propiedades de la traza

Vamos a aprovechar la visión del producto de matrices como una suma de matrices para deducir algunas de las propiedades de la traza de una matriz.

Definición 3.6. La traza de \mathbf{A} es la suma de los elementos de su diagonal principal, es decir:

$$\text{tr}(\mathbf{A}) = a_{11} + a_{22} + \dots + a_{kk}, \quad \text{donde } k = \min(m, n).$$

EJERCICIO 26. Sabiendo que los operadores selectores son lineales y que $\mathbf{A} \mathbf{B} = \sum_{k=1}^p (\mathbf{A}_{|(k,)})(_{(k,)}\mathbf{B})$, demuestre que

- (a) la traza es un operador lineal,
- (b) $\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{A}^\top)$,
- (c) $\text{tr}(\mathbf{A}^\top \mathbf{A}) = \text{tr}(\mathbf{A} \mathbf{A}^\top)$,
- (d) si \mathbf{A} y \mathbf{B} son de orden $m \times n$, entonces: $\text{tr}(\mathbf{A}^\top \mathbf{B}) = \text{tr}(\mathbf{B} \mathbf{A}^\top) = \text{tr}(\mathbf{A} \mathbf{B}^\top) = \text{tr}(\mathbf{B}^\top \mathbf{A})$.

Consecuentemente si \mathbf{A} es de orden $m \times n$ y \mathbf{C} de orden $n \times m$, tenemos que

$$\text{tr}(\mathbf{AC}) = \text{tr}(\mathbf{CA}). \quad (3.8)$$

Basta renombrar \mathbf{C} como \mathbf{B}^\top y aplicar (d) para verlo: $\text{tr}(\mathbf{AC}) = \text{tr}(\mathbf{AB}^\top) = \text{tr}(\mathbf{B}^\top \mathbf{A}) = \text{tr}(\mathbf{CA})$.

Por último, para \mathbf{A} , \mathbf{B} y \mathbf{C} se verifica que $\text{tr}(\mathbf{ABC}) = \text{tr}(\mathbf{CAB}) = \text{tr}(\mathbf{BCA})$, pues aplicando (3.8):

$$\text{tr}((\mathbf{AB})\mathbf{C}) = \text{tr}(\mathbf{C}(\mathbf{AB})) = \text{tr}((\mathbf{CA})\mathbf{B}) = \text{tr}(\mathbf{B}(\mathbf{CA})).$$

Parte II

Transformaciones elementales, métodos de eliminación y matriz inversa

Transformaciones elementales y métodos de eliminación

Como veremos a lo largo de este curso, hay un algoritmo que permite resolver gran cantidad de problemas: decidir si una matriz es invertible y, si lo es, encontrar su inversa; decidir si un sistema de ecuaciones lineales es resoluble y, si lo es, resolverlo; encontrar una base del complemento ortogonal de un subespacio de \mathbb{R}^n ; pasar de las ecuaciones paramétricas a las cartesianas (y viceversa); calcular determinantes, etc. De hecho, ¡prácticamente todos los problemas que veremos en este curso se puede resolver con un único algoritmo!

Este algoritmo se denomina *Método de Eliminación* y es sorprendentemente simple: *consiste en aplicar una secuencia de transformaciones sobre las columnas (o sobre las filas) una matriz*. Cada transformación en la secuencia (cada paso en la secuencia) realiza uno de estos dos tipos de operaciones:

Tipo I. Sumar un múltiplo de un vector a *otro* vector. Ejemplo: $[a; b;] \rightarrow [a; (b + \lambda a);]$.

Tipo II. Multiplicar un vector por un número *distinto de cero*. Ejemplo: $[a; b;] \rightarrow [a; (\lambda b);]$.

Ambas operaciones se denominan *transformaciones elementales*.¹

4.1. Transformaciones y matrices elementales

Denotaremos con “ τ ” al operador correspondiente a la transformación aplicada. Lo situaremos como subíndice a la derecha de la matriz para indicar que opera sobre las columnas: \mathbf{A}_τ . Por debajo de “ τ ” [entre corchetes] indicaremos los detalles de la transformación: $\tau_{[\text{detalles}]}$.

La matriz que resulta de aplicar una sola transformación elemental sobre \mathbf{I} tiene un nombre especial:

Definición 4.1. *Matriz elemental \mathbf{I}_τ es cualquier matriz obtenida tras aplicar una sola transformación elemental sobre las columnas de una matriz identidad.*

En esta sección veremos que $\mathbf{A}_\tau = \mathbf{A}(\mathbf{I}_\tau)$. Es decir, que mediante un producto de matrices se puede aplicar la transformación elemental τ a las columnas de \mathbf{A} . Antes debemos detenernos en las transformaciones (y matrices) elementales de tipo I y II.

4.1.1. Transformaciones elementales (y matrices elementales) de Tipo I

Transf. elemental de Tipo I: $\tau_{[(\lambda)^i+j]}$ suma λ veces el vector i -ésimo al vector j -ésimo (con $j \neq i$).
(solo modifica el vector cuyo índice aparece en última posición en el corchete).

Librería NAcAL para Python

```
T( (8,2,3) ) # Suma ocho veces el segundo vector al tercero
```

$$\tau_{[(8)2+3]}$$

¹En muchos manuales consideran el intercambio como una tercera transformación elemental, pero aquí no lo haremos.

Una **matriz elemental Tipo I** es la obtenida *tras aplicar una sola transformación elemental Tipo I* sobre las columnas de una matriz identidad:

$$\mathbf{I}_{[(\lambda)\mathbf{i}+\mathbf{j}]}$$

Por ejemplo, si sobre las columnas de la matriz identidad de orden 3 aplicamos la transformación $\tau_{[(\lambda)\mathbf{2}+\mathbf{3}]}$ (que suma “ λ veces” la segunda columna a la tercera) obtenemos:

$$\mathbf{I}_{[(\lambda)\mathbf{2}+\mathbf{3}]} = \begin{bmatrix} 1 & & \\ & 1 & \lambda \\ & & 1 \end{bmatrix};$$

que es como la matriz identidad de orden 3 salvo porque la segunda componente de su tercera columna ahora es λ (en lugar de cero).

Fíjese que la notación $[(\lambda)\mathbf{i}+\mathbf{j}]$ indica que sumamos λ veces la \mathbf{i} -ésima columna a la columna \mathbf{j} -ésima; pero también describe la matriz elemental correspondiente: en negrita aparecen los índices (fila \mathbf{i} -ésima, columna \mathbf{j} -ésima) de la componente de la matriz \mathbf{I} cuyo valor ahora es (λ) .

Librería NAcAL para Python

```
I(3) & T( (8,2,3) ) # Transf. de las columnas de la matriz identidad de orden 3
                  # (suma 8 veces la segunda columna a la tercera)
```

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 8 \\ 0 & 0 & 1 \end{bmatrix}$$

Aplicar $\tau_{[(\lambda)\mathbf{i}+\mathbf{j}]}$ sobre las columnas de \mathbf{I} solo cambia la \mathbf{j} -ésima columna. Las columnas de la nueva matriz son

$$\left(\mathbf{I}_{[(\lambda)\mathbf{i}+\mathbf{j}]} \right)_{|k} = \begin{cases} \text{cuando } k = j : \left(\mathbf{I}_{[(\lambda)\mathbf{i}+\mathbf{j}]} \right)_{|j} = \lambda \mathbf{I}_{|i} + \mathbf{I}_{|j} & \text{(nueva columna } j\text{-ésima)} \\ \text{cuando } k \neq j : \left(\mathbf{I}_{[(\lambda)\mathbf{i}+\mathbf{j}]} \right)_{|k} = \mathbf{I}_{|k} & \text{(el resto de columnas no cambian)} \end{cases} \quad (4.1)$$

donde $k = 1 : n$. Veamos qué pasa si aplicamos una transformación elemental Tipo I sobre \mathbf{A} .

Transformación de las columnas de \mathbf{A} : Si aplicamos $\tau_{[(\lambda)\mathbf{i}+\mathbf{j}]}$ a las columnas de \mathbf{A} obtenemos una matriz cuyas columnas son:

$$\left(\mathbf{A}_{[(\lambda)\mathbf{i}+\mathbf{j}]} \right)_{|k} = \begin{cases} \lambda \mathbf{A}_{|i} + \mathbf{A}_{|j} = \mathbf{A}(\lambda \mathbf{I}_{|i} + \mathbf{I}_{|j}) = \mathbf{A} \left(\mathbf{I}_{[(\lambda)\mathbf{i}+\mathbf{j}]} \right)_{|j} & \text{para } k = j \\ \mathbf{A}_{|k} = \mathbf{A}(\mathbf{I}_{|k}) = \mathbf{A} \left(\mathbf{I}_{[(\lambda)\mathbf{i}+\mathbf{j}]} \right)_{|k} & \text{para } k \neq j \end{cases} = \mathbf{A} \left(\mathbf{I}_{[(\lambda)\mathbf{i}+\mathbf{j}]} \right)_{|k},$$

donde $k = 1 : n$ y donde hemos usado (4.1).

Por tanto, aplicar sobre las columnas de \mathbf{A} la transformación $\tau_{[(\lambda)\mathbf{i}+\mathbf{j}]}$ es equivalente a multiplicar \mathbf{A} por la matriz elemental $\mathbf{I}_{[(\lambda)\mathbf{i}+\mathbf{j}]}$ de orden n por la derecha,

$$\mathbf{A}_{[(\lambda)\mathbf{i}+\mathbf{j}]} = \mathbf{A} \left(\mathbf{I}_{[(\lambda)\mathbf{i}+\mathbf{j}]} \right).$$

Por ejemplo, si $\mathbf{A} = \begin{bmatrix} 0 & 0 & c \\ 0 & b & c \\ a & b & c \end{bmatrix}$; al sumar “ λ veces” la segunda columna a la tercera obtenemos:

$$\mathbf{A}_{[(\lambda)\mathbf{2}+\mathbf{3}]} = \begin{bmatrix} 0 & 0 & (\lambda 0 + c) \\ 0 & b & (\lambda b + c) \\ a & b & (\lambda b + c) \end{bmatrix} = \begin{bmatrix} 0 & 0 & c \\ 0 & b & c \\ a & b & c \end{bmatrix} \begin{bmatrix} 1 & & \\ & 1 & \lambda \\ & & 1 \end{bmatrix} = \mathbf{A} \left(\mathbf{I}_{[(\lambda)\mathbf{2}+\mathbf{3}]} \right).$$

4.1.2. Transformaciones elementales (y matrices elementales) de Tipo II

Transf. elemental de Tipo II: $\tau_{[(\alpha)\mathbf{i}]}$ multiplica por α el i -ésimo vector (con $\alpha \neq 0$).

(solo modifica el vector cuyo índice aparece en última posición en el corchete).

```
T( (5, 1) ) # multiplica por 5 el primer vector
```

Librería NAcAL para Python

$$\tau_{[(5)\mathbf{1}]}$$

Una **matriz elemental Tipo II** es la obtenida tras aplicar una sola transformación elemental Tipo II sobre las columnas de una matriz identidad:

$$\mathbf{I}_{[(\alpha)\mathbf{i}]}$$

Si sobre las columnas de la matriz identidad de orden 3 aplicamos la transformación $\tau_{[(\alpha)\mathbf{2}]}$ obtenemos:

$$\mathbf{I}_{[(\alpha)\mathbf{2}]} = \begin{bmatrix} 1 & & \\ & \alpha & \\ & & 1 \end{bmatrix};$$

que es como la matriz $\mathbf{I}_{3 \times 3}$ salvo porque la segunda componente de la diagonal es α .

```
I(3) & T( (5,1) ) # Transf. de las columnas de la matriz identidad de orden 3
# (multiplica por 5 la primera columna)
```

Librería NAcAL para Python

$$\begin{bmatrix} 5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Aplicar $\tau_{[(\alpha)\mathbf{i}]}$ sobre las columnas de \mathbf{I} solo cambia la i -ésima columna. Las columnas de la nueva matriz son

$$\left(\mathbf{I}_{[(\alpha)\mathbf{i}]} \right)_{|k} = \begin{cases} \text{cuando } k = i : \left(\mathbf{I}_{[(\alpha)\mathbf{i}]} \right)_{|i} = \alpha \mathbf{I}_{|i} & \text{(nueva columna } i\text{-ésima)} \\ \text{cuando } k \neq i : \left(\mathbf{I}_{[(\alpha)\mathbf{i}]} \right)_{|k} = \mathbf{I}_{|k} & \text{(el resto de columnas no cambian)} \end{cases}, \text{ donde } k = 1 : n.$$

Es decir, la matriz $\mathbf{I}_{[(\alpha)\mathbf{i}]}$ es como \mathbf{I} salvo por la i -ésima componente de su diagonal, que es α .

Transformación de las columnas de \mathbf{A} : Si aplicamos $\tau_{[(\alpha)\mathbf{i}]}$ a las columnas de \mathbf{A} obtenemos una matriz $m \times n$ cuyas columnas son:

$$\left(\mathbf{A}_{[(\alpha)\mathbf{i}]}^{\tau} \right)_{|k} = \begin{cases} \alpha \mathbf{A}_{|k} = \mathbf{A}(\alpha \mathbf{I}_{|k}) & \text{si } k = i \\ \mathbf{A}_{|k} = \mathbf{A}(\mathbf{I}_{|k}) & \text{para el resto de columnas} \end{cases} = \mathbf{A} \left(\mathbf{I}_{[(\alpha)\mathbf{i}]}^{\tau} \right)_{|k}, \text{ donde } k = 1 : n.$$

De nuevo aplicar una transformación elemental (ahora de *Tipo II*) sobre las columnas de \mathbf{A} es equivalente a multiplicar dicha matriz \mathbf{A} por la correspondiente matriz elemental (de *Tipo II*) de orden n por la derecha.

$$\mathbf{A}_{[(\alpha)\mathbf{i}]}^{\tau} = \mathbf{A} \left(\mathbf{I}_{[(\alpha)\mathbf{i}]}^{\tau} \right).$$

Por ejemplo, si $\mathbf{A} = \begin{bmatrix} 0 & 0 & c \\ 0 & b & c \\ a & b & c \end{bmatrix}$ entonces al multiplicar la segunda columna por α resulta que:

$$\mathbf{A}_{[(\alpha)\mathbf{2}]}^{\tau} = \begin{bmatrix} 0 & 0 & c \\ 0 & \alpha b & c \\ a & \alpha b & c \end{bmatrix} = \begin{bmatrix} 0 & 0 & c \\ 0 & b & c \\ a & b & c \end{bmatrix} \begin{bmatrix} 1 & & \\ & \alpha & \\ & & 1 \end{bmatrix} = \mathbf{A} \left(\mathbf{I}_{[(\alpha)\mathbf{2}]}^{\tau} \right).$$

4.1.3. Transformación de las columnas de una matriz

Como al aplicar cualquier transformación elemental τ sobre las columnas de \mathbf{A} siempre se verifica que:

$$\mathbf{A}_{\tau} = \mathbf{A}(\mathbf{I}_{\tau}),$$

disponemos de una nueva regla de reescritura asociativa:

$$(\mathbf{AB})_{\tau} = \mathbf{AB}(\mathbf{I}_{\tau}) = \mathbf{A}(\mathbf{B}_{\tau}).$$

Por tanto no son necesarios los paréntesis. Basta escribir: \mathbf{AB}_{τ}



Recuerde que

- **multiplicar una columna por cero no es una transformación elemental;**
- **restar una columna a si misma tampoco lo es.**

4.1.4. Sobre las matrices elementales y su notación

Las matrices elementales (de orden n) se obtienen aplicando las correspondientes transformaciones elementales sobre la matriz identidad (de orden n). La notación empleada “casi” describe por completo las correspondientes matrices elementales (solo falta indicar el orden de la matriz identidad). Así,

- la matriz $\mathbf{I}_{[(-7)\mathbf{1}+2]}^{\tau}$ (de orden n) es como la matriz identidad (de orden n) salvo por la componente de la fila 1 y columna 2, que es un (-7) . Por tanto

$$\mathbf{I}_{[(-7)\mathbf{1}+2]}^{\tau} = \begin{bmatrix} 1 & -7 \\ 0 & 1 \end{bmatrix}, \text{ pero también } \mathbf{I}_{[(-7)\mathbf{1}+2]}^{\tau} = \begin{bmatrix} 1 & -7 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \text{ ó } \mathbf{I}_{[(-7)\mathbf{1}+2]}^{\tau} = \begin{bmatrix} 1 & -7 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \dots$$

- la matriz $\mathbf{I}_{\tau_{[(3)1]}}$ es como la matriz identidad (de orden n) salvo por que la primera componente de su diagonal es un (3) . Por tanto

$$\mathbf{I}_{\tau_{[(3)1]}} = [3]; \quad \text{pero también, } \mathbf{I}_{\tau_{[(3)1]}} = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}; \quad \text{ó } \mathbf{I}_{\tau_{[(3)1]}} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \text{ó...}$$

Fíjese que la matriz identidad es simultáneamente de **Tipo I** y de **Tipo II**: $\mathbf{I}_{\tau_{[(0)i+j]}} = \mathbf{I} = \mathbf{I}_{\tau_{[(1)i]}}$.

Nota 3. Es fácil comprobar que la transpuesta de una matriz elemental es otra matriz elemental.

- Para las matrices elementales de Tipo I resulta evidente que $\left(\mathbf{I}_{\tau_{[(\lambda)i+j]}}\right)^T = \mathbf{I}_{\tau_{[(\lambda)j+i]}}$,
- y las matrices elementales de Tipo II son simétricas.



Solo hay dos tipos de transformaciones elementales

de Tipo I. Suman a un vector un múltiplo de otro vector: $\tau_{[(\lambda)i+j]}$

de Tipo II. Multiplican un vector por un número distinto de cero: $\tau_{[(\alpha)i]}$

Llamamos *matriz elemental* a la resultante de aplicar una sola transformación elemental τ sobre las columnas de la matriz identidad, \mathbf{I}_{τ} . Solo hay dos tipos:

de Tipo I. $\mathbf{I}_{\tau_{[(\lambda)i+j]}}$, donde $j \neq i$

de Tipo II. $\mathbf{I}_{\tau_{[(\alpha)i]}}$, donde $\alpha \neq 0$.

Al aplicar una transformación elemental τ sobre las columnas de \mathbf{A} , se verifica que:

$$\mathbf{A}_{\tau} = \mathbf{A}(\mathbf{I}_{\tau}).$$

Consecuentemente tenemos una nueva propiedad asociativa:

$$(\mathbf{AB})_{\tau} = \mathbf{A}(\mathbf{B}_{\tau}) = \mathbf{AB}_{\tau}.$$

Además, la transpuesta de una matriz elemental también es elemental (y del mismo tipo).

EJERCICIO 27. Escriba las siguientes matrices elementales de orden 4.

(a) $\mathbf{I}_{\tau_{[(-7)1+3]}}$

(b) $\mathbf{I}_{\tau_{[(1)1+4]}}$

(c) $\mathbf{I}_{\tau_{[(3)2+1]}}$

(d) $\mathbf{I}_{\tau_{[(-10)3]}}$

EJERCICIO 28. ¿Qué operaciones realizan las matrices anteriores si multiplican por la derecha a \mathbf{B} ?

EJERCICIO 29. Sea la matriz $\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix}$, exprese las siguientes transformaciones elementales de \mathbf{A} como productos de \mathbf{A} por matrices elementales y escriba el resultado:

- (a) Multiplicar la primera columna por 3.
- (b) Restar la primera columna de la segunda.

4.1.5. Secuencias de transformaciones elementales. Parte I

Considere la aplicación de una secuencia de k transformaciones elementales sobre las columnas de \mathbf{A} :

$$\left(\left(\dots (\mathbf{A}_{\tau_1})_{\tau_2} \dots \right)_{\tau_{(k-1)}} \right)_{\tau_k}$$

Denotamos la aplicación dicha secuencia con $\mathbf{A}_{\tau_1 \dots \tau_k}$. Es fácil ver que dicha aplicación es equivalente al producto de \mathbf{A} por una secuencia de k matrices elementales:

$$\mathbf{A}_{\tau_1 \dots \tau_k} = \left(\left(\dots (\mathbf{A}_{\tau_1})_{\tau_2} \dots \right)_{\tau_{(k-1)}} \right)_{\tau_k} = \left(\left(\dots (\mathbf{A} \mathbf{I}_{\tau_1})_{\tau_2} \dots \right)_{\tau_{(k-1)}} \right)_{\tau_k} \mathbf{I}_{\tau_k} = \mathbf{A} (\mathbf{I}_{\tau_1}) \dots (\mathbf{I}_{\tau_k}).$$

En particular, si $\mathbf{A} = \mathbf{I}$ entonces

$$\mathbf{I}_{\tau_1 \dots \tau_k} = (\mathbf{I}_{\tau_1}) \dots (\mathbf{I}_{\tau_k}), \quad (4.2)$$

es decir, *la matriz $(\mathbf{I}_{\tau_1 \dots \tau_k})$ es producto de matrices elementales*; y por tanto

$$\mathbf{A}_{\tau_1 \dots \tau_k} = \mathbf{A} (\mathbf{I}_{\tau_1 \dots \tau_k}).$$

EJERCICIO 30. Demuestre que la aplicación de una secuencia de transformaciones elementales de las columnas es un *operador lineal*.



Puesto que la aplicación de una secuencia de transformaciones elementales es un **producto de matrices**:

$$\mathbf{A}_{\tau_1 \dots \tau_k} = \mathbf{A} (\mathbf{I}_{\tau_1}) \dots (\mathbf{I}_{\tau_k}) = \mathbf{A} (\mathbf{I}_{\tau_1 \dots \tau_k})$$

la aplicación de una secuencia de transformaciones elementales de las columnas es un **operador lineal**.

Notación Denotaremos una secuencia de k transformaciones de las *columnas* de \mathbf{A} con el siguiente esquema:

$$\mathbf{A} \xrightarrow{\tau_1} \mathbf{A}_{\tau_1} \xrightarrow{\tau_2} \mathbf{A}_{\tau_1 \tau_2} \xrightarrow{\tau_3} \dots \xrightarrow{\tau_k} \mathbf{A}_{\tau_1 \dots \tau_k} \quad \text{o agregando varios pasos} \quad \mathbf{A} \xrightarrow{\tau_1 \atop \tau_p} \mathbf{A}_{\tau_1 \dots \tau_p} \xrightarrow{\tau_{p+1} \atop \tau_k} \mathbf{A}_{\tau_1 \dots \tau_k}.$$

Por ejemplo, la sucesión de transformaciones $\mathbf{A}_{\tau_{[(3)2]}\tau_{[(2)1+2]}}$ donde $\mathbf{A} = \begin{bmatrix} 0 & 1 & 5 \\ 7 & 0 & 4 \end{bmatrix}$, se escribe como:

$$\begin{bmatrix} 0 & 1 & 5 \\ 7 & 0 & 4 \end{bmatrix} \xrightarrow{\tau_{[(3)2]}} \begin{bmatrix} 0 & 3 & 5 \\ 7 & 0 & 4 \end{bmatrix} \xrightarrow{\tau_{[(2)1+2]}} \begin{bmatrix} 0 & 3 & 5 \\ 7 & 14 & 4 \end{bmatrix}; \quad \text{o bien como} \quad \begin{bmatrix} 0 & 1 & 5 \\ 7 & 0 & 4 \end{bmatrix} \xrightarrow{\tau_{[(3)2]}\tau_{[(2)1+2]}} \begin{bmatrix} 0 & 3 & 5 \\ 7 & 14 & 4 \end{bmatrix}.$$

Alternativamente podemos expresar una secuencia de transformaciones elementales mediante productos de **matrices elementales**. Por ejemplo, para la secuencia anterior:

$$\begin{bmatrix} 0 & 1 & 5 \\ 7 & 0 & 4 \end{bmatrix}_{\tau_{[(3)2]}\tau_{[(2)1+2]}} = \begin{bmatrix} 0 & 1 & 5 \\ 7 & 0 & 4 \end{bmatrix} (\mathbf{I}_{\tau_{[(3)2]}}) (\mathbf{I}_{\tau_{[(2)1+2]}}) = \begin{bmatrix} 0 & 3 & 5 \\ 7 & 14 & 4 \end{bmatrix}.$$

T((2,3,4)) & T((-7,1)) & T((-1,3,1))

Librería NAcAL para Python

$\tau_{[(2)3+4]}\tau_{[(-7)1]}\tau_{[(-1)3+1]}$

4.1.6. Intercambios

Aunque solo hemos considerado como transformaciones elementales las de *Tipo I* y *II*, la mayoría de manuales de Álgebra Lineal consideran como tercer tipo el *intercambio*:

Intercambio: $\mathbf{I}_{[i \rightleftharpoons j]}^{\tau} \rightarrow$ intercambia los vectores i -ésimo y j -ésimo.

```
T( {1, 3} ) # intercambia los vectores primero y tercero
```

Librería NAcAL para Python

$$\mathbf{I}_{[1 \rightleftharpoons 3]}^{\tau}$$

Definición 4.2 (Matriz de intercambio). *Cualquier matriz obtenida tras intercambiar las columnas p y q de la matriz identidad de orden n :* $\mathbf{I}_{[p \rightleftharpoons q]}^{\tau}$

Puesto que en $\mathbf{I}_{[p \rightleftharpoons q]}^{\tau}$ solo han cambiado las columnas p y q de \mathbf{I} , se deduce que $\mathbf{I}_{[p \rightleftharpoons q]}^{\tau}$ es simétrica:

$$\left(\mathbf{I}_{[p \rightleftharpoons q]}^{\tau} \right)_{|j} = \begin{cases} \mathbf{I}_{|q} & \text{si } j = p \\ \mathbf{I}_{|p} & \text{si } j = q \\ \mathbf{I}_{|j} & \text{si } j \neq p, q \end{cases} \implies \left(\mathbf{I}_{[p \rightleftharpoons q]}^{\tau} \right)_{|j} = \begin{cases} 1 & \text{si } k = q \text{ y } j = p \\ 1 & \text{si } k = p \text{ y } j = q \\ 1 & \text{si } k = j \neq p, q \\ 0 & \text{en el resto de casos} \end{cases} = \left(\mathbf{I}_{[p \rightleftharpoons q]}^{\tau} \right)_{|k}^j,$$

es decir, transponiendo (intercambiando los índices k y j) la matriz no cambia. Por tanto:

toda matriz intercambio $\mathbf{I}_{[p \rightleftharpoons q]}^{\tau}$ es simétrica.

No consideramos el intercambio como una transformación elemental puesto que en realidad es una sucesión de transformaciones elementales de **Tipo I** más una transformación de **Tipo II** que multiplica una de las dos columnas intercambiadas por -1 . El siguiente ejercicio le pide que lo compruebe.

EJERCICIO 31.

- Mediante una secuencia de transformaciones elementales Tipo I y II transforme $\mathbf{I}_{3 \times 3}$ en la matriz $\mathbf{I}_{[1 \rightleftharpoons 2]}^{\tau}$.
- Mediante el producto de las matrices elementales correspondientes a los pasos dados en el apartado anterior obtenga la matriz intercambio $\mathbf{I}_{[i \rightleftharpoons j]}^{\tau}$ de orden n .

```
i = 2; j = 4 # indicamos quienes son j e i
I(4) & T( [(-1,j), (-1,j,i), (1,i,j), (-1,j,i)] ) # aplicación de lista de transf. elem.
# I(4) & T( {i,j} ) # es más sencillo aplicar intercambio
```

Librería NAcAL para Python

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$



$\mathbf{I}_{[i \rightleftharpoons j]}^{\tau}$ se puede factorizar en un producto de matrices elementales **Tipo I** y una matriz elemental **Tipo II**, (que multiplica por -1 una de las columnas intercambiadas) cuando $i \neq j$. Si $i = j$, entonces $\mathbf{I}_{[i \rightleftharpoons j]}^{\tau} = \mathbf{I}$.

Intercambio de las columnas de \mathbf{A} :

$$\left(\mathbf{A}_{[i \rightleftharpoons j]}^{\tau} \right)_{|k} = \begin{cases} \mathbf{A}_{|j} = \mathbf{A}_{|j} & \text{si } k = i \text{ (es decir, en la posición } i \text{ se coloca la columna } j) \\ \mathbf{A}_{|i} = \mathbf{A}_{|i} & \text{si } k = j \text{ (es decir, en la posición } j \text{ se coloca la columna } i) \\ \mathbf{A}_{|k} = \mathbf{A}_{|k} & \text{en el resto de casos se mantienen las columnas en su sitio} \end{cases} = \mathbf{A} \left(\mathbf{I}_{[i \rightleftharpoons j]}^{\tau} \right)_{|k}.$$

Si $\mathbf{A} = \begin{bmatrix} a & 0 & 0 \\ b & b & 0 \\ c & c & c \end{bmatrix}$, entonces $\mathbf{A}_{[1 \rightleftharpoons 2]}^{\tau} = \mathbf{A} \left(\mathbf{I}_{[1 \rightleftharpoons 2]}^{\tau} \right) = \begin{bmatrix} a & 0 & 0 \\ b & b & 0 \\ c & c & c \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ intercambia las columnas 1 y 2.

Librería NAcAL para Python

```
a, b, c, = sympy.symbols('a b c')           # definimos variables simbólicas
A = Matrix([ [a,0,0], [a,b,0], [a,b,c] ])    # definimos Matrix fila a fila
A & T( {1,2} )                               # intercambio de las dos primeras columnas
```

$$\begin{bmatrix} 0 & a & 0 \\ b & a & 0 \\ b & a & c \end{bmatrix}$$

4.1.7. Permutaciones

Mediante una sucesión intercambios se logra un reordenamiento o *permutación* de elementos de un sistema.

Permutación: $\tau_{[\mathfrak{S}]}$ \rightarrow reordena una lista mediante una sucesión de intercambios.

Definición 4.3 (Matriz permutación $\mathbf{I}_{[\mathfrak{S}]}^{\tau}$). *Cualquier matriz obtenida tras realizar un número arbitrario de intercambios entre las columnas de la matriz identidad.*

Librería NAcAL para Python

```
I(4) & T( [{1,3}, {4,2}, {3,2}] )           # ejemplo de matriz permutación
```

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Por definición, una matriz permutación se puede factorizar como producto de matrices intercambio. Aunque las matrices intercambio $\mathbf{I}_{[p \rightleftharpoons q]}^{\tau}$ son simétricas, las matrices permutación $\mathbf{I}_{[\mathfrak{S}]}^{\tau}$ no son simétricas en general:²

$$\left(\mathbf{I}_{[1 \rightleftharpoons 3]}^{\tau} \right) \left(\mathbf{I}_{[1 \rightleftharpoons 2]}^{\tau} \right) = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

²el producto de dos matrices simétricas es simétrico si y solo si dichas matrices conmutan, es decir, si $\mathbf{AB} = \mathbf{BA}$.

El intercambio, $\tau_{[i \rightleftharpoons j]}$, es una secuencia de transformaciones *Tipo I* junto a una de *Tipo II* (que multiplica por -1 uno de los vectores). Las matrices intercambio, $\mathbf{I}_{\tau_{[i \rightleftharpoons j]}}$, son simétricas y su cuadrado es la matriz identidad (es decir, “intercambiar dos veces los mismos vectores nos deja como al principio”).

Una secuencia de intercambios da lugar a un reordenamiento de los vectores. Llamamos matriz permutación al resultado de aplicar una sucesión de intercambios sobre las columnas de \mathbf{I} . En general las matrices permutación $\mathbf{I}_{\tau_{[\mathbf{S}]}}$ no son simétricas.

4.2. Eliminación por columnas

Definición 4.4. Llamamos **pivote** de una columna no nula, a su primer componente no nulo, y **posición de pivote** al índice de la fila en la que está el pivote.

Definición 4.5. Una matriz \mathbf{K} es **pre-escalonada** si las componentes a la derecha de cada pivote son nulas.

Ejemplo 8.

$$\mathbf{K} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 7 \\ 0 & 0 & 0 & 6 \\ 3 & 0 & 0 & 0 \\ 6 & 0 & 1 & 0 \end{bmatrix}$$

El pivote de la primera columna es 3, el de la tercera es 1, y el de la cuarta 7. La posición de pivote de la primera columna es 4 (cuarta fila), la de la tercera es 5 (quinta fila) y a de la última es 2 (segunda fila).

4.2.1. Método de eliminación (o eliminación de “Izquierda a derecha”)

El método de eliminación consiste en *pre-escalonar* una matriz mediante transformaciones elementales, es decir, en lograr una matriz en la que todas las componentes a la derecha de cada pivote son nulas:

$$\begin{bmatrix} 0 & 1 & 1 \\ 2 & -2 & 2 \\ 2 & -8 & 1 \end{bmatrix} \xrightarrow{[(-1)\mathbf{2}+\mathbf{3}]} \begin{bmatrix} 0 & 1 & 0 \\ 2 & -2 & 4 \\ 2 & -8 & 9 \end{bmatrix} \xrightarrow{\begin{matrix} [(1)\mathbf{1}+\mathbf{2}] \\ [(-2)\mathbf{1}+\mathbf{3}] \end{matrix}} \begin{bmatrix} 0 & 1 & 0 \\ 2 & 0 & 0 \\ 2 & -6 & 5 \end{bmatrix}$$

Toda matriz se puede *pre-escalonar* mediante transformaciones elementales. ¡Prácticamente el resto del curso descansa sobre en este teorema! Lo demostraremos por inducción: demostrando que si es cierto para cualquier matriz de n columnas, también lo es para cualquier matriz de $n + 1$ columnas.

Teorema 4.2.1 (Eliminación). Para toda \mathbf{A} existe una secuencia $\tau_1 \cdots \tau_k$ tal que $\mathbf{A}_{\tau_1 \cdots \tau_k}$ es pre-escalonada.

Demostración. Por inducción sobre el número de filas de \mathbf{A} .

PASO DE INDUCCIÓN: Supongamos que \mathbf{A} tiene m filas y que el resultado es cierto para matrices de $(m - 1)$ filas. Es decir, existe una sucesión $\tau_1 \cdots \tau_k$ de transformaciones elementales de las columnas tal que

$$\mathbf{A}_{\tau_1 \cdots \tau_k} = \begin{bmatrix} \mathbf{K} \\ b_{m1} & \cdots & b_{mn} \end{bmatrix}; \quad \text{y donde la submatriz } \mathbf{K} \text{ es pre-escalonada.}$$

Entonces $\mathbf{A}_{\tau_1 \cdots \tau_k}$ también es pre-escalonada si se da alguno de los siguientes supuestos en la última fila:

- *Supuesto 1.* La última fila es nula, es decir, $b_{mj} = 0$ para $j = 1 : n$.
- *Supuesto 2.* Son nulos los componentes de la última fila que están bajo las columnas nulas de \mathbf{K} (como en el caso de la segunda columna del Ejemplo 8).

- *Supuesto 3.* De los componentes de la última fila situados bajo las columnas nulas de \mathbf{K} , solo uno es distinto de cero, y las componentes a su derecha son nulas (como en la última fila del Ejemplo 8, donde 1 es el primer componente no nulo de aquellos con una columna de ceros por encima).

Si no se da ninguno de estos supuestos, entonces necesariamente se da el siguiente caso:

- como no se da el supuesto 1, la última fila es no nula.
- como no se da el supuesto 2, la última fila contiene algún pivote.
- como no se da el supuesto 3, a la derecha del primer pivote en la última fila hay coeficientes no nulos.

Si b_{mr} es el primer pivote de los situados en la última fila (i.e., de aquellos componentes situados bajo las columnas nulas de \mathbf{K} , es el situado más a la izquierda), podemos *eliminar* los componentes a su derecha restando de cada columna j -ésima (con $j \geq r$) la columna r -ésima multiplicada por $(\frac{b_{mj}}{b_{mr}})$. Así, aplicando la secuencia de transformaciones ³

$$\left[\left(\frac{-b_{mj}}{b_{mr}} \right) r + j \right] \quad \text{para todo } j \geq r,$$

obtenemos una matriz que cumple el *Supuesto 3*.

BASE DE INDUCCIÓN: Se prueba igual que en el paso de inducción, pero sin \mathbf{K} , es decir con una matriz \mathbf{A} con una sola fila y donde $\mathbf{A}_{|r}$ es la primera columna no nula. \square

Llamaremos a este Método “*Eliminación de izquierda a derecha*” o sencillamente “*Eliminación*”.

```
A = Matrix( [ [0,2,2], [1,-2,-8], [1,1,1] ] )
Elim(A,1)
```

Librería NAcAL para Python

Eliminación por columnas

$$\begin{bmatrix} 0 & 2 & 2 \\ 1 & -2 & -8 \\ 1 & 1 & 1 \end{bmatrix} \xrightarrow{[(-1)2+3]} \begin{bmatrix} 0 & 2 & 0 \\ 1 & -2 & -6 \\ 1 & 1 & 0 \end{bmatrix} \xrightarrow{\begin{matrix} [(2)1+2] \\ [(6)1+3] \end{matrix}} \begin{bmatrix} 0 & 2 & 0 \\ 1 & 0 & 0 \\ 1 & 3 & 6 \end{bmatrix}$$

El método de eliminación está implementado en **NAcAL**, pero es importante que usted sepa aplicarlo con lápiz y papel (o mejor aún, que sepa programarlo). ⁴

4.2.2. Método de eliminación Gaussiano

Definición 4.6. Decimos que la matriz \mathbf{L} es *escalonada*, si toda columna que precede a una no nula $\mathbf{L}_{|k}$ es no nula y su posición de pivote es anterior a la posición de pivote de $\mathbf{L}_{|k}$.

(Esta definición implica que a la derecha de cualquier pivote solo puede haber ceros).

En una matriz escalonada las columnas no nulas están a la izquierda y las nulas a la derecha (si las hubiere). Además, conforme nos movemos de izquierda a derecha, los pivotes aparecen cada vez más abajo, es decir, los pivotes describen una escalera descendente (de ahí el nombre de matriz “escalonada”).

Ejemplo 9.

$$\mathbf{L} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 7 & 0 & 0 & 0 \\ 6 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 6 & 1 & 0 \end{bmatrix}$$

El pivote de la primera columna es 7, el de la segunda es 3 y el de la tercera 1. La posición de pivote de la primera columna es 2 (segunda fila), la de la segunda es 4 (cuarta fila) y la de la tercera es 5 (quinta fila).

³que no modifican \mathbf{K} puesto que por encima del pivote b_{mr} todo son ceros

⁴Use el módulo solo para verificar los resultados que usted ha obtenido previamente usando lápiz y papel.

Fíjese que aunque toda matriz *escalonada* es *pre-escalonada*, las matrices *pre-escalonadas* no suelen ser *escalonadas* (Ejemplo 8 en la página 51).

Toda matriz se puede escalonar mediante transformaciones elementales tal como afirma el siguiente

Corolario 4.2.2 (Eliminación Gaussiana). *Para toda \mathbf{A} existen $\tau_1 \cdots \tau_k$ tales que $\mathbf{A}_{\tau_1 \dots \tau_k}$ es escalonada.*

Demostración. Basta pre-escalonar \mathbf{A} y reordenar sus columnas para que los pivotes describan una escalera descendente, dejando las columnas nulas a la derecha de las no nulas. \square

Dicho procedimiento se denomina *Método de Eliminación de Gaussiano*.

Librería NAcAL para Python

```

A = Matrix( [ [0,2,2], [1,-2,-8], [1,1,1] ] )
L = ElimG(A,1)                                     # Eliminación Gaussiana por columnas

```

$$\begin{bmatrix} 0 & 2 & 2 \\ 1 & -2 & -8 \\ 1 & 1 & 1 \end{bmatrix} \xrightarrow{[(-1)\mathbf{2}+\mathbf{3}]} \begin{bmatrix} 0 & 2 & 0 \\ 1 & -2 & -6 \\ 1 & 1 & 0 \end{bmatrix} \xrightarrow{\begin{matrix} \tau \\ [(2)\mathbf{1}+\mathbf{2}] \\ [(6)\mathbf{1}+\mathbf{3}] \end{matrix}} \begin{bmatrix} 0 & 2 & 0 \\ 1 & 0 & 0 \\ 1 & 3 & 6 \end{bmatrix} \xrightarrow{[\mathbf{1}=\mathbf{2}]} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 1 & 6 \end{bmatrix}$$

4.2.3. Método de eliminación Gauss-Jordan

La eliminación gaussiana se puede llevar más lejos, obteniendo una *forma escalonada reducida*.

Definición 4.7. *Decimos que una matriz \mathbf{R} es **escalonada reducida**, si es escalonada, todos los pivotes son “unos” y los componentes situados a derecha e izquierda de los pivotes son cero.*

Ejemplo 10.

$$\mathbf{R} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ \mathbf{1} & 0 & 0 & 0 \\ \frac{6}{7} & 0 & 0 & 0 \\ 0 & \mathbf{1} & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 \end{bmatrix}$$

Corolario 4.2.3 (Eliminación Gauss-Jordan). *Para toda \mathbf{A} existen $\tau_1 \cdots \tau_k$ tales que $\mathbf{A}_{\tau_1 \dots \tau_k}$ es escalonada reducida.*

Demostración. Basta escalonar la matriz, y con cada pivote eliminar el resto de componentes no nulos de su fila. Después se divide cada columna no nula por el valor de su pivote para lograr los “unos”. \square

Dicho procedimiento se denomina *Método de Eliminación Gauss-Jordan*.

Librería NAcAL para Python

```

A = Matrix( [ [0,2,2], [1,-2,-8], [1,1,1] ] )
R = ElimGJ(A,1)                                     # Forma escalonada reducida

```

$$\begin{bmatrix} 0 & 2 & 2 \\ 1 & -2 & -8 \\ 1 & 1 & 1 \end{bmatrix} \xrightarrow{\begin{matrix} \tau \\ [(-1)\mathbf{2}+\mathbf{3}] \\ [(2)\mathbf{1}+\mathbf{2}] \\ [(6)\mathbf{1}+\mathbf{3}] \end{matrix}} \begin{bmatrix} 0 & 2 & 0 \\ 1 & 0 & 0 \\ 1 & 3 & 6 \end{bmatrix} \xrightarrow{[\mathbf{1}=\mathbf{2}]} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 1 & 6 \end{bmatrix} \xrightarrow{\begin{matrix} \tau \\ [(2)\mathbf{1}] \\ [(-1)\mathbf{3}+\mathbf{1}] \\ [(6)\mathbf{2}] \\ [(-1)\mathbf{3}+\mathbf{2}] \end{matrix}} \begin{bmatrix} 4 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 6 \end{bmatrix} \xrightarrow{\begin{matrix} \tau \\ [(\frac{1}{4})\mathbf{1}] \\ [(\frac{1}{6})\mathbf{2}] \\ [(\frac{1}{6})\mathbf{3}] \end{matrix}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

☞ Cualquier matriz se puede *pre-escalonar*, *escalonar* o *escalonar y reducir* mediante una secuencia de transformaciones elementales.

- Una matriz \mathbf{K} es *pre-escalonada* si las componentes a la derecha de cada pivote son cero.
- Una matriz \mathbf{L} es *escalonada* si está pre-escalonada y los pivotes describen una escalera descendente (cada pivote está más bajo que el pivote de la columna anterior) y las columnas nulas (si las hubiere) están todas a la derecha de la matriz.
- Una matriz \mathbf{R} es *escalonada reducida* si es escalonada, cada pivote es un “1” y el resto de componentes de su fila son ceros (por ejemplo, la matriz identidad es escalonada reducida).

Apendices a la lección

4.A. Transformaciones elementales de las filas

☞ Casi todos los manuales de Álgebra Lineal aplican las transformaciones elementales principalmente sobre las filas. Aquí usaremos las transformaciones sobre las filas en la Página 72 para demostrar la Proposición 5.2.4; y serán fundamentales para diagonalizar matrices cuadradas tanto por semejanza como por congruencia (Parte VI del libro). Este apéndice trata las transformaciones sobre las filas de una matriz y su relación con las transformaciones sobre las columnas.

La primera observación es que siempre podemos aplicar una transformación elemental a las filas de \mathbf{A} sencillamente transformando las columnas de \mathbf{A}^\top , y transponiendo de nuevo el resultado. Así concluimos que

$$((\mathbf{A}^\top)_\tau)^\top = ((\mathbf{A}^\top)\mathbf{I}_\tau)^\top = ((\mathbf{I}_\tau)^\top)((\mathbf{A}^\top)^\top) = ((\mathbf{I}_\tau)^\top)\mathbf{A},$$

donde $(\mathbf{I}_\tau)^\top$ es una matriz elemental (por ser la transpuesta de una elemental —Nota 3 en la página 47).

Por tanto, al multiplicar *por la izquierda* de \mathbf{A} , por una matriz elemental, transformamos las filas de \mathbf{A} . Ello sugiere usar la notación ${}_\tau\mathbf{A}$ para la transformación τ de las filas de \mathbf{A} ;

$${}_\tau\mathbf{A} = (\mathbf{I}_\tau)^\top\mathbf{A}.$$

Consecuentemente ${}_\tau\mathbf{I}$ denota la transformación τ de las filas de \mathbf{I} ; es decir,⁵

$${}_\tau\mathbf{I} = (\mathbf{I}_\tau)^\top.$$

Así pues,

$${}_\tau\mathbf{A} = ({}_\tau\mathbf{I})\mathbf{A}.$$

Además, como el producto de matrices es asociativo y $\mathbf{I}_{\tau_1 \cdots \tau_k} = \mathbf{I}_{\tau_1} \cdots \mathbf{I}_{\tau_k}$, deducimos que

$$(\mathbf{I}_{\tau_1 \cdots \tau_k})^\top = (\mathbf{I}_{\tau_1} \cdots \mathbf{I}_{\tau_k})^\top = (\mathbf{I}_{\tau_k})^\top \cdots (\mathbf{I}_{\tau_1})^\top = {}_{\tau_k}\mathbf{I} \cdots {}_{\tau_1}\mathbf{I} = {}_{\tau_k \cdots \tau_1}\mathbf{I}. \quad (4.3)$$

Nótese que al transponer no solo cambiamos de lado los subíndices, sino que también invertimos el orden de la secuencia de transformaciones (de manera similar a cuando se transpone un producto de matrices).

Por todo lo anterior, denotaremos con la transposición a la operación de *invertir el orden de una secuencia de transformaciones elementales*:

$$(\tau_1 \cdots \tau_k)^\top = \tau_k \cdots \tau_1. \quad (4.4)$$

⁵Fíjese que nuevamente la transposición supone cambiar de lado el subíndice!

```
Tr = T( [(66,1,2), {3,2}, (4,1)], 'h' )
Tr
```

Librería NAcAL para Python

$$[(66)1+2] \overset{\tau}{[2 \Rightarrow 3]} \overset{\tau}{[(4)1]}$$

```
~Tr # La transposición invierte el orden
```

Librería NAcAL para Python

$$[(4)1] \overset{\tau}{[2 \Rightarrow 3]} \overset{\tau}{[(66)1+2]}$$

```
I(3) & Tr
```

Librería NAcAL para Python

$$\begin{bmatrix} 4 & 0 & 66 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

```
(~Tr) & I(3) # transpuesta de la matriz anterior
```

Librería NAcAL para Python

$$\begin{bmatrix} 4 & 0 & 0 \\ 0 & 0 & 1 \\ 66 & 1 & 0 \end{bmatrix}$$

Matrices simétricas. Por (4.3) y (4.4) sabemos que $(\mathbf{l}_{\tau_1 \dots \tau_k})^\top = \tau_k \dots \tau_1 \mathbf{l} = (\tau_1 \dots \tau_k)^\top \mathbf{l}$. Consecuentemente la matriz $\mathbf{l}_{\tau_1 \dots \tau_k}$ es simétrica si y solo si

$$(\tau_1 \dots \tau_k)^\top \mathbf{l} = \mathbf{l}_{\tau_1 \dots \tau_k}. \quad (4.5)$$

```
p = T([(9,2), (18,1), (-1,2,1), (frac(1,9),1), (-1,1,2), (frac(1,2),2)], 'h')
p
```

Librería NAcAL para Python

$$[(9)2] \overset{\tau}{[(18)1]} \overset{\tau}{[(-1)2+1]} \overset{\tau}{[(\frac{1}{9})1]} \overset{\tau}{[(-1)1+2]} \overset{\tau}{[(\frac{1}{2})2]}$$

```
(~p & I(2)) == (I(2) & p) # matriz simétrica si se verifica la igualdad
```

Librería NAcAL para Python

```
True
```

```
I(2) & p # veamos que efectivamente la matriz es simétrica
```

Librería NAcAL para Python

$$\begin{bmatrix} 2 & -1 \\ -1 & 5 \end{bmatrix}$$

Una secuencia de transformaciones elementales se comporta casi como un producto de matices: fíjese que aunque la secuencia $\tau_1 \cdots \tau_k$ no es un producto de matrices... ¡casi se puede interpretar como tal!... Tan solo falta que las transformaciones actúen sobre una matriz para saber el orden n de las matrices elementales, y entonces sí actúan como un producto de matrices elementales. De tal manera que al igual que $\mathbf{AB}^\top = (\mathbf{B}^\top)(\mathbf{A}^\top)$, también

$$(\mathbf{A}_{\tau_1 \cdots \tau_k})^\top = (\mathbf{A}(\mathbf{I}_{\tau_1 \cdots \tau_k}))^\top = (\tau_k \cdots \tau_1)(\mathbf{A}^\top) = (\tau_k \cdots \tau_1)(\mathbf{A}^\top) = (\tau_1 \cdots \tau_k)^\top(\mathbf{A}^\top).$$

Por (4.3) y (4.4) sabemos que

$$\tau_k \cdots \tau_1 \mathbf{I} = (\tau_1 \cdots \tau_k)^\top \mathbf{I} = (\mathbf{I}_{\tau_1 \cdots \tau_k})^\top.$$

No obstante, en general tanto

$$\tau_1 \cdots \tau_k \mathbf{I} \neq \mathbf{I}_{\tau_1 \cdots \tau_k} \quad \text{como} \quad \tau_k \cdots \tau_1 \mathbf{I} \neq \mathbf{I}_{\tau_1 \cdots \tau_k};$$

salvo cuando $\mathbf{I}_{\tau_1 \cdots \tau_k}$ es simétrica, en cuyo caso ya hemos visto que $\tau_k \cdots \tau_1 \mathbf{I} = \mathbf{I}_{\tau_1 \cdots \tau_k}$. (fíjese en el orden de las transformaciones tanto en las igualdades como en las desigualdades).

Por último, nótese que incluso cuando operamos con las filas, la notación sigue siendo asociativa:

$$\boxed{(\tau_1 \cdots \tau_k \mathbf{A})\mathbf{B} = \tau_1 \cdots \tau_k (\mathbf{AB})}.$$

Así que podemos escribir sencillamente $\tau_1 \cdots \tau_k \mathbf{AB}$.

Notación para una secuencia de transformaciones sobre las filas de una matriz. De igual modo que una secuencia de transformaciones de las columnas se puede expresar de varias formas, también disponemos de representaciones alternativas para una secuencia de transformaciones de las *filas* de una matriz. Por ejemplo:

$$\tau_{[(4)1][2 \rightleftharpoons 3][[(66)1+2]]} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \tau_{\substack{[(4)1] \\ [2 \rightleftharpoons 3] \\ [(66)1+2]}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 0 & 1 \\ 66 & 1 & 0 \end{bmatrix}.$$

es decir, $\tau_{[(4)1][2 \rightleftharpoons 3][[(66)1+2]]} \mathbf{A} = \left(\tau_{[(4)1]} \right) \left(\tau_{[2 \rightleftharpoons 3]} \right) \left(\tau_{[(66)1+2]} \right) \mathbf{A} = \tau_{\substack{[(4)1] \\ [2 \rightleftharpoons 3] \\ [(66)1+2]}} \mathbf{A}.$

Al operar sobre las filas, las primeras transformaciones en actuar son las que quedan más a la derecha de $\tau_{[(4)1][2 \rightleftharpoons 3][[(66)1+2]]}$ o más abajo de $\tau_{\substack{[(4)1] \\ [2 \rightleftharpoons 3] \\ [(66)1+2]}}$.

Además, para indicar que una secuencia de transformaciones actúa sobre las *filas* de una matriz, escribiremos las abreviaturas por debajo de las fechas. Por tanto, los pasos en el esquema de la transformación $\tau_{[(4)1][2 \rightleftharpoons 3][[(66)1+2]]} \mathbf{I}$ son los siguientes:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{\tau_{[(66)1+2]}} \begin{bmatrix} 1 & 0 & 0 \\ 66 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{\tau_{[2 \rightleftharpoons 3]}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 66 & 1 & 0 \end{bmatrix} \xrightarrow{\tau_{[(4)1]}} \begin{bmatrix} 4 & 0 & 0 \\ 0 & 0 & 1 \\ 66 & 1 & 0 \end{bmatrix}.$$

4.A.1. Matriz escalonada por filas

Definición 4.8. Una matriz está escalonada por filas si su transpuesta está escalonada por columnas.

Librería NAcAL para Python

```
A = Matrix([ [0,2,2], [1,-2,-8], [3,6,-6] ])
U = ElimGF(A,1) # Escalonamiento operando con las filas
```

$$\begin{bmatrix} 0 & 2 & 2 \\ 1 & -2 & -8 \\ 3 & 6 & -6 \end{bmatrix} \xrightarrow{[(-3)\tau_{2+3}]} \begin{bmatrix} 0 & 2 & 2 \\ 1 & -2 & -8 \\ 0 & 12 & 18 \end{bmatrix} \xrightarrow{\begin{matrix} [(-6)\tau_{1+3}] \\ [(1)\tau_{1+2}] \end{matrix}} \begin{bmatrix} 0 & 2 & 2 \\ 1 & 0 & -6 \\ 0 & 0 & 6 \end{bmatrix} \xrightarrow{[1\tau_{2-2}]} \begin{bmatrix} 1 & 0 & -6 \\ 0 & 2 & 2 \\ 0 & 0 & 6 \end{bmatrix}$$

(recuerde que en el paso intermedio primero actúa $[(1)\tau_{1+2}]$ y luego $[(-6)\tau_{1+3}]$)

4.B. Transformación elemental “espejo” de otra transformación

Usaremos las transformaciones “espejo” en la Lección 16 cuando veamos *matrices semejantes*. Si lo prefiere, puede posponer la lectura de esta sección hasta ese momento.

Resulta que podemos obtener una misma matriz elemental operando sobre las filas o sobre las columnas de la matriz identidad. Si la matriz elemental es de *Tipo I*, por ejemplo $\mathbf{I}_{[(\alpha)\tau_{2+3}]}$, es fácil verificar que

$$\mathbf{I}_{[(\alpha)\tau_{3+2}]} \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} = \begin{bmatrix} 1 & & \\ & 1 & \alpha \\ & & 1 \end{bmatrix} = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix}_{[(\alpha)\tau_{2+3}]}.$$

En general, para una misma matriz elemental de *Tipo I* tenemos que

$$\mathbf{I}_{[(\alpha)\tau_{k+j}]} = \mathbf{I}_{[(\alpha)\tau_{j+k}]}.$$

Fíjese en el *intercambio entre los índices j y k* en función de si se transforman las filas o las columnas de la matriz identidad. Es decir, la transformación elemental necesaria para crear una matriz elemental de *Tipo I* es distinta en función de si actuamos sobre las filas, o sobre las columnas de \mathbf{I} .

A falta de mejor nombre, llamaremos a cada una de estas transformaciones el “espejo” de la otra:

Definición 4.9. Llamamos *espejo* de τ (que denotamos con “ $esp(\tau)$ ”) a aquella transformación elemental que actuando por el otro lado de la matriz identidad arroja el mismo resultado; es decir

$$_{esp(\tau)}\mathbf{I} = \mathbf{I}_{\tau} \quad \text{ó} \quad \tau\mathbf{I} = \mathbf{I}_{esp(\tau)}.$$

El *espejo* de una transformación elemental *Tipo I* $\mathbf{I}_{[(\alpha)\tau_{j+k}]}$ es a la transformación que resulta de intercambiar los índices j y k , es decir,

$$esp\left(\mathbf{I}_{[(\alpha)\tau_{j+k}]}\right) = \mathbf{I}_{[(\alpha)\tau_{k+j}]}.$$

En cuanto a las transformaciones elementales *Tipo II* y los *intercambios*... ¡son su propio “espejo”! pues tanto las *matrices elementales Tipo II* como las *matrices intercambio* son simétricas ($\mathbf{I}_{\tau} = \tau\mathbf{I} = _{esp(\tau)}\mathbf{I}$). Por ejemplo

$$\mathbf{I}_{[(\alpha)\tau_{2}]} \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} = \begin{bmatrix} 1 & & \\ & \alpha & \\ & & 1 \end{bmatrix} = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix}_{[(\alpha)\tau_{2}]} \quad \text{y} \quad \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix}_{[2\tau_{=3}]} = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix}_{[2\tau_{=3}]}.$$

Además, el *espejo* de una sucesión de transformaciones elementales es la sucesión de transformaciones espejo:

$$esp(\tau_1) \dots esp(\tau_k) = esp(\tau_1 \cdots \tau_k);$$

es decir

$$esp(\tau_1) \dots esp(\tau_k) \mathbf{I} = esp(\tau_1 \cdots \tau_k) \mathbf{I} = \mathbf{I}_{\tau_1 \cdots \tau_k}$$

Librería NAcAL para Python

```
Tr = T((3,2,1)) & T((-2,3,2)) & T((7,1,3)) & T({2,3}) & T((20,1)) # Suc. transf.
A = I(3) & Tr # Transformación de las columnas de la Identidad
B = Tr.espejo() & I(3) # Transformación espejo sobre las filas de la Identidad
A == B # Verificación de que A y B son iguales
```

True

Matrices simétricas. Por la Ecuación 4.5 en la página 55, la matriz $\mathbf{I}_{\tau_1 \cdots \tau_k}$ es simétrica si y solo si

$$esp(\tau_1 \cdots \tau_k) \mathbf{I} = \mathbf{I}_{\tau_1 \cdots \tau_k} = (\tau_1 \cdots \tau_k)^T \mathbf{I}.$$

Librería NAcAL para Python

```
p = T([(9,2), (18,1), (-1,2,1), (frac(1,9),1), (-1,1,2), (frac(1,2),2)], 'h')
p
```

$$\begin{bmatrix} \tau \\ (9)2 \end{bmatrix} \begin{bmatrix} \tau \\ (18)1 \end{bmatrix} \begin{bmatrix} \tau \\ (-1)2+1 \end{bmatrix} \begin{bmatrix} \tau \\ (\frac{1}{9})1 \end{bmatrix} \begin{bmatrix} \tau \\ (-1)1+2 \end{bmatrix} \begin{bmatrix} \tau \\ (\frac{1}{2})2 \end{bmatrix}$$

Librería NAcAL para Python

```
(p.espejo() & I(2)) == (~p & I(2)) # matriz simétrica si se verifica la igualdad
```

True

Librería NAcAL para Python

```
p.espejo() & I(2) # veamos que efectivamente la matriz es simétrica
```

$$\begin{bmatrix} 2 & -1 \\ -1 & 5 \end{bmatrix}$$

Aún cuando $esp(\tau_1 \cdots \tau_k) \mathbf{I} = \mathbf{I}_{\tau_1 \cdots \tau_k} = (\tau_1 \cdots \tau_k)^T \mathbf{I}$ (por ser $\mathbf{I}_{\tau_1 \cdots \tau_k}$ simétrica), las tres sucesiones de transformaciones son en general distintas. En el ejemplo de más arriba: la sucesión $\tau_1 \cdots \tau_k$, la sucesión $esp(\tau_1 \cdots \tau_k)$ y la sucesión $(\tau_1 \cdots \tau_k)^T$ son respectivamente

$$\begin{bmatrix} \tau \\ (9)2 \end{bmatrix}, \begin{bmatrix} \tau \\ (18)1 \end{bmatrix}, \begin{bmatrix} \tau \\ (-1)2+1 \end{bmatrix}, \begin{bmatrix} \tau \\ (\frac{1}{9})1 \end{bmatrix}, \begin{bmatrix} \tau \\ (-1)1+2 \end{bmatrix}, \begin{bmatrix} \tau \\ (\frac{1}{2})2 \end{bmatrix} \quad \text{y} \quad \begin{bmatrix} \tau \\ (\frac{1}{2})2 \end{bmatrix}, \begin{bmatrix} \tau \\ (-1)2+1 \end{bmatrix}, \begin{bmatrix} \tau \\ (\frac{1}{9})1 \end{bmatrix}, \begin{bmatrix} \tau \\ (-1)1+2 \end{bmatrix}, \begin{bmatrix} \tau \\ (18)1 \end{bmatrix}, \begin{bmatrix} \tau \\ (9)2 \end{bmatrix}.$$

(si se fija en los índices de las transformaciones elementales de Tipo I de las dos primeras sucesiones, constatará que las tres sucesiones son distintas).

Espejo de la inversa de una transformación elemental. Cuando una transformación elemental deshace la operación realizada por otra se dice de ambas que “una es la inversa de la otra”⁶.

Por ejemplo, si sumamos el triple del vector \mathbf{x} al vector \mathbf{y} , podemos deshacer (o invertir) la operación restando el triple de \mathbf{x} a \mathbf{y} . Así, la transformación inversa de $\begin{bmatrix} \tau \\ (\alpha)j+i \end{bmatrix}$ es $\begin{bmatrix} \tau \\ (-\alpha)j+i \end{bmatrix}$.

De manera similar, si multiplicamos por 2 un vector, podemos deshacer (o invertir) la operación dividiendo dicho vector por 2. Así tenemos que transformación inversa de $\begin{bmatrix} \tau \\ (\alpha)j \end{bmatrix}$ es $\begin{bmatrix} \tau \\ (\frac{1}{\alpha})j \end{bmatrix}$.

Consecuentemente, empleando las transformaciones espejo concluimos que

$$\begin{bmatrix} \tau \\ (-\alpha)j+i \end{bmatrix} \mathbf{I} \begin{bmatrix} \tau \\ (\alpha)j+i \end{bmatrix} = \mathbf{I} \quad \text{y que} \quad \begin{bmatrix} \tau \\ (\frac{1}{\alpha})j \end{bmatrix} \mathbf{I} \begin{bmatrix} \tau \\ (\alpha)j \end{bmatrix} = \mathbf{I};$$

así, si denotamos con τ^{-1} la inversa de τ , tenemos que

$$esp(\tau)^{-1} \mathbf{I} \tau = esp(\tau^{-1}) \mathbf{I} \tau = \mathbf{I};$$

y en general, para una secuencia de transformaciones elementales⁷

$$esp(\tau_1 \cdots \tau_k)^{-1} \mathbf{I} \tau_1 \cdots \tau_k = \mathbf{I}.$$

Librería NAcAL para Python

```
Tr = T((3,2,1)) & T((-2,3,2)) & T((7,1,3)) & T({2,3}) & T((20,1))
(Tr**-1).espejo() & I(3) & Tr
```

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$



Veremos que la Lección 16 se basa en aplicar, a matrices *cuadradas*, transformaciones de la forma

$$\mathbf{S}^{-1} \mathbf{A} \mathbf{S} = (\mathbf{I}_{\tau_1 \cdots \tau_k})^{-1} \mathbf{A} (\mathbf{I}_{\tau_1 \cdots \tau_k}) = esp(\tau_1 \cdots \tau_k)^{-1} \mathbf{I} \mathbf{A} \mathbf{I}_{\tau_1 \cdots \tau_k} = esp(\tau_1 \cdots \tau_k)^{-1} \mathbf{A}_{\tau_1 \cdots \tau_k},$$

donde $\mathbf{S} = \mathbf{I}_{\tau_1 \cdots \tau_k}$.

La Lección 18 se basa en aplicar, a matrices *simétricas*, transformaciones de la forma

$$\mathbf{B}^T \mathbf{A} \mathbf{B} = (\mathbf{I}_{\tau_1 \cdots \tau_k})^T \mathbf{A} (\mathbf{I}_{\tau_1 \cdots \tau_k}) = (\tau_1 \cdots \tau_k)^T \mathbf{I} \mathbf{A} \mathbf{I}_{\tau_1 \cdots \tau_k} = (\tau_1 \cdots \tau_k)^T \mathbf{A}_{\tau_1 \cdots \tau_k},$$

donde $\mathbf{B} = \mathbf{I}_{\tau_1 \cdots \tau_k}$.

La Lección 17 es un caso especial en el que se aplican, a matrices *simétricas*, transformaciones de la forma

$$\mathbf{Q}^{-1} \mathbf{A} \mathbf{Q}$$

donde $\mathbf{Q} = \mathbf{I}_{\tau_1 \cdots \tau_k}$ es tal que $\mathbf{Q}^{-1} = \mathbf{Q}^T$, es decir, tales que $esp(\tau_1 \cdots \tau_k)^{-1} \mathbf{I} = (\tau_1 \cdots \tau_k)^T \mathbf{I}$; por lo que es un caso particular de lo tratado tanto de la Lección 16 como en la Lección 18.

⁶La Lección 5 trata sobre este tema.

⁷véase la Sección 5.A en la página 73 la justificación para la notación: $esp(\tau_1 \cdots \tau_k)^{-1}$.

Inversa de una matriz. Rango de una matriz

Matrices invertibles por un lado. Decimos que \mathbf{A} es *invertible por la derecha* si existe otra matriz \mathbf{B} tal que $\mathbf{AB} = \mathbf{I}$. Y decimos que \mathbf{A} *invertible por la izquierda* si existe otra matriz \mathbf{B} tal que $\mathbf{BA} = \mathbf{I}$. Esta lección se centra en las matrices *cuadradas* e invertibles por ambos lados a la vez...

5.1. Matrices invertibles

Definición 5.1 (Matriz invertible). Se dice que \mathbf{A} es *invertible* (o tiene inversa) si existe otra matriz \mathbf{B} del mismo orden tal que

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I}.$$

(nótese que $\mathbf{AB} = \mathbf{BA}$ implica que \mathbf{A} y \mathbf{B} son necesariamente *cuadradas*).

Pues bien, si \mathbf{A} es invertible su *inversa es única*:

Proposición 5.1.1 (La inversa es única). Si \mathbf{A} , \mathbf{B} y \mathbf{B}' verifican que

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I} \quad \text{y que} \quad \mathbf{AB}' = \mathbf{B}'\mathbf{A} = \mathbf{I}$$

entonces \mathbf{B} y \mathbf{B}' son iguales.

Demostración. En efecto:
$$\mathbf{B} = \mathbf{BI} = \mathbf{B}(\underbrace{\mathbf{AB}'}_{\mathbf{I}=\mathbf{AB}'}) = (\underbrace{\mathbf{BA}}_{\mathbf{BA}=\mathbf{I}})\mathbf{B}' = \mathbf{IB}' = \mathbf{B}'.$$

□

La unicidad nos permite dar una definición de “la” *matriz inversa* de \mathbf{A} :

Definición 5.2 (Matriz inversa). Si \mathbf{A} es invertible, denotaremos con \mathbf{A}^{-1} a la única matriz tal que

$$\mathbf{A}(\mathbf{A}^{-1}) = (\mathbf{A}^{-1})\mathbf{A} = \mathbf{I}.$$

Además diremos que \mathbf{A}^{-1} es *la matriz inversa* de \mathbf{A} .

Definición 5.3 (Matriz singular). Si una matriz cuadrada *no tiene inversa* se dice que es *singular*.

A continuación veremos varias propiedades de las matrices inversas. En particular, que el *producto de matrices invertibles* es invertible; que la transpuesta de una *matriz invertible* también es invertible; y que si una matriz tiene alguna fila o columna nula, entonces no es invertible.

EJERCICIO 32. Demuestre las siguientes proposiciones:

(a) **Proposición 5.1.2.** Si \mathbf{A} y \mathbf{B} (de orden n) tienen inversa entonces:
$$(\mathbf{AB})^{-1} = (\mathbf{B}^{-1})(\mathbf{A}^{-1}).$$

(b) **Proposición 5.1.3.** Si $\mathbf{A}_1, \dots, \mathbf{A}_n$ tienen inversa, entonces
$$(\mathbf{A}_1 \cdots \mathbf{A}_k)^{-1} = \mathbf{A}_k^{-1} \cdots \mathbf{A}_1^{-1}.$$

Pista. $(\mathbf{A}_1 \cdots \mathbf{A}_k)^{-1} = (\mathbf{A}_1(\mathbf{A}_2 \cdots \mathbf{A}_k))^{-1} = (\mathbf{A}_2 \cdots \mathbf{A}_k)^{-1}(\mathbf{A}_1^{-1}).$

- (c) **Proposición 5.1.4.** Sea \mathbf{A} de orden n y sea \mathbf{B} invertible de orden n . Entonces \mathbf{AB} es invertible si y solo si \mathbf{A} es invertible.
- (d) **Proposición 5.1.5.** Sea \mathbf{A} invertible de orden n y sea \mathbf{B} de orden n . Entonces \mathbf{AB} es invertible si y solo si \mathbf{B} es invertible.
- (e) **Proposición 5.1.6.** Si \mathbf{A} de orden n tiene inversa entonces la inversa de \mathbf{A}^\top es: $(\mathbf{A}^\top)^{-1} = (\mathbf{A}^{-1})^\top$.
- Pista. Recuerde que $\mathbf{B}^\top \mathbf{A}^\top = (\mathbf{AB})^\top$ y aplíquelo a $\mathbf{A}^\top (\mathbf{A}^{-1})^\top$ y a $(\mathbf{A}^{-1})^\top \mathbf{A}^\top$.
- (f) **Proposición 5.1.7.** Si alguna fila o columna de \mathbf{A} es nula entonces la matriz es singular.



\mathbf{A} es invertible si existe \mathbf{B} tal que $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$; su inversa \mathbf{B} es única y se denota con \mathbf{A}^{-1} .

- el producto de matrices invertibles es invertible: $(\mathbf{AB})^{-1} = \mathbf{B}^{-1} \mathbf{A}^{-1}$.
- la transpuesta de una matriz invertible es invertible: $(\mathbf{A}^\top)^{-1} = (\mathbf{A}^{-1})^\top$.
- Si alguna fila o columna de \mathbf{A} es nula la matriz es *singular* (no tiene inversa).

5.1.1. Inversa de las matrices (y las transformaciones) elementales.



¡No todas las matrices cuadradas son invertibles!

¿Qué hace que una matriz cuadrada sea invertible?... Todo tiene que ver con las matrices elementales. Veamos primero que las transformaciones elementales (y las matrices elementales) son invertibles.

Las transformaciones elementales son reversibles. Por ejemplo, sumar el doble de la primera columna a la tercera, y luego restar el doble de la primera columna de la tercera, nos deja como al principio:

$$\mathbf{A}_{[(2)\tau+3][(-2)\tau+3]} = \mathbf{A} \begin{pmatrix} \mathbf{I} \\ \tau_{[2)\tau+3]} \end{pmatrix} \begin{pmatrix} \mathbf{I} \\ (-2)\tau_{[2)\tau+3]} \end{pmatrix} = \mathbf{A} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \mathbf{A} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \mathbf{A}.$$

Algo similar ocurre cuando multiplicamos y dividimos una columna por $\alpha \neq 0$:

$$\mathbf{A}_{[(5)\tau][(\frac{1}{5})\tau]} = \mathbf{A} \begin{pmatrix} \mathbf{I} \\ \tau_{[(5)\tau]} \end{pmatrix} \begin{pmatrix} \mathbf{I} \\ (\frac{1}{5})\tau_{[(5)\tau]} \end{pmatrix} = \mathbf{A} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{5} & 0 \\ 0 & 0 & 1 \end{bmatrix} = \mathbf{A} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \mathbf{A}$$

En general,

$$\mathbf{A}_{[(\lambda)\tau+i+j][(-\lambda)\tau+i+j]} = \mathbf{A}_{[(-\lambda)\tau+i+j][(\lambda)\tau+i+j]} = \mathbf{A} \quad \text{y} \quad \mathbf{A}_{[(\alpha)\tau+i][(\frac{1}{\alpha})\tau+i]} = \mathbf{A}_{[(\frac{1}{\alpha})\tau+i][(\alpha)\tau+i]} = \mathbf{A},$$

es decir, para toda transformación elemental τ existe otra, que denotamos τ^{-1} , tal que

$$\mathbf{A}_{(\tau^{-1})\tau} = \mathbf{A}_{\tau(\tau^{-1})} = \mathbf{A}. \quad (5.1)$$

Recordando que $\mathbf{A}_{\tau_1\tau_2} = (\mathbf{A}_{\tau_1})_{\tau_2} = (\mathbf{A}_{\tau_1})\mathbf{I}_{\tau_2}$ podemos expresar (5.1) como

$$(\mathbf{A}_{\tau^{-1}})\mathbf{I}_\tau = (\mathbf{A}_\tau)\mathbf{I}_{\tau^{-1}} = \mathbf{A};$$

si en particular $\mathbf{A} = \mathbf{I}$ tenemos que

$$(\mathbf{I}_{\tau^{-1}})\mathbf{I}_\tau = (\mathbf{I}_\tau)\mathbf{I}_{\tau^{-1}} = \mathbf{I} \quad \implies \quad (\mathbf{I}_{\tau^{-1}}) = (\mathbf{I}_\tau)^{-1}.$$

Así pues, llegamos al siguiente

Teorema 5.1.8. Para toda matriz elemental \mathbf{I}_τ , existe otra matriz elemental $\mathbf{I}_{\tau^{-1}}$, tal que

$$\left(\mathbf{I}_\tau\right)\left(\mathbf{I}_{\tau^{-1}}\right)=\left(\mathbf{I}_{\tau^{-1}}\right)\left(\mathbf{I}_\tau\right)=\mathbf{I}.$$

En particular, si $i \neq j$ y $\alpha \neq 0$:

$$\left(\mathbf{I}_{\left[\begin{smallmatrix} \tau \\ [(-\lambda)i+j] \end{smallmatrix}\right]}\right)\left(\mathbf{I}_{\left[\begin{smallmatrix} \tau \\ [(\lambda)i+j] \end{smallmatrix}\right]}\right)=\left(\mathbf{I}_{\left[\begin{smallmatrix} \tau \\ [(\lambda)i+j] \end{smallmatrix}\right]}\right)\left(\mathbf{I}_{\left[\begin{smallmatrix} \tau \\ [(-\lambda)i+j] \end{smallmatrix}\right]}\right)=\mathbf{I} \quad \text{y} \quad \left(\mathbf{I}_{\left[\begin{smallmatrix} \tau \\ \left[\left(\frac{1}{\alpha}\right)i \end{smallmatrix}\right]}\right]}\right)\left(\mathbf{I}_{\left[\begin{smallmatrix} \tau \\ [(\alpha)i] \end{smallmatrix}\right]}\right)=\left(\mathbf{I}_{\left[\begin{smallmatrix} \tau \\ [(\alpha)i] \end{smallmatrix}\right]}\right)\left(\mathbf{I}_{\left[\begin{smallmatrix} \tau \\ \left[\left(\frac{1}{\alpha}\right)i \end{smallmatrix}\right]}\right]}\right)=\mathbf{I}.$$

Como toda transformación elemental es invertible, también lo es cualquier sucesión de transformaciones elementales (por ejemplo un intercambio). Es decir, se verifica la siguiente

Proposición 5.1.9. Si \mathbf{A} es producto de matrices elementales, entonces existe \mathbf{B} (también producto de matrices elementales) tal que $\mathbf{AB} = \mathbf{I} = \mathbf{BA}$ (es decir, si $\mathbf{A} = \mathbf{I}_{\tau_1 \dots \tau_k}$ entonces es invertible).

Pero hay que tener en cuenta que la última transformación realizada ha de ser la primera en ser invertida (como cuando uno se pone los calcetines y luego los zapatos... para descalzarse se empieza por los zapatos y se finaliza por los calcetines); así, si se realizan una serie de transformaciones elementales en determinado orden, se deshacen en el orden inverso —véase la Proposición 5.1.3.

☞ Cualquier matriz $\mathbf{E} = \mathbf{I}_{\tau_1 \dots \tau_k}$ es invertible por ser producto de matrices elementales, en particular,

$$\left(\mathbf{I}_{\tau_1 \dots \tau_k}\right)\left(\mathbf{I}_{\tau_k^{-1} \dots \tau_1^{-1}}\right)=\left(\mathbf{I}_{\tau_1}\right) \cdots\left(\mathbf{I}_{\tau_k}\right)\left(\mathbf{I}_{\tau_k^{-1}}\right) \cdots\left(\mathbf{I}_{\tau_1^{-1}}\right)=\mathbf{I}_{\tau_1 \dots \tau_k \tau_k^{-1} \dots \tau_1^{-1}}=\mathbf{I};$$

por lo que podemos denotar $\tau_k^{-1} \dots \tau_1^{-1}$ como $(\tau_1 \dots \tau_k)^{-1}$. Así pues,

$$\mathbf{E}^{-1}=\mathbf{I}_{\tau_k^{-1} \dots \tau_1^{-1}}=\mathbf{I}_{\left(\tau_1 \dots \tau_k\right)^{-1}}.$$

Librería NAcAL para Python

```
Tr = T([ (3,2), (5,2,3), (10,1,2) ], 'h'); Tr # Secuencia de transformaciones
# el parámetro 'h' es para mostrar
# la secuencia en 'h' horizontal
```

$$\begin{smallmatrix} \tau \\ [(3)2] \end{smallmatrix} \begin{smallmatrix} \tau \\ [(5)2+3] \end{smallmatrix} \begin{smallmatrix} \tau \\ [(10)1+2] \end{smallmatrix}$$

```
Tr**(-1) # Inversa de la secuencia de transf.
```

$$\begin{smallmatrix} \tau \\ [(-10)1+2] \end{smallmatrix} \begin{smallmatrix} \tau \\ [(-5)2+3] \end{smallmatrix} \begin{smallmatrix} \tau \\ \left[\left(\frac{1}{3}\right)2 \right] \end{smallmatrix}$$

```
A = I(3) & Tr # A: producto de elementales
B = I(3) & Tr**(-1) # B es la inversa de A
A*B # Comprobación de que son inversas
```

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

5.1.2. Matrices pre-escaladas con inversa.



Como primer paso para encontrar un algoritmo que calcule la inversa, demostraremos que una matriz \mathbf{K} , de orden n , pre-escalada y *sin columnas nulas*, se puede transformar en la matriz identidad \mathbf{I} mediante una secuencia de transformaciones elementales

$$\mathbf{K}_{\tau_1 \dots \tau_k} = \mathbf{K}(\mathbf{I}_{\tau_1 \dots \tau_k}) = \mathbf{I} \quad (\text{Teorema 5.1.10}).$$

Consecuentemente, \mathbf{K}^{-1} es el producto de las correspondientes matrices elementales; $\mathbf{K}^{-1} = \mathbf{I}_{\tau_1 \dots \tau_k}$.

Teorema 5.1.10. Si una matriz \mathbf{K} , de orden n , es pre-escalada y sin columnas nulas, existen k transformaciones elementales tales que $\mathbf{K}_{\tau_1 \dots \tau_k} = \mathbf{I}$.

Demostración. Toda matriz pre-escalada sin columnas nulas tiene pivotes en todas sus columnas; si además es *cuadrada*, entonces tiene pivotes en todas sus *filas* (pues cada pivote ocupa una fila y columna distintas).

Por tanto, basta aplicar las k transformaciones τ_1, \dots, τ_k de la eliminación Gauss-Jordan (Corolario 4.2.3 en la página 53) que transforman \mathbf{K} en su forma escalonada reducida $\mathbf{K}_{\tau_1 \dots \tau_k}$; que en este caso resulta ser la matriz identidad de orden n (pues \mathbf{K} , de orden n , tiene n pivotes). \square

Librería NAcAL para Python

```
K = Matrix( [[0,5,0], [2,0,0], [6,9,3]] )
R = ElimGJ(K,1)
```

Forma Escalonada Reducida

$$\begin{bmatrix} 0 & 5 & 0 \\ 2 & 0 & 0 \\ 6 & 9 & 3 \end{bmatrix} \xrightarrow{[1 \leftrightarrow 2]} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 6 & 9 & 3 \end{bmatrix} \xrightarrow{\begin{bmatrix} (-3)3+1 \\ (-2)3+2 \end{bmatrix}} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 3 \end{bmatrix} \xrightarrow{\begin{bmatrix} (\frac{1}{5})1 \\ (\frac{1}{5})2 \\ (\frac{1}{3})3 \end{bmatrix}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Corolario 5.1.11. Si \mathbf{K} es *pre-escalada* y *cuadrada*, las siguientes propiedades son equivalentes

1. \mathbf{K} no tiene columnas nulas
2. \mathbf{K} es un producto de matrices elementales
3. \mathbf{K} tiene inversa

Demostración.

1. \Rightarrow 2. Si \mathbf{K} no tiene columnas nulas, existen transf. elementales $\tau_1 \dots \tau_k$ tales que $\mathbf{K}_{\tau_1 \dots \tau_k} = \mathbf{I}$; así que

$$\begin{aligned} \mathbf{I} &= \mathbf{K}_{\tau_1 \dots \tau_k} = \mathbf{K}(\mathbf{I}_{\tau_1 \dots \tau_k}), \\ \mathbf{I}(\mathbf{I}_{\tau_1 \dots \tau_k})^{-1} &= \mathbf{K}(\mathbf{I}_{\tau_1 \dots \tau_k})(\mathbf{I}_{\tau_1 \dots \tau_k})^{-1}; \quad \text{multiplicando ambos lados por } (\mathbf{I}_{\tau_1 \dots \tau_k})^{-1}, \\ (\mathbf{I}_{\tau_1 \dots \tau_k})^{-1} &= \mathbf{K}, \end{aligned}$$

donde $(\mathbf{I}_{\tau_1 \dots \tau_k})^{-1} = \mathbf{I}_{(\tau_1 \dots \tau_k)^{-1}} = \mathbf{I}_{\tau_k^{-1} \dots \tau_1^{-1}}$ es un producto de matrices elementales.

2. \Rightarrow 3. ¡Ya se sabe! Proposición 5.1.9 en la página anterior.

3. \Rightarrow 1. ¡Ya se sabe!... si tuviera columnas nulas sería singular por la Proposición 5.1.7 en la página 62. \square


```
(R.TrC)**-1 # inversa de la secuencia de Transf. de las
             # Columnas en el proceso de eliminación K --> R
```

$$\begin{matrix} \tau & \tau & \tau & \tau & \tau & \tau \\ [(3)3] & [(2)2] & [(5)1] & [(2)3+2] & [(3)3+1] & [1=2] \end{matrix}$$

(compare con la salida del bloque de código anterior y fíjese cómo la primera transformación inversa deshace la última del bloque anterior, la segunda deshace la penúltima, etc.)

```
I(K.n) & ( R.TrC )**-1 # al aplicar la inversa de las
                       #Transf. sobre I(3) recuperamos K
```

$$\begin{bmatrix} 0 & 5 & 0 \\ 2 & 0 & 0 \\ 6 & 9 & 3 \end{bmatrix}$$

(K.n es el número de columnas de \mathbf{K} , así que $I(K.n)$ es la matriz identidad de orden 3)

5.1.3. Matrices invertibles



El siguiente corolario caracteriza las matrices con inversa. Además, nos dice cómo verificar si \mathbf{A} es invertible:

- \mathbf{A} es invertible si y solo si tiene alguna forma pre-escalada sin columnas nulas.

Corolario 5.1.12 (Caracterización de las matrices invertibles). *Dada \mathbf{A} de orden n , las siguientes propiedades son equivalentes*

1. Al (pre)escalar \mathbf{A} se obtiene una matriz sin columnas nulas.
2. \mathbf{A} es producto de matrices elementales.
3. \mathbf{A} tiene inversa.

Demostración.

1. \Rightarrow 2. Por el Teorema 4.2.1 sabemos que existen $\tau_1 \cdots \tau_k$ tales que $\mathbf{A}_{\tau_1 \cdots \tau_k} = \mathbf{K}$ es pre-escalada; y si \mathbf{K} no tiene columnas nulas, entonces es producto de p matrices elementales, $\mathbf{K} = \mathbf{I}_{\tau'_1 \cdots \tau'_p}$:

$$\begin{aligned} \mathbf{A}(\mathbf{I}_{\tau_1 \cdots \tau_k}) &= (\mathbf{I}_{\tau'_1 \cdots \tau'_p}) \\ \mathbf{A}(\mathbf{I}_{\tau_1 \cdots \tau_k})(\mathbf{I}_{\tau_1 \cdots \tau_k})^{-1} &= (\mathbf{I}_{\tau'_1 \cdots \tau'_p})(\mathbf{I}_{\tau_1 \cdots \tau_k})^{-1} \\ \mathbf{A} &= (\mathbf{I}_{\tau'_1 \cdots \tau'_p})(\mathbf{I}_{(\tau_1 \cdots \tau_k)^{-1}}) = \mathbf{I}_{\tau'_1 \cdots \tau'_p \tau_k^{-1} \cdots \tau_1^{-1}}, \end{aligned}$$

2. \Rightarrow 3. ¡Ya se sabe! Proposición 5.1.9 en la página 63

3. \Rightarrow 1. Sean $\tau_1 \cdots \tau_k$ tales que $\mathbf{A}(\mathbf{I}_{\tau_1 \cdots \tau_k}) = \mathbf{K}$ es pre-escalada. Entonces, puesto que $\mathbf{I}_{\tau_1 \cdots \tau_k}$ es invertible, \mathbf{A} es invertible si y solo si \mathbf{K} es invertible (Proposición 5.1.4 en la página 62).

□

Por tanto, para saber si una matriz cuadrada es singular o si es invertible, basta con aplicar la eliminación y verificar si la forma pre-escalada encontrada tiene columnas nulas o no.

```
A = Matrix( [ [2,-1,0], [-1,2,-1], [0,-1,2] ] )
K = Elim(A,1) # una forma pre-escalada de A
```

$$\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \xrightarrow{\begin{smallmatrix} \tau_{(2)2} \\ [(1)1+2] \end{smallmatrix}} \begin{bmatrix} 2 & 0 & 0 \\ -1 & 3 & -1 \\ 0 & -2 & 2 \end{bmatrix} \xrightarrow{\begin{smallmatrix} \tau_{(3)3} \\ [(1)2+3] \end{smallmatrix}} \begin{bmatrix} 2 & 0 & 0 \\ -1 & 3 & 0 \\ 0 & -2 & 4 \end{bmatrix}$$

(Como \mathbf{A} es cuadrada y la forma pre-escalada \mathbf{K} no tiene columnas nulas, \mathbf{A} es invertible.)

Podemos ver las transformaciones que hemos aplicado para llegar a la forma pre-escalada \mathbf{K} con:

```
Tr = K.TrC # Transformaciones aplicadas a las columnas
T(Tr, 'h') # Las representaremos en 'h'horizontal
```

$$\begin{smallmatrix} \tau_{(2)2} \\ [(1)1+2] \end{smallmatrix} \begin{smallmatrix} \tau_{(3)3} \\ [(1)2+3] \end{smallmatrix}$$

Aplicado la inversa de dichas transformaciones sobre \mathbf{K} recuperamos la matriz original \mathbf{A} :

```
Matrix(K) & ( Tr**-1 ) # es A
```

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

Es decir, $\mathbf{A} = (\mathbf{I}_{\tau_1' \dots \tau_p'})_{\tau_k^{-1} \dots \tau_1^{-1}}$ donde $\mathbf{I}_{\tau_1' \dots \tau_p'} = \mathbf{K}$ es pre-escalada, como en la demostración anterior:

$$\mathbf{K}_{[(1)2+3][(\frac{1}{3})3][(-1)1+2][(\frac{1}{2})2]} = \begin{bmatrix} 2 & 0 & 0 \\ -1 & 3 & 0 \\ 0 & -2 & 4 \end{bmatrix}_{[(-1)2+3][(\frac{1}{3})3][(-1)1+2][(\frac{1}{2})2]} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} = \mathbf{A}.$$



La siguiente proposición nos ofrece una nueva característica de las matrices NO invertibles:

- Si \mathbf{A} tiene columnas que son combinación lineal del resto, entonces es singular.

EJERCICIO 33. Demuestre la siguiente proposición.

Proposición 5.1.13. Sea \mathbf{A} de orden n ; si alguna de sus columnas es combinación lineal del resto (o si alguna de sus filas es combinación lineal del resto), entonces \mathbf{A} es singular.



Y ahora veamos que si \mathbf{A} y \mathbf{B} son de orden n y $\mathbf{AB} = \mathbf{I}$, entonces \mathbf{A} y \mathbf{B} son una la inversa de la otra.

Este resultado es de gran importancia, pues nos dice que [para comprobar que una matriz cuadrada es invertible, basta ver si existe inversa por uno cualquiera de sus lados](#).

Como nos apoyaremos en la eliminación para su demostración, no hemos podido ver antes este resultado.

Proposición 5.1.14. Para \mathbf{A} y \mathbf{B} de orden n se verifican las siguientes propiedades:

- Si \mathbf{B} es singular, entonces \mathbf{AB} es singular.
- Si \mathbf{A} es singular, entonces \mathbf{AB} es singular.
- Si $\mathbf{AB} = \mathbf{I}$, entonces tanto \mathbf{A} como \mathbf{B} son invertibles y $\mathbf{B} = \mathbf{A}^{-1}$.

Demostración. Vayamos punto por punto:

- Si \mathbf{B} es singular, la última columna de cualquiera de sus formas escalonadas es nula. Consecuentemente si \mathbf{E} es invertible y tal que \mathbf{BE} es escalonada, entonces la última columna de \mathbf{ABE} es nula. Y por tanto \mathbf{ABE} es singular, pero como \mathbf{E} es invertible, necesariamente \mathbf{AB} es singular.
- Como \mathbf{A} es singular su transpuesta también. Por tanto $\mathbf{B}^T \mathbf{A}^T$ es singular (punto anterior). Pero como $\mathbf{B}^T \mathbf{A}^T = (\mathbf{AB})^T$ es singular, entonces su transpuesta \mathbf{AB} también.
- Como \mathbf{I} es invertible, por el punto anterior, \mathbf{A} tiene que ser invertible, es decir, existe \mathbf{A}^{-1} . Multiplicando ambos lados por \mathbf{A}^{-1} obtenemos el resultado deseado: $\mathbf{A}^{-1}(\mathbf{AB}) = \mathbf{A}^{-1}\mathbf{I} \Rightarrow \mathbf{B} = \mathbf{A}^{-1}$.

□



¡Ya tenemos las piezas para diseñar un algoritmo que encuentre la inversa de una matriz! (lo hace buscando la inversa por el lado derecho mediante eliminación por columnas).

La [eliminación](#) permite encontrar una forma pre-escalonada de toda matriz (Teorema 4.2.1 en la página 51)

$$\mathbf{A}_{\tau_1 \dots \tau_p} = \mathbf{K},$$

y si \mathbf{K} no tiene columnas nulas, aplicando Gauss-Jordan, se puede continuar con las transformaciones elementales hasta reducir \mathbf{K} hasta \mathbf{I} (Teorema 5.1.10 en la página 64)

$$\mathbf{K}_{\tau_{(p+1)} \dots \tau_k} = \mathbf{I}.$$

Por tanto, la sucesión de k transformaciones elementales $\tau_1 \dots \tau_p, \tau_{(p+1)} \dots \tau_k$ transforma \mathbf{A} en \mathbf{I}

$$\mathbf{A}_{\tau_1 \dots \tau_k} = \mathbf{A}(\mathbf{I}_{\tau_1 \dots \tau_k}) = \mathbf{I};$$

es decir, si \mathbf{K} no tiene columnas nulas, entonces $\mathbf{A}^{-1} = \mathbf{I}_{\tau_1 \dots \tau_k}$; y si las tiene entonces \mathbf{A}^{-1} no existe.

Librería NAcAL para Python

```
a, b, c, d = sympy.symbols('a b c d') # definimos las variables simbólicas
A = Matrix([[a,b], [c,d]])           # definimos la matriz
R = ElimGJ(A,1)                       # Eliminación hasta obtener la Identidad
Tr = R.TrC                            # Transformaciones de las Columnas
```

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \xrightarrow{\left[\left(-\frac{b}{a}\right)1+2 \right]} \begin{bmatrix} a & 0 \\ c & d - \frac{bc}{a} \end{bmatrix} \xrightarrow{\left[\left(-\frac{ac}{ad-bc}\right)2+1 \right]} \begin{bmatrix} a & 0 \\ 0 & d - \frac{bc}{a} \end{bmatrix} \xrightarrow{\left[\left(\frac{1}{a}\right)1 \right]} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

(asumiendo que $a \neq 0$ y que $ad - bc \neq 0$)

Ahora ya sabemos que aplicando las transformaciones sobre las columnas de **I** obtenemos la inversa

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \xrightarrow{\left[\left(-\frac{b}{a}\right)^{1+2} \right] \left[\left(-\frac{ac}{ad-bc}\right)^{2+1} \right] \left[\left(\frac{1}{a}\right)^1 \right] \left[\left(\frac{a}{ad-bc}\right)^2 \right]} = \begin{bmatrix} \frac{d}{ad-bc} & -\frac{b}{ad-bc} \\ -\frac{c}{ad-bc} & \frac{a}{ad-bc} \end{bmatrix}$$

Librería NAcAL para Python

A * (I(2) & Tr)

comprobación de que I(2) & Tr es la inversa de A

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



...Pero aplicar k transformaciones elementales $\tau_1 \cdots \tau_k$ a las columnas de **A** hasta tener **I**, y después aplicar dichas transformaciones a las columnas de la matriz identidad, **I** $_{\tau_1 \cdots \tau_k}$, puede ser bastante pesado.

¿Hay alguna una forma de realizar ambas operaciones a la vez: la transformación **A** $_{\tau_1 \cdots \tau_k} = \mathbf{I}$ y la transformación **I** $_{\tau_1 \cdots \tau_k} = \mathbf{A}^{-1}$?

¡Si! y es sencillísimo... basta “alargar” las columnas de **A** con **I** y entonces aplicar $\tau_1 \cdots \tau_k$ a $\begin{bmatrix} \mathbf{A} \\ \mathbf{I} \end{bmatrix}$.

5.1.4. Secuencias de transformaciones elementales. Parte II

Apilando **A** y **B** (ambas con el mismo número de columnas), es decir, alargando las columnas de **A** poniendo las de **B** por debajo $\begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix}$; podemos aplicar una transformación elemental τ a las columnas de esta matriz “alargada”, y el resultado es equivalente a transformar tanto **A** como **B** y apilar luego las matrices resultantes:

$$\begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix} \xrightarrow{\tau} \begin{bmatrix} \mathbf{A}_\tau \\ \mathbf{B}_\tau \end{bmatrix} = \begin{bmatrix} \mathbf{A}_\tau \\ \mathbf{B}_\tau \end{bmatrix},$$

de manera que la columna j -ésima $\begin{bmatrix} \mathbf{A}_\tau \\ \mathbf{B}_\tau \end{bmatrix}_{|j}$ es la concatenación de las columnas $(\mathbf{A}_\tau)_{|j}$ y $(\mathbf{B}_\tau)_{|j}$.

Lo mismo ocurre si aplicamos una secuencia de k transformaciones elementales:

$$\begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix}_{\tau_1 \cdots \tau_k} = \begin{bmatrix} \mathbf{A}_{\tau_1 \cdots \tau_k} \\ \mathbf{B}_{\tau_1 \cdots \tau_k} \end{bmatrix}.$$

Por tanto, si antes de realizar las transformaciones “pegamos” por debajo de **A** = $\begin{bmatrix} 0 & 1 & 5 \\ 7 & 0 & 4 \end{bmatrix}$ la matriz **I** de orden 3 y aplicamos algunas transformaciones:

$$\begin{bmatrix} \mathbf{A} \\ \mathbf{I} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 5 \\ 7 & 0 & 4 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{[(-5)\tau_{2+3}]} \begin{bmatrix} 0 & 1 & 0 \\ 7 & 0 & 4 \\ 1 & 0 & 0 \\ 0 & 1 & -5 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{[(\tau_7)\tau_3]} \begin{bmatrix} 0 & 1 & 0 \\ 7 & 0 & 0 \\ 1 & 0 & -4 \\ 0 & 1 & -35 \\ 0 & 0 & 7 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{\tau_1 \cdots \tau_k} \\ \mathbf{I}_{\tau_1 \cdots \tau_k} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{\tau_1 \cdots \tau_k} \\ \mathbf{I}_{\tau_1 \cdots \tau_k} \end{bmatrix},$$

entonces sin necesidad de multiplicar sabemos por una parte que

$$\mathbf{I}_{\tau_1 \cdots \tau_k} = \begin{bmatrix} 1 & 0 & -4 \\ 0 & 1 & -35 \\ 0 & 0 & 7 \end{bmatrix} : \quad \text{y por la otra que} \quad \mathbf{A}_{\tau_1 \cdots \tau_k} = \begin{bmatrix} 0 & 1 & 0 \\ 7 & 0 & 0 \end{bmatrix}.$$

Así pues, es fácil encontrar la inversa de \mathbf{A} : basta calcular la forma escalonada reducida de la matriz que resulta de apilar \mathbf{A} con la matriz identidad que tiene el mismo número de columnas:

Ejemplo 11.

$$\left[\begin{array}{c|c} \mathbf{A} & \mathbf{I} \end{array} \right] = \left[\begin{array}{cc|cc} 5 & 0 & 1 & 0 \\ 2 & -2 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{array} \right] \xrightarrow{[(1)2+1]} \left[\begin{array}{cc|cc} 5 & 0 & 1 & 0 \\ 0 & -2 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{array} \right] \xrightarrow{\begin{array}{c} [(\frac{1}{5})1] \\ [(-\frac{1}{2})2] \end{array}} \left[\begin{array}{cc|cc} 1 & 0 & 1/5 & 0 \\ 0 & 1 & 0 & 1 \\ 1/5 & 0 & 0 & 0 \\ 1/5 & -1/2 & 0 & 0 \end{array} \right] = \left[\begin{array}{c|c} \mathbf{A}_{\tau_1 \tau_2 \tau_3} & \mathbf{I}_{\tau_1 \tau_2 \tau_3} \end{array} \right].$$

Puesto que $\mathbf{A}_{\tau_1 \tau_2 \tau_3} = \mathbf{I}$, entonces $\mathbf{I}_{\tau_1 \tau_2 \tau_3} = \begin{bmatrix} 1/5 & 0 \\ 1/5 & -1/2 \end{bmatrix} = \mathbf{A}^{-1}$.



Método para encontrar la inversa de una matriz (cuadrada)

1. Con transformaciones elementales de las columnas encuentre una forma escalonada reducida de $\left[\begin{array}{c|c} \mathbf{A} & \mathbf{I} \end{array} \right]$.
2. Si en la matriz obtenida $\left[\begin{array}{c|c} \mathbf{R} & \mathbf{I}_{\tau_1 \dots \tau_k} \end{array} \right]$ la submatriz \mathbf{R} tiene columnas nulas, entonces \mathbf{A} no tiene inversa.
3. En caso contrario, $\mathbf{I}_{\tau_1 \dots \tau_k}$ es la inversa de \mathbf{A} .

Librería NAcAL para Python

```
A = Matrix([ [1, -1, -3], [0, 1, 0], [0, -1, 1] ])
InvMat(A,1) # Calcula la inversa por eliminación Gauss-Jordan
```

$$\left[\begin{array}{ccc|ccc} 1 & -1 & -3 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 3 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array} \right] \xrightarrow{[(1)1+2], [(3)1+3]} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 & -1 & 1 \\ 1 & 1 & 3 & 1 & 1 & 3 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array} \right] \xrightarrow{[(1)3+2]} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 4 & 3 & 1 & 4 & 3 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 \end{array} \right]$$

5.1.5. Inversa de una matriz triangular



Veamos primero que una matriz triangular no puede tener ceros en la diagonal principal y ser invertible.

Por ejemplo, la siguiente matriz no es invertible, pues es imposible encontrar un vector (x, y, z) tal que

$$\begin{bmatrix} 1 & a & b \\ 0 & c & d \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \mathbf{I}_{|2}.$$

Por una parte z es necesariamente cero, y por otra, con la primera y segunda columnas no podemos obtener la segunda componente del vector del lado derecho... Ahora vayamos con la demostración general.

Proposición 5.1.15. Una matriz triangular invertible no tiene ceros en la diagonal principal.

Demostración. Si \mathbf{A} es triangular superior (si *las componentes por debajo de la diagonal son nulas*). ¿Puede haber un cero en a_{11} ? (en la primera posición de la diagonal). Puesto que las componentes por debajo de

la diagonal son cero, si además a_{11} también fuera cero, entonces la primera columna sería nula, lo cual es imposible dado que \mathbf{A} es invertible. **Por tanto a_{11} no puede ser cero.**

¿Puede haber un *primer* cero en la posición j -ésima de la diagonal? Por ser el *primer* cero de la diagonal, las anteriores componentes de la diagonal serían distintas de cero. Así que aplicando eliminación, podemos anular todas las componentes por encima de a_{jj} ; pero como todas las componentes por debajo también son cero (por ser la matriz triangular superior), anularíamos la columna; algo imposible, pues \mathbf{A} es invertible. **Así que a_{jj} tampoco puede ser cero.** (La demostración es semejante para una triangular inferior) \square

Si $\tau_1 \cdots \tau_k$ es una sucesión de transformaciones elementales de *Tipo I* de un proceso de eliminación de “izquierda a derecha”, la matriz $\mathbf{I}_{\tau_1 \cdots \tau_k}$ es triangular superior unitaria, pues solo cambian las componentes a la derecha de la diagonal, ya que ninguna columna es multiplicada por un escalar. Así tenemos el siguiente

Teorema 5.1.16. *La aplicación de proceso de eliminación de “izquierda a derecha” sobre una matriz identidad da lugar a una matriz triangular superior unitaria e invertible.*

Ahora considere una matriz \mathbf{U} triangular superior unitaria. Para calcular su inversa basta aplicar la eliminación de “izquierda a derecha” mediante una sucesión de transformaciones $\tau_1 \cdots \tau_k$ que eliminen todas las componentes a la derecha de la diagonal:

$$\begin{bmatrix} \mathbf{U} \\ \mathbf{I} \end{bmatrix} \xrightarrow[\text{(de "izquierda a derecha")}]{{\tau_1 \cdots \tau_k}} \begin{bmatrix} \mathbf{I} \\ \mathbf{I}_{\tau_1 \cdots \tau_k} \end{bmatrix} \quad \text{por tanto} \quad \mathbf{I}_{\tau_1 \cdots \tau_k} = \mathbf{U}^{-1} \text{ es triangular superior unitaria.}$$

Así pues, tenemos el siguiente

Corolario 5.1.17. *Si \mathbf{U} es cuadrada y triangular superior unitaria, \mathbf{U}^{-1} también es triangular superior unitaria.*

$$\text{Por ejemplo: } \begin{bmatrix} 1 & b & c \\ 0 & 1 & e \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow[\begin{smallmatrix} [(-b)\mathbf{1}+2] \\ [(-c)\mathbf{1}+3] \end{smallmatrix}]{\tau} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & e \\ 0 & 0 & 1 \\ 1 & -b & -c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{[(-e)\mathbf{2}+3]} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -b & be-c \\ 0 & 1 & -e \\ 0 & 0 & 1 \end{bmatrix}.$$

Si \mathbf{U} no es unitaria, aunque si invertible, tras escalar \mathbf{U} por eliminación, será necesario dividir cada columna por el valor de su correspondiente pivote (última operación en la **eliminación Gauss-Jordan**). Consecuentemente la inversa de \mathbf{U} es triangular superior (por el proceso de eliminación), pero no es unitaria (pues el último paso de la eliminación Gauss-Jordan modifica la diagonal). Así hemos demostrado la siguiente

Proposición 5.1.18. *Si \mathbf{U} es triangular superior e invertible, \mathbf{U}^{-1} también es triangular superior.*

Por ejemplo:

$$\begin{bmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & f \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow[\begin{smallmatrix} [(-\frac{b}{a})\mathbf{1}+2] \\ [(-\frac{c}{a})\mathbf{1}+3] \end{smallmatrix}]{\tau} \begin{bmatrix} a & 0 & 0 \\ 0 & d & e \\ 0 & 0 & f \\ 1 & -\frac{b}{a} & -\frac{c}{a} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{[(-\frac{e}{d})\mathbf{2}+3]} \begin{bmatrix} a & 0 & 0 \\ 0 & d & 0 \\ 0 & 0 & f \\ 1 & -\frac{b}{a} & \frac{be-cd}{ad} \\ 0 & 1 & -\frac{e}{d} \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{\begin{smallmatrix} [(\frac{1}{a})\mathbf{1}] \\ [(\frac{1}{d})\mathbf{2}] \\ [(\frac{1}{f})\mathbf{3}] \end{smallmatrix}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{a} & -\frac{b}{ad} & \frac{be-cd}{adf} \\ 0 & \frac{1}{d} & -\frac{e}{df} \\ 0 & 0 & \frac{1}{f} \end{bmatrix}$$

Para terminar esta sección, como la inversa de transpuesta es la transpuesta de la inversa (Proposición 5.1.6 en la página 62), llegamos a los siguientes corolarios:

Corolario 5.1.19. *Si \mathbf{L} es cuadrada y triangular inferior unitaria, \mathbf{L}^{-1} también.*

Corolario 5.1.20. *Si \mathbf{L} es triangular inferior e invertible, \mathbf{L}^{-1} también.*

5.2. Rango de una matriz



Vimos (Corolario 5.1.12) que si una matriz es invertible, ninguna de sus formas pre-escaladas tiene columnas nulas. Una generalización de este resultado da lugar a la definición de **rango** que usaremos aquí.

Considere una matriz \mathbf{A} tal que $\mathbf{A} = \mathbf{KB}$, donde \mathbf{K} es una matriz pre-escalada. Puesto que las columnas de \mathbf{A} son combinación lineal de las de \mathbf{K} , resulta que la **posición de pivote** de cada una de las columnas no nulas de \mathbf{A} coincide con la posición de pivote de *alguna* (solo una) de las columnas de \mathbf{K} .

El motivo es que la posición de pivote de la columna no nula $\mathbf{A}_{|j} = \mathbf{KB}_{|j}$ es, de entre aquellas columnas de \mathbf{K} multiplicadas por coeficientes no nulos de $\mathbf{B}_{|j}$, la posición del pivote que se encuentre más “arriba”, es decir, la de menor posición (recuerde que como \mathbf{K} está escalonada *cada pivote en \mathbf{K} ocupa una fila distinta*).

Para verlo, nótese que si \mathbf{K} tiene n columnas entonces, para toda lista γ de índices entre 1 y n , no repetidos y ordenados de menor a mayor, la submatriz $\mathbf{K}_{|\gamma}$ es pre-escalada (pues como mucho habrán desaparecido algunas columnas, por lo que sigue sin haber elementos distintos de cero a la derecha de cada pivote). En particular, sea γ la lista formada por los índices h tales que $\mathbf{K}_{|h}$ y $_{h|}\mathbf{B}_{|j}$ son distintos de cero, es decir $\gamma = (h \mid \mathbf{K}_{|h} \neq \mathbf{0} \text{ y } _{h|}\mathbf{B}_{|j} \neq 0)$; y sea ν la lista de índices de 1 a n que no están en γ . Entonces, por la Ecuación 3.7 sabemos que

$$\mathbf{A}_{|j} = \mathbf{KB}_{|j} = (\mathbf{K}_{|\gamma})(_{\gamma|}\mathbf{B}_{|j}) + \underbrace{(\mathbf{K}_{|\nu})(_{\nu|}\mathbf{B}_{|j})}_0 = (\mathbf{K}_{|\gamma})(_{\gamma|}\mathbf{B})_{|j}.$$

Sea i la *menor posición* de los pivotes en $\mathbf{K}_{|\gamma}$, entonces (por ser preescalada) el pivote con dicha posición es el único elemento distinto de cero en la fila $_{i|}(\mathbf{K}_{|\gamma})$. Así, como el sub-vector $(_{\gamma|}\mathbf{B})_{|j}$ no tiene elementos nulos, necesariamente $_{i|}\mathbf{A}_{|j} = _{i|}(\mathbf{K}_{|\gamma})(_{\gamma|}\mathbf{B})_{|j} \neq 0$ es el pivote de la columna $\mathbf{A}_{|j}$.

Resumiendo: sabemos que si $\mathbf{A} = \mathbf{KB}$ (con \mathbf{K} pre-escalada), la posición de pivote de cada una de las columnas no nulas de \mathbf{A} coincide con la posición de pivote de *alguna* (solo una) de las columnas de \mathbf{K} . Por tanto, *el conjunto de posiciones de pivote en \mathbf{K} contiene al conjunto de posiciones de pivote en \mathbf{A}* (pues todas las posiciones de pivote en \mathbf{A} están en \mathbf{K}).

Demostremos ahora que las formas pre-escaladas de una matriz \mathbf{A} tienen el mismo número de pivotes:

Supongamos que $\mathbf{E} = \mathbf{I}_{\tau_1 \dots \tau_k}$ y $\mathbf{E}' = \mathbf{I}_{\tau'_1 \dots \tau'_p}$ son matrices tales que $\mathbf{AE} = \mathbf{K}$ y $\mathbf{AE}' = \mathbf{K}'$ son pre-escaladas. Entonces las posiciones de los pivotes de \mathbf{K} y \mathbf{K}' son coincidentes. Veámoslo.

Como $\mathbf{A} = \mathbf{KE}^{-1}$, tenemos que

$$\mathbf{K}' = \mathbf{AE}' = \mathbf{KE}^{-1}\mathbf{E}';$$

por tanto las columnas de $\mathbf{K}' = \mathbf{K}(\mathbf{E}^{-1}\mathbf{E}')$ son combinación lineal de las de la matriz pre-escalada \mathbf{K} ; así que *el conjunto de posiciones de pivote en \mathbf{K} contiene al de \mathbf{K}'* . Pero también $\mathbf{K} = \mathbf{K}'(\mathbf{E}'^{-1}\mathbf{E})$, por tanto las columnas de \mathbf{K} son combinación lineal de las de la matriz pre-escalada \mathbf{K}' ; así que *el conjunto de posiciones de pivote en \mathbf{K}' contiene al de \mathbf{K}* . Consecuentemente, ambos conjuntos son iguales. Y como ambas matrices están pre-escaladas, es decir, como en cada matriz cada pivote ocupa una posición distinta, necesariamente ambas matrices tienen el mismo número de pivotes.

Este resultado nos permite dar la siguiente definición:

Definición 5.4 (Rango de una matriz). *El rango de \mathbf{A} es el número de columnas no nulas de cualquiera de sus formas pre-escaladas. Denotamos el rango de \mathbf{A} con $\text{rg}(\mathbf{A})$.*

Es decir, el rango de \mathbf{A} es el número de pivotes de cualquiera de sus formas pre-escaladas.

5.2.1. Algunas propiedades del rango de una matriz

Proposición 5.2.1. Si \mathbf{E} es invertible, entonces $\text{rg}(\mathbf{A}) = \text{rg}(\mathbf{AE})$.

Demostración. Con ambas matrices se pueden obtener idéntica forma pre-escalónada $\mathbf{A}_{\tau_1 \dots \tau_k}$. Veámoslo: sea $\mathbf{E} = \mathbf{I}_{\tau'_1 \dots \tau'_p}$, entonces $\mathbf{A}_{\tau_1 \dots \tau_k} = (\mathbf{AEE}^{-1})_{\tau_1 \dots \tau_k} = (\mathbf{AE})_{(\tau'_1 \dots \tau'_p)^{-1} \tau_1 \dots \tau_k}$. \square

Proposición 5.2.2. El rango de \mathbf{AB} es menor o igual que el de \mathbf{B} .

Demostración. Sea $\mathbf{E} = \mathbf{I}_{\tau_1 \dots \tau_k}$ tal que $\mathbf{BE} = \mathbf{L}$ es escalónada. Entonces

- Aplicando la proposición anterior tenemos que en particular: $\text{rg}(\mathbf{AB}) = \text{rg}((\mathbf{AB})\mathbf{E})$.
- Por otra parte tenemos que $(\mathbf{AB})\mathbf{E} = \mathbf{A}(\mathbf{BE}) = \mathbf{AL}$, y si k es el número de columnas nulas de la matriz escalónada \mathbf{L} , las últimas k columnas de \mathbf{AL} son necesariamente nulas.
- Si realizamos transformaciones elementales sobre las columnas de \mathbf{AL} de manera que no modificamos las k últimas, obtendremos una forma escalónada con al menos k columnas nulas. Por tanto, el rango de \mathbf{AL} es menor o igual que el rango de \mathbf{L} , por lo que finalmente concluimos que

$$\text{rg}(\mathbf{AB}) = \text{rg}(\mathbf{ABE}) = \text{rg}(\mathbf{AL}) \leq \text{rg}(\mathbf{L}) = \text{rg}(\mathbf{B}). \quad \square$$

Proposición 5.2.3. Si \mathbf{E} es invertible, entonces $\text{rg}(\mathbf{EA}) = \text{rg}(\mathbf{A})$.

Demostración. La proposición anterior nos indica que $\text{rg}(\mathbf{EA}) \leq \text{rg}(\mathbf{A})$, pero también nos indica que $\text{rg}(\mathbf{E}^{-1}(\mathbf{EA})) \leq \text{rg}(\mathbf{EA})$; por tanto $\text{rg}(\mathbf{A}) \leq \text{rg}(\mathbf{EA})$. Por tanto, $\text{rg}(\mathbf{EA}) = \text{rg}(\mathbf{A})$. \square

Aplicando eliminación *gaussiana* por filas¹ a una matriz escalónada reducida \mathbf{R} (anulando las componentes por debajo de los pivotes y dejando las filas nulas por debajo de las no nulas), obtenemos una matriz \mathbf{S} escalónada reducida cuyos pivotes están en la diagonal principal. Así, tanto \mathbf{S} como su transpuesta son diagonales (con idéntica diagonal); consecuentemente $\text{rg}(\mathbf{S}) = \text{rg}(\mathbf{S}^T)$:

Librería NAcAL para Python

```
R = Matrix( [ [1,0,0,0], [-1,0,0,0], [0,1,0,0], [0,0,1,0], [0,0,2,0] ] )
S = ElimGF(R,1) # Eliminación Gaussiana por filas
~S # transpuesta de S
```

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & 0 \end{bmatrix} \xrightarrow{[(1)\mathbf{I}+2]} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & 0 \end{bmatrix} \xrightarrow{[(-2)\mathbf{I}+5]} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{\begin{smallmatrix} \mathbf{I} \\ [3 \rightleftharpoons 4] \\ [2 \rightleftharpoons 3] \end{smallmatrix}} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \mathbf{S}$$

$$\mathbf{S}^T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Aprovechemos este hecho para demostrar las dos siguientes proposiciones.

¹Véase el Apéndice 4.A en la página 54

Proposición 5.2.4. Si \mathbf{R} es escalonada reducida, entonces $\text{rg}(\mathbf{R}) = \text{rg}(\mathbf{R}^\top)$.

Demostración. Puesto que podemos encontrar una matriz invertible $\mathbf{E} = \tau_1 \cdots \tau_k \mathbf{I}$ tal que $\mathbf{ER} = \mathbf{S}$ es diagonal. Así, por las proposiciones 5.2.3 y 5.2.1, y puesto que \mathbf{E}^\top también es invertible:

$$\text{rg}(\mathbf{R}) = \text{rg}(\mathbf{ER}) = \text{rg}(\mathbf{S}) = \text{rg}(\mathbf{S}^\top) = \text{rg}(\mathbf{R}^\top \mathbf{E}^\top) = \text{rg}(\mathbf{R}^\top). \quad \square$$

Proposición 5.2.5. $\text{rg}(\mathbf{A}) = \text{rg}(\mathbf{A}^\top)$.

Demostración. Sea $\mathbf{E} = \mathbf{I}_{\tau_1 \cdots \tau_k}$ tal que $\mathbf{AE} = \mathbf{R}$ es escalonada reducida, entonces (aplicando la Proposición 5.2.3 en la última igualdad) tenemos: $\text{rg}(\mathbf{A}) = \text{rg}(\mathbf{AE}) = \text{rg}(\mathbf{R}) = \text{rg}(\mathbf{R}^\top) = \text{rg}(\mathbf{E}^\top \mathbf{A}^\top) = \text{rg}(\mathbf{A}^\top)$. \square

Proposición 5.2.6. El rango de \mathbf{AB} es menor o igual que el de \mathbf{A} .

Demostración. $\text{rg}(\mathbf{AB}) = \text{rg}(\mathbf{B}^\top \mathbf{A}^\top) \leq \text{rg}(\mathbf{A}^\top) = \text{rg}(\mathbf{A})$. \square

Librería NAcAL para Python


```
Elim(A).rango    # rango es una atributo de la clase Elim
A.rango()        # rango() es un método de la clase Matrix
```

Rango completo por columnas (por filas). Una matriz cuyas formas pre-escaladas tienen pivote en todas las columnas se dice que es *de rango completo por columnas*; y si tienen pivote en todas las filas se dice que es *de rango completo por filas*.

Rango completo. Una matriz que es de rango completo tanto por filas como por columnas (y que por tanto es cuadrada) se dice que es *de rango completo*.

Así pues, por el Corolario 5.1.11, *una matriz de rango completo es invertible*, pues es cuadrada y sin columnas nulas (por tener todas pivote).

Apendices a la lección

 Vamos a aclarar algo sobre la notación empleada en la Sección 4.B en la página 57. Recuerde que puede posponer el estudio de las transformaciones elementales “espejo” hasta que llegue a la Lección 16.

5.A. “Espejo” de la inversa de una secuencia de transformaciones elementales

Fíjese que ${}_{\text{esp}(\tau^{-1})}\mathbf{I} = \mathbf{I}_{\tau^{-1}} = (\mathbf{I}_\tau)^{-1} = ({}_{\text{esp}(\tau)}\mathbf{I})^{-1} = ({}_{\text{esp}(\tau)}\mathbf{I})^{-1}$; es decir, que $\text{esp}(\tau^{-1}) = (\text{esp}(\tau))^{-1}$.

Como $\text{esp}(\tau_1 \cdots \tau_k) = \text{esp}(\tau_1) \cdots \text{esp}(\tau_k)$, y como $(\tau_1 \cdots \tau_k)^{-1} = \tau_k^{-1} \cdots \tau_1^{-1}$; tenemos que $\text{esp}((\tau_1 \cdots \tau_k)^{-1}) = \text{esp}(\tau_k^{-1} \cdots \tau_1^{-1}) = \text{esp}(\tau_k^{-1}) \cdots \text{esp}(\tau_1^{-1}) = (\text{esp}(\tau_k))^{-1} \cdots (\text{esp}(\tau_1))^{-1} = (\text{esp}(\tau_1 \cdots \tau_k))^{-1}$; así que podemos escribir sencillamente:

$$\boxed{\text{esp}(\tau_1 \cdots \tau_k)^{-1}}$$

Librería NAcAL para Python

```
Tr = T((3,2,1)) & T((-2,3,2)) & T((7,1,3)) & T({2,3}) & T((20,1)) # Suc. transf.
(Tr**-1).espejo().t == (Tr.espejo()**-1).t # ¿son iguales?
```

True

Parte III

Subespacios y resolución de sistemas de ecuaciones lineales

Espacios vectoriales y funciones lineales



El propósito de esta lección es introducir terminología **abstracta** que usaremos a partir de ahora. De paso recapitularemos lo visto hasta aquí empleando la nueva terminología.

Comenzaremos con los espacios vectoriales y las funciones lineales. Luego veremos formas de definir espacios más pequeños que uno dado (los subespacios) y espacios más grandes a partir de otros subespacios: mediante productos cartesianos y con funciones cuya imagen está contenida en un espacio vectorial.

Casi cualquier espacio vectorial con el que se encuentre en otras disciplinas será fruto de un encadenamiento de espacios vectoriales de funciones y/o productos cartesianos, o subespacios de estos.

6.1. Espacios vectoriales

Para cualesquiera *números reales* a , b y c (y empleando la suma y producto habituales) se verifica que:

- | | |
|----------------------------------|---------------------------|
| 1. $a + b = b + a$. | 5. $a(b + c) = ab + ac$. |
| 2. $a + (b + c) = (a + b) + c$. | 6. $(a + b)c = ac + bc$. |
| 3. $a + 0 = a$. | 7. $a(bc) = (ab)c$. |
| 4. $a + (-a) = 0$. | 8. $1a = a$. |

También sabemos que si definimos la suma y el producto por escalares para \mathbb{R}^n como:

$$(a + b)_{|i} = a_{|i} + b_{|i} \quad \text{y} \quad (\lambda b)_{|i} = \lambda(b_{|i}),$$

se verifican las siguientes propiedades (Proposición 1.2.1 en la página 9):

- | | |
|--|---|
| 1. $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$ | 5. $\lambda(\mathbf{a} + \mathbf{b}) = \lambda\mathbf{a} + \lambda\mathbf{b}$ |
| 2. $\mathbf{a} + (\mathbf{b} + \mathbf{c}) = (\mathbf{a} + \mathbf{b}) + \mathbf{c}$ | 6. $(\lambda + \varphi)\mathbf{a} = \lambda\mathbf{a} + \varphi\mathbf{a}$ |
| 3. $\mathbf{a} + \mathbf{0} = \mathbf{a}$ | 7. $\lambda(\varphi\mathbf{a}) = (\lambda\varphi)\mathbf{a}$ |
| 4. $\mathbf{a} + (-\mathbf{a}) = \mathbf{0}$ | 8. $1\mathbf{a} = \mathbf{a}$ |

Y si definimos la suma y el producto por escalares para $\mathbb{R}^{m \times n}$ (*matrices de orden m por n*) como

$$(\mathbf{A} + \mathbf{B})_{|j} = \mathbf{A}_{|j} + \mathbf{B}_{|j} \quad \text{y} \quad (\lambda\mathbf{A})_{|j} = \lambda(\mathbf{A}_{|j}),$$

se verifican las siguientes propiedades (Proposición 1.5.1 en la página 19):

- | | |
|--|---|
| 1. $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$ | 5. $\lambda(\mathbf{A} + \mathbf{B}) = \lambda\mathbf{A} + \lambda\mathbf{B}$ |
| 2. $\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$ | 6. $(\lambda + \varphi)\mathbf{A} = \lambda\mathbf{A} + \varphi\mathbf{A}$ |
| 3. $\mathbf{A} + \mathbf{0} = \mathbf{A}$ | 7. $\lambda(\varphi\mathbf{A}) = (\lambda\varphi)\mathbf{A}$ |
| 4. $\mathbf{A} + (-\mathbf{A}) = \mathbf{0}$ | 8. $1\mathbf{A} = \mathbf{A}$ |

A primera vista es evidente que subyace una estructura común a \mathbb{R} , a \mathbb{R}^n y al conjunto de matrices $\mathbb{R}^{m \times m}$ (con sus respectivas operaciones de suma y producto por escalares). Esta estructura se denomina **espacio vectorial**.

Definición 6.1. Un espacio vectorial es un conjunto \mathcal{V} de objetos¹ junto con dos operaciones que denominamos: **suma** y **producto por escalares**.

En cuanto a los elementos de \mathcal{V} :

- se denominan **vectores** y los denotaremos genéricamente con letras minúsculas en cursiva y con una flecha por encima: $\vec{x}, \vec{y}, \vec{z}, \dots$ ²

En cuanto a las operaciones:

- la **suma** asocia cualquier par de elementos de \mathcal{V} con otro elemento de \mathcal{V} .

Dicho de otro modo, la suma de dos vectores, \vec{x} e \vec{y} , de \mathcal{V} es otro vector de \mathcal{V} que denotamos $\vec{x} + \vec{y}$. Esto se representa esquemáticamente con³

$$\begin{aligned} _ + _ : \mathcal{V} \times \mathcal{V} &\longrightarrow \mathcal{V} \\ (\vec{x}, \vec{y}) &\longmapsto \vec{x} + \vec{y} \end{aligned}$$

(Puesto que sumando vectores de \mathcal{V} se obtienen elementos \mathcal{V} , se dice que \mathcal{V} es *cerrado para la suma*).

La operación suma debe verificar las siguientes cuatro propiedades:

1. $\vec{x} + \vec{y} = \vec{y} + \vec{x}$
 2. $\vec{x} + (\vec{y} + \vec{z}) = (\vec{x} + \vec{y}) + \vec{z}$
 3. Existe un vector,⁴ que denotamos con $\vec{0}$ (vector nulo) tal que $\vec{x} + \vec{0} = \vec{x}$, para todo $\vec{x} \in \mathcal{V}$.
 4. Para cada $\vec{x} \in \mathcal{V}$ existe un vector, que denotamos con $-\vec{x}$ (vector opuesto) tal que $\vec{x} + (-\vec{x}) = \vec{0}$.
- el **producto por escalares** asocia cualquier par, formado por un un escalar y un vector de \mathcal{V} , con un vector de \mathcal{V} .

Dicho de otro modo, el producto de $a \in \mathbb{R}$ por $\vec{x} \in \mathcal{V}$ es otro vector de \mathcal{V} que denotamos $a \cdot \vec{x}$. Esto se representa esquemáticamente con

$$\begin{aligned} _ \cdot _ : \mathbb{R} \times \mathcal{V} &\longrightarrow \mathcal{V} \\ (a, \vec{y}) &\longmapsto a \cdot \vec{y} \end{aligned}$$

(normalmente omitimos el “punto” y escribimos sencillamente $a\vec{x}$, que llamaremos *múltiplo* de \vec{x}).

La operación producto por un escalar debe verificar las siguientes cuatro propiedades:

5. $a(\vec{x} + \vec{y}) = a\vec{x} + a\vec{y}$
6. $(a + b)\vec{x} = a\vec{x} + b\vec{x}$
7. $(ab)\vec{x} = a(b\vec{x})$
8. $1\vec{x} = \vec{x}$

Por tanto, y recordando lo visto en la Lección 1, sabemos que el conjunto de números reales \mathbb{R} , el conjunto de listas ordenadas \mathbb{R}^n y el conjunto de matrices $\mathbb{R}^{m \times n}$ (junto con las operaciones de suma y producto por escalares definidas respectivamente en a cada conjunto) son espacios vectoriales.

EJERCICIO 34. Demuestre que

- (a) si $\vec{z} + \vec{x} = \vec{x}$ para cualquier \vec{x} , entonces necesariamente $\vec{z} = \vec{0}$.
- (b) si $\vec{z} + \vec{x} = \vec{0}$ entonces $\vec{z} = -\vec{x}$.

¹objetos del mismo tipo: o números, o listas de números, o matrices, o funciones, o polinomios, o variables aleatorias, etc.

²por tanto, según qué tipo de objetos contenga \mathcal{V} , resultará que \vec{x} será un número, o una lista de números, etc.

³Véase la Sección 6.A.1 en la página 88 del apéndice a esta lección.

⁴y por tanto $\mathcal{V} \neq \emptyset$

Por tanto, en todo espacio vectorial los vectores $\vec{0}$ y, para cada \vec{x} , el vector opuesto son únicos.

Notación. En el caso particular de \mathbb{R}^n usamos los símbolos \mathbf{a} , \mathbf{x} , $\mathbf{0}$, etc. para denotar las *listas ordenadas de números reales* (i.e., los vectores de \mathbb{R}^n); y en el caso de $\mathbb{R}^{m \times n}$ usamos \mathbf{A} , \mathbf{X} , $\mathbf{0}$, etc. para las matrices (i.e., los vectores de $\mathbb{R}^{m \times n}$). Pero en general denotamos los vectores de un espacio vectorial genérico \mathcal{V} con \vec{a} , \vec{x} , $\vec{0}$, etc.

Para denotar espacios vectoriales genéricos usaremos caracteres como: \mathcal{C} , \mathcal{N} , \mathcal{X} , \mathcal{V} , \mathcal{W} ...

☞ Fíjese que la definición de espacio vectorial es abstracta, pues nada se dice sobre la naturaleza de los objetos que constituyen el conjunto \mathcal{V} , como tampoco sobre la definición de las operaciones *suma de vectores* y *producto de un escalar por un vector*; tan solo se enumeran las propiedades que deben verificar las operaciones.

Consecuentemente, para definir un espacio vectorial concreto necesitaremos indicar qué elementos constituyen el conjunto; y definir las operaciones (suma y producto por escalares) de manera que verifiquen las propiedades indicadas. Implícitamente así lo hicimos en la Lección 1. (véase el ejercicio a continuación del recuadro)

EJERCICIO 35. Considere el conjunto de números reales *positivos*, \mathbb{R}^+ , junto con las siguientes definiciones: suma de \vec{x} e \vec{y} es xy ; y el producto de un escalar por el vector \vec{x} es x^c , es decir,

$$\begin{aligned} _ + _ : \mathbb{R}^+ \times \mathbb{R}^+ &\longrightarrow \mathbb{R} & \text{y} & & _ \cdot _ : \mathbb{R} \times \mathbb{R}^+ &\longrightarrow \mathbb{R} \\ (\vec{x}, \vec{y}) &\longmapsto xy & & & (c, \vec{x}) &\longmapsto x^c \end{aligned}$$

Demuestre que \mathbb{R}^+ junto con las operaciones indicadas es un espacio vectorial. ¿Quién es $\vec{0}$ en este caso?

6.1.1. Algunas propiedades

Hay algunas propiedades muy elementales que se cumplen en cualquier espacio vectorial \mathcal{V} y que se deducen de las ocho propiedades descritas más arriba: para el vector nulo $\vec{0} \in \mathcal{V}$ y cualquier escalar a :

$$a\vec{0} = a(\vec{0} + \vec{0}) = a\vec{0} + a\vec{0}.$$

Restando $-(a\vec{0})$ a ambos lados deducimos que $\vec{0} = a\vec{0}$. Y para el escalar 0 y cualquier vector $\vec{x} \in \mathcal{V}$:

$$0\vec{x} = (0 + 0)\vec{x} = 0\vec{x} + 0\vec{x}$$

Restando $-(0\vec{x})$ a ambos lados deducimos que $\vec{0} = 0\vec{x}$. Además, $(-1)\vec{x} = -\vec{x}$, puesto que

$$\vec{x} + (-1)\vec{x} = 1\vec{x} + (-1)\vec{x} = (1 - 1)\vec{x} = 0\vec{x} = \vec{0}.$$

Por último, si $a\vec{v} = \vec{0}$, hay dos casos posibles. O bien $a = 0$; o si es distinto de cero, multiplicando ambos lados por a^{-1} , tenemos $a^{-1}(a\vec{v}) = a^{-1}\vec{0}$. Desarrollando el lado izquierdo llegamos a $a^{-1}(a\vec{v}) = (a^{-1}a)\vec{v} = 1\vec{v} = \vec{v}$; y el lado derecho es $\vec{0}$, pues $\vec{0}$ por cualquier escalar es $\vec{0}$. Así llegamos a que si $a \neq 0$ entonces $\vec{v} = \vec{0}$.

Acabamos de demostrar el siguiente

Teorema 6.1.1. Si \mathcal{V} es un espacio vectorial

1. $\vec{0} = a\vec{0}$ para $\vec{0} \in \mathcal{V}$ y $a \in \mathbb{R}$.
2. $\vec{0} = 0\vec{x}$ para $\vec{x} \in \mathcal{V}$ y $0 \in \mathbb{R}$.
3. $(-1)\vec{x} = -\vec{x}$ para todo $\vec{x} \in \mathcal{V}$.
4. Si $a\vec{v} = \vec{0}$ o bien $a = 0$ o bien $\vec{v} = \vec{0}$.

Particularizando a los espacios \mathbb{R}^n y $\mathbb{R}^{m \times n}$ tenemos que:

En \mathbb{R}^n

1. $\mathbf{0} = a\mathbf{0}$ para $\mathbf{0} \in \mathbb{R}^n$ y $a \in \mathbb{R}$.
2. $\mathbf{0} = 0\mathbf{x}$ para $\mathbf{x} \in \mathbb{R}^n$ y $0 \in \mathbb{R}$.
3. $(-1)\mathbf{x} = -\mathbf{x}$ para todo $\mathbf{x} \in \mathbb{R}^n$.
4. Si $a\mathbf{x} = \mathbf{0}$ o bien $a = 0$ o bien $\mathbf{x} = \mathbf{0}$.

En $\mathbb{R}^{m \times n}$

1. $\mathbf{0} = a\mathbf{0}$ para $\mathbf{0} \in \mathbb{R}^{m \times n}$ y $a \in \mathbb{R}$.
2. $\mathbf{0} = 0\mathbf{A}$ para $\mathbf{A} \in \mathbb{R}^{m \times n}$ y $0 \in \mathbb{R}$.
3. $(-1)\mathbf{A} = -\mathbf{A}$ para todo $\mathbf{A} \in \mathbb{R}^{m \times n}$.
4. Si $a\mathbf{A} = \mathbf{0}$ o bien $a = 0$ o bien $\mathbf{A} = \mathbf{0}$.



Hemos visto que el conjunto de matrices m por n ($\mathbb{R}^{m \times n}$), las listas ordenadas de n números (\mathbb{R}^n) y el conjunto de números reales \mathbb{R} , comparten una estructura común: la estructura de *espacio vectorial*.

Dicha estructura consiste en un conjunto de objetos junto con las operaciones de *suma* y *producto por escalares* que verifican las mismas propiedades que ya vimos para la suma y producto para escalares en la Lección 1. Los elementos de un espacio vectorial \mathcal{V} se denominan *vectores* de \mathcal{V} .

En la Definición 2.2 de la Página 26 definimos las *combinaciones lineales* de vectores de \mathbb{R}^n . Ahora daremos una definición de combinación lineal para vectores de un espacio vectorial genérico \mathcal{V} :

Definición 6.2 (Combinación lineal). Sea $[\vec{v}_1; \vec{v}_2; \dots \vec{v}_n;]$ un sistema de n vectores de \mathcal{V} . Llamamos *combinación lineal* a cualquier suma de múltiplos de dichos vectores:

$$a_1 \vec{v}_1 + a_2 \vec{v}_2 + \dots + a_n \vec{v}_n$$

donde los “ a_i ” son los coeficientes de la combinación lineal.



En este curso trabajaremos principalmente con el espacio vectorial \mathbb{R}^n , pero debe saber que existen otros conjuntos que (con la definición de suma y producto por escalares adecuada) poseen estructura de espacio vectorial. Usted se encontrará con ellos en otras asignaturas (véase el final de la Sección 6.3.3).

6.2. Funciones lineales



El segundo término abstracto de la lección son las funciones lineales...

A lo largo de estas lecciones hemos indicado en numerosas ocasiones que ciertos operadores eran *lineales*.⁵ En cada una de esas ocasiones lo que hemos visto es un ejemplo de función lineal.⁶

Definición 6.3 (Función lineal). Sean \mathcal{D} y \mathcal{V} dos espacios vectoriales. Decimos que la función $f: \mathcal{D} \rightarrow \mathcal{V}$ es *lineal* si satisface las siguientes propiedades:

1. Para todo $\vec{x}, \vec{y} \in \mathcal{D}$, $f(\vec{x} + \vec{y}) = f(\vec{x}) + f(\vec{y})$.
2. Para todo $\vec{x} \in \mathcal{D}$ y para todo $\alpha \in \mathbb{R}$, $f(\alpha \vec{x}) = \alpha f(\vec{x})$.



Repasemos las operaciones empleadas hasta el momento y veremos que casi todas son funciones lineales...

6.2.1. Ejemplos de funciones lineales que ya hemos usado

Ejemplo 12. Las definiciones de suma y producto por escalares dadas en la Sección 1.2 en la página 6 convierten a la función $f(\mathbf{x}) = \mathbf{x}_i$ que selecciona la componente i -ésima de un vector de \mathbb{R}^n en una función

⁵Consideraremos como sinónimos los términos: función lineal, aplicación lineal, transformación lineal u operador lineal.

⁶Se supone que usted conoce las funciones, pero dispone del Apéndice 6.A en la página 87 con todo lo necesario.

lineal. Podemos describir la función “*selector de una componente*”, $f(\mathbf{x}) = \mathbf{x}_{|i}$, con el siguiente esquema:

$$\begin{array}{ccc} -|i : \mathbb{R}^n & \longrightarrow & \mathbb{R} \\ \mathbf{x} & \longmapsto & x_i \end{array}$$

Y sabemos que dicha función (que selecciona la i -ésima componente, $\mathbf{x} \rightarrow \mathbf{x}_{|i}$) es lineal puesto que:

$$\begin{aligned} (\mathbf{x} + \mathbf{y})_{|i} &= \mathbf{x}_{|i} + \mathbf{y}_{|i} \\ (\lambda \mathbf{x})_{|i} &= \lambda(\mathbf{x}_{|i}). \end{aligned}$$

EJERCICIO 36. Describa con un esquema (al modo de ejemplo anterior) las siguientes funciones lineales, e indique (como en el ejemplo anterior) por qué son funciones lineales:

- (a) La selección de la columna j -ésima de una matriz: $\mathbf{A} \rightarrow \mathbf{A}_{|j}$ (Página 19)
- (b) La selección de la componente de la fila i -ésima y columna j -ésima de una matriz: $\mathbf{A} \rightarrow {}_i|\mathbf{A}_{|j}$ (Página 20)
- (c) La transposición de una matriz: $\mathbf{A} \rightarrow \mathbf{A}^\top$ (Página 20)
- (d) La selección de la fila i -ésima de una matriz: $\mathbf{A} \rightarrow {}_i|\mathbf{A}$ (Página 21)
- (e) El producto de una matriz por el vector \mathbf{b} por su derecha: $\mathbf{A} \rightarrow \mathbf{A}\mathbf{b}$ (Página 28)
- (f) El producto de un vector por una matriz \mathbf{A} a su izquierda: $\mathbf{b} \rightarrow \mathbf{A}\mathbf{b}$ (Página 26)
- (g) El producto de una matriz por el vector \mathbf{a} por su izquierda: $\mathbf{B} \rightarrow \mathbf{a}\mathbf{B}$ (Página 29)
- (h) El producto de un vector por una matriz \mathbf{B} a su derecha: $\mathbf{a} \rightarrow \mathbf{a}\mathbf{B}$ (Página 30)
- (i) El producto de una matriz por una matriz \mathbf{B} a su derecha: $\mathbf{A} \rightarrow \mathbf{A}\mathbf{B}$ (Página 33)
- (j) El producto de una matriz por una matriz \mathbf{A} a su izquierda: $\mathbf{B} \rightarrow \mathbf{A}\mathbf{B}$ (Página 33)
- (k) Una transformación elemental τ de las columnas de una matriz: $\mathbf{A} \rightarrow \mathbf{A}_\tau$ (Página 46)
- (l) Una transformación elemental τ de las filas de una matriz: $\mathbf{A} \rightarrow {}_\tau\mathbf{A}$ (Página 54)

🔗 ¡Un título alternativo para las primeras lecciones podría haber sido... “*Ejemplos de funciones lineales*”!

6.2.2. Algunas propiedades de las funciones lineales

A continuación vamos a ver algunos resultados básicos de las funciones lineales. Empecemos con la *composición* y la *inversa* de funciones lineales:

EJERCICIO 37. Demuestre las siguientes proposiciones (ambas demostraciones son muy similares):

- (a) **Proposición 6.2.1.** La composición $g \circ f$ de dos funciones lineales $f: \mathcal{D} \rightarrow \mathcal{V}$ y $g: \mathcal{V} \rightarrow \mathcal{W}$ es una función lineal.
- (b) **Proposición 6.2.2.** Si $f: \mathcal{D} \rightarrow \mathcal{V}$ es lineal e invertible, su inversa $f^{-1}: \mathcal{V} \rightarrow \mathcal{D}$ también es lineal.

¡En las primeras lecciones también nos hemos topado con composiciones de funciones lineales! Por ejemplo:

- La selección de la componente de la fila i -ésima y columna j -ésima de una matriz $f(\mathbf{A}) = {}_i|\mathbf{A}_{|j}$:

$${}_i| - |j : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$$

es la composición de la función que selecciona la i -ésima fila de la matriz, con la función que selecciona la componente j -ésima de la fila.

$${}_i| - |j : \mathbb{R}^{m \times n} \xrightarrow{{}_i| -} \mathbb{R}^n \xrightarrow{-|j} \mathbb{R}$$

(fíjese que del mismo modo es la composición de la función que selecciona la j -ésima columna de la matriz, con la función que selecciona la componente i -ésima de la columna.)

- El producto de dos matrices por un vector $\mathbf{A}\mathbf{B}\mathbf{x}$ es un ejemplo de composición de funciones lineales; pues si $f_{\mathbf{A}}(\mathbf{x}) = \mathbf{A}\mathbf{x}$ y $f_{\mathbf{B}}(\mathbf{x}) = \mathbf{B}\mathbf{x}$; entonces

$$[f_{\mathbf{A}} \circ f_{\mathbf{B}}](\mathbf{x}) = f_{\mathbf{A}}(f_{\mathbf{B}}(\mathbf{x})) = f_{\mathbf{A}}(\mathbf{B}\mathbf{x}) = \mathbf{A}\mathbf{B}\mathbf{x} = f_{\mathbf{AB}}(\mathbf{x}).$$

Fíjese que aunque en el ejemplo anterior el orden en el que aparecen compuestas las funciones coincide con el orden en el que aparecen las matrices en el producto, ocurre lo contrario cuando se multiplica por la izquierda: si $g_{\mathbf{A}}(\mathbf{x}) = \mathbf{x}\mathbf{A}$ y $g_{\mathbf{B}}(\mathbf{x}) = \mathbf{x}\mathbf{B}$, entonces

$$[g_{\mathbf{A}} \circ g_{\mathbf{B}}](\mathbf{x}) = g_{\mathbf{A}}(g_{\mathbf{B}}(\mathbf{x})) = g_{\mathbf{A}}(\mathbf{x}\mathbf{B}) = \mathbf{x}\mathbf{B}\mathbf{A} = g_{\mathbf{BA}}(\mathbf{x}).$$

Otro ejemplo es el producto de varias matrices: \mathbf{ABX} .


- Una sucesión de k transformaciones elementales de las columnas $\mathbf{A}_{\tau_1 \dots \tau_k}$ es una composición de k funciones lineales; y lo mismo una sucesión de transformaciones de las filas. . . o también una sucesión de transformaciones elementales tanto de las filas como de las columnas (puesto que las transformaciones elementales sobre \mathbf{A} se pueden representar como un producto de varias matrices).

Como ejemplos de funciones lineales invertibles con las que ya hemos trabajado podemos indicar:

- Las transformaciones elementales.
- Si \mathbf{A} es invertible, entonces $f_{\mathbf{A}}$ es invertible; en particular $(f_{\mathbf{A}})^{-1} = f_{\mathbf{A}^{-1}}$.

Explicación: por una parte $f_{\mathbf{A}}(f_{\mathbf{A}^{-1}}(\mathbf{x})) = \mathbf{A}\mathbf{A}^{-1}\mathbf{x} = \mathbf{x}$ y por otra $f_{\mathbf{A}^{-1}}(f_{\mathbf{A}}(\mathbf{y})) = \mathbf{A}^{-1}\mathbf{A}\mathbf{y} = \mathbf{y}$.
(Proposición 6.A.1 en la página 89 del apéndice de la lección).

6.3. Creación de nuevos espacios vectoriales

 Comencemos con la creación de espacios vectoriales dentro de espacios vectoriales.

6.3.1. Subespacios

Algunos subconjuntos de un espacio vectorial \mathcal{V} (y con las mismas operaciones de \mathcal{V}) son espacios vectoriales por sí mismos. Dichos subconjuntos se denominan *subespacios*.

Definición 6.4. *Un subconjunto \mathcal{W} no vacío de un espacio vectorial \mathcal{V} es un subespacio de \mathcal{V} si es cerrado para la suma y el producto por escalares. Es decir, si para todo \vec{x} e \vec{y} en \mathcal{W} y todo escalar a ,*

1. $\vec{x} + \vec{y}$ esta en el subespacio \mathcal{W}
2. $a\vec{x}$ esta en el subespacio \mathcal{W}

Así, para demostrar que \mathcal{W} es un subespacio de \mathcal{V} basta comprobar que es cerrado tanto para la suma como para el producto por escalares. Alternativamente, también basta demostrar que *es cerrado para las combinaciones lineales*, pues si para todo \vec{x} e \vec{y} en \mathcal{W} y para todo $a, b \in \mathbb{R}$

$$a\vec{x} + b\vec{y} \text{ esta en el subespacio } \mathcal{W}$$

entonces con $a = b = 1$ satisfacemos la primera condición de la definición y con $b = 0$ la segunda. Por tanto, decir que *el conjunto es cerrado para las combinaciones lineales de cualquier para de vectores* es equivalente a decir que el conjunto es cerrado para la suma y el producto por escalares.

EJERCICIO 38. ¿Es subespacio el conjunto de combinaciones lineales de n vectores de un espacio vectorial?

Fíjese que si W es un subconjunto de un espacio vectorial \mathcal{V} ; sabremos inmediatamente que W no es un subespacio si $\vec{0}$ no está en W , o si está \vec{x} pero no $-\vec{x}$.



Un subespacio es un espacio vectorial que está contenido dentro de otro espacio vectorial.

Para comprobar que \mathcal{W} es subespacio basta verificar si es cerrado para la suma y el producto por escalares (es decir, que es cerrado para las combinaciones lineales).

Ejemplos de subespacios

- Los vectores de \mathbb{R}^3 cuya primera componente es cero
Si sumamos dos vectores con primera componente nula, obtenemos otro vector con la primera componente nula; y lo mismo ocurre con cualquier múltiplo de un vector con primera componente nula.
- Las funciones reales ($\mathbb{R} \rightarrow \mathbb{R}$) que además son continuas (véase la Sección 6.3.3 en la página siguiente).
Si sumamos dos funciones reales y continuas, obtenemos otra función real y continua; y lo mismo ocurre si multiplicamos una función real y continua por un escalar.
- Polinomios ($\mathbb{R}[x]$).
- Variables aleatorias con esperanza cero: su suma tiene esperanza cero y cualquier múltiplo también.
- Variables aleatorias con distribución normal (gaussiana).



Nótese que como todo conjunto está contenido en sí mismo, todo espacio vectorial es un *subespacio* de sí mismo. Por tanto, para cualquier espacio vectorial \mathcal{V} , son subespacios tanto \mathcal{V} , como el subconjunto $\{\vec{0}\}$ que solo contiene el vector nulo (pues obviamente es cerrado para la suma y el producto por escalares).

Intersección de subespacios

La intersección de subespacios es un subespacio:

Teorema 6.3.1. Sean \mathcal{V}_1 y \mathcal{V}_2 subespacios de \mathcal{S} , entonces la intersección de $\mathcal{W} = \mathcal{V}_1 \cap \mathcal{V}_2$ también lo es.

Demostración. Sabemos que la intersección es no vacía pues el $\vec{0}$ está en ambos subespacios. Así que basta probar que la intersección es cerrada para las combinaciones lineales de dos vectores. Para ello tomamos a y b de \mathbb{R} y \vec{x} e \vec{y} de \mathcal{W} . Como \mathcal{W} es la intersección, entonces \vec{x} e \vec{y} pertenecen tanto a \mathcal{V}_1 como a \mathcal{V}_2 .

- Por ser \mathcal{V}_1 un subespacio (i.e., cerrado para las comb. lineales) sabemos que $a\vec{x} + b\vec{y} \in \mathcal{V}_1$.
- Por ser \mathcal{V}_2 un subespacio (i.e., cerrado para las comb. lineales) sabemos que $a\vec{x} + b\vec{y} \in \mathcal{V}_2$.

Por tanto, si \vec{x} y \vec{y} pertenecen a \mathcal{W} , entonces $a\vec{x} + b\vec{y} \in \mathcal{V}_1 \cap \mathcal{V}_2 = \mathcal{W}$. □

Sin embargo, la unión de subespacios no suele ser cerrada para la suma (piense en la unión del conjunto de vectores de \mathbb{R}^3 cuya primera componente es cero y el conjunto de vectores de \mathbb{R}^3 cuya última componente es cero; si sumamos un vector del primer grupo con uno del segundo, generalmente obtendremos un vector en el que no son cero ni la primera, ni la última componentes (al sumar nos “salimos” del conjunto).

6.3.2. Espacios vectoriales de productos cartesianos



Y ahora una forma de construir espacios vectoriales “más grandes” a partir de otros espacios vectoriales.

El *Producto Cartesiano* de dos conjuntos A y B , que denotamos con $A \times B$, es el conjunto de pares ordenados (a, b) donde $a \in A$ y $b \in B$:

$$A \times B = \{ (a, b) \mid a \in A \text{ y } b \in B \}.$$

Si \mathcal{A} y \mathcal{B} son espacios vectoriales, definimos la suma de pares ordenados del conjunto $\mathcal{A} \times \mathcal{B}$ como el par ordenado formado por la suma de las primeras componentes y la suma de las segundas:

$$(\vec{a}, \vec{b}) + (\vec{c}, \vec{d}) = (\vec{a} + \vec{c}, \vec{b} + \vec{d}); \quad (6.1)$$

donde es importante hacer notar que el símbolo “+” de la izquierda representa la suma de pares ordenados; pero que los símbolos “+” de la derecha representan las sumas de los vectores de los respectivos espacios vectoriales. Por ejemplo, si consideramos pares del producto Cartesiano $\mathbb{R}^n \times \mathbb{R}^{p \times q}$, entonces

$$(\mathbf{x}, \mathbf{A}) + (\mathbf{y}, \mathbf{B}) = (\mathbf{x} + \mathbf{y}, \mathbf{A} + \mathbf{B})$$

evidentemente los tres tipos de sumas, lo son de objetos muy distintos.

Y definimos el producto de un escalar λ por un par ordenado como el par ordenado que resulta de multiplicar cada componente por el escalar λ

$$\lambda(\vec{a}, \vec{b}) = (\lambda\vec{a}, \lambda\vec{b}); \quad (6.2)$$

siguiendo con los productos el mismo criterio que con las sumas en la definición anterior. Por ejemplo

$$\lambda(\mathbf{x}, \mathbf{A}) = (\lambda\mathbf{x}, \lambda\mathbf{A}).$$

EJERCICIO 39. Demuestre la siguiente

Proposición 6.3.2. *El producto cartesiano $\mathcal{A} \times \mathcal{X}$ de los espacios vectoriales \mathcal{A} y \mathcal{X} , junto con la suma de pares ordenados (6.1) y el producto de un par ordenado por un escalar (6.2) es un espacio vectorial.*

La generalización a k conjuntos A, B, \dots, M del *Producto Cartesiano* (que denotamos con $A \times B \times \dots \times M$) es el conjunto de k -tuplas ordenadas (a, b, \dots, m) donde $a \in A$, $b \in B, \dots$ y $m \in M$:

$$A \times B \times \dots \times M = \{ (a, b, \dots, m) \mid a \in A, b \in B, \dots \text{ y } m \in M \}.$$

Pero alternativamente podemos considerar esta generalización como el producto cartesiano del primer conjunto con el producto cartesiano de los $n - 1$ restantes, es decir:


$$A \times B \times \dots \times M = A \times (B \times \dots \times M).$$

Ejemplo 13.

$$A \times B \times C \times D = A \times (B \times (C \times D)).$$

Empleando dicha visión alternativa podemos aprovechar la Proposición 6.3.2 para saber que el producto cartesiano $\mathcal{A} \times \mathcal{B} \times \dots \times \mathcal{X}$ de los espacios vectoriales $\mathcal{A}, \mathcal{B}, \dots, \mathcal{X}$, junto con la suma de pares ordenados (6.1) y el producto de un par ordenado por un escalar (6.2) es un espacio vectorial.

6.3.3. Espacios vectoriales de funciones

 La suma de funciones y el producto de un escalar por una función son potentes herramientas para crear nuevos espacios vectoriales. Algunos de ellos aparecerán en otras asignaturas.

Definamos la suma y el producto por escalares para funciones con el mismo dominio X y cuyas imágenes están contenidas en el mismo espacio vectorial \mathcal{V} :

Definición 6.5. Se define la suma de dos funciones $f: X \rightarrow \mathcal{V}$ y $g: X \rightarrow \mathcal{V}$ como la función

$$[f + g]: X \longrightarrow \mathcal{V} \quad \vec{x} \longmapsto f(\vec{x}) + g(\vec{x}) \quad (6.3)$$

Definición 6.6. Se define producto de un escalar α por una función $f: X \rightarrow \mathcal{V}$ como la función

$$[\alpha \cdot f]: X \longrightarrow \mathcal{V} \quad \vec{x} \longmapsto \alpha f(\vec{x}) \quad (6.4)$$

Con \mathcal{V}^D denotaremos al conjunto de funciones cuyo dominio es D y cuya imagen está contenida en \mathcal{V} :

$$\mathcal{V}^D = \{f \mid f: D \rightarrow \mathcal{V}\}. \quad (6.5)$$

EJERCICIO 40. Demuestre la siguiente

Proposición 6.3.3. *El conjunto \mathcal{V}^D de funciones cuyo dominio es D y cuya imagen está contenida en \mathcal{V} junto con la suma (6.3) y el producto por escalares (6.4), es un espacio vectorial.*

Ejemplos de este tipo de espacios que probablemente encontrará en futuras asignaturas

... aunque quizá no le indiquen explícitamente que está trabajando con espacios vectoriales:

- $\mathbb{R}^{\mathbb{R}}$: funciones reales de variable real: $f: \mathbb{R} \rightarrow \mathbb{R}$.
- $\mathbb{R}^{(\mathbb{R}^n)}$: funciones reales de varias variables reales: $f: \mathbb{R}^n \rightarrow \mathbb{R}$.
- $\mathbb{R}^{\mathbb{N}}$: sucesiones infinitas de números reales (con índice sobre los números naturales): $\mathbf{a}: \mathbb{N} \rightarrow \mathbb{R}$.
- $\mathbb{R}^{\mathbb{Z}}$: sucesiones infinitas de números reales (con índice sobre los números enteros): $\mathbf{a}: \mathbb{Z} \rightarrow \mathbb{R}$.
- \mathbb{R}^{Ω} : variables aleatorias (donde Ω es el conjunto de sucesos elementales): $X: \Omega \rightarrow \mathbb{R}$.
- $(\mathbb{R}^{\Omega})^{\mathbb{Z}}$: series temporales: $\mathbf{Y}: \mathbb{Z} \rightarrow \mathbb{R}^{\Omega}$.
- Etcetera.

6.3.4. Subespacio de funciones lineales

EJERCICIO 41. Demuestre las siguientes proposiciones (ambas demostraciones son muy similares):

- (a) **Proposición 6.3.4.** *La suma de dos funciones lineales es una función lineal.*
 (b) **Proposición 6.3.5.** *El producto de un escalar por una función lineal es una función lineal.*

Por tanto, concluimos que el conjunto de funciones lineales que van del espacio vectorial \mathcal{D} al espacio vectorial \mathcal{V} es un subespacio del espacio vectorial $\mathcal{V}^{\mathcal{D}}$ (el conjunto de funciones cuyo dominio es \mathcal{D} y cuya imagen está contenida en el espacio vectorial \mathcal{V}).

Por tanto, cerramos la lección con el siguiente

Corolario 6.3.6. *Dados \mathcal{D} y \mathcal{V} , el conjunto de funciones lineales $f: \mathcal{D} \rightarrow \mathcal{V}$ es un subespacio de $\mathcal{V}^{\mathcal{D}}$.*

(hemos citado algunos ejemplos en la Sección 6.3.1)

6.4. Generalizaciones o visiones alternativas (★)



Hay una estrecha relación entre subespacios y funciones lineales que no se suele contar.

6.4.1. Subespacios y funciones lineales

Proposición 6.4.1. *Si \mathcal{V} y \mathcal{W} son dos espacios vectoriales y $f \subset \mathcal{V} \times \mathcal{W}$ las siguientes propiedades son equivalentes:*

1. $f: \mathcal{V} \rightarrow \mathcal{W}$ es una aplicación lineal
2. f es un subespacio vectorial de $\mathcal{V} \times \mathcal{W}$ suplementario⁷ de $\{\vec{0}\} \times \mathcal{W}$.

Demostración. 1. \Rightarrow 2. Veamos en primer lugar que f es un subespacio vectorial de $\mathcal{V} \times \mathcal{W}$.

⁷ es decir $\mathcal{V} \times \mathcal{W} = f \oplus (\{\vec{0}\} \times \mathcal{W})$. (véase la Definición 10.2 en la página 123)

1. Por ser f una función cuyo dominio es \mathcal{V} y cuya imagen está contenida en \mathcal{W} , cualquier elemento de f es un par ordenado cuya primera componente pertenece a \mathcal{V} y cuya segunda componente pertenece a \mathcal{W} . Por tanto $f \subset \mathcal{V} \times \mathcal{W}$.
2. Como \mathcal{V} es un espacio vectorial $\mathcal{V} \neq \emptyset$. Luego, $\text{dom}(f) = \mathcal{V} \neq \emptyset$ y por consiguiente $f \neq \emptyset$.
3. Si (\vec{v}_1, \vec{w}_1) y (\vec{v}_2, \vec{w}_2) pertenecen a f y $\alpha_1, \alpha_2 \in \mathbb{R}$ entonces, como $f(\vec{v}_1) = \vec{w}_1$ y $f(\vec{v}_2) = \vec{w}_2$ y f es lineal, tendremos que

$$f(\alpha_1 \vec{v}_1 + \alpha_2 \vec{v}_2) = \alpha_1 f(\vec{v}_1) + \alpha_2 f(\vec{v}_2) = \alpha_1 \vec{w}_1 + \alpha_2 \vec{w}_2$$

Luego $(\alpha_1 \vec{v}_1 + \alpha_2 \vec{v}_2, \alpha_1 \vec{w}_1 + \alpha_2 \vec{w}_2) \in f$ y por consiguiente

$$\alpha(\vec{u}_1, \vec{w}_1) + \beta(\vec{u}_2, \vec{w}_2) = (\alpha_1 \vec{v}_1 + \alpha_2 \vec{v}_2, \alpha_1 \vec{w}_1 + \alpha_2 \vec{w}_2) \in f$$

Veamos en segundo lugar que f y $\{\vec{0}\} \times \mathcal{W}$ son subespacios suplementarios:

1. Si $(\vec{v}, \vec{w}) \in f \cap (\{\vec{0}\} \times \mathcal{W})$ tendremos que $\vec{w} = f(\vec{v})$ y que $\vec{v} = \vec{0}$. Luego, por ser f lineal, necesariamente $\vec{w} = \vec{0}$. Por tanto

$$f \cap (\{\vec{0}\} \times \mathcal{W}) = (\vec{0}, \vec{0})$$

2. Dado $(\vec{v}, \vec{w}) \in \mathcal{V} \times \mathcal{W}$, como $\mathcal{V} = \text{dom}(f)$, tendremos que

$$(\vec{v}, \vec{w}) = (\vec{v}, f(\vec{v})) + (\vec{0}, \vec{w} - f(\vec{v})) \in f + (\{\vec{0}\} \times \mathcal{W})$$

Por tanto $\mathcal{V} \times \mathcal{W} = f \oplus (\{\vec{0}\} \times \mathcal{W})$.

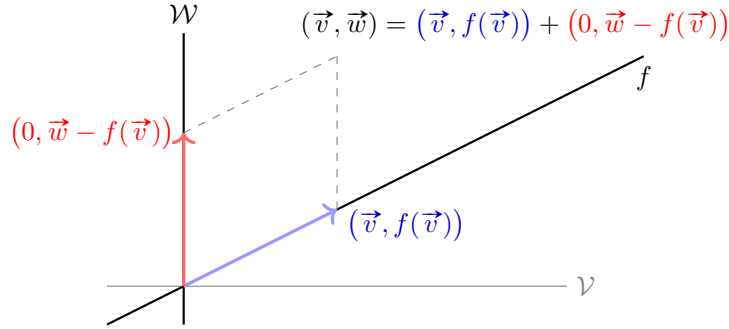


Figura 6.1: Representación esquemática de que $\mathcal{V} \times \mathcal{W} = f \oplus (\{\vec{0}\} \times \mathcal{W})$.

1. \Leftarrow 2. Veamos primero que f es una función cuyo dominio es \mathcal{V} y cuya imagen está contenida en \mathcal{W} .
 1. Como $f \subset \mathcal{V} \times \mathcal{W}$, los elementos de f son pares ordenados de primeras componentes en \mathcal{V} y de segundas en \mathcal{W} .
 2. Si (\vec{v}, \vec{w}_1) y (\vec{v}, \vec{w}_2) pertenecen a f , por ser f un subespacio vectorial de $\mathcal{V} \times \mathcal{W}$, tendremos que su diferencia $(\vec{0}, \vec{w}_1 - \vec{w}_2) \in f$. Como dicha diferencia pertenece al subespacio $\{\vec{0}\} \times \mathcal{W}$ que es suplementario de f , necesariamente⁸ $\vec{w}_1 - \vec{w}_2 = \vec{0}$. Por tanto $\vec{w}_1 = \vec{w}_2$, con lo que es imposible encontrar dos pares distintos de f cuyas primeras componentes sean coincidentes.

⁸ $f \cap (\{\vec{0}\} \times \mathcal{W}) = (\vec{0}, \vec{0})$

3. De momento ya sabemos que f es una función cuyo dominio está contenido en \mathcal{V} y cuya imagen está contenida en \mathcal{W} . Veamos que $\mathcal{V} = \text{dom}(f)$.

Si $\vec{v} \in \mathcal{V}$, como $\{\vec{0}\} \times \mathcal{W}$ y f son suplementarios, el vector $(\vec{v}, \vec{0})$ se descompone en suma de un vector (\vec{x}, \vec{y}) de f y otro $(\vec{0}, \vec{z})$ de $\{\vec{0}\} \times \mathcal{W}$.

$$(\vec{v}, \vec{0}) = (\vec{x}, \vec{y}) + (\vec{0}, \vec{z})$$

Como en dicha descomposición necesariamente $\vec{x} = \vec{v}$, concluimos que $\vec{v} \in \text{dom}(f)$.

Ya solo falta comprobar que $f : \mathcal{V} \rightarrow \mathcal{W}$ es lineal.

Si $\vec{v}_1, \vec{v}_2 \in \mathcal{V}$ y $\alpha_1, \alpha_2 \in \mathbb{R}$, como $(\vec{v}_1, f(\vec{v}_1))$ y $(\vec{v}_2, f(\vec{v}_2))$ pertenecen al subespacio f , tendremos que

$$(\alpha_1 \vec{v}_1 + \alpha_2 \vec{v}_2, \alpha_1 f(\vec{v}_1) + \alpha_2 f(\vec{v}_2)) = \alpha_1 (\vec{v}_1, f(\vec{v}_1)) + \alpha_2 (\vec{v}_2, f(\vec{v}_2)) \in f$$

Luego $f(\alpha_1 \vec{v}_1 + \alpha_2 \vec{v}_2) = \alpha_1 f(\vec{v}_1) + \alpha_2 f(\vec{v}_2)$.

□

Apendices a la lección

Se suele asumir que los estudiantes conocen las funciones, pero creo oportuno añadir el siguiente apéndice.

6.A. Funciones

Definición 6.7. Una función es un conjunto de pares ordenados en los que no existen dos pares distintos que tengan su primeras componentes iguales.

Dada una función f , se llama *dominio* de f , al conjunto de primeras componentes de los pares de f .

$$\text{dom}(f) = \{x \mid \text{existe } y \text{ tal que } (x, y) \in f\}.$$

Dada una función f , se llama *imagen* de f , al conjunto de segundas componentes de f .

$$\text{imagen}(f) = \{y \mid \text{existe } x \text{ tal que } (x, y) \in f\}.$$

Ejemplos.

Ejemplo 14. El siguiente conjunto (en el que las segundas componentes son el cuadrado de las primeras) es una función

$$f = \{(x, x) \mid x \in \mathbb{Z}\} = \{(0, 0), (1, 1), (-1, 1), (2, 4), (-2, 4), (3, 9), (-3, 9), \dots\}$$

donde \mathbb{Z} es el conjunto de números enteros.

Sin embargo, este otro conjunto (en el que las primeras componentes son el cuadrado de las segundas) NO es función:

$$\{(x^2, x) \mid x \in \mathbb{Z}\} = \{(0, 0), (1, 1), (1, -1), (4, 2), (4, -2), (9, 3), (9, -3), \dots\}$$

pues hay pares distintos que tienen la misma primera componente; por ejemplo $(1, 1)$ y $(1, -1)$.

Ejemplo 15 (Función nula). La *función nula* es aquella cuyas segundas componentes son todas nulas, es decir

$$n = \{(x, 0) \mid x \in \text{dom}(n)\} = \{(x, 0), (y, 0), (z, 0), \dots\};$$

(donde con el símbolo “0” denotamos el elemento *nulo* del conjunto que contiene la *imagen* de la función —véase un ejemplo de función nula más abajo).

Ejemplo 16 (Función identidad). La *función identidad* es aquella cuyas segundas componentes son iguales a las primeras, es decir

$$id = \{(x, x) \mid x \in \text{dom}(id)\} = \{(x, x), (y, y), (z, z), \dots\}.$$

6.A.1. Notación

Son habituales distintas formas de notación relacionadas con las funciones.

Por ejemplo, la expresión

$$f: X \rightarrow Y \quad \text{o equivalentemente} \quad X \xrightarrow{f} Y$$

es una forma abreviada de expresar lo siguiente:

1. f es una función
2. $\text{dom}(f) = X$
3. $\text{imagen}(f) \subset Y$

Por otra parte, el esquema

$$\begin{array}{l} f: X \longrightarrow Y \\ x \longmapsto \text{expresión de } x \end{array}$$

es una forma abreviada de expresar el siguiente conjunto de pares:

$$f = \{(x, \text{expresión de } x) \mid x \in X\}.$$

Así,

- el Ejemplo 14 en la página anterior se puede expresar como
$$\begin{array}{l} f: \mathbb{Z} \longrightarrow \mathbb{N} \\ x \longmapsto x^2 \end{array}$$
- la función *nula* que asigna a todo número real el vector nulo de \mathbb{R}^2 se puede expresar como

$$\begin{array}{l} n: \mathbb{R} \longrightarrow \mathbb{R}^2 \\ x \longmapsto (0, 0,) \end{array}, \quad \text{donde } \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ es el elemento } \textit{nulo} \text{ en } \mathbb{R}^2.$$

- y podemos expresar la función *identidad* en \mathbb{R}^3 como
$$\begin{array}{l} id: \mathbb{R}^3 \longrightarrow \mathbb{R}^3 \\ \mathbf{x} \longmapsto \mathbf{x} \end{array}$$

Uso de la notación funcional. Posiblemente esté más familiarizado con la siguiente notación funcional

$$f(x) = y$$

que equivale a escribir $(x, y) \in f$.

Así, el Ejemplo 14 en la página anterior se puede expresar como:

$$f(x) = x^2, \quad \text{donde } x \in \mathbb{Z};$$

pues indica que cada par (x, x^2) pertenece a f . Seguramente alguna vez haya tenido que representar gráficamente los pares (x, x^2) de esta función, y sabrá que describen una parábola.

De igual modo, expresamos la función *nula* que asigna a todo número real la matriz nula de $\mathbb{R}^{2 \times 2}$:

$$f(x) = \mathbf{0}_{2 \times 2}, \quad \text{donde } x \in \mathbb{R} \quad \text{y} \quad \mathbf{0}_{2 \times 2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Y podemos expresar la función *identidad* en \mathbb{R}^3 como:

$$f(\mathbf{x}) = \mathbf{x}, \quad \text{donde } \mathbf{x} \in \mathbb{R}^3.$$

Por tanto, empleando la **notación funcional** podemos expresar la imagen de f como:

$$\text{imagen}(f) = \{y \mid \text{existe } x \in \text{dom}(f) \text{ tal que } f(x) = y\}. \quad (6.6)$$

A veces se usa la notación $f(X)$ para indicar la imagen de $f: X \rightarrow Y$ (donde X es el dominio):

$$f(X) = \{y \mid \text{existe } x \in X \text{ tal que } f(x) = y\}.$$

6.A.2. Invertibilidad

Decimos que una función es invertible si al invertir el orden de los pares se obtiene una función. En tal caso dicha función se denota f^{-1} , y se llama *función inversa* de f .

Fíjese que de los tres ejemplos anteriores, únicamente la función identidad es invertible. En los otros dos, al invertir el orden de los pares se obtienen pares distintos pero que comparten la misma primera componente (si invertimos el orden de los pares de la función nula, todos tendrán como primera componente la matriz nula $\mathbf{0}$; y ya vimos qué pasaba al invertir los pares ordenados del Ejemplo 14 en la página 87).

Puesto que intercambiando dos veces el orden de los pares recuperamos los pares originales, tenemos que

$$(f^{-1})^{-1} = f.$$

También es evidente que $\text{dom}(f^{-1}) = \text{imagen}(f)$, y que $\text{imagen}(f^{-1}) = \text{dom}(f)$.

Proposición 6.A.1. Si f y g son dos funciones tales que

$$g(f(x)) = x \text{ para todo } x \in \text{dom}(f) \quad \text{y} \quad f(g(y)) = y \text{ para todo } y \in \text{dom}(g)$$

entonces f es invertible y su inversa es g .

Demostración. Puesto que f y g son funciones basta comprobar que $(x, y) \in f$ si y solo si $(y, x) \in g$.

Por una parte

$$(x, y) \in f \Rightarrow y = f(x) \Rightarrow g(y) = g(f(x)) = x \Rightarrow (y, x) \in g$$

y por otra

$$(y, x) \in g \Rightarrow g(y) = x \Rightarrow f(g(y)) = f(x) = y \Rightarrow (x, y) \in f$$

□

6.A.3. Composición de funciones

Si tenemos dos funciones f y g tales que el dominio de g contiene a la imagen de f , llamamos *composición* de f y g al conjunto de pares

$$g \circ f = \{(x, z) \mid \text{existe } z \text{ de modo que } (x, y) \in f \text{ y } (y, z) \in g\}.$$

La composición de funciones se escribe usando la notación funcional como

$$[g \circ f](x) = g(f(x)).$$

pues $z = [g \circ f](x)$ siempre que $y = f(x)$ y $z = g(y)$.

Así, resulta evidente que si f es invertible entonces

$$[f^{-1} \circ f](x) = f^{-1}(f(x)) = x, \quad \text{para todo } x \in \text{dom}(f).$$

Igualmente evidente es que

$$[f \circ f^{-1}](y) = f(f^{-1}(y)) = y, \quad \text{para todo } y \in \text{dom}(f^{-1}).$$

6.A.4. Los sistemas (listas ordenadas) son funciones



Iniciamos la Lección 1 haciendo uso del concepto “intuitivo” de *lista ordenada*. Para poder hablar con propiedad de conceptos tales como *el conjunto de vectores de \mathbb{R}^3* o *el conjunto de matrices de orden n* necesitamos contar con una definición formal de *sistema* (lista ordenada).

Un *sistema* (o lista ordenada) de n elementos es una función cuyo dominio es el conjunto de números naturales menores o iguales a n y mayores que cero, es decir, cuyo dominio es el conjunto de índices $\{i \mid 1 \leq i \leq n\}$.

Así, el sistema de tres elementos $[\mathbf{a}; 5; \mathbf{B}]$ es la función con dominio $\{1, 2, 3\}$ consistente en el conjunto de pares:

$$\{(1, \mathbf{a}), (2, 5), (3, \mathbf{B})\};$$

y el vector $(0, -2, \pi,)$ es la función (el conjunto de pares):

$$\{(1, 0), (2, -2), (3, \pi)\}.$$

Visto así, \mathbb{R}^3 es el conjunto de funciones de la forma $\mathbf{v}: \{1, 2, 3\} \rightarrow \mathbb{R}$, es decir, el conjunto de funciones cuyo dominio es $\{1, 2, 3\}$ y cuya imagen está contenida en \mathbb{R} . De hecho el símbolo \mathbb{R}^3 es (siguiendo la notación de la ecuación 6.5 en la página 85) una forma abreviada de escribir $\mathbb{R}^{\{1,2,3\}}$, donde

$$\mathbb{R}^{\{1,2,3\}} = \left\{ \mathbf{v} \mid \mathbf{v}: \{1, 2, 3\} \rightarrow \mathbb{R} \right\}.$$

Listas vacías. Una lista vacía (un sistema vacío) es cualquier función $[\]: \emptyset \rightarrow X$ cuyo dominio es el conjunto vacío \emptyset ; consecuentemente dicha función es *el conjunto vacío*, pues no contiene ningún par:⁹

$$[\] \equiv \emptyset. \tag{6.7}$$

⁹esta equivalencia matemática no se da en el lenguaje de programación Python puesto que la implementación informática de conjuntos y listas es diferente, así que en Python un objeto lista (`list`) y un objeto conjunto (`set`) nunca pueden ser lo mismo.

Resolviendo $Ax = 0$

7.1. Sistemas de ecuaciones lineales



La Parte III del curso trata sobre sistemas de ecuaciones lineales. Como veremos enseguida, los sistemas de ecuaciones lineales están íntimamente relacionados con los subespacios de \mathbb{R}^n .

Una advertencia. Quizá conozca algún método de resolución de sistemas. Mi experiencia es que generalmente los alumnos saben ejecutar unos pasos a ciegas para obtener una solución, pero sin entender bien el método... y con alguna frecuencia no son capaces de ejecutar el método correctamente.

El método que veremos es deliberadamente distinto del que se cuenta habitualmente. Con ello pretendo, entre otras cosas, que el estudiante no aplique ciegamente una batería de recetas (mal aprendidas en muchos casos). Así pues, *por el momento olvide lo que sabe y trate de entender el método que expondré aquí.*

Por suerte el método es muy sencillo, aplicable a otros problemas que veremos más adelante... y para mayor fortuna ¡usted ya lo ha empleado en la Lección 5 para invertir matrices!

Llamamos *ecuación lineal* a aquella que se puede escribir de la siguiente forma

$$a_1x_1 + \cdots + a_nx_n = b,$$

donde b y los *coeficientes* a_1, \dots, a_n denotan números fijos y x_1, \dots, x_n son las variables, es decir, etiquetas para ser reemplazadas por números.

Llamamos *solución* a los valores que, reemplazando a las variables, hacen cierta la igualdad.

Sistemas de ecuaciones lineales Se llama sistema de ecuaciones lineales a una colección de ecuaciones en la que cada variable es reemplazada por idéntico valor en todas las ecuaciones donde aparece. Por ejemplo,

$$\begin{cases} x + y - 2z = 1 \\ x - 2y + z = 2 \\ -2x + y + z = 3 \end{cases}$$

es un sistema de tres ecuaciones con tres variables (o incógnitas) x , y y z . La llave indica que cada variable debe ser reemplazada por el mismo valor en cada una de las tres ecuaciones, es decir, si asignamos el valor $x = 5$, lo hacemos para las tres ecuaciones a la vez. No hay un orden ni en la disposición de las ecuaciones ni en las sumas dentro de cada ecuación; de manera que el anterior sistema también lo podemos escribir como

$$\begin{cases} x + z - 2y = 2 \\ y + x - 2z = 1 \\ z + y - 2x = 3 \end{cases}$$

pero esta libertad en la notación tradicional no ayuda a trabajar con ellos.

7.1.1. Notación matricial de los sistemas de ecuaciones lineales

Aquí emplearemos la notación matricial en lugar de la tradicional. Ello nos permitirá aprovechar fácilmente toda la potencia de los conceptos de espacio vectorial.

El primer paso para usar la notación matricial requiere tomar una decisión inicial y arbitraria. Hemos de establecer quién es la primera variable, quién la segunda y quién la tercera; y también decidir qué ecuación será la primera, cuál la segunda y cuál la tercera. La decisión que tomemos no es importante, pero una vez tomada, estamos obligados a mantenerla.

Volviendo al ejemplo anterior, podemos decidir que el orden de las variables es, primero x , luego y y por último z ; así definimos el *vector de incógnitas*: $\mathbf{x} = (x, y, z)$. También podemos establecer que las ecuaciones serán ordenadas tal como aparecen dentro de la llave en la primera versión del ejemplo. Siguiendo este criterio, generamos una *matriz de coeficientes* cuya primera fila contiene los coeficientes de la primera ecuación (y en el orden que hayamos establecido para las variables), cuya segunda fila contiene los coeficientes de la segunda ecuación (respetando el mismo orden de las variables), etc.

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & -2 \\ 1 & -2 & 1 \\ -2 & 1 & 2 \end{bmatrix}.$$


El orden de las ecuaciones también determina el orden de las componentes del vector que contiene los números que aparecen a la derecha de cada ecuación. En el ejemplo anterior será $\mathbf{b} = (1, 2, 3)$. A este vector lo denominamos *vector del lado derecho*.

Así, con la siguiente ecuación matricial:

$$\mathbf{Ax} = \mathbf{b},$$

denotamos de manera compacta un sistema de ecuaciones, y donde la variable \mathbf{x} es una etiqueta para ser reemplazada por un vector. Así, con

$$\begin{bmatrix} 1 & 1 & -2 \\ 1 & -2 & 1 \\ -2 & 1 & 2 \end{bmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \quad \text{denotamos} \quad \begin{cases} x + y - 2z = 1 \\ x - 2y + z = 2 \\ -2x + y + z = 3 \end{cases}.$$

 La ecuación matricial $\mathbf{Ax} = \mathbf{b}$ define un *sistema de ecuaciones lineales*.

Decimos que el vector \mathbf{z} es una *solución* del sistema de ecuaciones $\mathbf{Ax} = \mathbf{b}$ si reemplazar \mathbf{x} por \mathbf{z} hace cierta la igualdad; y llamamos “*resolver*” un sistema $\mathbf{Ax} = \mathbf{b}$ a encontrar el *conjunto* de sus soluciones.

7.2. Sistemas de ecuaciones lineales homogéneos

Cuando el vector del lado derecho es un vector nulo el sistema de ecuaciones se dice que es *homogéneo*. En esta Lección 7 solo nos ocuparemos de los sistemas de ecuaciones que son homogéneos:

$$\mathbf{Ax} = \mathbf{0}.$$

Fíjese que el lado izquierdo de un sistema de ecuaciones está formado por el producto \mathbf{Ax} , es decir, *el lado izquierdo es una combinación lineal* de las columnas de la matriz de coeficientes. Por tanto, el sistema de más arriba nos está preguntando ¿qué combinaciones lineales de las columnas de \mathbf{A} son vectores nulos? Evidentemente si reemplazamos el vector de incógnitas \mathbf{x} por $\mathbf{0}$, tenemos que

$$\mathbf{A}\mathbf{0} = \mathbf{0}.$$

Pero aparte de la solución trivial $\mathbf{x} = \mathbf{0}$, ¿existen otros vectores $\mathbf{x} \neq \mathbf{0}$ para los que la combinación lineal \mathbf{Ax} es el vector nulo? Veamos un par de ejemplos. Imagine que la matriz de coeficientes es la matriz identidad de orden 3, entonces tenemos el sistema $\mathbf{Ix} = \mathbf{0}$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Puesto que el lado izquierdo del sistema es

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix},$$

y tiene que ser igual al vector del lado derecho (que es $\mathbf{0}$); no hay más solución que aquella en la que todas las componentes de \mathbf{x} son nulas (es decir, la *solución trivial* $\mathbf{x} = \mathbf{0}$).

Sin embargo es posible encontrar soluciones no triviales para el siguiente sistema $\mathbf{Ax} = \mathbf{b}$:

$$\begin{bmatrix} 1 & 1 & -2 \\ 1 & -2 & 1 \\ -2 & 1 & 2 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix};$$

por ejemplo $\mathbf{x} = (2, 2, 2)$. Piense un poco y seguro que es capaz de encontrar muchísimas más.

7.3. Espacio nulo de una matriz $\mathcal{N}(\mathbf{A})$

Los ejemplos anteriores muestran que el conjunto de soluciones de algunos sistemas de ecuaciones lineales homogéneos *únicamente* contiene el vector nulo $\mathbf{0}$ (la *solución trivial*), pero que para otros sistemas el conjunto de soluciones *también contiene* vectores distintos de $\mathbf{0}$.

Vamos a dar un nombre corto al “conjunto de soluciones del un sistema homogéneo $\mathbf{Ax} = \mathbf{0}$ ”:

Definición 7.1. Sea \mathbf{A} de orden $m \times n$. Denominamos **espacio nulo de \mathbf{A}** (que denotamos con $\mathcal{N}(\mathbf{A})$) al subconjunto de vectores de \mathbb{R}^n que son solución del sistema $\mathbf{Ax} = \mathbf{0}$, es decir ¹

$$\mathcal{N}(\mathbf{A}) = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{Ax} = \mathbf{0}\}.$$

En el primer ejemplo de la página anterior vimos que el *espacio nulo de \mathbf{I}* solo contiene el vector $\mathbf{0}$, es decir, $\mathcal{N}(\mathbf{I}) = \{\mathbf{0}\}$. En el segundo ejemplo, $\mathcal{N}(\mathbf{A})$, contiene infinitos vectores (además del $\mathbf{0}$).

Ahora bien, en todos los casos el *espacio nulo* de una matriz de n columnas es un subespacio de \mathbb{R}^n .

EJERCICIO 42. Demuestre la siguiente

Proposición 7.3.1. Para cualquier \mathbf{A} de orden $m \times n$, el espacio nulo $\mathcal{N}(\mathbf{A})$ es un subespacio de \mathbb{R}^n .

Pista. Basta comprobar que el conjunto es cerrado para las combinaciones lineales.

Es importante subrayar que aunque el conjunto de soluciones de cualquier sistema homogéneo es un subespacio, *el conjunto de soluciones de un sistema NO homogéneo $\mathbf{Ax} = \mathbf{b}$ (con $\mathbf{b} \neq \mathbf{0}$) nunca es un subespacio*, pues el vector nulo $\mathbf{0}$ no pertenece al conjunto de soluciones.

¹Esta forma de caracterizar los elementos de un conjunto como soluciones de un sistema de ecuaciones se denomina *ecuación cartesiana*.

7.4. Resolución de un Sistema Homogéneo por eliminación

Puesto que $\mathcal{N}(\mathbf{A})$ es un subespacio, sabemos que si conocemos algunas soluciones del sistema $\mathbf{A}\mathbf{x} = \mathbf{0}$, también son solución sus combinaciones lineales, pero ¿cómo encontrar todas las soluciones? es decir ¿cómo calcular $\mathcal{N}(\mathbf{A})$?

Revisitando el método de eliminación. Vimos en la Sección 5.1.4 que aplicando una secuencia de transformaciones elementales $\tau_1 \cdots \tau_k$ a las columnas de la matriz que resulta de apilar \mathbf{A} con \mathbf{I} (con idéntico número de columnas) tenemos:

$$\begin{bmatrix} \mathbf{A} \\ \mathbf{I} \end{bmatrix}_{\tau_1 \cdots \tau_k} = \begin{bmatrix} \mathbf{A}_{\tau_1 \cdots \tau_k} \\ \mathbf{I}_{\tau_1 \cdots \tau_k} \end{bmatrix} = \begin{bmatrix} \mathbf{A}(\mathbf{I}_{\tau_1 \cdots \tau_k}) \\ \mathbf{I}_{\tau_1 \cdots \tau_k} \end{bmatrix} = \begin{bmatrix} \mathbf{AE} \\ \mathbf{E} \end{bmatrix}, \quad \text{donde } \mathbf{E} = \mathbf{I}_{\tau_1 \cdots \tau_k}.$$

Ahora fijémonos en la columna j -ésima de la matriz resultante

$$\begin{bmatrix} \mathbf{AE} \\ \mathbf{E} \end{bmatrix}_{|j} = \begin{pmatrix} \mathbf{A}(\mathbf{E}_{|j}) \\ \mathbf{E}_{|j} \end{pmatrix} \quad (\text{que es la concatenación } (\mathbf{A}(\mathbf{E}_{|j})) \# \mathbf{E}_{|j});$$

es decir, para cada columna tenemos que el subvector de la parte inferior, $\mathbf{E}_{|j}$, contiene los coeficientes (“la receta”) de la combinación lineal de las columnas de \mathbf{A} que usamos para calcular (“cocinar”) la parte superior de la columna, $\mathbf{AE}_{|j}$.

Ejemplo 17. Apliquemos la eliminación (de izquierda a derecha) tras apilar la matriz

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

con la matriz identidad de orden 4. Fíjese cómo la parte inferior de cada columna, de cada matriz en cada uno de los pasos de la eliminación, indica cómo calcular la parte superior de esa misma columna:

$$\begin{bmatrix} \mathbf{A} \\ \mathbf{I} \end{bmatrix} = \begin{bmatrix} 1 & 2 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 2 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow[\begin{smallmatrix} [(-2)1+2] \\ [(-1)1+4] \end{smallmatrix}]{\tau} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & -2 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{[(-1)2+3]} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & -2 & 2 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{\tau_1 \tau_2 \tau_3} \\ \mathbf{I}_{\tau_1 \tau_2 \tau_3} \end{bmatrix}$$

Por ejemplo, la 1ª columna de la 1ª matriz, la 2ª columna de la 2ª matriz, y la 3ª columna de la 3ª matriz, verifican respectivamente que:

$$\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \mathbf{A} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}; \quad \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \mathbf{A} \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \end{pmatrix}; \quad \text{ó} \quad \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \mathbf{A} \begin{pmatrix} 2 \\ -1 \\ 1 \\ 0 \end{pmatrix}.$$

Y no puede ser de otra manera, ya que la relación $(\mathbf{A}_{\tau_1 \cdots \tau_k})_{|j} = \mathbf{A}(\mathbf{I}_{\tau_1 \cdots \tau_k})_{|j}$ se cumple en todas y cada una de las columnas, de todas y cada una de las matrices en todos y cada uno de los pasos de eliminación.



En todo momento a lo largo de la secuencia de transformaciones

$$\begin{bmatrix} \mathbf{A} \\ \mathbf{I} \end{bmatrix} \rightarrow \begin{bmatrix} \mathbf{A} \\ \mathbf{I} \end{bmatrix}_{\tau_1} \rightarrow \dots \rightarrow \begin{bmatrix} \mathbf{A} \\ \mathbf{I} \end{bmatrix}_{\tau_1 \dots \tau_k} = \begin{bmatrix} \mathbf{A}_{\tau_1 \dots \tau_k} \\ \mathbf{I}_{\tau_1 \dots \tau_k} \end{bmatrix}.$$

la parte inferior de cada columna indica qué combinación lineal de las columnas de \mathbf{A} hemos usado para calcular la parte superior de la columna: $(\mathbf{A}_{\tau_1 \dots \tau_k})_{|j} = \mathbf{A}(\mathbf{I}_{\tau_1 \dots \tau_k})_{|j}; \quad h = 1 : k.$

Por el Teorema 4.2.2 sabemos que para cualquier \mathbf{A} existen k transformaciones elementales tales que $\mathbf{A}_{\tau_1 \dots \tau_k} = \mathbf{K}$ es pre-escalónada. Por tanto

$$\begin{bmatrix} \mathbf{A} \\ \mathbf{I} \end{bmatrix}_{\tau_1 \dots \tau_k} = \begin{bmatrix} \mathbf{A}_{\tau_1 \dots \tau_k} \\ \mathbf{I}_{\tau_1 \dots \tau_k} \end{bmatrix} = \begin{bmatrix} \mathbf{A}(\mathbf{I}_{\tau_1 \dots \tau_k}) \\ \mathbf{I}_{\tau_1 \dots \tau_k} \end{bmatrix} = \begin{bmatrix} \mathbf{A}\mathbf{E} \\ \mathbf{E} \end{bmatrix} = \begin{bmatrix} \mathbf{K} \\ \mathbf{E} \end{bmatrix}; \quad \text{donde } \mathbf{E} = \mathbf{I}_{\tau_1 \dots \tau_k}.$$

Solo caben dos posibilidades, que \mathbf{K} tenga columnas nulas o que no tenga:

Cuando \mathbf{K} tiene columnas nulas: si la columna $\mathbf{K}_{|j}$ es nula sabemos que

$$\mathbf{K}_{|j} = \mathbf{A}(\mathbf{E}_{|j}) = \mathbf{0} \quad \text{y por lo tanto} \quad \mathbf{E}_{|j} \in \mathcal{N}(\mathbf{A}).$$

Es decir, que si \mathbf{K} tiene columnas nulas, los vectores que aparecen por debajo de las columnas nulas son *soluciones al sistema* $\mathbf{A}\mathbf{x} = \mathbf{0}$ (así como todas las combinaciones lineales de dichos vectores).

Así pues, buscar soluciones no nulas del sistema $\mathbf{A}\mathbf{x} = \mathbf{0}$ es tan sencillo como alargar \mathbf{A} poniendo la matriz identidad por debajo, y aplicar transformaciones elementales a las columnas de la matriz por bloques hasta pre-escalónar el bloque superior. Si la forma pre-escalónada \mathbf{K} tiene columnas nulas, los vectores que hay debajo son soluciones no nulas de $\mathbf{A}\mathbf{x} = \mathbf{0}$. Siguiendo la nomenclatura de G. Strang, llamaremos *soluciones especiales* a estas soluciones encontradas por eliminación.

Volviendo al Ejemplo 17 en la página anterior, sabemos que los vectores

$$\begin{pmatrix} 2 \\ -1 \\ 1 \\ 0 \end{pmatrix} \quad \text{y} \quad \begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

son soluciones *especiales* del sistema $\mathbf{A}\mathbf{x} = \mathbf{0}$. Y también sabemos que es solución cualquier combinación lineal de ambos vectores. Por ejemplo, si sumamos ambos vectores obtenemos una nueva solución

$$\begin{bmatrix} 1 & 2 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 2 & 0 & 1 \end{bmatrix} \left(\begin{pmatrix} 2 \\ -1 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right) = \begin{bmatrix} 1 & 2 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 2 & 0 & 1 \end{bmatrix} \begin{pmatrix} 1 \\ -1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Cuando \mathbf{K} no tiene columnas nulas: cuando \mathbf{K} es de rango completo por columnas (i.e., no tiene columnas nulas) podemos continuar con la eliminación Gauss-Jordan hasta alcanzar una forma escalónada reducida $\mathbf{A}\mathbf{E} = \mathbf{R}$. Como cada columna tiene un pivote y cada pivote es el único elemento no nulo de su fila, es evidente que $\mathbf{R}\mathbf{x} = \mathbf{0}$ si y solo si $\mathbf{x} = \mathbf{0}$. Así pues tenemos que

$$\mathbf{A}\mathbf{x} = \mathbf{0} \quad \Rightarrow \quad \mathbf{A}\mathbf{E}\mathbf{E}^{-1}\mathbf{x} = \mathbf{0} \quad \Rightarrow \quad \mathbf{R}(\mathbf{E}^{-1}\mathbf{x}) = \mathbf{0} \quad \Rightarrow \quad \mathbf{E}^{-1}\mathbf{x} = \mathbf{0} \quad \Rightarrow \quad \mathbf{E}\mathbf{E}^{-1}\mathbf{x} = \mathbf{E}\mathbf{0} \quad \Rightarrow \quad \mathbf{x} = \mathbf{0}.$$

Así que la única solución es $\mathbf{x} = \mathbf{0}$ (la solución trivial).

Y ahora nos puede surgir una duda, ¿existen otras soluciones que no sean combinación lineal de las *soluciones especiales* que hemos encontrado mediante la eliminación?

La Proposición 7.4.3 demuestra que no; es decir, que *todas las soluciones son combinación lineal de las soluciones especiales*.

Pero antes necesitamos demostrar un resultado previo:

EJERCICIO 43. Demuestre el siguiente

Lema 7.4.1. Si \mathbf{K} es pre-escalada, entonces $\mathbf{K}\mathbf{x} = \mathbf{0}$ si y solo si son nulos los coeficientes x_j correspondientes a las columnas no nulas de \mathbf{K} .

Veamos ahora el resultado fundamental de esta lección...

Proposición 7.4.2. Si \mathbf{A} tiene soluciones especiales, todo vector de $\mathcal{N}(\mathbf{A})$ es combinación lineal de ellas.

Demostración. Sea $\mathbf{E} = \mathbf{I}_{\tau_1 \dots \tau_k}$ tal que $\mathbf{AE} = \mathbf{K}$ es pre-escalada (con n columnas). Puesto que \mathbf{E} es invertible, para cada $\mathbf{x} \in \mathcal{N}(\mathbf{A})$ existe un único vector \mathbf{y} tal que

$$\mathbf{E}\mathbf{y} = \mathbf{x};$$

concretamente el vector $\mathbf{y} = (\mathbf{E}^{-1})\mathbf{x}$. Por tanto, como $\mathbf{Ax} = \mathbf{0}$, entonces $\mathbf{AEy} = \mathbf{Ky} = \mathbf{0}$.

Sea $\boldsymbol{\rho} = (\rho_1, \dots, \rho_r)$ la lista de índices de las r columnas no nulas de \mathbf{K} (de sus r columnas con pivote). Por el Lema 7.4.1 sabemos que la condición necesaria y suficiente para que $\mathbf{Ky} = \mathbf{0}$ es que las componentes del vector \mathbf{y} que multiplican a las columnas con pivote sean nulas, es decir², $\boldsymbol{\rho}|\mathbf{y} = \mathbf{0}$.

Si $\boldsymbol{\nu} \in \mathbb{R}^{n-r}$ es lista de índices de las columnas nulas de \mathbf{K} , entonces cada índice $j = 1 : n$ está en una y solo una de las dos listas: $\boldsymbol{\rho}$ ó $\boldsymbol{\nu}$. Así, usando la Ecuación 3.7 en la página 40 tenemos que para cualquier \mathbf{x} de $\mathcal{N}(\mathbf{A})$

$$\mathbf{x} = \mathbf{Ey} = (\mathbf{E}_{|\boldsymbol{\rho}})(\boldsymbol{\rho}|\mathbf{y}) + (\mathbf{E}_{|\boldsymbol{\nu}})(\boldsymbol{\nu}|\mathbf{y}) = (\mathbf{E}_{|\boldsymbol{\nu}})(\boldsymbol{\nu}|\mathbf{y})$$

por ser nulo el vector $\boldsymbol{\rho}|\mathbf{y}$. Como $\mathbf{E}_{|\boldsymbol{\nu}}$ es una matriz (la submatriz formada por las soluciones especiales) y como $\boldsymbol{\nu}|\mathbf{y}$ es un vector, todo vector \mathbf{x} de $\mathcal{N}(\mathbf{A})$ es combinación lineal de las soluciones especiales. \square

Por tanto tenemos el siguiente

Corolario 7.4.3. O todas las soluciones de $\mathbf{Ax} = \mathbf{0}$ son combinaciones lineales de las soluciones especiales, o bien, la forma pre-escalada de \mathbf{A} no tiene columnas nulas y la única solución es $\mathbf{0}$, es decir:

$$\mathcal{N}(\mathbf{A}) = \{\text{combinaciones lineales de las soluciones especiales}\} \quad \text{ó} \quad \mathcal{N}(\mathbf{A}) = \{\mathbf{0}\}.$$

Ahora ya podemos terminar el Ejemplo 17 en la página 94 concluyendo que el espacio nulo de \mathbf{A} (es decir, el conjunto de soluciones del sistema $\mathbf{Ax} = \mathbf{0}$) es:

$$\mathcal{N}(\mathbf{A}) = \left\{ \mathbf{v} \in \mathbb{R}^4 \left| \exists \mathbf{p} \in \mathbb{R}^2, \mathbf{v} = \begin{bmatrix} 2 & -1 \\ -1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{p} \right. \right\}.$$

²Repase la Sección 3.B.3 en la página 40

Es decir, $\mathcal{N}(\mathbf{A})$ es el subconjunto de vectores \mathbf{v} de \mathbb{R}^4 que verifican que, para cada \mathbf{v} , existen dos números reales $\mathbf{p}_{|1}$ y $\mathbf{p}_{|2}$ tales que $\mathbf{p}_{|1}$ veces la primera solución especial más $\mathbf{p}_{|2}$ veces la segunda es igual a \mathbf{v} .³

Nomenclatura usada por G. Strang Siguiendo la nomenclatura de G. Strang, a las columnas de \mathbf{A} que acaban siendo nulas en la forma pre-escalada \mathbf{K} se denominarán *columnas libres* y el resto (i.e. las que mantienen un pivote) se llamarán *columnas pivote*.⁴

Las variables (o incógnitas) correspondientes a columnas pivote, se denominarán *variables pivote* (o variables endógenas), el resto se denominarán *variables libres* (o variables exógenas); y las columnas que quedan debajo de las columnas nulas de \mathbf{K} se denominarán *soluciones especiales* (realmente no son especiales, pero así es como las llama G. Strang).



Algoritmo para resolver $\mathbf{Ax} = \mathbf{0}$

1. Aplicamos la eliminación: $\begin{bmatrix} \mathbf{A} \\ \mathbf{I} \end{bmatrix} \xrightarrow{\tau_1, \dots, \tau_k} \begin{bmatrix} \mathbf{K} \\ \mathbf{E} \end{bmatrix}$, donde $\mathbf{K} = \mathbf{A}_{\tau_1 \dots \tau_k}$ y $\mathbf{E} = \mathbf{I}_{\tau_1 \dots \tau_k}$.
2. Si hay *soluciones especiales* (columnas de \mathbf{E} bajo las columnas nulas de \mathbf{K})
 - Solución completa: $\mathcal{N}(\mathbf{A}) = \{\text{combinaciones lineales de las } \textit{soluciones especiales}\}$
3. Si no hay *soluciones especiales* (si \mathbf{K} no tiene columnas nulas)
 - Solución completa: $\mathcal{N}(\mathbf{A}) = \{\mathbf{0}\}$

```
A = Matrix( [ [1,2,0,1], [0,1,1,0], [1,2,0,1] ] )
H = Homogenea(A, 1)
```

Librería NAcAL para Python

Resuelve la ecuación homogenea

$$\begin{bmatrix} 1 & 2 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 2 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{\begin{matrix} [(-2)1+2] \\ [(-1)1+4] \end{matrix}} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & -2 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{[(-1)2+3]} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & -2 & 2 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

³ Esta forma de caracterizar los elementos \mathbf{v} de un subconjunto como aquellos que para los que es posible encontrar parámetros \mathbf{p} que permitan expresar cada \mathbf{v} como una combinación lineal se denomina *ecuación paramétrica*.

Fíjese que con las *ecuaciones paramétricas* es muy fácil obtener elementos del conjunto, basta elegir dos valores p_1 y p_2 cualesquiera para obtener una solución. Por el contrario, las *ecuaciones cartesianas* (Definición 7.1) nos sirven para verificar si un vector \mathbf{x} pertenece o no al conjunto, basta verificar que \mathbf{Ax} es cero. ¡Cada tipo de ecuación sirve para una cosa distinta, unas para generar ejemplos, y las otras para verificar la pertenencia al conjunto!

⁴ Nótese que la clasificación indicada por esta nomenclatura depende de la secuencia de transformaciones $\tau_1 \dots \tau_k$ empleada pues, en general, una columna acabará siendo nula o no dependiendo de dicha secuencia.

Resolviendo $\mathbf{Ax} = \mathbf{b}$

En esta lección veremos que el método de eliminación también nos permite deducir si es resoluble $\mathbf{Ax} = \mathbf{b}$ y, cuando lo es, encontrar el conjunto de vectores \mathbf{x} tales que \mathbf{Ax} es igual a \mathbf{b} .

8.1. Eliminación sobre la matriz ampliada

Considere el sistema *NO* homogéneo $\mathbf{Ax} = \mathbf{b}$ (es decir, con $\mathbf{b} \neq \mathbf{0}$). Podemos re-escribir el sistema para que tenga una apariencia de *sistema homogéneo*:

$$\mathbf{Ax} = \mathbf{b} \iff \mathbf{Ax} - \mathbf{b} = \mathbf{0} \iff [\mathbf{A} \mid -\mathbf{b}] \begin{pmatrix} \mathbf{x} \\ 1 \end{pmatrix} = \mathbf{0}. \quad (8.1)$$

El nuevo vector de incógnitas, $(\mathbf{x}|1) = (x_1, \dots, x_n, 1)$, es la concatenación del vector \mathbf{x} con el vector (1) . Por tanto el nuevo vector de incógnitas tiene como restricción que *su última componente debe ser un 1*. La matriz $[\mathbf{A} \mid -\mathbf{b}]$, que es la matriz que obtenemos al *añadir* el vector opuesto de \mathbf{b} a la lista de columnas de \mathbf{A} , se denomina *matriz de coeficientes “ampliada”*. Dibujamos una línea vertical para recordar que la parte izquierda de la matriz corresponde a la *matriz de coeficientes* del sistema $\mathbf{Ax} = \mathbf{b}$ y que la columna de la derecha es el *opuesto* del *vector del lado derecho*... ¡fíjese en el cambio de signo!

El método de resolución que vamos a ver intenta resolver el “pseudo-sistema homogéneo” (8.1) buscando soluciones que tengan un 1 en su última componente (es decir, *trataremos de encontrar algún vector de $\mathcal{N}([\mathbf{A} \mid -\mathbf{b}])$ que tenga un 1 en la última componente*).

☞ Las implicaciones “si y solo si” de más arriba indican que encontrar soluciones de $\mathbf{Ax} = \mathbf{b}$ es equivalente a encontrar *vectores del espacio nulo de la matriz ampliada que tengan un 1 como última componente*.

La mecánica consiste en aplicar el método de eliminación como en la lección anterior, pero usando ahora la *matriz de coeficientes “ampliada”*. Si logramos anular la última columna de la matriz de coeficientes “ampliada”, entonces el conjunto de soluciones del sistema no es vacío. ¡Si además logramos mantener el 1 de la última componente de la solución especial correspondiente a la última columna, entonces tendremos a la vista una solución de $\mathbf{Ax} = \mathbf{b}$!

Para resolver por eliminación el sistema $\mathbf{Ax} = \mathbf{d}$ (de n incógnitas), “apilamos” la matriz ampliada $[\mathbf{A} \mid -\mathbf{b}]$, de $(n+1)$ columnas, con una matriz identidad de orden $(n+1)$ por debajo:

$$\left[\begin{array}{c|c} \mathbf{A} & -\mathbf{d} \\ \hline \mathbf{I} & \\ \hline \end{array} \right] \quad \text{donde} \quad \mathbf{I}_{(n+1) \times (n+1)} = \left[\begin{array}{c|c} \mathbf{I} & \\ \hline \mathbf{0} & 1 \\ \hline \end{array} \right];$$

y distinguimos los bloques correspondientes a las distintas partes con líneas horizontales y verticales. Después pre-escalamos la matriz como hicimos en la lección anterior con los sistemas homogéneos.

Probemos con el siguiente ejemplo de sistema de ecuaciones “sin solución”:

Ejemplo 18. Considere el sistema $\mathbf{Ax} = \mathbf{d}$ con $\mathbf{A} = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 2 & 4 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{bmatrix}$ y $\mathbf{d} = \begin{pmatrix} 5 \\ -1 \\ 1 \\ -1 \end{pmatrix}$.

$$\left[\begin{array}{ccc|c} \mathbf{A} & -\mathbf{d} \\ \mathbf{I} & \\ & 1 \end{array} \right] = \left[\begin{array}{ccc|c} 1 & 1 & 2 & -5 \\ 2 & 2 & 4 & 1 \\ 3 & 1 & 4 & -1 \\ 4 & 1 & 5 & 1 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 1 \end{array} \right] \xrightarrow{\begin{array}{l} \tau \\ [(-1)1+2] \\ [(-2)1+3] \\ [(5)1+4] \end{array}} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 2 & 0 & 0 & 11 \\ 3 & -2 & -2 & 14 \\ 4 & -3 & -3 & 21 \\ \hline 1 & -1 & -2 & 5 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 1 \end{array} \right] \xrightarrow{\begin{array}{l} \tau \\ [(-1)2+3] \\ [(7)2+4] \end{array}} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 2 & 0 & 0 & 11 \\ 3 & -2 & 0 & 0 \\ 4 & -3 & 0 & 0 \\ \hline 1 & -1 & -1 & -2 \\ 0 & 1 & -1 & 7 \\ 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 1 \end{array} \right]$$

Como no encontramos un vector de $\mathcal{N}([\mathbf{A}|\mathbf{b}])$ cuya última componente sea un **1**, el conjunto de soluciones de $\mathbf{Ax} = \mathbf{d}$ es vacío (entonces se dice que el sistema es incompatible, o que **NO tiene solución**).

Ejemplo 19. Ahora considere el sistema $\mathbf{Ax} = \mathbf{b}$ con $\mathbf{A} = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 2 & 4 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{bmatrix}$ y $\mathbf{b} = \begin{pmatrix} 5 \\ 10 \\ 1 \\ -1 \end{pmatrix}$. Aplicando el método de eliminación sobre la matriz ampliada tenemos

$$\left[\begin{array}{ccc|c} 1 & 1 & 2 & -5 \\ 2 & 2 & 4 & -10 \\ 3 & 1 & 4 & -1 \\ 4 & 1 & 5 & 1 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 1 \end{array} \right] \xrightarrow{\begin{array}{l} \tau \\ [(-1)1+2] \\ [(-2)1+3] \\ [(5)1+4] \end{array}} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ 3 & -2 & -2 & 14 \\ 4 & -3 & -3 & 21 \\ \hline 1 & -1 & -2 & 5 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 1 \end{array} \right] \xrightarrow{\begin{array}{l} \tau \\ [(-1)2+3] \\ [(7)2+4] \end{array}} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ 3 & -2 & 0 & 0 \\ 4 & -3 & 0 & 0 \\ \hline 1 & -1 & -1 & -2 \\ 0 & 1 & -1 & 7 \\ 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 1 \end{array} \right]$$

Hemos encontrado dos soluciones especiales (los vectores por debajo de las dos columnas nulas), de las cuales la segunda tiene un **1** como última componente. Así pues, por las implicaciones “si y solo si” de la Ecuación 8.1 sabemos que el sistema $\mathbf{Ax} = \mathbf{b}$ es resoluble y que $\mathbf{x} = (-2, 7, 0)$ es una solución, pues

$$[\mathbf{A}|\mathbf{b}] \begin{pmatrix} \mathbf{x} \\ 1 \end{pmatrix} = \mathbf{0} \iff \mathbf{Ax} - \mathbf{d} = \mathbf{0} \iff \mathbf{Ax} = \mathbf{d}.$$

¿Una solución o infinitas soluciones?

Fíjese en la última matriz obtenida en el proceso de eliminación del último ejemplo. Sumando cualquier múltiplo de la tercera columna a la cuarta obtenemos otro vector de $\mathcal{N}([\mathbf{A}|\mathbf{b}])$ con un 1 como última componente. Es decir, el conjunto de vectores de $\mathcal{N}([\mathbf{A}|\mathbf{b}])$ con un 1 en la última componente es

$$\left\{ \mathbf{y} \in \mathbb{R}^4 \mid \text{existe } p \in \mathbb{R} \text{ tal que } \mathbf{y} = \begin{pmatrix} -2 \\ 7 \\ 0 \\ 1 \end{pmatrix} + p \begin{pmatrix} -1 \\ -1 \\ 1 \\ 0 \end{pmatrix} \right\}$$

Consecuentemente (Ecuación 8.1) el conjunto de soluciones del sistema $\mathbf{Ax} = \mathbf{b}$ es

$$\left\{ \mathbf{v} \in \mathbb{R}^3 \mid \text{existe } p \in \mathbb{R} \text{ tal que } \mathbf{v} = \begin{pmatrix} -2 \\ 7 \\ 0 \end{pmatrix} + p \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix} \right\}$$

Las ecuaciones paramétricas permiten obtener ejemplos de soluciones; basta asignar valores arbitrarios a los parámetros (en este caso a p). Obtenga alguna otra solución \mathbf{v} y compruebe que $\mathbf{A}\mathbf{v} = \mathbf{b}$.

Una vez visto el funcionamiento en un ejemplo, veamos el caso general de sistema $\mathbf{A}\mathbf{x} = \mathbf{b}$ con solución.

Considere un sistema $\mathbf{A}\mathbf{x} = \mathbf{b}$ que tiene solución y donde r es el rango de la matriz de coeficientes \mathbf{A} de n columnas. Sean $[\mathbf{A}|\mathbf{b}]_{\tau_1 \dots \tau_k}$ una forma pre-escalada y $\mathbf{E} = \mathbf{I}_{\tau_1 \dots \tau_k}$.

Sabemos que el espacio nulo $\mathcal{N}([\mathbf{A}|\mathbf{b}])$ contiene todas las combinaciones lineales de las soluciones especiales. Como en la demostración de la Proposición 7.4.2 en la página 96, sea $\boldsymbol{\nu} = (\nu_1, \dots, \nu_{n-r}, n+1,)$ la lista de índices correspondientes a las columnas nulas de la forma pre-escalada (sabemos que el último índice es $n+1$ puesto que $\mathbf{A}\mathbf{x} = \mathbf{b}$ tiene solución).

Por conveniencia expresaré $\boldsymbol{\nu}$ como concatenación de dos listas: $\boldsymbol{\nu} = \boldsymbol{\alpha} \# (n+1,)$; donde $\boldsymbol{\alpha} = (\nu_1, \dots, \nu_{n-r},)$ es la lista con los $n-r$ primeros índices, y la segunda lista, $(n+1,)$, tan solo contiene el último índice. Entonces:

$$\mathcal{N}([\mathbf{A}|\mathbf{b}]) = \left\{ \mathbf{y} \in \mathbb{R}^{(n+1)} \mid \text{existen } \mathbf{x} \in \mathbb{R}^{n-r} \text{ y } z \in \mathbb{R} \text{ tales que } \mathbf{y} = (\mathbf{E}_{|\boldsymbol{\alpha}})\mathbf{x} + (\mathbf{E}_{|n+1})z \right\};$$

y como todos los vectores de $\mathcal{N}([\mathbf{A}|\mathbf{b}])$ tienen un 1 en su última componente y la única solución especial con un 1 en la última componente¹ es $\mathbf{E}_{|n+1}$, necesariamente $z = 1$. En consecuencia,

$$\mathcal{N}([\mathbf{A}|\mathbf{b}]) = \left\{ \mathbf{v} \in \mathbb{R}^{(n+1)} \mid \text{existe } \mathbf{x} \in \mathbb{R}^{n-r} \text{ tal que } \mathbf{v} = (\mathbf{E}_{|\boldsymbol{\alpha}})\mathbf{x} + \mathbf{E}_{|n+1} \right\}. \quad (8.2)$$

De esto se deduce que *cuando $\mathbf{A}\mathbf{x} = \mathbf{b}$ es resoluble*:

1. El conjunto de soluciones del sistema $\mathbf{A}\mathbf{x} = \mathbf{b}$ es el resultado de quitar la última componente de las soluciones especiales de la Ecuación (8.2).
2. Si la forma pre-escalada de \mathbf{A} tiene columnas nulas (si $\text{rg}(\mathbf{A}) < n$) entonces el sistema tiene infinitas soluciones. Y si por el contrario $\mathcal{N}(\mathbf{A}) = \{\mathbf{0}\}$, entonces existe un único vector solución... (el vector cuyas n componentes son las n primeras componentes de la solución especial $\mathbf{E}_{|n+1}$).

Algoritmo de resolución de un sistema no homogéneo

Aplicando el método de eliminación de izquierda a derecha para pre-escalar, logramos:

$$\left[\begin{array}{c|c} \mathbf{A} & -\mathbf{b} \\ \hline \mathbf{I} & \\ \hline & \mathbf{1} \end{array} \right]_{\tau_1 \dots \tau_k \dots \tau_p} = \left[\begin{array}{c|c} \mathbf{K} & \mathbf{c} \\ \hline \mathbf{I}_{\tau_1 \dots \tau_k} & \mathbf{w} \\ \hline & \mathbf{d} \end{array} \right], \quad (8.3)$$

donde $[\mathbf{K}|\mathbf{c}]$ es una matriz pre-escalada de la matriz de coeficientes ampliada $[\mathbf{A}|\mathbf{b}]$; y donde la única componente distinta de cero de la última fila es \mathbf{d} (en la última columna).

Si se siguen los pasos de la demostración del Teorema 4.2.1 en la página 51, entonces \mathbf{d} siempre será igual a 1, pues no se habrán empleado transformaciones de Tipo II. Pero si usted opera como lo hace el módulo NAcAL (que multiplica algunas columnas por números no nulos para evitar operar con fracciones en la medida de lo posible), entonces \mathbf{d} podrá tomar cualquier valor no nulo.

¹recuerde que la última componente del resto de vectores $\mathbf{E}_{|j}$ con $j \neq (n+1)$ es un 0

Tras la eliminación en (8.3), solo caben dos posibilidades:

1. Si $\mathbf{c} = \mathbf{0}$, entonces el sistema $\mathbf{Ax} = \mathbf{b}$ tiene solución. Si $d \neq 1$, basta dividir la última columna por d para obtener una solución especial cuya última componente es un 1.

$$\left[\begin{array}{c|c} \mathbf{K} & \mathbf{0} \\ \hline \mathbf{I}_{\tau_1 \dots \tau_k} & \mathbf{w} \\ \hline & d \end{array} \right] \xrightarrow{\left[\left(\frac{1}{d} \right)^{\tau_{(n+1)}} \right]} \left[\begin{array}{c|c} \mathbf{K} & \mathbf{0} \\ \hline \mathbf{I}_{\tau_1 \dots \tau_k} & \mathbf{s} \\ \hline & 1 \end{array} \right],$$

donde \mathbf{s} es tal que $\mathbf{As} = \mathbf{b}$. Al vector \mathbf{s} se le suele llamar “solución particular” (aunque no tiene nada de particular, tan solo que lo hemos encontrado aplicando la eliminación).

2. Si $\mathbf{c} \neq \mathbf{0}$, entonces el sistema $\mathbf{Ax} = \mathbf{b}$ NO tiene solución, pues $\mathbf{b} \notin \mathcal{C}(\mathbf{A})$.

Por tanto, *el sistema tiene solución si y solo si logramos transformar $-\mathbf{b}$ en el vector $\mathbf{0}$* .



Mediante la eliminación podemos saber si $\mathbf{A} \mathbf{x} = \mathbf{b}$ es resoluble (y encontrar una solución cuando lo es).

Algoritmo para resolver $\mathbf{Ax} = \mathbf{b}$

Pre-escalamos (y luego dividimos la última columna por su última componente, d)^a:

$$\left[\begin{array}{c|c} \mathbf{A} & -\mathbf{b} \\ \hline \mathbf{I} & \\ \hline & 1 \end{array} \right] \xrightarrow{\tau_1 \dots \tau_k \dots \tau_p} \left[\begin{array}{c|c} \mathbf{K} & \mathbf{f} \\ \hline \mathbf{I}_{\tau_1 \dots \tau_k} & \mathbf{w} \\ \hline & d \end{array} \right] \xrightarrow{\left[\left(\frac{1}{d} \right)^{\tau_{(n+1)}} \right]} \left[\begin{array}{c|c} \mathbf{K} & \mathbf{c} \\ \hline \mathbf{I}_{\tau_1 \dots \tau_k} & \mathbf{s} \\ \hline & 1 \end{array} \right]$$

donde $\mathbf{K} = \mathbf{A}_{\tau_1 \dots \tau_k}$ está pre-escalada; y por tanto: $\mathbf{K}_{|j} = \mathbf{0} \implies (\mathbf{I}_{\tau_1 \dots \tau_k})_{|j} \in \mathcal{N}(\mathbf{A})$.

Solo caben dos posibilidades:

- Si $\mathbf{c} \neq \mathbf{0}$ el sistema $\mathbf{Ax} = \mathbf{b}$ NO tiene solución.
- Si $\mathbf{c} = \mathbf{0}$, entonces $\mathbf{As} = \mathbf{b}$, y el conjunto de todas las soluciones es

$$\{ \mathbf{x} \in \mathbb{R}^n \mid \text{existe } \mathbf{v} \in \mathcal{N}(\mathbf{A}) \text{ tal que } \mathbf{v} = \mathbf{s} + \mathbf{y} \}.$$

Por tanto, cuando \mathbf{K} no tiene columnas nulas ($\mathcal{N}(\mathbf{A}) = \{\mathbf{0}\}$) la única solución es \mathbf{s} .

^aNótese que d siempre es igual a 1 si se aplica el algoritmo de la demostración del Teorema 4.2.1 en la página 51, pero si también se emplean transformaciones Tipo II, entonces d podría tomar cualquier valor no nulo.

Decimos que un sistema de ecuaciones es *incompatible* si el conjunto de soluciones es *vacío*; y *compatible* en caso contrario. A su vez, los sistemas compatibles se dividen en *determinados* cuando el vector solución es único, e *indeterminado* cuando el conjunto de soluciones contiene más de un vector.

“Resolver un sistema de ecuaciones” consiste en indicar que el conjunto de soluciones es vacío cuando el sistema es *incompatible*; indicar la única solución cuando el sistema es *compatible determinado*; y expresar el conjunto de soluciones mediante ecuaciones paramétricas cuando el sistema es *compatible indeterminado*.

```
A = Matrix( [ [7,-2,0], [6,4,2] ] ); b = Vector([4,12])
SEL( A, b, 1 )           # resuelve el Sistema de Ecuaciones Lineales
```

$$\begin{array}{c|c} \begin{bmatrix} 7 & -2 & 0 & -4 \\ 6 & 4 & 2 & -12 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} & \xrightarrow{\begin{matrix} [(\frac{7}{2})2] \\ [(2)1+2] \\ [(\frac{7}{4})4] \\ [(4)1+4] \end{matrix}} \begin{bmatrix} 7 & 0 & 0 & 0 \\ 6 & 40 & 2 & -60 \\ 1 & 2 & 0 & 4 \\ 0 & 7 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 7 \end{bmatrix} & \xrightarrow{\begin{matrix} [(\frac{20}{3})3] \\ [(-1)2+3] \\ [(2)4] \\ [(3)2+4] \end{matrix}} \begin{bmatrix} 7 & 0 & 0 & 0 \\ 6 & 40 & 0 & 0 \\ 1 & 2 & -2 & 14 \\ 0 & 7 & -7 & 21 \\ 0 & 0 & 20 & 0 \\ 0 & 0 & 0 & 14 \end{bmatrix} & \xrightarrow{[(\frac{7}{14})4]} \begin{bmatrix} 7 & 0 & 0 & 0 \\ 6 & 40 & 0 & 0 \\ 1 & 2 & -2 & 1 \\ 0 & 7 & -7 & \frac{3}{2} \\ 0 & 0 & 20 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}
 \end{array}$$

$$\left\{ v \in \mathbb{R}^3 \mid \exists p \in \mathbb{R}^1, v = \begin{pmatrix} 1 \\ \frac{3}{2} \\ 0 \end{pmatrix} + \begin{bmatrix} -2 \\ -7 \\ 20 \end{bmatrix} p \right\}.$$

(Como multiplicamos la cuarta columna por 7 y por 2, al final debemos dividir la cuarta columna por 14.)

8.1.1. El Teorema de de Rouché-Frobenius



Del procedimiento anterior se deduce que los rangos de la matriz de coeficientes \mathbf{A} y de la matriz ampliada $[\mathbf{A}|\mathbf{b}]$ nos clasifican los posibles casos en cuanto al número de soluciones del sistema $\mathbf{Ax} = \mathbf{b}$.

Sabemos que tras aplicar el método de eliminación sobre las columnas de la matriz de coeficientes ampliada

$$[\mathbf{A}|\mathbf{b}] \xrightarrow{\tau_1 \cdots \tau_k} [\mathbf{K}|\mathbf{c}]$$

solo se pueden dar dos casos: que la última columna sea nula ($\mathbf{c} = \mathbf{0}$), o que no lo sea ($\mathbf{c} \neq \mathbf{0}$):

1. Si $\mathbf{c} \neq \mathbf{0}$, entonces la matriz ampliada tiene una **columna pivote** adicional, es decir,

$$\text{rg}([\mathbf{A}|\mathbf{b}]) > \text{rg}(\mathbf{A}) \iff \text{el sistema NO tiene solución.}$$

2. Si $\mathbf{c} = \mathbf{0}$, entonces

$$\text{rg}([\mathbf{A}|\mathbf{b}]) = \text{rg}(\mathbf{A}) \iff \text{el sistema es resoluble.}$$

Si además \mathbf{A} es de rango completo por columnas, entonces la forma pre-escalada \mathbf{K} no tiene columnas nulas (es decir, $\mathcal{N}(\mathbf{A}) = \{\mathbf{0}\}$), por lo que el sistema tiene una única solución.

Pero si el rango es menor que n entonces el sistema tiene infinitas soluciones, ya que a una solución \mathbf{x} se le puede sumar cualquier vector de $\mathcal{N}(\mathbf{A})$ para obtener otra solución.

El anterior resultado se resume en el siguiente

Teorema 8.1.1 (Rouché-Frobenius). *Un sistema de ecuaciones lineales $\mathbf{Ax} = \mathbf{b}$ con n incógnitas tiene solución si y solo si el rango de la matriz de coeficientes \mathbf{A} es igual al rango de la matriz de coeficientes ampliada $[\mathbf{A}|\mathbf{b}]$. En particular:*

1. Si $\text{rg}(\mathbf{A}) = n$, la solución es única.
2. En caso contrario hay infinitas soluciones.

Si además \mathbf{A} es de rango completo por filas (si su rango es igual a su número de filas m), el sistema tiene solución para cualquier vector $\mathbf{b} \in \mathbb{R}^m$; pues al pre-escalar la matriz ampliada se anulan todas las componentes de $-\mathbf{b}$, por estar a la derecha de \mathbf{A} y haber un pivote en cada fila de \mathbf{A} .

Así, para el caso de matrices cuadradas, tenemos el siguiente

Corolario 8.1.2. Para toda matriz *cuadrada* \mathbf{A} de orden n , las siguientes afirmaciones son equivalentes:

1. El rango de \mathbf{A} es n (ó \mathbf{A} es de rango completo).
2. $\mathbf{A}\mathbf{x} = \mathbf{0}$ si y solo si $\mathbf{x} = \mathbf{0}$.
3. $\mathbf{A}\mathbf{x} = \mathbf{b}$ tiene solución única.
4. \mathbf{A} no es singular.
5. \mathbf{A} es invertible.
6. \mathbf{A} es producto de matrices elementales: $\mathbf{A} = \mathbf{I}_{\tau_1 \dots \tau_k}$.
7. $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ es la única solución a $\mathbf{A}\mathbf{x} = \mathbf{b}$

8.2. Espacio columna de una matriz

Considere el conjunto de todas las combinaciones lineales de las columnas de \mathbf{A} de orden m por n

$$\left\{ \mathbf{b} \in \mathbb{R}^n \mid \text{existen } x_1, \dots, x_n \in \mathbb{R} \text{ tales que } \mathbf{b} = x_1 \mathbf{A}_{|1} + \dots + x_n \mathbf{A}_{|n} \right\}$$

o bien, usando la notación matricial,

$$\{ \mathbf{b} \in \mathbb{R}^n \mid \text{existe } \mathbf{x} \in \mathbb{R}^n \text{ tal que } \mathbf{b} = \mathbf{A}\mathbf{x} \}.$$

Puesto que este subconjunto de \mathbb{R}^n contiene todas las combinaciones lineales de las columnas, obviamente es cerrado para las combinaciones lineales. Por tanto, este conjunto es un *subespacio de \mathbb{R}^n* . Lo denominamos *espacio columna* de \mathbf{A} , y lo denotamos por $\mathcal{C}(\mathbf{A})$. Por ejemplo

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}; \quad \Rightarrow \quad \mathcal{C}(\mathbf{A}) = \left\{ \mathbf{b} \in \mathbb{R}^2 \mid \text{existen } x, y, z \in \mathbb{R} \text{ tales que } \mathbf{b} = x \begin{pmatrix} 1 \\ 1 \end{pmatrix} + y \begin{pmatrix} 1 \\ 0 \end{pmatrix} + z \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}.$$

Evidentemente

$$\mathbf{A}\mathbf{x} = \mathbf{b} \text{ tiene solución si y solo si } \mathbf{b} \in \mathcal{C}(\mathbf{A}).$$

El espacio columna y la eliminación

El subespacio $\mathcal{C}(\mathbf{A})$ es cerrado para la suma y el producto por escalares, pero precisamente ese es el tipo de operaciones que realizan las transformaciones elementales de las columnas. Por tanto, al aplicar transformaciones elementales sobre las columnas de \mathbf{A} obtenemos nuevas matrices cuyas columnas pertenecen a $\mathcal{C}(\mathbf{A})$. Así, el espacio columna de $\mathbf{A}_{\tau_1 \dots \tau_k}$ está contenido en el espacio columna de \mathbf{A} :

$$\mathcal{C}(\mathbf{A}_{\tau_1 \dots \tau_k}) \subset \mathcal{C}(\mathbf{A});$$

pero como las transformaciones elementales son “reversibles”, mediante una sucesión de transformaciones elementales se puede retornar de $\mathbf{A}_{\tau_1 \dots \tau_k}$ a \mathbf{A} , pues $\mathbf{A} = (\mathbf{A}_{\tau_1 \dots \tau_k})_{\tau_k^{-1} \dots \tau_1^{-1}}$. Por tanto, también ocurre que

$$\mathcal{C}(\mathbf{A}) \subset \mathcal{C}(\mathbf{A}_{\tau_1 \dots \tau_k}).$$

Por tanto

$$\mathcal{C}(\mathbf{A}) = \mathcal{C}(\mathbf{A}_{\tau_1 \dots \tau_k}).$$

¡Al aplicar la eliminación, todas las matrices que aparecen en el proceso tienen el mismo espacio columna!

Así pues, concluimos que

$$\text{Si } \mathbf{B} = \mathbf{A}\mathbf{E} \text{ y } \mathbf{E} = \mathbf{I}_{\tau_1 \dots \tau_k}, \text{ entonces } \mathcal{C}(\mathbf{A}) = \mathcal{C}(\mathbf{B}).$$

8.3. Generalizaciones o visiones alternativas (★)

Resolviendo varios sistemas a la vez El método de resolución que hemos visto permite (con una mínima variación) resolver simultáneamente sistemas que compartan la misma matriz de coeficientes. Por ejemplo, podemos tratar de resolver simultáneamente los ejemplos 18 y 19:

$$\left[\begin{array}{ccc|cc} \mathbf{A} & -\mathbf{b} & -\mathbf{d} \\ \mathbf{I} & & \\ & 1 & 1 \end{array} \right] = \left[\begin{array}{ccc|cc} 1 & 1 & 2 & -5 & -5 \\ 2 & 2 & 4 & -10 & 1 \\ 3 & 1 & 4 & -1 & -1 \\ 4 & 1 & 5 & 1 & 1 \\ \hline 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{array} \right] \rightarrow \left[\begin{array}{ccc|cc} 1 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 11 \\ 3 & -2 & -2 & 14 & 14 \\ 4 & -3 & -3 & 21 & 21 \\ \hline 1 & -1 & -2 & 5 & 5 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{array} \right] \rightarrow \left[\begin{array}{ccc|cc} 1 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 11 \\ 3 & -2 & 0 & 0 & 0 \\ 4 & -3 & 0 & 0 & 0 \\ \hline 1 & -1 & -1 & -2 & -2 \\ 0 & 1 & -1 & 7 & 7 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{array} \right]$$

Es decir, $\mathbf{Ax} = \mathbf{d}$ NO tiene solución pero $\mathbf{Ax} = \mathbf{b}$ si es resoluble, y el conjunto de todas las soluciones es

$$\left\{ \mathbf{x} \in \mathbb{R}^3 \left| \text{existe } a \in \mathbb{R} \text{ tal que } \mathbf{x} = \begin{pmatrix} -2 \\ 7 \\ 0 \end{pmatrix} + a \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix} \right. \right\}.$$

(Tenga en cuenta que aquí no puede usar la penúltima columna para eliminar componentes de la última).

Independencia, base y dimensión

9.1. Combinaciones lineales de vectores de \mathcal{V}

[Recuerde las definiciones de *Espacio Vectorial* (Pag. 78) y de *Subespacio* (Pag. 82)]

En esta lección vamos a tratar con *sistemas* de vectores de un subespacio genérico \mathcal{V} :

$$Z = [\vec{z}_1; \vec{z}_2; \dots \vec{z}_n], \quad \text{con } \vec{z}_i \in \mathcal{V};$$

a dichos sistemas los denotaremos con caracteres del tipo: $\mathbf{A}, \mathbf{B}, \dots \mathbf{X}, \mathbf{Y}, \mathbf{Z}$. De nuevo podemos emplear el operador selector para denotar el j -ésimo vector de un sistema $Z = [\vec{z}_1; \dots \vec{z}_n]$:

$$Z_{|j} = \vec{z}_j; \quad \text{por tanto } Z = [Z_{|1}; Z_{|2}; \dots Z_{|n}].$$

Fíjese que las matrices son un caso particular, pues \mathbf{A} de orden m por n es un sistema de n vectores de \mathbb{R}^m :

$$\mathbf{A} = [\mathbf{A}_{|1}; \mathbf{A}_{|2}; \dots \mathbf{A}_{|n}], \quad \text{con } \mathbf{A}_{|j} \in \mathbb{R}^m.$$

Librería NAcAL para Python

```
A = Matrix( [ [1,0,0], [-1,0,1] ] )
A.sis() # Devuelve el Sistema correspondiente a las columnas de A
```

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix} = \left[\begin{pmatrix} 1 \\ -1 \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right].$$

9.1.1. Extendiendo la notación matricial a los espacios vectoriales

Vamos a extender la notación matricial a subespacios genéricos¹ empezando por el...

Producto de un sistema de n vectores de \mathcal{V} por un vector de \mathbb{R}^n

Definimos el producto $Z\mathbf{a}$ de un *sistema de n vectores de \mathcal{V} por un vector de \mathbb{R}^n* como la combinación lineal de los vectores de Z , cuyos coeficientes son las componentes de \mathbf{a} :

$$Z\mathbf{a} = (Z_{|1})a_1 + \dots + (Z_{|n})a_n.$$

Por tanto $Z\mathbf{a}$ es un vector de \mathcal{V} . En el caso particular del producto $\mathbf{A}\mathbf{x}$ obtenemos un vector de \mathbb{R}^m , pero en el caso general $Z\mathbf{a}$, y dependiendo de la naturaleza de \mathcal{V} , el vector resultante pudiera ser una función, una matriz, un polinomio, una variable aleatoria, etc.... todo depende de la naturaleza de \mathcal{V} .

¹subespacios genéricos pero de dimensión finita.

Ejemplo 20. Veamos un ejemplo de un sistema de tres matrices por un vector de \mathbb{R}^3 :

```

A = Matrix([ [ 1, 1, 1], [1, 0, 0] ])
B = Matrix([ [ 4, 4, 4], [0, 4, 0] ])
C = Matrix([ [ 2, 0, 1], [0, 0, 1] ])
Z = Sistema( [ A, B, C] )           # Sistema de Matrices
a = Vector ( [ 1, 0, 1] )
Z*a                                # Sistema por Vector (comb. lineal de matrices)

```

Librería NAcAL para Python

$$Za = \left[\begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}; \begin{bmatrix} 4 & 4 & 4 \\ 0 & 4 & 0 \end{bmatrix}; \begin{bmatrix} 2 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}; \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right] = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 2 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 1 & 2 \\ 1 & 0 & 1 \end{bmatrix}.$$

El siguiente ejercicio muestra que el producto Za verifica las *propiedades de linealidad*: la primera respecto a la suma, $Z(b + c) = Zb + Zc$, y la segunda respecto al producto por un escalar, $Z(\lambda b) = \lambda(Zb)$.

EJERCICIO 44. Demuestre las siguientes proposiciones.

(a) **Proposición 9.1.1.** Sean Z , un sistema de n vectores de \mathcal{V} , y sean b y c vectores de \mathbb{R}^n ; entonces

$$Z(b + c) = Zb + Zc.$$

(b) **Proposición 9.1.2.** Sean Z , un sistema de n vectores de \mathcal{V} , el escalar λ y b un vector de \mathbb{R}^n ; entonces

$$Z(\lambda b) = \lambda(Zb).$$

Podemos extender aún más la notación matricial.

Sistema de n vectores de \mathcal{V} por una matriz de $\mathbb{R}^{n \times p}$

Si Z es un sistema de n vectores y A una matriz de n filas y p columnas; entonces denotamos por ZA al sistema de vectores

$$ZA = [Z(A_{|1}); \dots Z(A_{|p})];$$

es decir, el vector j -ésimo del sistema ZA es

$$(ZA)_{|j} = Z(A_{|j});$$

(compare esta expresión con la Definición 3.1). Así que podemos escribir sin ambigüedad el vector $ZA_{|j}$.

Continuando con el ejemplo de más arriba

```

D = Matrix([ Vector([0,2,0]), Vector([1,0,1]) ])
Z*D                                # Sistema por Matriz (Sistema de Matrices)

```

Librería NAcAL para Python

$$ZD = \left[\begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}; \begin{bmatrix} 4 & 4 & 4 \\ 0 & 4 & 0 \end{bmatrix}; \begin{bmatrix} 2 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}; \begin{bmatrix} 0 & 1 \\ 2 & 0 \\ 0 & 1 \end{bmatrix} \right] = \left[\begin{bmatrix} 8 & 8 & 8 \\ 0 & 8 & 0 \end{bmatrix}; \begin{bmatrix} 3 & 1 & 2 \\ 1 & 0 & 1 \end{bmatrix}; \right].$$

Definición 9.1. Sea $Z_{|\gamma}$ el sistema formado por los vectores de Z indicados en γ , es decir, $Z_{|\gamma} = Z(I_{|\gamma})$.

Veamos que también se verifica la asociatividad para el producto: $Z(\mathbf{B}\mathbf{c}) = (\mathbf{ZB})\mathbf{c}$ y $Z(\mathbf{B}\mathbf{C}) = (\mathbf{ZB})\mathbf{C}$:

EJERCICIO 45. Demuestre las siguientes proposiciones.

(a) **Proposición 9.1.3.** Sean Z un sistema de m vectores de \mathcal{V} , la matriz \mathbf{B} $m \times n$, y el vector \mathbf{c} de \mathbb{R}^n , entonces:

$$Z(\mathbf{B}\mathbf{c}) = (\mathbf{ZB})\mathbf{c}.$$

(b) **Proposición 9.1.4.** Sean Z un sistema de p vectores de \mathcal{V} , y \mathbf{B} $p \times q$ y \mathbf{C} $q \times n$, entonces $Z(\mathbf{B}\mathbf{C}) = (\mathbf{ZB})\mathbf{C}$.

Por lo que podemos escribir sin ambigüedad el vector \mathbf{ZBc} y el sistema \mathbf{ZBC} .



Al extender la notación matricial para expresar combinaciones lineales en espacios vectoriales genéricos, hemos logrado algunos resultados sobre combinaciones lineales de sistemas de vectores de \mathcal{V} que nos serán de utilidad.

$$\begin{array}{ll} \blacksquare Z(\mathbf{b} + \mathbf{c}) = Z\mathbf{b} + Z\mathbf{c} & \blacksquare Z(\mathbf{B}\mathbf{c}) = (\mathbf{ZB})\mathbf{c} \\ \blacksquare Z(\lambda\mathbf{b}) = \lambda(Z\mathbf{b}) & \blacksquare Z(\mathbf{B}\mathbf{C}) = (\mathbf{ZB})\mathbf{C} \end{array}$$

9.2. Sistemas generadores

En la Sección 8.2 definimos el espacio columna de la matriz \mathbf{A} como el conjunto de todas las combinaciones lineales de sus columnas:

$$\mathcal{C}(\mathbf{A}) = \left\{ \mathbf{y} \in \mathbb{R}^n \mid \text{existen } x_1, \dots, x_n \in \mathbb{R} \text{ tales que } \mathbf{y} = (\mathbf{A}_{|1})x_1 + \dots + (\mathbf{A}_{|n})x_n \right\}.$$

o usando notación matricial

$$\mathcal{C}(\mathbf{A}) = \{ \mathbf{y} \in \mathbb{R}^n \mid \text{existe } \mathbf{x} \in \mathbb{R}^n \text{ tal que } \mathbf{y} = \mathbf{Ax} \}.$$

Ahora vamos a generalizar esta idea a cualquier espacio vectorial \mathcal{V} .

Definición 9.2 ($\mathcal{L}(Z)$: subespacio engendrado por un sistema Z). Dado un sistema $Z = [\vec{z}_1; \dots; \vec{z}_n]$ de n vectores del espacio vectorial \mathcal{V} , el subespacio engendrado por dicho sistema es

$$\mathcal{L}(Z) = \left\{ \vec{v} \in \mathcal{V} \mid \text{existen } a_1, \dots, a_n \in \mathbb{R} \text{ tales que } \vec{v} = \vec{z}_1 a_1 + \dots + \vec{z}_n a_n \right\},$$

es decir $\mathcal{L}(Z)$ es el conjunto de combinaciones lineales de $Z_{|j}$ con $j = 1 : n$; o usando notación matricial

$$\mathcal{L}(Z) = \left\{ \vec{v} \in \mathcal{V} \mid \text{existe } \mathbf{a} \in \mathbb{R}^n \text{ tal que } \vec{v} = Z\mathbf{a} \right\}. \quad (9.1)$$

Definición 9.3 (Sistema generador de \mathcal{V}). Decimos que Z es un sistema generador de \mathcal{V} si $\mathcal{V} = \mathcal{L}(Z)$.

Ejemplo 21.

- Las columnas de \mathbf{A} son un sistema generador de $\mathcal{C}(\mathbf{A})$; es decir $\mathcal{L}(\mathbf{A}) = \mathcal{C}(\mathbf{A})$.

Librería NAcAL para Python

```
A = Matrix([ Vector([1,0,0]), Vector([-1,0,1]) ])
SubEspacio( A.sis() ) # SubEspacio generado por el Sistema de columnas de A
```

$$\left\{ \mathbf{v} \in \mathbb{R}^3 \mid \exists \mathbf{p} \in \mathbb{R}^2, \mathbf{v} = \begin{bmatrix} 1 & -1 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{p} \right\} = \{ \mathbf{v} \in \mathbb{R}^3 \mid \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \mathbf{v} = (0) \}$$

(Ecuaciones paramétricas a la izquierda y cartesianas a la derecha)

- Las soluciones especiales de $\mathbf{Ax} = \mathbf{0}$ son un *sistema generador* de $\mathcal{N}(\mathbf{A})$.

```
C = Matrix([ [1,0,0], [1,1,1] ])
Homogenea(C,1).sgen      # Sistema generador del espacio nulo de C
```

Librería NAcAL para Python

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{[(-1)\mathbf{2}+\mathbf{3}]} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\left[\begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} \right];$$

```
Homogenea(C).enulo      # Espacio nulo de C
```

Librería NAcAL para Python

$$\left\{ \mathbf{v} \in \mathbb{R}^3 \mid \exists \mathbf{p} \in \mathbb{R}^1, \mathbf{v} = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} \mathbf{p} \right\} = \left\{ \mathbf{v} \in \mathbb{R}^3 \mid \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \mathbf{v} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\}$$



$\mathcal{L}(\mathbf{Z})$ es el conjunto de combinaciones lineales de los vectores de \mathbf{Z} ; y puesto que

- la suma de dos combinaciones lineales $\mathbf{Zb} + \mathbf{Zc}$ es la combinación lineal $\mathbf{Z}(\mathbf{b} + \mathbf{c})$
- el producto de una combinación lineal por un escalar $(\mathbf{Zb})\lambda$ es la combinación lineal $\mathbf{Z}(\mathbf{b}\lambda)$

$\mathcal{L}(\mathbf{Z})$ es un *subespacio*, pues es cerrado para la suma y el producto por escalares.

Además, puesto que $\mathcal{L}(\mathbf{Y})$ contiene *únicamente* las combinaciones lineales de $\mathbf{Y} = [\vec{y}_1; \dots; \vec{y}_n]$, el subespacio $\mathcal{L}(\mathbf{Y})$ *está contenido en cualquier otro subespacio que también contenga los vectores* $\vec{y}_1, \dots, \vec{y}_n$; es decir, $\mathcal{L}(\mathbf{Y})$ es el subespacio más pequeño que contiene los vectores $\vec{y}_1, \dots, \vec{y}_n$ (véase la Sección 9.6).

Un sistema de vectores \mathbf{Z} es un sistema generador de \mathcal{V} si: $\mathcal{V} = \mathcal{L}(\mathbf{Z})$

Ahora ya podemos establecer un criterio para saber si dos sistemas \mathbf{Y} y \mathbf{Z} generan el mismo subespacio:

Sea \mathbf{Y} un sistema de n vectores; puesto que $\mathcal{L}(\mathbf{Y}) \subset \mathcal{L}(\mathbf{Z})$ si y solo si $\mathbf{Y}_{|i} \in \mathcal{L}(\mathbf{Z})$, $i = 1 : n$, y puesto que dos conjuntos A y B son iguales si y solo si $A \subset B$ y $B \subset A$, llegamos al siguiente

Corolario 9.2.1. Sean \mathbf{Y} con n vectores y \mathbf{Z} con k vectores; entonces $\mathcal{L}(\mathbf{Y}) = \mathcal{L}(\mathbf{Z})$ si y sólo si

$$\mathbf{Y}_{|i} \in \mathcal{L}(\mathbf{Z}) \text{ para todo } i = 1 : n \quad \text{y} \quad \mathbf{Z}_{|j} \in \mathcal{L}(\mathbf{Y}) \text{ para todo } j = 1 : k.$$

Definición 9.4. Diremos que dos sistemas generadores son *equivalentes* si generan el mismo espacio.

```
A = Matrix([ [2,-1,0], [-1,2,-1] ])
B = ElimGJ(Matrix(A))
SubEspacio( A.sis() ) == SubEspacio( B.sis() ) # ¿Son equivalentes?
```

Librería NAcAL para Python

True

9.3. Transformaciones elementales sobre sistemas de vectores

Al igual que con las matrices, podemos aplicar transformaciones elementales sobre los sistemas de vectores:

- (Transformación Tipo I sobre el sistema de vectores) Si $i \neq j$ entonces

$$Z = [\vec{z}_1; \dots \vec{z}_i; \dots \vec{z}_j; \dots \vec{z}_n] \xrightarrow{[(\alpha)\vec{i}+\vec{j}]} [\vec{z}_1; \dots (\alpha\vec{z}_i + \vec{z}_j); \dots \vec{z}_j; \dots \vec{z}_n] = Z_{[(\alpha)\vec{i}+\vec{j}]} = Z \begin{pmatrix} \mathbf{I} \\ \tau_{[(\alpha)\vec{i}+\vec{j}]} \end{pmatrix}$$

- (Transformación Tipo II sobre el sistema de vectores) Si $\alpha \neq 0$ entonces

$$Z = [\vec{z}_1; \dots \vec{z}_i; \dots \vec{z}_n] \xrightarrow{[(\alpha)\vec{i}]} [\vec{z}_1; \dots (\alpha\vec{z}_i); \dots \vec{z}_n] = Z_{[(\alpha)\vec{i}]} = Z \begin{pmatrix} \mathbf{I} \\ \tau_{[(\alpha)\vec{i}]} \end{pmatrix}$$

Sistemas equivalentes

Así, si mediante una sucesión de transformaciones elementales τ_k, \dots, τ_n transformamos Z en un nuevo sistema $Y = Z_{\tau_1 \dots \tau_k}$, puesto que por una parte, $Y_{|j} = Z \mathbf{E}_{|j}$ y por otra $Z_{|j} = Y(\mathbf{E}^{-1})_{|j}$; por el Corolario 9.2.1 concluimos que ambos sistemas son equivalentes:

$$\mathcal{L}(Z) = \mathcal{L}(Z\mathbf{E}) = \mathcal{L}(Y), \quad \text{donde } \mathbf{E} = \mathbf{I}_{\tau_1 \dots \tau_k}. \quad (9.2)$$

Consideremos ahora una transformación de naturaleza distinta: *quitar vectores nulos de un sistema*:

- (Quitando vectores nulos) $[\vec{z}_1; \vec{0}; \dots \vec{z}_i; \vec{0}; \dots \vec{z}_n; \vec{0}] \xrightarrow{\text{Quitando vectores nulos}} [\vec{z}_1; \dots \vec{z}_i; \dots \vec{z}_n].$

Por el Corolario 9.2.1

$$\mathcal{L}([\vec{z}_1; \vec{0}; \dots \vec{z}_i; \vec{0}; \dots \vec{z}_n; \vec{0}]) = \mathcal{L}([\vec{z}_1; \dots \vec{z}_i; \dots \vec{z}_n]).$$

Es mas, podemos generalizar el resultado con la siguiente

Proposición 9.3.1. Si $Z = [\vec{z}_1; \dots \vec{z}_n]$ es un sistema de vectores entonces

$$\mathcal{L}([\vec{z}_1; \dots \vec{z}_{i-1}; \vec{z}_i; \vec{z}_{i+1}; \dots \vec{z}_n]) = \mathcal{L}([\vec{z}_1; \dots \vec{z}_{i-1}; \vec{z}_{i+1}; \dots \vec{z}_n]).$$

si, y solo si, \vec{z}_i es combinación lineal del resto de vectores de Z .

Demostración. Basta aplicar el Corolario 9.2.1. □



Como consecuencia, y volviendo al espacio generado por las columnas de una matriz \mathbf{A} :

Si \mathbf{K} es una forma pre-escalada de \mathbf{A} y si ρ es la lista de índices de columnas no nulas de \mathbf{K} , entonces

$$\mathcal{C}(\mathbf{A}) = \mathcal{C}(\mathbf{K}) = \mathcal{C}(\mathbf{K}_{|\rho}),$$

donde $\mathbf{K}_{|\rho}$ es la submatriz correspondiente a las columnas no nulas de \mathbf{K} (véase Apéndice 1.B en la página 22).

Por tanto, las columnas no nulas de cualquier forma pre-escalada de \mathbf{A} generan $\mathcal{C}(\mathbf{A})$.

Observación. Como $\vec{0} \in \mathcal{V}$, y como quitando los vectores nulos de un sistema generador se engendra el mismo espacio, tenemos que $\mathcal{L}([\vec{0}; \vec{0}; \vec{0}]) = \mathcal{L}([\vec{0}; \vec{0}]) = \mathcal{L}([\vec{0}]) = \mathcal{L}([\]) = \mathcal{L}(\emptyset) = \{\vec{0}\}$; es decir, un sistema (o lista) vacío es el sistema generador del subespacio $\{\vec{0}\}$ (véase la Sección 9.6).

9.4. Sistemas linealmente dependientes vs independientes

EJERCICIO 46. Demuestre la siguiente...

Proposición 9.4.1. Dado un sistema $Z = [\vec{z}_1; \dots; \vec{z}_n]$ de n vectores de \mathcal{V} , existe $i \in \{1, \dots, n\}$ tal que z_i es combinación lineal del resto de vectores de Z si, y solo si, existe $\mathbf{a} \neq \mathbf{0}$ en \mathbb{R}^n tal que $Z\mathbf{a} = \vec{0}$.

Esto da lugar a la definición de *sistema linealmente dependiente* y de *sistema linealmente independiente*:

Definición 9.5. Diremos que el sistema Z de n vectores de \mathcal{V}

- es linealmente **dependiente** si existe $\mathbf{a} \in \mathbb{R}^n$ no nulo tal que, $Z\mathbf{a} = \vec{0}$.
- es linealmente **independiente** si ocurre lo contrario,² por tanto, $Z\mathbf{a} = \vec{0}$ si y solo si $\mathbf{a} = \mathbf{0}$.

(véase también la Sección 9.6.)

Definición 9.6. Sea Z de n vectores de \mathcal{V} . Denominamos *espacio nulo* de Z al subconjunto $\mathcal{N}(Z)$ de vectores \mathbf{x} de \mathbb{R}^n tales que $Z\mathbf{x} = \vec{0}$, es decir

$$\mathcal{N}(Z) = \{\mathbf{x} \in \mathbb{R}^n \mid Z\mathbf{x} = \vec{0}\}.$$

Las anteriores definiciones implican que Z es linealmente independiente si y solo si $\mathcal{N}(Z) = \{\mathbf{0}\}$.

Ejemplos de sistemas linealmente independientes

- Las columnas de una matriz \mathbf{A} *invertible* constituyen un sistema linealmente independiente, puesto que

$$\mathbf{A}\mathbf{x} = \mathbf{0} \quad \Leftrightarrow \quad \mathbf{A}^{-1}\mathbf{A}\mathbf{x} = \mathbf{A}^{-1}\mathbf{0} \quad \Leftrightarrow \quad \mathbf{x} = \mathbf{0}.$$

- Si, de un sistema Z de vectores linealmente independiente, quitamos uno cualquiera de sus vectores, el sistema resultante sigue siendo linealmente independiente.
- Por tanto, cualquier sistema formado por una selección de columnas de una matriz invertible (sin repetición) $\mathbf{E}_{|\gamma}$ es un sistema linealmente independiente.
- Como las soluciones especiales son una selección de columnas de una matriz invertible \mathbf{E} , *las soluciones especiales constituyen un sistema linealmente independiente*.

Ahora que hemos definido los sistemas linealmente independientes, ya podemos enunciar la siguiente proposición sobre matrices de rango completo por columnas (¡son equivalencias vistas anteriormente!):

Proposición 9.4.2. Para $\mathbf{A} = [\mathbf{v}_1; \dots; \mathbf{v}_n]$ de n vectores de \mathbb{R}^m , las siguientes propiedades son equivalentes:

- El rango de $\mathbf{A} = [\mathbf{v}_1; \dots; \mathbf{v}_n]$ es n .
- La combinación lineal $\mathbf{A}\mathbf{x}$ es $\mathbf{0}$ si y solo si $\mathbf{x} = \mathbf{0}$.
- $\mathcal{N}(\mathbf{A}) = \{\mathbf{0}\}$.
- Las columnas de \mathbf{A} son linealmente independientes.

9.5. Bases y dimensión

Hemos visto que en un sistema de vectores linealmente dependiente es posible retirar algún vector sin reducir el espacio generado. Pero si el sistema es linealmente independiente, retirar cualquiera de los vectores reduce el espacio engendrado por ellos. Un sistema que tenga el tamaño justo para generar el subespacio \mathcal{V} sin que sobre (sin que podamos retirar) ningún vector tiene un nombre especial:

Definición 9.7. Diremos que el sistema B es **una base** de un subespacio \mathcal{V} si simultáneamente

1. B genera el subespacio, es decir, si $\mathcal{L}(B) = \mathcal{V}$.
2. B es linealmente independiente, es decir, $B\mathbf{x} = \vec{0}$ si y solo si $\mathbf{x} = \mathbf{0}$.

²Para ver por qué un sistema vacío, $Z = []$, es linealmente independiente véase la Sección 9.6.

Es decir, una base B de \mathcal{V} es un sistema de vectores suficientemente reducido para ser linealmente independiente pero suficiente grande como para generar el subespacio \mathcal{V} .

Ejemplo 22. Las soluciones *especiales* encontradas al resolver $\mathbf{A}\mathbf{x} = \mathbf{0}$ constituyen una base de $\mathcal{N}(\mathbf{A})$.

Justificación: Por la Proposición 7.4.2 en la página 96 sabemos que las soluciones especiales son un sistema generador de $\mathcal{N}(\mathbf{A})$ y, puesto que son columnas una matriz invertible, son linealmente independientes.

Un importante resultado es que todas las bases de un subespacio \mathcal{V} tienen el mismo número de vectores. Usaremos la siguiente proposición para demostrarlo:

Proposición 9.5.1. Si X y Z son dos sistemas con p y q vectores de \mathcal{V} respectivamente, donde $p > q$; tales que $\mathcal{L}(X) \subset \mathcal{L}(Z)$ entonces X es linealmente dependiente.

Demostración. Como cada X_{I_i} es combinación lineal de los vectores del sistema Z , existe \mathbf{M} tal que

$$X_{I_j} = ZM_{I_j} \text{ para } j = 1 : p; \text{ es decir, tal que } X = ZM.$$

Como \mathbf{M} tiene más columnas que filas, existe $\mathbf{a} \neq \mathbf{0}$ tal que $\mathbf{Ma} = \mathbf{0}$; y por tanto $X\mathbf{a} = Z\mathbf{Ma} = Z\mathbf{0} = \vec{0}$. \square

Así pues, *dos bases de un mismo subespacio tienen el mismo número de vectores*; pues si una base tuviera más que la otra, por la Proposición 9.5.1, sería un sistema dependiente, y por tanto ya no sería una base.

Definición 9.8. Decimos que un subespacio \mathcal{V} es de dimensión finita cuando tiene una base con un número finito de vectores. En tal caso, llamamos *dimensión de \mathcal{V}* al número de vectores de cualquiera de sus bases.

En la sección anterior vimos un procedimiento para generar una base quitando vectores. Veamos otro procedimiento para completar una base añadiendo vectores.

Proposición 9.5.2. Si $[\vec{z}_1; \dots; \vec{z}_n;]$ es un sistema linealmente independiente y $\vec{z}_{n+1} \notin \mathcal{L}([\vec{z}_1; \dots; \vec{z}_n;])$ entonces $[\vec{z}_1; \dots; \vec{z}_n; \vec{z}_{n+1};]$ es linealmente independiente.

Demostración. Supongamos que $a_1\vec{z}_1 + \dots + a_n\vec{z}_n + a_{n+1}\vec{z}_{n+1} = \vec{0}$. Como $\vec{z}_{n+1} \notin \mathcal{L}([\vec{z}_1; \dots; \vec{z}_n;])$, necesariamente $a_{n+1} = 0$. Y como $[\vec{z}_1; \dots; \vec{z}_n;]$ es linealmente independiente, también $a_1, \dots, a_n = 0$. \square

Corolario 9.5.3. Cualquier subespacio \mathcal{W} , de un espacio \mathcal{V} de dimensión finita, tiene dimensión finita menor o igual a la dimensión de \mathcal{V} .

Demostración. Sea \mathcal{W} un subespacio de \mathcal{V} . Sabemos que existen sistemas de vectores de \mathcal{W} linealmente independientes. Por ejemplo el vacío. Por otra parte, cualquier sistema de vectores linealmente independientes tiene como mucho tantos vectores como la dimensión del espacio. De todos los posibles sistemas linealmente independientes de \mathcal{W} , tomemos un sistema M con el mayor número posible de vectores. Entonces $\mathcal{L}(M) \subset \mathcal{W}$. Veamos que si suponemos que existe $\vec{v} \in \mathcal{W}$ tal que $\vec{v} \notin \mathcal{L}(M)$, llegamos a una contradicción. Basta añadir \vec{v} al sistema M para obtener un sistema linealmente independiente formado por vectores de \mathcal{W} . Pero esto contradice que M sea un sistema linealmente independiente con el mayor número posible de vectores. \square

Corolario 9.5.4. Si \mathcal{W} es un subespacio de \mathcal{V} de la misma dimensión de \mathcal{V} , entonces \mathcal{W} y \mathcal{V} son iguales.

Demostración. Tomemos una base B de \mathcal{W} . Entonces sabemos que $\mathcal{L}(B) = \mathcal{W}$. Para demostrar que $\mathcal{L}(B) = \mathcal{V}$, supongamos lo contrario. Supongamos que existe $\vec{v} \in \mathcal{V}$ tal que $\vec{v} \notin \mathcal{L}(B)$. Entonces podríamos incluir \vec{v} en B y obtendríamos un sistema linealmente independiente con más elementos que la dimensión de \mathcal{V} . \square

Corolario 9.5.5. Cualquier sistema B de vectores de \mathcal{W} linealmente independiente cuyo número de vectores coincida con la dimensión de \mathcal{W} es necesariamente es una base.

9.5.1. Eliminación “de izquierda a derecha” y sistemas “acoplados” de vectores



Vamos a diseñar un algoritmo para encontrar una base de $\mathcal{L}(Z)$. La mecánica es sencilla y se basa en la Ecuación 9.2, es decir, $\mathcal{L}(Z) = \mathcal{L}(Z_{\tau_1 \dots \tau_k})$

- consiste en aplicar la eliminación de “izquierda a derecha” sobre Z , y observar qué vectores de Z no se anulan. Aquellos vectores de Z que no se anulen constituyen una base de $\mathcal{L}(Z)$.

Aunque el procedimiento es sencillo, demostrar por qué funciona no es inmediato. Aprovecharemos para introducir notación y nuevas operaciones; a cambio lograremos una comprensión más profunda de las propiedades de los sistemas obtenidos en el proceso de eliminación (“de izquierda a derecha”).

Notación. Denotaremos con $Y_{|(1:s)}$ al subsistema de $Y = [\vec{y}_1; \dots; \vec{y}_n]$ formado por sus primeros s vectores

$$Y_{|(1:s)} = [\vec{y}_1; \dots; \vec{y}_s] = [\vec{y}_j \mid j < s] \quad \text{con } s \leq n,$$

y asumiremos que $\mathcal{L}(Z_{|(1:0)}) = \mathcal{L}([\]) = \{\vec{0}\}$; (véase la Sección 9.6 en la página 118 y (a pie de página) la Nota 6 en la página 4).

Definición 9.9 (Sistemas acoplados). Los sistemas Y y Z , de n vectores cada uno, están **acoplados** si

$$\mathcal{L}([\vec{y}_1; \dots; \vec{y}_i]) = \mathcal{L}([\vec{z}_1; \dots; \vec{z}_i]) \quad \text{para } i = 1 : n,$$

es decir, si

$$\mathcal{L}(Y_{|(1:i)}) = \mathcal{L}(Z_{|(1:i)}) \quad \text{para } i = 1 : n.$$

Proposición 9.5.6 (Eliminación “de izquierda a derecha” en Z). Si transformamos $Z = [\vec{z}_1; \dots; \vec{z}_n]$ mediante una secuencia $\tau_1 \dots \tau_k$ de transformaciones elementales “de izquierda a derecha”

$$Z_{\tau_1 \dots \tau_k} = ZE = Y; \quad \text{donde } E = I_{\tau_1 \dots \tau_k}, \text{ es triangular superior unitaria,}^3 \text{ es decir,}$$

$$Y = [\vec{y}_1; \dots; \vec{y}_n] = [\vec{z}_1; \dots; \vec{z}_n] \begin{bmatrix} 1 & e_{12} & e_{13} & \cdots & e_{1n} \\ & 1 & e_{23} & \cdots & e_{2n} \\ & & 1 & \cdots & e_{3n} \\ & & & \ddots & \vdots \\ & & & & 1 \end{bmatrix} = ZE,$$

entonces Y y Z están acoplados, es decir, $\mathcal{L}(Y_{|(1:i)}) = \mathcal{L}(Z_{|(1:i)})$ para $i = 1 : n$.

Demostración. Por una parte, como E es triangular superior unitaria, cada vector $Y_{|j} = ZE_{|j}$ es el vector $Z_{|j}$ más una combinación lineal de los vectores $Z_{|k}$ con $k < j$, y por tanto $\mathcal{L}(Y_{|(1:j)}) \subset \mathcal{L}(Z_{|(1:j)})$ para $j = 1 : n$.

Por otra parte, como $Z = YE^{-1}$, y como E^{-1} también es triangular superior unitaria (Corolario 5.1.17), usando el mismo argumento del párrafo anterior concluimos que $\mathcal{L}(Z_{|(1:j)}) \subset \mathcal{L}(Y_{|(1:j)})$ para $j = 1 : n$. \square

Notación. Sean $Y = [\vec{y}_1; \dots; \vec{y}_n]$ y $Z = [\vec{z}_1; \dots; \vec{z}_k]$ dos sistemas de vectores de \mathcal{V} . Denotaremos con $Y \# Z$ al sistema de $n + k$ vectores resultante de concatenar Y (de n vectores) y Z (de k vectores)

$$Y \# Z = [\vec{y}_1; \dots; \vec{y}_n; \vec{z}_1; \dots; \vec{z}_k].$$

³por el Teorema 5.1.16 en la página 70

Lema 9.5.7. Sean $Y = [\vec{y}_1; \dots; \vec{y}_n;]$ y $Z = [\vec{z}_1; \dots; \vec{z}_k;]$ dos sistemas de vectores de \mathcal{V} tales que $\mathcal{L}(Y) = \mathcal{L}(Z)$. Entonces para cualquier sistema $X = [\vec{x}_1; \dots; \vec{x}_r;]$ de vectores de \mathcal{V} se verifica que

$$\mathcal{L}(Y \# X) = \mathcal{L}(Z \# X).$$

Demostración. Hay que demostrar que $\{\vec{z}_1, \dots, \vec{z}_k, \vec{x}_1, \dots, \vec{x}_r\} \subset \mathcal{L}(Y \# X)$ y que $\{\vec{y}_1, \dots, \vec{y}_n, \vec{x}_1, \dots, \vec{x}_r\} \subset \mathcal{L}(Z \# X)$. Como ambos contenidos se prueban de la misma manera, solo comprobaremos el primero.

Por un lado tenemos que $\{\vec{z}_1, \dots, \vec{z}_k\} \subset \mathcal{L}(Z) = \mathcal{L}(Y) \subset \mathcal{L}(Y \# X)$. Por otro que $\{\vec{x}_1, \dots, \vec{x}_r\} \subset \mathcal{L}(Y \# X)$. Luego $\{\vec{z}_1, \dots, \vec{z}_k\} \cup \{\vec{x}_1, \dots, \vec{x}_r\} \subset \mathcal{L}(Y \# X)$. \square

Corolario 9.5.8. Si $Y = [\vec{y}_1; \dots; \vec{y}_n;]$ y $Z = [\vec{z}_1; \dots; \vec{z}_n;]$ son dos sistemas acoplados y \vec{z}_k es combinación de los vectores que le anteceden en la lista Z

$$\vec{z}_k \in \mathcal{L}(Z_{|(1:k-1)}) = \mathcal{L}([\vec{z}_j \mid j < k]), \quad \text{entonces} \quad (9.3)$$

los sistemas resultantes tras quitar el k -ésimo vector, $[\vec{z}_j \mid j \neq k]$ y $[\vec{y}_j \mid j \neq k]$, también están acoplados.

Demostración. El resultado es trivial si $k = n$. Así que supongamos que $k < n$. Puesto que Z y Y están acoplados, $\mathcal{L}(Y_{|(1:i)}) = \mathcal{L}(Z_{|(1:i)})$ para todo $i < k$, y como consecuencia de (9.3) tendremos que

$$\mathcal{L}(Y_{|(1:k-1)}) = \mathcal{L}(Z_{|(1:k-1)}) \stackrel{(9.3)}{=} \mathcal{L}(Z_{|(1:k)}) = \mathcal{L}(Y_{|(1:k)})$$

luego aplicando el Lema 9.5.7 (en $\stackrel{*}{=}$) tendremos que para todo $k < i$

$$\begin{aligned} \mathcal{L}([\vec{y}_j \mid j \neq k]) &= \mathcal{L}(Y_{|(1:k-1)} \# Y_{|(k+1:i)}) \stackrel{*}{=} \mathcal{L}(Y_{|(1:k)} \# Y_{|(k+1:i)}) = \mathcal{L}(Y_{|(1:i)}) = \mathcal{L}(Z_{|(1:i)}) \\ &= \mathcal{L}(Z_{|(1:k)} \# Z_{|(k+1:i)}) \\ &\stackrel{*}{=} \mathcal{L}(Z_{|(1:k-1)} \# Z_{|(k+1:i)}) = \mathcal{L}([\vec{z}_j \mid j \neq k]). \end{aligned}$$

\square

Corolario 9.5.9. Si Y y Z son dos sistemas acoplados de n vectores y $\vec{y}_s = \vec{0}$ entonces los sistemas $[\vec{y}_j \mid j \neq s]$ y $[\vec{z}_j \mid j \neq s]$ también están acoplados.

Definición 9.10. Si $[\vec{y}_1; \dots; \vec{y}_n;]$ y $[\vec{z}_1; \dots; \vec{z}_n;]$ son dos sistemas acoplados diremos que el par (\vec{y}_s, \vec{z}_s) es superfluo si $\vec{y}_s = \vec{0}$. En tal caso diremos que $[\vec{y}_j \mid j \neq s]$ y $[\vec{z}_j \mid j \neq s]$ son los sistemas acoplados resultantes de quitar el par superfluo (\vec{y}_s, \vec{z}_s) .

Corolario 9.5.10. Si $[\vec{y}_1; \dots; \vec{y}_n;]$ y $[\vec{z}_1; \dots; \vec{z}_n;]$ son sistemas acoplados, también lo son $[\vec{y}_j \mid \vec{y}_j \neq \vec{0}]$ y $[\vec{z}_j \mid \vec{y}_j \neq \vec{0}]$.

Demostración. Basta quitar los pares superfluos correspondientes a los ceros en el sistema $[\vec{y}_1; \dots; \vec{y}_n;]$. \square



El anterior corolario nos muestra una forma de encontrar una base de $\mathcal{L}(Z)$ entre los vectores de Z : Basta seleccionar el subsistema con los vectores de $Z_{|j}$ que no se anulen tras aplicar el método de eliminación $Z_{\tau_1 \dots \tau_k}$, es decir

$$[Z_{|j} \mid Z\mathbf{E}_{|j} \neq \vec{0}], \quad \text{donde} \quad \mathbf{E} = \mathbf{I}_{\tau_1 \dots \tau_k}.$$

Así pues, tenemos otro ejemplo de base de un subespacio:

Ejemplo 23. Las columnas no nulas de una forma escalonada de \mathbf{A} constituyen una base de $\mathcal{C}(\mathbf{A})$.

Justificación: Sea $\mathbf{A}_{\tau_1 \dots \tau_k}$ una forma escalonada de \mathbf{A} , por (9.2) sabemos que $\mathcal{L}(\mathbf{A}) = \mathcal{L}(\mathbf{A}_{\tau_1 \dots \tau_k})$; así, al seleccionar las columnas no nulas de $\mathbf{A}_{\tau_1 \dots \tau_k}$ creamos un sistema linealmente independiente.

Encontrando dos bases de $\mathcal{C}(\mathbf{A})$ y una base de $\mathcal{N}(\mathbf{A})$

Concluimos que aplicando la eliminación sobre la matriz resultante de concatenar las n columnas de \mathbf{A} con las columnas de la matriz identidad de orden n , encontramos dos bases de $\mathcal{C}(\mathbf{A})$ y una base de $\mathcal{N}(\mathbf{A})$:

$$\begin{bmatrix} \mathbf{A} \\ \mathbf{I} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 2 \\ 1 & 2 & 1 & 3 \\ 2 & 4 & 3 & 8 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{[(-2)\mathbf{3}+4]} \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 2 & 1 & 1 \\ 2 & 4 & 3 & 2 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{\begin{matrix} \tau \\ [(-2)\mathbf{1}+2] \\ [(-1)\mathbf{1}+3] \\ [(-1)\mathbf{1}+4] \end{matrix}} \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 2 & 0 & 1 & 0 \\ 1 & -2 & -1 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{K} \\ \mathbf{E} \end{bmatrix};$$

$\mathcal{C}(\mathbf{A}) = \mathcal{L}([\mathbf{K}_{|1}; \mathbf{K}_{|3}])$ es de dimensión 2 por ser $\mathbf{K}_{|1}$ y $\mathbf{K}_{|3}$ linealmente independientes; y por el Corolario 9.5.10 sabemos que $\mathcal{C}(\mathbf{A}) = \mathcal{L}([\mathbf{A}_{|1}; \mathbf{A}_{|3}])$ y como $[\mathbf{A}_{|1}; \mathbf{A}_{|3}]$ tiene dos vectores es una base de $\mathcal{C}(\mathbf{A})$. Por otra parte, ya vimos (Página 113) que las soluciones especiales $\mathbf{E}_{|2}$ y $\mathbf{E}_{|4}$ son una base de $\mathcal{N}(\mathbf{A})$.

Es evidente que, con este procedimiento, el número de columnas que se anulan (la dimensión de $\mathcal{N}(\mathbf{A})$) más el número de columnas que no se anulan (la dimensión de $\mathcal{C}(\mathbf{A})$) es el número total de columnas de \mathbf{A} . Por tanto podemos enunciar el siguiente

Corolario 9.5.11. $\dim \mathcal{C}(\mathbf{A}) + \dim \mathcal{N}(\mathbf{A})$ es igual al número de columnas de \mathbf{A} .

<pre>A = Matrix([[0,0,1,2], [1,2,1,3], [2,4,3,8]]); SubEspacio(A.sis()).base # una base del SubEspacio</pre>	Librería NAcAL para Python
$\begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}; \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix};$	
<pre>SubEspacio(A.sis()).dim # dimensión del SubEspacio</pre> <p>2</p>	Librería NAcAL para Python
<pre>Homogenea(A).enulo.dim # dimensión del Espacio Nulo</pre> <p>2</p>	Librería NAcAL para Python

Pero nótese que, en general, ambos son subespacios de espacios distintos: $\mathcal{C}(\mathbf{A}) \subset \mathbb{R}^m$ y $\mathcal{N}(\mathbf{A}) \subset \mathbb{R}^n$.

Pero el anterior resultado no se limita a sistemas de vectores de \mathbb{R}^m . De manera análoga, si mediante eliminación anulamos todo vector $Z_{|j}$ que sean combinación lineal de los vectores de Z que estén a su izquierda; es decir si $Y = Z_{\tau_1 \dots \tau_k}$, donde $E = I_{\tau_1 \dots \tau_k}$ y

$$\begin{bmatrix} Z \\ I \end{bmatrix} \xrightarrow{\tau_1 \dots \tau_k} \begin{bmatrix} Y \\ E \end{bmatrix} = \begin{bmatrix} \vec{y}_1; & \vec{0}; & \vec{y}_3; & \vec{0}; & \dots & \vec{y}_n; \\ 1 & e_{12} & e_{13} & e_{14} & \dots & e_{1n} \\ & 1 & e_{23} & e_{24} & \dots & e_{2n} \\ & & 1 & e_{34} & \dots & e_{3n} \\ & & & 1 & \dots & e_{3n} \\ & & & & \ddots & \vdots \\ & & & & & 1 \end{bmatrix}; \quad (9.4)$$

el número total de **vectores no anulados** más el número total de **soluciones especiales** es igual al número de vectores del sistema Z . Por tanto, también es cierto el siguiente

Corolario 9.5.12. $\dim \mathcal{L}(Z) + \dim \mathcal{N}(Z)$ es igual al número de vectores de Z .

Pero nótese que, en general, ambos son subespacios de espacios distintos: $\mathcal{L}(Z) \subset \mathcal{V}$ y $\mathcal{N}(A) \subset \mathbb{R}^n$.



Si tenemos un sistema de vectores y aplicamos el método de eliminación (de “izquierda a derecha”), los sucesivos sistemas que van apareciendo tienen la propiedad de que los subespacios generados por los j primeros vectores de los distintos sistemas generan el mismo espacio. Es decir, para Z , y la sucesión de transformaciones elementales τ_k, \dots, τ_n tal que $E = I_{\tau_1 \dots \tau_k}$ es triangular superior unitaria:

$$\text{si } Y = Z_{\tau_1 \dots \tau_k} = ZE, \text{ entonces } Y \text{ y } Z \text{ están acoplados.}$$

es decir,

$$\mathcal{L}([Z_{|1}; \dots Z_{|j}]) = \mathcal{L}([Y_{|1}; \dots Y_{|j}]) \quad \text{para } j = 1 : n.$$

Si algunos vectores de $Y = Z_{\tau_1 \dots \tau_k}$ acaban anulándose (tras la eliminación de “izquierda a derecha”) los correspondientes vectores de Z se pueden quitar y el sistema resultante continúa generando el mismo subespacio (Corolario 9.5.10).

$$\mathcal{L}(Z) = \mathcal{L}(Z_{|\rho}) \quad \text{donde } \rho = (i \mid (Z_{\tau_1 \dots \tau_k})_{|i} \neq \vec{0}).$$

Hemos usado esta idea para encontrar una base de $\mathcal{C}(A)$ entre las columnas de A pre-escalando la matriz mediante eliminación (de “izquierda a derecha”).

Recordando la Proposición 5.2.5, completemos el Corolario 8.1.2 con cuatro nuevas afirmaciones:

Corolario 9.5.13. Para toda matriz cuadrada A de orden n , las siguientes afirmaciones son equivalentes:

1. El rango de A es n (ó A es de rango completo).
2. $Ax = 0$ si y solo si $x = 0$.
3. $Ax = b$ tiene solución única.
4. A no es singular.
5. A es invertible.
6. $A = I_{\tau_1 \dots \tau_k}$. (prod. de mat. elementales)
7. Las columnas de A son linealmente indep.
8. Las filas de A son linealmente independientes.
9. Las columnas de A forman una base de \mathbb{R}^n .
10. Las filas de A forman una base de \mathbb{R}^n .

9.6. Generalizaciones o visiones alternativas (★)

Conjunto generador de un subespacio

En ocasiones es necesario definir el subespacio engendrado por un *conjunto* Z de vectores en lugar del espacio engendrado por un *sistema* Z de vectores.⁴

Definición 9.11. Dado un conjunto Z de vectores de \mathcal{V} , el subespacio $\mathcal{L}(Z)$ de \mathcal{V} engendrado por Z es

$$\mathcal{L}(Z) = \left\{ \vec{v} \in \mathcal{V} \mid \text{existen } a_1, \dots, a_n \in \mathbb{R} \text{ y } \vec{z}_1, \dots, \vec{z}_n \in Z \text{ tales que } \vec{v} = a_1 \vec{z}_1 + \dots + a_n \vec{z}_n \right\}.$$

En un recuadro la Sección 9.2 se dió una segunda caracterización para $\mathcal{L}(Z)$ que, según para qué, resulta más operativa; y que enunciamos de nuevo en relación al *conjunto* generador Z :

Dado un subconjunto Z de vectores de \mathcal{V} , el subespacio $\mathcal{L}(Z)$ de \mathcal{V} engendrado por Z es el subespacio más pequeño de \mathcal{V} que contiene a Z .

Sistema (o conjunto) vacío como sistema generador. Con esta caracterización se hace evidente que el subespacio de \mathcal{V} engendrado por el conjunto vacío es el subespacio $\{\vec{0}\}$ de \mathcal{V} que únicamente contiene el vector nulo $\vec{0} \in \mathcal{V}$; pues es el menor subespacio de \mathcal{V} que contiene el conjunto vacío:

$$\mathcal{L}(\emptyset) = \mathcal{L}([\]) = \{\vec{0}\};$$

recuerde que un *sistema vacío* es equivalente al conjunto vacío (Ecuación 6.7 en la página 90) y que todo subespacio contiene al vector nulo.

Observación. Esto nos indica que la notación $\mathcal{L}([\])$ es incompleta por no indicar de quien es el subespacio que estamos generando con $[\]$. Lo correcto sería escribir algo como⁵

$$\mathcal{L}_{\mathcal{V}}([\]) = \{\vec{0}\} \subset \mathcal{V} \quad \text{ó} \quad \mathcal{L}_{\mathbb{R}^n}([\]) = \{\mathbf{0}\} \subset \mathbb{R}^n;$$

pero arrastrar una notación de la forma $\mathcal{L}_{\mathcal{V}}(X)$, donde se indica explícitamente el espacio vectorial \mathcal{V} , sería pesado y normalmente innecesario, pues sólo pueden surgir dudas cuando el sistema generador es vacío.

Dependencia e independencia lineal Con la última caracterización de subespacio engendrado por un conjunto también podemos caracterizar los conjuntos de vectores linealmente dependientes e independientes.

Diremos que subconjunto Z de vectores de \mathcal{V} , es linealmente dependiente si es posible quitar algún vector de Z y que el conjunto resultante Z^ siga engendrando el mismo subespacio es decir, que $\mathcal{L}(Z) = \mathcal{L}(Z^*)$.*

Un conjunto de vectores que no es linealmente dependiente se dice que es *linealmente independiente*. Por tanto un sistema vacío, $Z = [\]$, es linealmente independiente. Consecuentemente $[\]$ es una base del subespacio $\{\vec{0}\}$, de lo cual se deduce que $\dim(\{\vec{0}\}) = 0$.

Sistemas de vectores y funciones lineales

Las proposiciones 9.1.1 y 9.1.2 implican que $Z\mathbf{x}$ es una función lineal; lo que sugiere el uso de la siguiente:

Notación. Sea Z un sistema de n vectores de \mathcal{V} , denotaremos con f_Z a la función lineal:

$$\begin{aligned} f_Z: \mathbb{R}^n &\longrightarrow \mathcal{V} \\ \mathbf{x} &\longmapsto Z\mathbf{x} \end{aligned}$$

⁴lo necesitaremos en la Sección 11.2; y también es necesario cuando se trabaja con espacios vectoriales de dimensión infinita.

⁵de hecho lo correcto sería escribir $\mathcal{L}_{(\mathcal{V}, +, \cdot)}([\])$; pues para denotar un espacio vectorial es necesario indicar no solo el conjunto “ \mathcal{V} ”, sino las operaciones suma “ $+$ ” y producto por escalares “ \cdot ” (véase el Ejercicio 35 en la página 79).

Proposición 9.6.1. Para cualquier función lineal $f: \mathbb{R}^n \rightarrow \mathcal{V}$, existe Z de n vectores de \mathcal{V} tal que $f(\mathbf{x}) = Z\mathbf{x}$ para todo $\mathbf{x} \in \mathbb{R}^n$.

Demostración.

$$\begin{aligned}
 f(\mathbf{x}) &= f(\mathbf{I}\mathbf{x}) \\
 &= f(x_1\mathbf{I}_{|1} + \cdots + x_n\mathbf{I}_{|n}) && \text{por la definición de matriz por vector de } \mathbb{R}^n \\
 &= x_1f(\mathbf{I}_{|1}) + \cdots + x_nf(\mathbf{I}_{|n}) && \text{por ser } f \text{ una función lineal} \\
 &= [f(\mathbf{I}_{|1}); \dots f(\mathbf{I}_{|n})]\mathbf{x} && \text{por la definición de sistema por vector de } \mathbb{R}^n \\
 &= Z\mathbf{x} && \text{donde } Z = [f(\mathbf{I}_{|1}); \dots f(\mathbf{I}_{|n})].
 \end{aligned}$$

□

Nótese que tanto $Z\mathbf{B}\mathbf{c}$ como $Z\mathbf{B}\mathbf{C}$ son composiciones de funciones lineales, ya que por una parte: si Z es un sistema de m vectores de \mathcal{V} , si \mathbf{B} es de orden m por n , y si $\mathbf{c} \in \mathbb{R}^n$; entonces

$$[f_Z \circ f_{\mathbf{B}}](\mathbf{c}) = f_Z(f_{\mathbf{B}}(\mathbf{c})) = f_Z(\mathbf{B}\mathbf{c}) = Z\mathbf{B}\mathbf{c} = f_{Z\mathbf{B}}(\mathbf{c}),$$

es decir,

$$f_{Z\mathbf{B}}(\mathbf{c}): \mathbb{R}^n \xrightarrow{f_{\mathbf{B}}} \mathbb{R}^m \xrightarrow{f_Z} \mathcal{V}.$$

Y por otra parte, si \mathbf{B} y \mathbf{C} ,

$$[f_Z \circ f_{\mathbf{B}}](\mathbf{C}) = f_Z(f_{\mathbf{B}}(\mathbf{C})) = f_Z(\mathbf{B}\mathbf{C}) = Z\mathbf{B}\mathbf{C} = f_{Z\mathbf{B}}(\mathbf{C});$$

es decir,

$$f_{Z\mathbf{B}}(\mathbf{C}): \mathbb{R}^{q \times n} \xrightarrow{f_{\mathbf{B}}} \mathbb{R}^{m \times n} \xrightarrow{f_Z} \mathcal{V}^n,$$

donde \mathcal{V}^n es el conjunto de sistemas de n vectores de \mathcal{V} .

EJERCICIO 47. Compruebe que el espacio engendrado por un sistema Z es la imagen de f_Z .

EJERCICIO 48. Considere \mathbf{A} , y compruebe que $\mathcal{C}(\mathbf{A}) = \text{imagen}(f_{\mathbf{A}})$, donde $f_{\mathbf{A}}: \mathbb{R}^n \rightarrow \mathbb{R}^m$, $\mathbf{x} \mapsto \mathbf{A}\mathbf{x}$.

Definición 9.12. Llamamos núcleo (o espacio nulo) de la función lineal $f: \mathcal{V} \rightarrow \mathcal{W}$, al siguiente conjunto

$$\mathcal{N}(f) = \{\vec{v} \in \mathcal{V} \mid f(\vec{v}) = \vec{0}\}.$$

EJERCICIO 49. Demuestre la siguiente

Proposición 9.6.2. El núcleo de una función lineal es un subespacio.

Fíjese que la Proposición 7.3.1 es un caso particular en el que la función es $f_{\mathbf{A}}: \mathbb{R}^n \rightarrow \mathbb{R}^m$, $\mathbf{x} \mapsto \mathbf{A}\mathbf{x}$.

EJERCICIO 50. Demuestre la siguiente

Proposición 9.6.3. La imagen de una función lineal es un subespacio.

EJERCICIO 51. Demuestre la siguiente

Proposición 9.6.4. Sean las funciones $f: \mathcal{V} \rightarrow \mathcal{W}$ y $g: \mathcal{W} \rightarrow \mathcal{Y}$, entonces la imagen de $g \circ f$ está contenida en la imagen de g .

EJERCICIO 52. Demuestre el siguiente

Corolario 9.6.5. Sean las funciones $f: \mathcal{V} \rightarrow \mathcal{W}$ y $g: \mathcal{W} \rightarrow \mathcal{Y}$, y donde f es invertible, entonces $\text{imagen}(g \circ f) = \text{imagen}(g)$.

Sistemas linealmente independientes y coordenadas

Proposición 9.6.6. *Un sistema Z de n vectores es independiente si y solo si f_Z es invertible.*

Demostración. Si Z es independiente, tenemos que comprobar que el conjunto de pares $(Z\mathbf{x}, \mathbf{x})$ tales que $\mathbf{x} \in \mathbb{R}^n$ es función. Supongamos que $Z\mathbf{x} = Z\mathbf{y}$, entonces $Z(\mathbf{x} - \mathbf{y}) = \mathbf{0}$. Como Z es independiente, necesariamente $(\mathbf{x} - \mathbf{y}) = \mathbf{0}$; es decir $\mathbf{x} = \mathbf{y}$.

Supongamos el conjunto de pares $(Z\mathbf{x}, \mathbf{x})$ tales que $\mathbf{x} \in \mathbb{R}^n$ es función, es decir, $(f_Z)^{-1}$; y que $Z\mathbf{x} = \mathbf{0}$. Como también $Z\mathbf{0} = \mathbf{0}$, los pares $(Z\mathbf{x}, \mathbf{x})$ y $(Z\mathbf{0}, \mathbf{0})$ pertenecen a $(f_Z)^{-1}$. Como las primeras componentes son nulas, necesariamente \mathbf{x} tiene que ser $\mathbf{0}$. \square

Cuando Z es independiente, la función inversa $(f_Z)^{-1}: \mathcal{L}(Z) \rightarrow \mathbb{R}^n$ se conoce como *función de coordenadas respecto a Z* ; y la denotaremos del siguiente modo:

$$\vec{v}_{/Z} = (f_Z)^{-1}(\vec{v}) \quad \Rightarrow \quad -_{/Z}: \mathcal{V} \rightarrow \mathbb{R}^n;$$

por tanto $\vec{v}_{/Z}$ (las coordenadas de \vec{v} respecto de Z) es *el vector de \mathbb{R}^n con los coeficientes de la única combinación lineal* de los vectores del sistema Z que es igual a \vec{v} . Es decir, $\mathbf{x} = \vec{v}_{/Z}$ es el *único vector de \mathbb{R}^n* tal que

$$Z\mathbf{x} = \vec{v}; \quad \text{pues} \quad Z(\vec{v}_{/Z}) = \vec{v}.$$

Como la función *función de coordenadas respecto a Z* es lineal, sabemos que

$$(\vec{x} + \vec{y})_{/Z} = \vec{x}_{/Z} + \vec{y}_{/Z} \quad \text{y} \quad (a\vec{x})_{/Z} = a(\vec{x}_{/Z}).$$

EJERCICIO 53. Demuestre la siguiente

Proposición 9.6.7. *Si B es una base de \mathcal{V} , entonces para cualquier función lineal $f: \mathcal{V} \rightarrow \mathcal{W}$ existe un sistema Z de vectores de \mathcal{W} tal que $f(\vec{v}) = Z(\vec{v}_{/B})$.*

Los cuatro subespacios fundamentales de una matriz \mathbf{A}

10.1. Suma de subespacios

Sean \mathcal{A} y \mathcal{B} subespacios de \mathcal{V} . Su *suma*, que escribimos como $\mathcal{A} + \mathcal{B}$, es el conjunto de vectores de \mathcal{V} que se pueden escribir como $\vec{a} + \vec{b}$ con $\vec{a} \in \mathcal{A}$ y $\vec{b} \in \mathcal{B}$. Veamos que la suma $\mathcal{A} + \mathcal{B}$ es un subespacio:

Proposición 10.1.1. *Si \mathcal{A} y \mathcal{B} son subespacios de \mathcal{V} , entonces la suma $\mathcal{A} + \mathcal{B}$ también es un subespacio.*

Demostración. $\mathcal{A} + \mathcal{B}$ es no vacío por ser no vacío ni \mathcal{A} ni \mathcal{B} (pues ambos contienen $\vec{0}$). Sean $\vec{a}_1, \vec{a}_2 \in \mathcal{A}$ y $\vec{b}_1, \vec{b}_2 \in \mathcal{B}$. Si $\vec{p} = \vec{a}_1 + \vec{b}_1$ y $\vec{q} = \vec{a}_2 + \vec{b}_2$, entonces $\vec{p}, \vec{q} \in \mathcal{A} + \mathcal{B}$. Veamos que cualquier combinación lineal también pertenece a $\mathcal{A} + \mathcal{B}$. Sean $x, y \in \mathbb{R}$, entonces

$$x(\vec{p}) + y(\vec{q}) = x(\vec{a}_1 + \vec{b}_1) + y(\vec{a}_2 + \vec{b}_2) = (x(\vec{a}_1) + y(\vec{a}_2)) + (x(\vec{b}_1) + y(\vec{b}_2)) \in \mathcal{A} + \mathcal{B},$$

pues $(x(\vec{a}_1) + y(\vec{a}_2)) \in \mathcal{A}$ y $(x(\vec{b}_1) + y(\vec{b}_2)) \in \mathcal{B}$. □

Librería NAcAL para Python

```
A = Sistema([Vector([1, 0, 1, 0]), Vector([0, -1, 0, -1])])
B = Sistema([Vector([1, 1, 1, 1]), Vector([1, 0, 0, 0])])
SubEspacio(A) + SubEspacio(B) # SubEspacio suma de los SubEspacios L(A) y L(B)
```

$$\left\{ \mathbf{v} \in \mathbb{R}^4 \mid \exists \mathbf{p} \in \mathbb{R}^3, \mathbf{v} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix} \mathbf{p} \right\} = \{ \mathbf{v} \in \mathbb{R}^4 \mid [0 \quad -1 \quad 0 \quad 1] \mathbf{v} = (0) \}$$

Proposición 10.1.2. *Sean $\mathbf{A} = [\vec{a}_1; \dots; \vec{a}_p]$ y $\mathbf{B} = [\vec{b}_1; \dots; \vec{b}_q]$ dos sistemas de vectores de \mathcal{V} . Y sea \mathbf{X} un sistema de s vectores de \mathbb{R}^p e \mathbf{Y} un sistema de s vectores de \mathbb{R}^q ; tales que*

$$[\mathbf{A} \# \mathbf{B}] \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} = [\vec{0}; \dots; \vec{0}], \quad \text{es decir} \quad \mathcal{N}(\mathbf{A} \# \mathbf{B}) = \mathcal{C} \left(\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \right),$$

donde $\mathbf{A} \# \mathbf{B}$ es el sistema de $p+q$ vectores resultante de concatenar \mathbf{A} y \mathbf{B} . Entonces $\mathcal{L}(\mathbf{A}\mathbf{X}) = \mathcal{L}(\mathbf{A}) \cap \mathcal{L}(\mathbf{B})$.

Si además la matriz “apilada” $\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix}$ es de rango completo por columnas, y si \mathbf{A} y \mathbf{B} son linealmente independientes, también lo es el sistema $\mathbf{A}\mathbf{X}$ de vectores de \mathcal{V} .

Demostración. Para ver que $\mathcal{L}(\mathbf{A}\mathbf{X}) = \mathcal{L}(\mathbf{A}) \cap \mathcal{L}(\mathbf{B})$ tenemos que demostrar la pertenencia en ambos sentidos:

1. Demostremos que $\mathbf{A}\mathbf{X}_{|j} \in \mathcal{L}(\mathbf{A}) \cap \mathcal{L}(\mathbf{B})$.

$$[\mathbf{A} \# \mathbf{B}] \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix}_{|j} = \vec{0}; \Rightarrow \mathbf{A}\mathbf{X}_{|j} + \mathbf{B}\mathbf{Y}_{|j} = \vec{0}; \Rightarrow \mathbf{A}\mathbf{X}_{|j} = -\mathbf{B}\mathbf{Y}_{|j}; \Rightarrow \mathbf{A}\mathbf{X}_{|j} \in \mathcal{L}(\mathbf{A}) \cap \mathcal{L}(\mathbf{B}),$$

pues $\mathbf{A}\mathbf{X}_{|j}$ también es una combinación lineal de los vectores de \mathbf{B} .

2. Demostremos que si $\vec{v} \in \mathcal{L}(\mathbf{A}) \cap \mathcal{L}(\mathbf{B})$ entonces $\vec{v} \in \mathcal{L}(\mathbf{A}\mathbf{X})$, es decir, $\vec{v} = (\mathbf{A}\mathbf{X})\mathbf{z}$ para algún $\mathbf{z} \in \mathbb{R}^s$.

Sea $\vec{v} \in \mathcal{L}(\mathbf{A}) \cap \mathcal{L}(\mathbf{B})$, entonces existen \mathbf{x} e \mathbf{y} tales que $\vec{v} = \mathbf{A}\mathbf{x}$ y $\vec{v} = -\mathbf{B}\mathbf{y}$. Por tanto $\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} = \vec{0}$; es decir $\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \in \mathcal{N}(\mathbf{A} \# \mathbf{B}) = \mathcal{C}\left(\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix}\right)$. Así pues, existe $\mathbf{z} \in \mathbb{R}^s$ tal que $\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \mathbf{z}$. Consecuentemente $\mathbf{x} = \mathbf{X}\mathbf{z}$ y por tanto $\vec{v} = \mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{X}\mathbf{z}$.

Demostremos que si además la matriz “apilada” $\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix}$ es de rango completo por columnas, y si \mathbf{A} y \mathbf{B} son linealmente independientes, también es linealmente independiente el sistema $\mathbf{A}\mathbf{X}$ de vectores de \mathcal{V} .

Supongamos que $(\mathbf{A}\mathbf{X})\mathbf{z} = \vec{0}$. Entonces $\mathbf{A}(\mathbf{X}\mathbf{z}) = \vec{0}$, y como \mathbf{A} es linealmente independiente, $\mathbf{X}\mathbf{z} = \vec{0}$. Por otra parte, como $\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \mathbf{z} \in \mathcal{N}(\mathbf{A} \# \mathbf{B})$, tendremos que $\vec{0} = \mathbf{A}(\mathbf{X}\mathbf{z}) + \mathbf{B}(\mathbf{Y}\mathbf{z}) = \mathbf{B}(\mathbf{Y}\mathbf{z})$; y al ser \mathbf{B} linealmente independiente $\mathbf{Y}\mathbf{z} = \vec{0}$. Pero entonces tenemos que $\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \mathbf{z} = \begin{pmatrix} \vec{0} \\ \vec{0} \end{pmatrix}$; y como $\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix}$ es de rango completo por columnas, necesariamente $\mathbf{z} = \vec{0}$. \square

Consecuentemente, si \mathbf{A} y \mathbf{B} son linealmente independientes y si las columnas de $\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix}$ son una base de $\mathcal{N}(\mathbf{A} \# \mathbf{B})$, entonces $\mathbf{A}\mathbf{X} = [\mathbf{A}\mathbf{X}_{|1}; \dots \mathbf{A}\mathbf{X}_{|s}]$ es una base de $\mathcal{L}(\mathbf{A}) \cap \mathcal{L}(\mathbf{B})$.

Corolario 10.1.3. Si \mathcal{A} y \mathcal{B} son subespacios de \mathcal{V} entonces $\dim(\mathcal{A} + \mathcal{B}) = \dim(\mathcal{A}) + \dim(\mathcal{B}) - \dim(\mathcal{A} \cap \mathcal{B})$.

Demostración. (Esta demostración muestra un método para encontrar una base de $\mathcal{A} \cap \mathcal{B}$)

Por lo de arriba sabemos que si tomamos una base \mathbf{A} de \mathcal{A} y una base \mathbf{B} de \mathcal{B} , las dimensiones de $\mathcal{L}(\mathbf{A}) \cap \mathcal{L}(\mathbf{B})$ y de $\mathcal{N}(\mathbf{A} \# \mathbf{B})$ son iguales; i.e. coinciden con el número de “soluciones especiales” del sistema $[\mathbf{A} \# \mathbf{B}]\mathbf{z} = \vec{0}$.

Sean $\mathbf{A} = [\vec{a}_1; \dots; \vec{a}_p]$ y $\mathbf{B} = [\vec{b}_1; \dots; \vec{b}_q]$ bases de \mathcal{A} y \mathcal{B} respectivamente; y consideremos el sistema $\mathbf{A} \# \mathbf{B} = [\vec{a}_1; \dots; \vec{a}_p; \vec{b}_1; \dots; \vec{b}_q]$ resultante de concatenar \mathbf{A} y \mathbf{B} . Evidentemente $\mathcal{L}(\mathbf{A} \# \mathbf{B}) = \mathcal{A} + \mathcal{B}$. Aplicando la eliminación “de izquierda a derecha” sobre el sistema $\mathbf{A} \# \mathbf{B}$ (como en la Proposición 9.5.6 o la Ecuación 9.4 en la página 117) observaremos el siguiente esquema:

$$\left[\begin{array}{c} [\mathbf{A} \# \mathbf{B}] \\ \mathbf{I} \end{array} \right]_{\tau_1 \dots \tau_k} = \left[\begin{array}{c} [\mathbf{A} \# \mathbf{B}] \mathbf{I}_{\tau_1 \dots \tau_k} \\ \mathbf{I}_{\tau_1 \dots \tau_k} \end{array} \right] = \left[\begin{array}{c} [\mathbf{A} \# \mathbf{B}] \mathbf{E} \\ \mathbf{E} \end{array} \right]; \quad \text{donde } \mathbf{E} = \mathbf{I}_{\tau_1 \dots \tau_k}.$$

Con la eliminación al sistema $\mathbf{A} \# \mathbf{B}$ habremos anulado todos los vectores que eran combinación lineal de los que les anteceden. Por tanto, los vectores no nulos de $[\mathbf{A} \# \mathbf{B}]_{\tau_1 \dots \tau_k}$ constituyen una base de $\mathcal{A} + \mathcal{B}$ (Corolario 9.5.10). Así, si tras la eliminación se anulan s vectores de $\mathbf{A} \# \mathbf{B}$, entonces $\boxed{\dim(\mathcal{A} + \mathcal{B}) = p + q - s}$.

De propina podemos construir una base de $\mathcal{A} \cap \mathcal{B}$. Sea $\mathbf{X} =_{(1:p)} \mathbf{E}$ la submatriz formada por las p primeras filas de \mathbf{E} , e \mathbf{Y} la submatriz formada por las q últimas. Entonces, si ν es la lista con los índices de los s vectores de $\mathbf{A} \# \mathbf{B}$ que se han anulado en el proceso de eliminación, tenemos que

$$\left([\mathbf{A} \# \mathbf{B}]_{\tau_1 \dots \tau_k} \right)_{|\nu} = [\mathbf{A} \# \mathbf{B}] \mathbf{E}_{|\nu} = [\vec{0}; \dots \vec{0};], \quad \text{con } \mathbf{E}_{|\nu} = \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix}_{|\nu} \Rightarrow \mathbf{A}\mathbf{X}_{|\nu} + \mathbf{B}\mathbf{Y}_{|\nu} = [\vec{0}; \dots \vec{0};].$$

Así, por la Proposición 10.1.2 sabemos que $\dim \mathcal{A} \cap \mathcal{B} = s$ y que \mathbf{AX}_{ν} es una base de $\mathcal{L}(\mathcal{A}) \cap \mathcal{L}(\mathcal{B})$. \square

EJERCICIO 54. Encuentre con NAcAL una base de la intersección de los SubEspacios generados por:

```
A = Sistema([Vector([1, 0, 1, 0]), Vector([0, -1, 0, -1])])
B = Sistema([Vector([1, 1, 1, 1]), Vector([1, 0, 0, 0])])
```

Y ahora verifiquemos que se cumple que $\dim(\mathcal{A} + \mathcal{B}) = \dim(\mathcal{A}) + \dim(\mathcal{B}) - \dim(\mathcal{A} \cap \mathcal{B})$:

Librería NAcAL para Python

```
DimSuma      = (SubEspacio(A) + SubEspacio(B)).dim # dimensión Suma A + B
DimA         = SubEspacio(A).dim                  # dimensión A
DimB         = SubEspacio(B).dim                  # dimensión B
DimIntersec  = (SubEspacio(A) & SubEspacio(B)).dim # dimensión Intersecc
print( DimSuma, '=', DimA, '+', DimB, '-', DimIntersec)
```

```
3 = 2 + 2 - 1
```

10.1.1. Suma directa y subespacios suplementarios

Definición 10.1 (Suma directa). Si \mathcal{A} y \mathcal{B} son subespacios de \mathcal{V} , y su intersección es $\mathcal{A} \cap \mathcal{B} = \vec{0}$; entonces la suma de los subespacios \mathcal{A} y \mathcal{B} se denomina suma directa y se denota con $\mathcal{A} \oplus \mathcal{B}$.

Definición 10.2 (Espacios suplementarios). Si \mathcal{A} y \mathcal{B} son subespacios de \mathcal{V} , y su suma directa es todo el espacio, $\mathcal{A} \oplus \mathcal{B} = \mathcal{V}$, entonces decimos que \mathcal{A} y \mathcal{B} son subespacios suplementarios.

Observación. Espacios suplementarios a uno dado hay muchos. Viendo la Figura 6.1 en la página 86 es fácil darse cuenta que cualquier función lineal $g: \mathcal{V} \rightarrow \mathcal{W}$ es un espacio suplementario de $\{\vec{0}\} \times \mathcal{W}$.

Veamos que si $\mathcal{V} = \mathcal{A} \oplus \mathcal{B}$, cualquier vector $\vec{v} \in \mathcal{V}$ se puede descomponer de manera única en la suma de dos vectores, uno que pertenece a \mathcal{A} y otro que pertenece a \mathcal{B} .

Proposición 10.1.4. Si $\vec{v} = \vec{a} + \vec{b}$ con $\vec{a} \in \mathcal{A}$, $\vec{b} \in \mathcal{B}$ y $\mathcal{A} \cap \mathcal{B} = \{\vec{0}\}$, entonces \vec{a} y \vec{b} son únicos.

Demostración. Imagine que $\vec{v} = \vec{a} + \vec{b}$ y $\vec{v} = \vec{a}_* + \vec{b}_*$, con $\vec{a}, \vec{a}_* \in \mathcal{A}$ y $\vec{b}, \vec{b}_* \in \mathcal{B}$. Entonces

$$\vec{v} - \vec{v} = (\vec{a} + \vec{b}) - (\vec{a}_* + \vec{b}_*) = \underbrace{(\vec{a} - \vec{a}_*)}_{\in \mathcal{A}} - \underbrace{(\vec{b} + \vec{b}_*)}_{\in \mathcal{B}} = \vec{0} \implies (\vec{a} - \vec{a}_*) = -(\vec{b} + \vec{b}_*) \in \mathcal{B}.$$

Es decir, $(\vec{a} - \vec{a}_*) \in \mathcal{A} \cap \mathcal{B} = \{\vec{0}\}$, y por tanto $(\vec{a} - \vec{a}_*) = \vec{0} = (\vec{b} - \vec{b}_*)$, es decir, $\vec{a} = \vec{a}_*$ y $\vec{b} = \vec{b}_*$. \square

EJERCICIO 55. Sea $\mathcal{V} = \mathcal{A} \oplus \mathcal{B}$ y considere $f: \mathcal{V} \longrightarrow \mathcal{A} \times \mathcal{B}$, tal que $\vec{v} = \vec{a} + \vec{b}$ con $\vec{a} \in \mathcal{A}$ y $\vec{b} \in \mathcal{B}$.

$$\vec{v} \longmapsto (\vec{a}, \vec{b})$$

- (a) Demuestre que f es lineal.
- (b) Considere ahora una función parecida a la anterior, salvo que ahora a cada vector \vec{v} le asocia únicamente la primera de las componentes de los anteriores pares cartesianos; es decir, a cada \vec{v} le asocia la componente \vec{a} correspondiente al subespacio \mathcal{A} ; por tanto $f_{\mathcal{A}}: \mathcal{V} \longrightarrow \mathcal{A}$ de tal manera que $\vec{v} - \vec{a} \in \mathcal{B}$.

$$\vec{v} \longmapsto \vec{a}$$

Demuestre que $f_{\mathcal{A}}$ es lineal.¹

- (c) (Opcional) Demuestre que $\mathcal{N}(f_{\mathcal{A}}) = \mathcal{B}$.

¹Esta función $f_{\mathcal{A}}$ es la proyección sobre el subespacio \mathcal{A} de \mathcal{V} paralela a \mathcal{B} . En la Lección 12 trataremos del caso particular en el que \mathcal{A} y \mathcal{B} son perpendiculares.

10.2. Los cuatro subespacios fundamentales de una matriz \mathbf{A}



En esta sección veremos que toda matriz \mathbf{A} , de orden m por n , define una descomposición de \mathbb{R}^m y de \mathbb{R}^n en subespacios suplementarios. En particular:

- \mathbb{R}^n queda descompuesto en el *espacio nulo* $\mathcal{N}(\mathbf{A})$ y en el *espacio fila* $\mathcal{C}(\mathbf{A}^\top)$:

$$\mathbb{R}^n = \mathcal{N}(\mathbf{A}) \oplus \mathcal{C}(\mathbf{A}^\top).$$

- \mathbb{R}^m queda descompuesto en el *espacio columna* $\mathcal{C}(\mathbf{A})$ y en el *espacio nulo por la izquierda* $\mathcal{N}(\mathbf{A}^\top)$:

$$\mathbb{R}^m = \mathcal{C}(\mathbf{A}) \oplus \mathcal{N}(\mathbf{A}^\top).$$

El profesor Strang (2007) se refiere a estos cuatro subespacios: $\mathcal{N}(\mathbf{A})$, $\mathcal{C}(\mathbf{A})$, $\mathcal{N}(\mathbf{A}^\top)$ y $\mathcal{C}(\mathbf{A}^\top)$, con el nombre de “los cuatro subespacios fundamentales de la matriz \mathbf{A} ”.

La notación indica que los nuevos espacios $\mathcal{C}(\mathbf{A}^\top)$ y $\mathcal{N}(\mathbf{A}^\top)$ son el espacio columna y el nulo de \mathbf{A}^\top .

10.2.1. El espacio fila

El espacio fila de \mathbf{A} es el conjunto de las combinaciones lineales de las filas:

$$\mathcal{L}(\text{filas de } \mathbf{A}) = \mathcal{L}(\text{columnas de } \mathbf{A}^\top) = \mathcal{C}(\mathbf{A}^\top),$$

que es un subespacio de \mathbb{R}^n , pues cada fila tiene n componentes (cada uno situado en una columna diferente).

Como las filas de \mathbf{A} son las columnas de \mathbf{A}^\top , tenemos que

$$\begin{aligned} \mathcal{C}(\mathbf{A}^\top) &= \{\mathbf{x} \mid \text{existe } \mathbf{y} \in \mathbb{R}^m : \mathbf{x} = (\mathbf{A}^\top)\mathbf{y}\} \\ &= \{\mathbf{x} \mid \text{existe } \mathbf{y} \in \mathbb{R}^m : \mathbf{x} = \mathbf{y}\mathbf{A}\} \end{aligned} \quad \text{ya que } (\mathbf{A}^\top)\mathbf{y} = \mathbf{y}\mathbf{A};$$

es decir, el *espacio fila* $\mathcal{C}(\mathbf{A}^\top)$ está formado por los vectores que se pueden describir como combinación lineal de las filas de \mathbf{A} . Como el rango de una matriz no cambia al transponerla (Proposición 5.2.5), tenemos que

$$\dim \mathcal{C}(\mathbf{A}) = \text{rg}(\mathbf{A}) = \text{rg}(\mathbf{A}^\top) = \dim \mathcal{C}(\mathbf{A}^\top).$$

Es decir, para una matriz \mathbf{A} de rango r tenemos que $\dim \mathcal{C}(\mathbf{A}^\top) = r$.

Fíjese que si \mathbf{A} es de orden $m \times n$, como $\dim \mathcal{N}(\mathbf{A}) = n - r$, resulta que $\dim \mathcal{C}(\mathbf{A}^\top) + \dim \mathcal{N}(\mathbf{A}) = n$, que es igual a la dimensión de todo el espacio \mathbb{R}^n . Así, si probamos que $\mathcal{N}(\mathbf{A}) \cap \mathcal{C}(\mathbf{A}^\top) = \{\mathbf{0}\}$ (i.e., que la dimensión de la intersección es cero), por la Proposición 10.1.4 habremos demostrado que $\mathbb{R}^n = \mathcal{C}(\mathbf{A}^\top) \oplus \mathcal{N}(\mathbf{A})$:

Proposición 10.2.1. Para toda \mathbf{A} se verifica que $\mathcal{N}(\mathbf{A}) \cap \mathcal{C}(\mathbf{A}^\top) = \{\mathbf{0}\}$.

Demostración. Para todo vector $\mathbf{y} \in \mathcal{C}(\mathbf{A}^\top)$ (es decir, de la forma $\mathbf{y} = \mathbf{z}\mathbf{A}$) y para todo vector $\mathbf{x} \in \mathcal{N}(\mathbf{A})$ (es decir, tal que $\mathbf{A}\mathbf{x} = \mathbf{0}$) el producto punto $\mathbf{y} \cdot \mathbf{x}$ es nulo:

$$\mathbf{y} \cdot \mathbf{x} = \mathbf{z}\mathbf{A}\mathbf{y} = \mathbf{z} \cdot \mathbf{0} = 0.$$

Por tanto, si \mathbf{v} pertenece simultáneamente a los dos subespacios, entonces

$$\mathbf{v} \cdot \mathbf{v} = \sum v_i^2 = 0 \Rightarrow v_i = 0;$$

es decir, si $\mathbf{v} \in \mathcal{N}(\mathbf{A}) \cap \mathcal{C}(\mathbf{A}^\top)$ entonces necesariamente $\mathbf{v} = \mathbf{0}$. □

Puesto que $\mathcal{N}(\mathbf{A}) \cap \mathcal{C}(\mathbf{A}^\top) = \{\mathbf{0}\}$, por la Proposición 10.1.4 concluimos que $\mathbb{R}^n = \mathcal{C}(\mathbf{A}^\top) \oplus \mathcal{N}(\mathbf{A})$.

10.2.2. El espacio nulo por la izquierda

Fíjese que $\mathcal{N}(\mathbf{A}^\top)$ es el conjunto de soluciones del sistema $(\mathbf{A}^\top) \mathbf{y} = \mathbf{0}$, que nos pregunta ¿qué combinaciones lineales de las columnas de \mathbf{A}^\top son iguales a cero? es decir ¿qué combinaciones lineales de las *filas* de \mathbf{A} son iguales a cero? Por tanto, $\mathcal{N}(\mathbf{A}^\top)$ son las soluciones del siguiente *sistema de ecuaciones homogéneo*

$$\mathbf{y}\mathbf{A} = \mathbf{0},$$

lo que justifica el nombre de *espacio nulo por la izquierda* (pues el vector de incógnitas \mathbf{y} multiplica por la izquierda). Así, definimos el *espacio nulo por la izquierda* de \mathbf{A} (de orden m por n) como

$$\mathcal{N}(\mathbf{A}^\top) = \{\mathbf{y} \mid \mathbf{y}\mathbf{A} = \mathbf{0}\}.$$

Si llamamos \mathbf{B} a \mathbf{A}^\top , aplicando el resultado de la sección anterior tenemos que $\mathbb{R}^m = \mathcal{C}(\mathbf{B}^\top) \oplus \mathcal{N}(\mathbf{B})$.

Es decir, $\boxed{\mathbb{R}^m = \mathcal{C}(\mathbf{A}) \oplus \mathcal{N}(\mathbf{A}^\top)}$.

10.2.3. El gran esquema

Así, dada \mathbf{A} de rango r , los espacios \mathbb{R}^m y \mathbb{R}^n quedan descompuestos según el siguiente esquema:

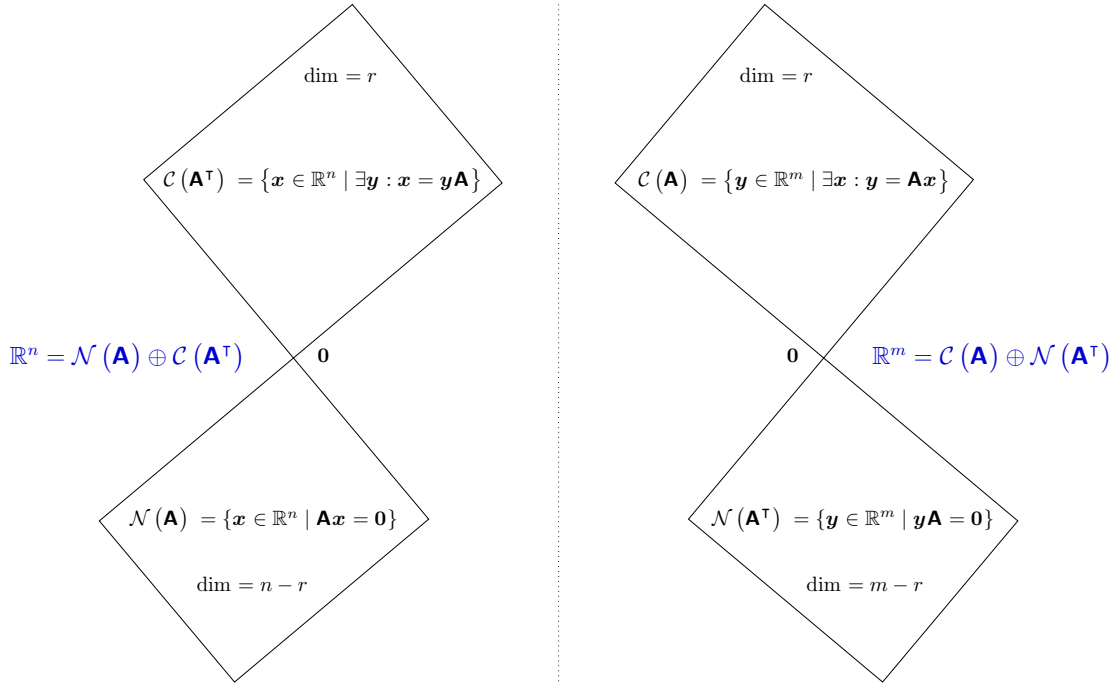


Figura 10.1: Esquema de descomposición de los cuatro subespacios fundamentales de una matriz (parte 1)

Ejemplo 24. Para $\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 2 & 2 & 0 \\ 0 & -1 & -1 & 0 \end{bmatrix}$ de rango 1, comprobemos el Corolario 10.1.3, es decir, que

$$3 = \dim \mathbb{R}^m = \dim \mathcal{N}(\mathbf{A}^\top) + \dim \mathcal{C}(\mathbf{A}) - \dim (\mathcal{N}(\mathbf{A}^\top) \cap \mathcal{C}(\mathbf{A})) = 2 + 1 - 0$$

$$4 = \dim \mathbb{R}^n = \dim \mathcal{N}(\mathbf{A}) + \dim \mathcal{C}(\mathbf{A}^\top) - \dim (\mathcal{N}(\mathbf{A}) \cap \mathcal{C}(\mathbf{A}^\top)) = 3 + 1 - 0$$

```
A = Matrix([ [0,1,1,0], [0,2,2,0],[0,-1,-1,0] ])
print( A.rg() )           # Rango de A
```

Librería NAcAL para Python

1

```
Cola = SubEspacio(A.sis()); # Espacio Columna de A (en R3)
print(Cola.dim); Cola
```

Librería NAcAL para Python

1

$$\left\{ v \in \mathbb{R}^3 \mid \exists p \in \mathbb{R}^1, v = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} p \right\} = \left\{ v \in \mathbb{R}^3 \mid \begin{bmatrix} -2 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} v = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\}$$

Librería NAcAL para Python

```
NulAT = SubEspacio( ~A ); # Espacio Nulo Izda de A (en R3)
print(NulAT.dim); NulAT
```

2

$$\left\{ v \in \mathbb{R}^3 \mid \exists p \in \mathbb{R}^2, v = \begin{bmatrix} -2 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} p \right\} = \{ v \in \mathbb{R}^3 \mid \begin{bmatrix} -1 & -2 & 1 \end{bmatrix} v = (0,) \}$$

Librería NAcAL para Python

```
Fila = SubEspacio( (~A).sis() ); # Espacio Fila de A (en R4)
print(Fila.dim); Fila
```

1

$$\left\{ v \in \mathbb{R}^4 \mid \exists p \in \mathbb{R}^1, v = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} p \right\} = \left\{ v \in \mathbb{R}^4 \mid \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} v = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right\}$$

Librería NAcAL para Python

```
NulA = SubEspacio( A ); # Espacio Nulo de A (en R4)
print(NulA.dim); NulA
```

3

$$\left\{ v \in \mathbb{R}^4 \mid \exists p \in \mathbb{R}^3, v = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} p \right\} = \{ v \in \mathbb{R}^4 \mid \begin{bmatrix} 0 & 1 & 1 & 0 \end{bmatrix} v = (0,) \}$$

Librería NAcAL para Python

```
print( (Cola + NulAT).dim ) # Dim. del SubEspacio suma (suma es todo R3)
print( (Cola & NulAT).dim ) # Dim. de la intersección (que solo contiene el cero)
```

3

0

(son espacios suplementarios)

```
print( (Fila + NulA).dim ) # Dim. del SubEspacio suma (suma es todo  $\mathbb{R}^4$ )
print( (Fila & NulA).dim ) # Dim. de la intersección (que solo contiene el cero)
```

4

0

(son espacios suplementarios)

10.3. Encontrando bases para los espacios $\mathcal{C}(\mathbf{A}^\top)$ y $\mathcal{N}(\mathbf{A}^\top)$

Sabemos que aplicando la eliminación *sobre las columnas* de \mathbf{A} podemos encontrar bases para los espacios *columna* $\mathcal{C}(\mathbf{A})$ y *nulo* $\mathcal{N}(\mathbf{A})$. Por tanto, sabemos que aplicando la eliminación *sobre las columnas* de \mathbf{A}^\top encontraremos bases para los nuevos subespacios $\mathcal{C}(\mathbf{A}^\top)$ y $\mathcal{N}(\mathbf{A}^\top)$. ¡Pero seguir esta estrategia supondría que, para encontrar bases de los cuatro espacios, necesitamos aplicar el método de eliminación dos veces! una para \mathbf{A} y otra para \mathbf{A}^\top .

¿Es posible encontrar bases para cada uno de los cuatro subespacios aplicando únicamente eliminación sobre las *columnas* de \mathbf{A}^\top ?... ¡Afortunadamente sí!

La siguiente proposición nos da la pista necesaria para hacerlo.

Proposición 10.3.1. Sean \mathbf{A} de orden m por n y \mathbf{E} invertible de orden n , entonces $\mathcal{N}(\mathbf{A}^\top) = \mathcal{N}((\mathbf{AE})^\top)$, es decir, dado $\mathbf{y} \in \mathbb{R}^m$ se verifica que

$$\mathbf{yA} = \mathbf{0} \iff \mathbf{yAE} = \mathbf{0}.$$

Demostración. Veamos las implicaciones en uno y otro sentido.

- Si $\mathbf{y} \in \mathcal{N}(\mathbf{A}^\top) \implies \mathbf{yA} = \mathbf{0} \implies \mathbf{yAE} = \mathbf{0E} \implies \mathbf{y(AE)} = \mathbf{0} \implies \mathbf{y} \in \mathcal{N}((\mathbf{AE})^\top)$
- Si $\mathbf{y} \in \mathcal{N}((\mathbf{AE})^\top) \implies \mathbf{y(AE)} = \mathbf{0} \implies \mathbf{yAEE}^{-1} = \mathbf{0E}^{-1} \implies \mathbf{yA} = \mathbf{0} \implies \mathbf{y} \in \mathcal{N}(\mathbf{A}^\top).$

□

Consecuentemente, si \mathbf{A} es de orden m por n y \mathbf{E} es invertible y de orden n , entonces

- un sistema de vectores formado por varias filas $[\mathbf{A}_{i_1}; \dots \mathbf{A}_{i_r}]$ es linealmente independiente si y sólo si el sistema de vectores $[\mathbf{AE}_{i_1}; \dots \mathbf{AE}_{i_r}]$ es linealmente independiente.
- la dimensión del espacio fila de \mathbf{A} es igual a la dimensión del espacio fila de \mathbf{AE} .

Así pues, podemos aplicar transformaciones elementales a las columnas de una matriz para obtener otra más simple donde verificar qué filas son combinación lineal de las filas que la anteceden y cuales forman un sistema linealmente independiente.

El siguiente lema indica que las filas con pivote de cualquier forma pre-escalada son linealmente independientes.

EJERCICIO 56. Demuestre que

Lema 10.3.2. Si \mathbf{K} está pre-escalada, las filas con pivote son linealmente independientes. Además, cada fila sin pivote es combinación lineal de las filas con pivote que la preceden.

Por tanto, la selección de las filas con pivote de una forma pre-escalada \mathbf{K} de la matriz \mathbf{A} forma una base del espacio fila de la matriz \mathbf{K} (que generalmente es distinto del espacio fila de \mathbf{A}). Veamos que la misma selección aplicada a las filas de \mathbf{A} forma una base del espacio fila de \mathbf{A} .

Encontrando una base para el espacio fila de \mathbf{A}

Sea $\mathbf{E} = \mathbf{I}_{\tau_1 \dots \tau_k}$ tal que \mathbf{AE} es pre-escalada; y sea $\varphi = (\varphi_1, \dots, \varphi_r)$ la lista de índices de las filas con pivote. Entonces la matriz ${}_{\varphi}\mathbf{I}$ selecciona las filas pivote de \mathbf{AE} que forman una base del espacio fila de \mathbf{AE} :

$$({}_{(1,2,4)}\mathbf{I})(\mathbf{AE}) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ -2 & -4 & 0 & 0 \\ 4 & 4 & 0 & 0 \\ -1 & -1 & 1 & 0 \\ -6 & -7 & 3 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -2 & -4 & 0 & 0 \\ -1 & -1 & 1 & 0 \end{bmatrix} = {}_{(1,2,4)}(\mathbf{AE}).$$

Como ${}_{\varphi}\mathbf{I}$ selecciona una base del espacio fila de \mathbf{AE} , entonces para cualquier vector \mathbf{u} existe un único vector \mathbf{x} (una única combinación lineal de las filas con pivote) tal que

$$\mathbf{x}({}_{\varphi}\mathbf{AE}) = \mathbf{uAE}.$$

Pero entonces también la ecuación $\mathbf{x}({}_{\varphi}\mathbf{A}) = \mathbf{uA}$ tiene una única solución ya que por ser \mathbf{E} invertible

$$\mathbf{x}({}_{\varphi}\mathbf{A}) = \mathbf{uA} \iff \mathbf{x}({}_{\varphi}\mathbf{AE}) = \mathbf{uAE}.$$

Luego ${}_{\varphi}\mathbf{A}$ (la misma selección, pero de las filas de \mathbf{A}) selecciona una base del espacio de fila de \mathbf{A} .

Consecuentemente, la dimensión del espacio fila de una matriz coincide con la dimensión del espacio fila de cualquiera de sus formas pre-escaladas, que a su vez coincide con el rango.



Algoritmo para encontrar una base de $\mathcal{C}(\mathbf{A}^T)$

- Pre-escalamos \mathbf{A}
- Identificamos los índices φ de las filas con pivote de la forma escalonada \mathbf{K}
- Las filas de ${}_{\varphi}\mathbf{A}$ forman una base $\mathcal{C}(\mathbf{A}^T)$

Ejemplo 25.

$$\begin{bmatrix} \mathbf{A} \\ \mathbf{I} \end{bmatrix} = \begin{bmatrix} 1 & -2 & -1 & -1 \\ -2 & 0 & 1 & 2 \\ 4 & -4 & -3 & -4 \\ -1 & 1 & 1 & 1 \\ -6 & 5 & 5 & 6 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{\begin{matrix} [(2)\mathbf{1}+2] \\ [(1)\mathbf{1}+3] \\ [(1)\mathbf{1}+4] \end{matrix}} \begin{bmatrix} 1 & 0 & 0 & 0 \\ -2 & -4 & -1 & 0 \\ 4 & 4 & 1 & 0 \\ -1 & -1 & 0 & 0 \\ -6 & -7 & -1 & 0 \\ \hline 1 & 2 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{\begin{matrix} [(4)\mathbf{3}] \\ [(-1)\mathbf{2}+3] \end{matrix}} \begin{bmatrix} 1 & 0 & 0 & 0 \\ -2 & -4 & 0 & 0 \\ 4 & 4 & 0 & 0 \\ -1 & -1 & 1 & 0 \\ -6 & -7 & 3 & 0 \\ \hline 1 & 2 & 2 & 1 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{K} \\ \mathbf{E} \end{bmatrix}$$

De este proceso de eliminación, deducimos que:

$$\mathcal{C}(\mathbf{A}) = \mathcal{L}([\mathbf{K}_{|1}; \mathbf{K}_{|2}; \mathbf{K}_{|3};]), \quad \mathcal{C}(\mathbf{A}^T) = \mathcal{L}([{}_{|1}\mathbf{A}; {}_{|2}\mathbf{A}; {}_{|4}\mathbf{A};]) \quad \text{y} \quad \mathcal{N}(\mathbf{A}) = \mathcal{L}([{}_{|4}\mathbf{E};]).$$

Encontrando una base para el espacio nulo por la izquierda.

Nos falta un método para encontrar una base de $\mathcal{N}(\mathbf{A}^T)$.

Por la Proposición 10.3.1 sabemos que $\mathcal{N}(\mathbf{A}^T) = \mathcal{N}((\mathbf{AE})^T)$. Así que la estrategia es simplificar al máximo alguna forma pre-escalada de \mathbf{A} , y mirar ahí qué combinaciones de las filas son nulas. Por tanto

aplicaremos la eliminación Gauss-Jordan hasta llegar a una forma escalonada *reducida* \mathbf{R} (donde los pivotes son unos con ceros a derecha e izquierda). Continuando el ejemplo anterior:

$$\begin{aligned}
 \mathbf{A} = \begin{bmatrix} 1 & -2 & -1 & -1 \\ -2 & 0 & 1 & 2 \\ 4 & -4 & -3 & -4 \\ -1 & 1 & 1 & 1 \\ -6 & 5 & 5 & 6 \end{bmatrix} &\xrightarrow[\tau]{\begin{matrix} [(2)1+2] \\ [(1)1+3] \\ [(1)1+4] \\ [(4)3] \\ [(-1)2+3] \end{matrix}} \begin{bmatrix} 1 & 0 & 0 & 0 \\ -2 & -4 & 0 & 0 \\ 4 & 4 & 0 & 0 \\ -1 & -1 & 1 & 0 \\ -6 & -7 & 3 & 0 \end{bmatrix} \xrightarrow[\tau]{\begin{matrix} [(2)1] \\ [(-1)2+1] \end{matrix}} \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & -4 & 0 & 0 \\ 4 & 4 & 0 & 0 \\ -1 & -1 & 1 & 0 \\ -5 & -7 & 3 & 0 \end{bmatrix} \\
 &\xrightarrow[\tau]{\begin{matrix} [(1)3+1] \\ [(1)3+2] \end{matrix}} \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & -4 & 0 & 0 \\ 4 & 4 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -2 & -4 & 3 & 0 \end{bmatrix} \xrightarrow[\tau]{\begin{matrix} [(\frac{1}{2})1] \\ [(-\frac{1}{4})2] \end{matrix}} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 2 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 1 & 3 & 0 \end{bmatrix} = \mathbf{R}.
 \end{aligned}$$

Ahora es inmediato ver en \mathbf{R} que cada fila sin pivote es combinación de las filas con pivote que la anteceden:

$$\begin{cases} {}_3|\mathbf{R} = 2({}_1|\mathbf{R}) - 1({}_2|\mathbf{R}) \\ \text{y} \\ {}_5|\mathbf{R} = -1({}_1|\mathbf{R}) + 1({}_2|\mathbf{R}) + 3({}_3|\mathbf{R}) \end{cases} \Rightarrow \begin{matrix} (-2, & 1, & 1, & 0, & 0,) \\ (1, & -1, & 0, & -3, & 1,) \end{matrix} \mathbf{R} = \mathbf{0}$$

Puesto que $\mathcal{N}(\mathbf{A}^\top) = \mathcal{N}(\mathbf{R}^\top)$, las mismas relaciones se verifican en la matriz original \mathbf{A} ;

$$\begin{cases} {}_3|\mathbf{A} = 2({}_1|\mathbf{A}) - 1({}_2|\mathbf{A}) \\ \text{y} \\ {}_5|\mathbf{A} = -1({}_1|\mathbf{A}) + 1({}_2|\mathbf{A}) + 3({}_3|\mathbf{A}) \end{cases} \Rightarrow \begin{matrix} (-2, & 1, & 1, & 0, & 0,) \\ (1, & -1, & 0, & -3, & 1,) \end{matrix} \mathbf{A} = \mathbf{0}$$

👉 En el proceso de eliminación $\mathbf{A} \rightarrow \mathbf{R}$ “se ven” bases para los cuatro subespacios fundamentales de \mathbf{A}

Parte IV

Ortogonalidad

Vectores ortogonales. Subespacios ortogonales

Una importante aplicación de las matemáticas es la medición de distancias y ángulos. Los productos escalares definidos en espacios vectoriales crean el entorno que permite estas mediciones.

Comenzamos con los entornos más familiares: \mathbb{R}^2 y \mathbb{R}^3 . Luego generalizaremos a \mathbb{R}^n con el producto punto. (En la Sección 11.3 generalizamos aún más mediante productos escalares definidos en espacios vectoriales).

11.1. Geometría

11.1.1. Longitud de un vector en \mathbb{R}^2 y en \mathbb{R}^3

El cálculo de la longitud de un vector \mathbf{x} en \mathbb{R}^2 o de un vector \mathbf{y} en \mathbb{R}^3 se basa en el teorema de Pitágoras. En ambos casos el vector se corresponde con la hipotenusa de un triángulo rectángulo, de manera que el cuadrado de la longitud de la hipotenusa (el cuadrado de la longitud del vector) es igual a la suma del cuadrado de las longitudes de los catetos. Veámoslo.

En el caso de \mathbb{R}^2 , la longitud de cada cateto es el valor absoluto de cada componente, $|x_i|$. Por tanto

$$\|\mathbf{x}\|^2 = x_1^2 + x_2^2 = \mathbf{x} \cdot \mathbf{x} \quad \Rightarrow \quad \|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2} = \sqrt{\mathbf{x} \cdot \mathbf{x}};$$

donde $\|\mathbf{x}\|$ denota la longitud o *norma* de un vector \mathbf{x} y donde $|x_i|^2 = x_i^2$.

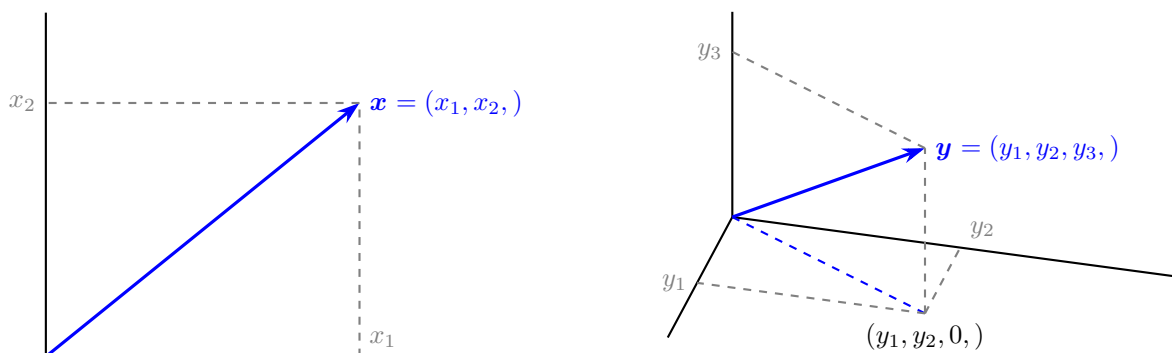


Figura 11.1: Longitud de un vector en \mathbb{R}^2 y en \mathbb{R}^3

En \mathbb{R}^3 , la longitud del cateto vertical es el valor absoluto de la tercera componente; pero la longitud del cateto horizontal ha de ser calculada. Como el cateto horizontal es la hipotenusa de un triángulo rectángulo (como en la figura en \mathbb{R}^2), tenemos que el cuadrado de la longitud de dicho cateto es $(y_1^2 + y_2^2)$:

$$\|\mathbf{y}\|^2 = (y_1^2 + y_2^2) + y_3^2 = \mathbf{y} \cdot \mathbf{y} \quad \Rightarrow \quad \|\mathbf{y}\| = \sqrt{y_1^2 + y_2^2 + y_3^2} = \sqrt{\mathbf{y} \cdot \mathbf{y}}.$$

Así, la longitud de un vector (en \mathbb{R}^2 o \mathbb{R}^3) es la raíz cuadrada de la suma de los cuadrados de sus componentes.

11.1.2. Ángulo formado por dos vectores tanto en \mathbb{R}^2 como en \mathbb{R}^3

Un modo de indicar el grado de abertura del ángulo formado entre dos segmentos unidos por un vértice es mediante la función *coseno*; que toma valores entre 1 y -1 . Cuando la abertura es nula (cuando el ángulo es cero) el coseno toma el valor 1 (marcado con un rombo sobre el eje horizontal de la figura de la izquierda). El coseno disminuye a medida que el ángulo se abre... cuando el ángulo es recto su coseno toma el valor 0 (figura central); y conforme la abertura aumenta el coseno continúa disminuyendo, hasta que se alcanza la abertura máxima, en cuyo caso el coseno alcanza su valor mínimo valor, es decir -1 (figura de la derecha).

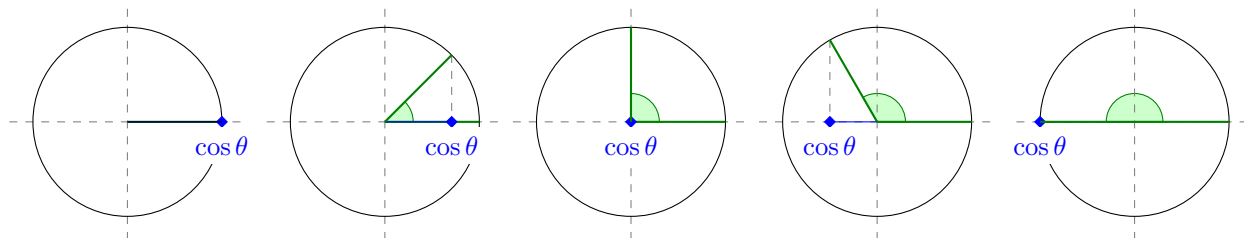


Figura 11.2: Relación entre el ángulo θ y su coseno (el círculo exterior es de radio uno)

Ahora piense que tiene tres listones de madera y que ninguno de ellos tiene longitud mayor que la suma de los otros dos. Entonces con ellos se puede formar un triángulo cuyos ángulos están unívocamente determinados por las longitudes de los listones de madera. Es decir, conociendo los lados de un triángulo podemos calcular sus ángulos. Para esto echaremos mano, una vez más, del teorema de Pitágoras.

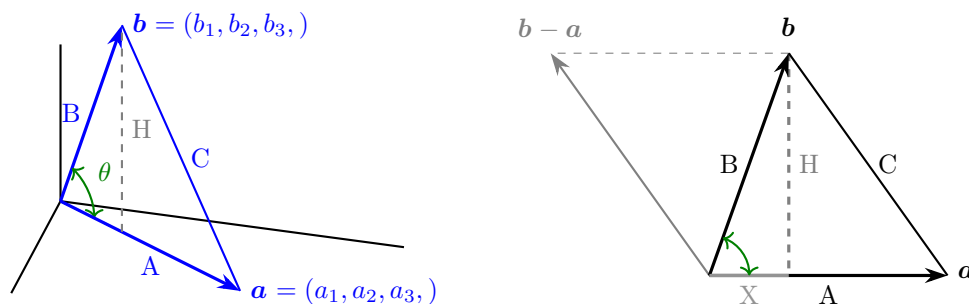


Figura 11.3: Triángulo formado por $\mathbf{0}$ y dos vectores \mathbf{a} y \mathbf{b} no nulos.

A la izquierda de la Figura 11.3 podemos ver un triángulo formado por los segmentos A, B y C (donde asociamos el segmento A con el vector \mathbf{a} y el segmento B con el vector \mathbf{b}). Para poder calcular el ángulo formado por los segmentos A y B (que es el mismo que el formado por los vectores \mathbf{a} y \mathbf{b}), dividimos el triángulo en dos triángulos rectángulos mediante el segmento vertical H. A la derecha aparece esquemáticamente representado el mismo triángulo; pero mostrando que la longitud del segmento C es igual a la longitud del vector $\mathbf{b} - \mathbf{a}$. Además, también se muestra el cateto X de uno de los triángulos rectángulos resultantes de dividir el triángulo ABC con el segmento vertical H.

En todo *triángulo rectángulo*, el coseno del ángulo θ formado por cada uno de los catetos con la hipotenusa es el ratio entre las longitudes de dicho cateto y la hipotenusa.¹ Por tanto,

$$\cos \theta = \frac{X}{B} \quad (11.1)$$

¹Como en un triángulo rectángulo ningún ángulo tiene más de 90 grados, el coseno del ángulo de cualquiera de sus vértices siempre es mayor o igual a cero

donde con X denotamos la longitud del cateto X y con B denotamos la longitud de la hipotenusa B .

Puesto que el ángulo formado por los vectores \mathbf{a} y \mathbf{b} es el mismo que el formado por los segmentos X y B , solo necesitamos calcular el ratio X/B . Por el teorema de Pitágoras, sabemos que

$$X^2 + H^2 = B^2. \quad (11.2)$$

Y si nos fijamos en el segundo triángulo rectángulo, también sabemos que

$$\begin{aligned} (A - X)^2 + H^2 &= C^2 \\ A^2 - 2AX + X^2 + H^2 &= C^2 \end{aligned} \quad \text{desarrollando el cuadrado.} \quad (11.3)$$

Restando (11.2) de (11.3) tenemos

$$\begin{aligned} A^2 - 2AX &= C^2 - B^2 \\ -2AX &= C^2 - B^2 - A^2 \\ X &= (A^2 + B^2 - C^2)/2A \end{aligned}$$

Y sustituyendo en (11.1) llegamos a que $\cos \theta = \frac{A^2 + B^2 - C^2}{2AB}$

Así, como en la figura asociamos el segmento A con el vector \mathbf{a} y el segmento B con \mathbf{b} , podemos sustituir A y B por las longitudes de los correspondientes vectores asociados y C por la longitud de $\mathbf{b} - \mathbf{a}$.

$$\begin{aligned} A = \|\mathbf{a}\| &= \sqrt{a_1^2 + a_2^2 + a_3^2} \\ B = \|\mathbf{b}\| &= \sqrt{b_1^2 + b_2^2 + b_3^2} \\ C = \|\mathbf{b} - \mathbf{a}\| &= \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + (b_3 - a_3)^2} \end{aligned}$$

Por tanto, operando llegamos a

$$\cos \theta = \frac{2(a_1b_1 + a_2b_2 + a_3b_3)}{2\sqrt{a_1^2 + a_2^2 + a_3^2}\sqrt{b_1^2 + b_2^2 + b_3^2}} = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\|\|\mathbf{b}\|}; \quad (\text{con } \mathbf{a} \neq \mathbf{0} \neq \mathbf{b}),$$

donde hemos sustituido la suma de productos de las componentes por el *producto punto*. Cuando $\cos \theta = 0$ los vectores son perpendiculares (véase la Figura 11.2), es decir, \mathbf{a} y \mathbf{b} son perpendiculares cuando $\mathbf{a} \cdot \mathbf{b} = 0$.

11.1.3. Generalizando al espacio euclídeo \mathbb{R}^n

Comenzamos la generalización con el concepto de perpendicularidad en el espacio euclídeo^a \mathbb{R}^n .

^aPara una definición de *espacio euclídeo* véase la Sección 11.3.4.

Definición 11.1. Decimos que \mathbf{a} y \mathbf{b} son ortogonales o perpendiculares ($\mathbf{a} \perp \mathbf{b}$) cuando $\mathbf{a} \cdot \mathbf{b} = 0$.

Consecuentemente, $\mathbf{0}$ es ortogonal a todos los vectores (incluido él mismo).

La generalización del concepto de longitud a vectores del espacio vectorial \mathbb{R}^n es inmediata.

Definición 11.2. La longitud (o norma) de un vector \mathbf{a} es la raíz cuadrada de $\mathbf{a} \cdot \mathbf{a}$:

$$\text{longitud de } \mathbf{a} = \|\mathbf{a}\| = \sqrt{\mathbf{a} \cdot \mathbf{a}}.$$

Por ejemplo, la longitud de $\mathbf{x} = \begin{pmatrix} 6 \\ 0 \\ -2 \\ 3 \end{pmatrix}$ es $\sqrt{\mathbf{x} \cdot \mathbf{x}} = \sqrt{\begin{pmatrix} 6 \\ 0 \\ -2 \\ 3 \end{pmatrix} \cdot \begin{pmatrix} 6 \\ 0 \\ -2 \\ 3 \end{pmatrix}} = \sqrt{6^2 + 0^2 + (-2)^2 + 3^2} = \sqrt{49} = 7$.

Puesto que $\mathbf{x} \cdot \mathbf{x} = 0$ si y solo si $\mathbf{x} = \mathbf{0}$, resulta que: $\|\mathbf{x}\| = 0$ si y solo si $\mathbf{x} = \mathbf{0}$.

Multiplicando \mathbf{x} por λ obtenemos un múltiplo de \mathbf{x} cuya longitud es $|\lambda|$ veces su longitud original.

$$\|\lambda \mathbf{x}\| = \sqrt{\lambda^2 (\mathbf{x} \cdot \mathbf{x})} = |\lambda| \cdot \|\mathbf{x}\|.$$

Definición 11.3 (Vector unitario). *Se dice que un vector es unitario si tiene longitud uno.*

Es sencillo obtener un múltiplo unitario (i.e., de norma uno) de $\mathbf{x} \neq \mathbf{0}$; basta dividir el vector por su longitud. Por ejemplo, si $\mathbf{x} = (1, 1, 1, 1)$ entonces $\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}} = \sqrt{4} = 2$. Así, el vector

$$\mathbf{y} = \frac{1}{\|\mathbf{x}\|} \mathbf{x} = \frac{1}{2} \mathbf{x} = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right) \text{ es unitario. ¡Compruébelo!}$$

Una vez definidas la perpendicularidad y la longitud en \mathbb{R}^n ; vamos a generalizar la propiedad más importante que verifican los vectores perpendiculares: el **Teorema de Pitágoras**.

(recuerde las propiedades del *producto punto* en la Página 25)

Teorema 11.1.1 (Teorema de Pitágoras en \mathbb{R}^n). *Sean \mathbf{x} e \mathbf{y} dos vectores de \mathbb{R}^n ; entonces $\mathbf{x} \cdot \mathbf{y} = 0$ (son perpendiculares) si y solo si*

$$\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2.$$

Demostración. Si \mathbf{x} e \mathbf{y} son perpendiculares, entonces $\mathbf{x} \cdot \mathbf{y} = 0$ y

$$\begin{aligned} \|\mathbf{x} + \mathbf{y}\|^2 &= (\mathbf{x} + \mathbf{y}) \cdot (\mathbf{x} + \mathbf{y}) = \mathbf{x} \cdot (\mathbf{x} + \mathbf{y}) + \mathbf{y} \cdot (\mathbf{x} + \mathbf{y}) \\ &= (\mathbf{x} + \mathbf{y}) \cdot \mathbf{x} + (\mathbf{x} + \mathbf{y}) \cdot \mathbf{y} \\ &= \mathbf{x} \cdot \mathbf{x} + \mathbf{y} \cdot \mathbf{x} + \mathbf{x} \cdot \mathbf{y} + \mathbf{y} \cdot \mathbf{y} \\ &= \mathbf{x} \cdot \mathbf{x} + 2(\mathbf{x} \cdot \mathbf{y}) + \mathbf{y} \cdot \mathbf{y} \\ &= \mathbf{x} \cdot \mathbf{x} + 0 + \mathbf{y} \cdot \mathbf{y} = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2. \end{aligned}$$

□

Fíjese que en el anterior teorema, \mathbf{x} e \mathbf{y} corresponden a los catetos y el vector suma $\mathbf{x} + \mathbf{y}$ a la hipotenusa.

Para generalizar el *coseno del ángulo formado por dos vectores* a \mathbb{R}^n , antes necesitamos probar que $\mathbf{a} \cdot \mathbf{b}$ es siempre menor en valor absoluto que $\|\mathbf{a}\| \|\mathbf{b}\|$, de manera que $\frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|}$ solo toma valores entre -1 y 1 . Dicho resultado se denomina *Desigualdad de Cauchy-Schwarz*.

Proposición 11.1.2 (Desigualdad de Cauchy-Schwarz). *Para \mathbf{a} y \mathbf{b} de \mathbb{R}^n se verifica que $|\mathbf{a} \cdot \mathbf{b}| \leq \|\mathbf{a}\| \|\mathbf{b}\|$.*

Demostración. Caso trivial: \mathbf{a} y \mathbf{b} son linealmente dependientes. Entonces, o bien $\mathbf{a} = \mathbf{0}$ o bien existe $\lambda \in \mathbb{R}$ tal que $\lambda \mathbf{a} = \mathbf{b}$. Si ocurre lo primero $\mathbf{a} \cdot \mathbf{b} = 0$ y $\|\mathbf{a}\| \|\mathbf{b}\| = 0$. Si ocurre lo segundo

$$|\mathbf{a} \cdot \mathbf{b}| = |\mathbf{a} \cdot \lambda \mathbf{a}| = |\lambda| \cdot \|\mathbf{a}\|^2 = |\lambda| \cdot \|\mathbf{a}\| \|\mathbf{a}\| = \|\mathbf{a}\| \|\mathbf{b}\|.$$

Caso no trivial: \mathbf{a} y \mathbf{b} son linealmente independientes. En tal caso veamos que existe un vector \mathbf{h} ortogonal a \mathbf{a} tal que $\mathcal{L}([\mathbf{a}; \mathbf{h}]) = \mathcal{L}([\mathbf{a}; \mathbf{b}])$.

Para encontrarlo definimos el vector como $\mathbf{h} = \mathbf{b} + \alpha \mathbf{a}$; así garantizamos que $\mathcal{L}([\mathbf{a}; \mathbf{h}]) = \mathcal{L}([\mathbf{a}; \mathbf{b}])$. Además vamos a imponer la condición de que $\mathbf{a} \perp \mathbf{h}$; así

$$0 = \mathbf{a} \cdot \mathbf{h} = \mathbf{a} \cdot (\mathbf{b} + \alpha \mathbf{a}) = \mathbf{a} \cdot \mathbf{b} + \alpha \mathbf{a} \cdot \mathbf{a} \quad \text{por tanto} \quad \boxed{\alpha = -\frac{\mathbf{a} \cdot \mathbf{b}}{\mathbf{a} \cdot \mathbf{a}}}.$$

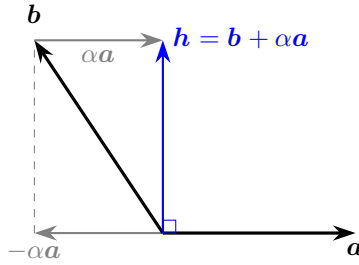


Figura 11.4: Representación esquemática donde \mathbf{h} es combinación lineal de \mathbf{a} y \mathbf{b} y perpendicular a \mathbf{a} .

$$\text{Por tanto, } \|\mathbf{b}\|^2 = \mathbf{b} \cdot \mathbf{b} = (\mathbf{h} - \alpha \mathbf{a}) \cdot (\mathbf{h} - \alpha \mathbf{a}) = \mathbf{h} \cdot \mathbf{h} - 2\alpha \mathbf{h} \cdot \mathbf{a} + \alpha^2 \mathbf{a} \cdot \mathbf{a} = \|\mathbf{h}\|^2 + \alpha^2 \|\mathbf{a}\|^2.$$

(Fíjese en el esquema de la figura: \mathbf{b} es la hipotenusa y \mathbf{h} y $-\alpha \mathbf{a}$ son los catetos de un triángulo rectángulo).

Como \mathbf{a} y \mathbf{b} son linealmente independientes, necesariamente $\mathbf{h} \neq \mathbf{0}$, es decir, $\|\mathbf{h}\| > 0$; y entonces

$$\|\mathbf{b}\|^2 = \|\mathbf{h}\|^2 + \alpha^2 \|\mathbf{a}\|^2 > \alpha^2 \|\mathbf{a}\|^2 = \frac{(\mathbf{a} \cdot \mathbf{b})^2}{(\mathbf{a} \cdot \mathbf{a})^2} \|\mathbf{a}\|^2 = \frac{(\mathbf{a} \cdot \mathbf{b})^2}{(\|\mathbf{a}\|^2)^2} \|\mathbf{a}\|^2 = \frac{(\mathbf{a} \cdot \mathbf{b})^2}{\|\mathbf{a}\|^2},$$

de donde $(\|\mathbf{a}\|^2)(\|\mathbf{b}\|^2) > (\mathbf{a} \cdot \mathbf{b})^2$ y por consiguiente $\|\mathbf{a}\| \|\mathbf{b}\| > |\mathbf{a} \cdot \mathbf{b}|$. □

Fíjese que, en un extremo tenemos $\mathbf{a} \cdot \mathbf{b} = 0$ cuando $\mathbf{a} \perp \mathbf{b}$; y en el otro $|\mathbf{a} \cdot \mathbf{b}| = \|\mathbf{a}\| \|\mathbf{b}\|$ cuando $\mathbf{b} = \lambda \mathbf{a}$.



Ahora, y gracias a que sabemos que se verifica la desigualdad de Cauchy-Schwarz, también podemos generalizar a \mathbb{R}^n el coseno del ángulo formado por dos vectores no nulos.

Definición 11.4. El coseno del ángulo θ formado por los vectores no nulos \mathbf{a} y \mathbf{b} de \mathbb{R}^n es

$$\cos \theta = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|}; \quad -1 \leq \cos \theta \leq 1.$$

Con esta generalización, la parte derecha de la Figura 11.3 se puede reinterpretar como una representación esquemática de un triángulo formado por vectores de \mathbb{R}^n .



Por experiencia sabemos que la distancia entre dos puntos \mathbf{a} y \mathbf{c} , siempre es menor o igual que la suma de las distancias del camino indirecto que va de \mathbf{a} a \mathbf{b} y luego de \mathbf{b} a \mathbf{c} . Para extender el concepto de *distancia* debemos comprobar que ocurre lo mismo en \mathbb{R}^n . Para ello demostraremos la *desigualdad triangular*.

Proposición 11.1.3 (Desigualdad triangular). Sean \mathbf{a} y \mathbf{b} vectores de \mathbb{R}^n , entonces $\|\mathbf{a} + \mathbf{b}\| \leq \|\mathbf{a}\| + \|\mathbf{b}\|$.

Demostración.

$$\begin{aligned}\|\mathbf{a} + \mathbf{b}\|^2 &= (\mathbf{a} + \mathbf{b}) \cdot (\mathbf{a} + \mathbf{b}) = \mathbf{a} \cdot \mathbf{a} + 2(\mathbf{a} \cdot \mathbf{b}) + \mathbf{b} \cdot \mathbf{b} \\ &= \|\mathbf{a}\|^2 + 2(\mathbf{a} \cdot \mathbf{b}) + \|\mathbf{b}\|^2 \\ &\leq \|\mathbf{a}\|^2 + 2\|\mathbf{a}\|\|\mathbf{b}\| + \|\mathbf{b}\|^2 \quad \text{puesto que } |\mathbf{a} \cdot \mathbf{b}| \leq \|\mathbf{a}\|\|\mathbf{b}\|; \text{ (Prop. 11.1.2)} \\ &= (\|\mathbf{a}\| + \|\mathbf{b}\|)^2\end{aligned}$$

Consecuentemente $\|\mathbf{a} + \mathbf{b}\| \leq \|\mathbf{a}\| + \|\mathbf{b}\|$. □

La *Desigualdad Triangular* con normas de vectores da lugar a la noción de *distancia entre vectores de \mathbb{R}^n* :

$$\text{distancia entre } \mathbf{x} \text{ y } \mathbf{y}: d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| = \|\mathbf{y} - \mathbf{x}\| = d(\mathbf{y}, \mathbf{x}),$$

que verifica otra *Desigualdad Triangular* (ahora referida a las distancias):

$$d(\mathbf{a}, \mathbf{c}) = \|\mathbf{a} - \mathbf{c}\| = \|(\mathbf{a} - \mathbf{b}) + (\mathbf{b} - \mathbf{c})\| \leq \|\mathbf{a} - \mathbf{b}\| + \|\mathbf{b} - \mathbf{c}\| = d(\mathbf{a}, \mathbf{b}) + d(\mathbf{b}, \mathbf{c}).$$

Puesto que $\|\mathbf{x}\| = 0$ si y solo si $\mathbf{x} = \mathbf{0}$, tenemos que $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{a} - \mathbf{b}\| = 0$ si y solo si $\mathbf{a} = \mathbf{b}$.

Vectores alineados. Cuando $\mathbf{y} = \lambda \mathbf{x}$ (cuando \mathbf{y} es un múltiplo de \mathbf{x}) el producto punto entre ambos es

$$\mathbf{y} \cdot \mathbf{x} = \lambda \mathbf{x} \cdot \mathbf{x} = \lambda \|\mathbf{x}\|^2.$$

En tal caso (y si $\lambda \neq 0$) el ángulo que forman \mathbf{x} e \mathbf{y} es cero o bien 180° (es decir 0 ó π radianes) ya que:

$$\cos \theta = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} = \frac{\lambda \|\mathbf{x}\|^2}{|\lambda| \cdot \|\mathbf{x}\|\|\mathbf{x}\|} = \frac{\lambda}{|\lambda|} = \pm 1,$$

y entonces se dice que los vectores están *alineados*.

Fíjese que cuando \mathbf{a} y \mathbf{b} están alineados: si $\lambda > 0$ entonces $\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\|\|\mathbf{b}\|$; y si por el contrario $\lambda < 0$, entonces $\mathbf{a} \cdot \mathbf{b} = -\|\mathbf{a}\|\|\mathbf{b}\|$.

Consecuentemente, cuando \mathbf{a} y \mathbf{b} están alineados y $\lambda > 0$ (cuando ambos “apuntan” en el mismo sentido) resulta que: $\|\mathbf{a} + \mathbf{b}\| = \|\mathbf{a}\| + \|\mathbf{b}\|$ (véase la demostración de la Desigualdad Triangular).

11.2. Subespacios ortogonales en \mathbb{R}^n

Definición 11.5 (Conjunto ortogonal al subconjunto A de vectores de \mathbb{R}^n). El conjunto de todos los vectores de \mathbb{R}^n que son perpendiculares a los vectores de $A \subset \mathbb{R}^n$ se denomina conjunto ortogonal (o perpendicular) a A , y se denota con A^\perp :

$$A^\perp = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} \cdot \mathbf{a} = 0 \text{ para cada } \mathbf{a} \in A\}$$

EJERCICIO 57. Demuestre que A^\perp es un subespacio.

¡Fíjese que aunque A no sea un subespacio, el conjunto ortogonal A^\perp sí lo es!

EJERCICIO 58.

- Demuestre que si B es un subconjunto de vectores de \mathbb{R}^n y $A \subset B$ entonces $B^\perp \subset A^\perp$.
- Demuestre que para todo subconjunto A de vectores de \mathbb{R}^n se verifica que $A^\perp = \mathcal{L}(A)^\perp$.
(donde $\mathcal{L}(A)$ corresponde a la Definición 9.11 en la página 118)
- Demuestre que para todo subespacio \mathcal{A} de \mathbb{R}^n se verifica que $\mathcal{A} \cap \mathcal{A}^\perp = \{\mathbf{0}\}$.

Proposición 11.2.1 ($\mathcal{A} \subset \mathbb{R}^m$ y \mathcal{A}^\perp son **suplementarios**).² Sean $\mathcal{A} \subset \mathbb{R}^m$ y \mathcal{A}^\perp , entonces $\mathcal{A} \oplus \mathcal{A}^\perp = \mathbb{R}^m$.

Demostración. Sea el sistema \mathbf{A} de r vectores una base de \mathcal{A} . Por el apartado (b) del ejercicio anterior sabemos que $\mathcal{A}^\perp = \mathcal{L}(\mathbf{A})^\perp = \mathcal{A}^\perp$.

Si llamamos \mathbf{A} al sistema \mathbf{A} de vectores linealmente independientes de \mathbb{R}^m , entonces \mathcal{A}^\perp es el subconjunto de vectores perpendiculares a las columnas de \mathbf{A} ; es decir $\mathcal{A}^\perp = \{\mathbf{x} \in \mathbb{R}^m \mid \mathbf{x}\mathbf{A} = \mathbf{0}\} = \mathcal{N}(\mathbf{A}^\top)$.

Consecuentemente, $\dim \mathcal{A}^\perp = \dim \mathcal{N}(\mathbf{A}^\top) = m - r$. Como $\dim \mathcal{A} = r$, como $\dim \mathcal{A}^\perp = m - r$ y como $\dim \mathcal{A} \cap \mathcal{A}^\perp = \dim \{\mathbf{0}\} = 0$, por la Proposición 10.1.4 en la página 123, necesariamente $\mathcal{A} \oplus \mathcal{A}^\perp = \mathbb{R}^m$. \square

Observación. Hay muchos subespacios suplementarios a uno dado: viendo la Figura 6.1 en la página 86 es fácil darse cuenta que cualquier función lineal $g: \mathcal{V} \rightarrow \mathcal{W}$ es un espacio suplementario de $\{\vec{0}\} \times \mathcal{W}$. Sin embargo, dado un subespacio, el correspondiente subespacio suplementario y ortogonal es *único*. En la citada figura corresponde al eje horizontal: $(\{\vec{0}\} \times \mathcal{W})^\perp = \mathcal{V} \times \{\vec{0}\}$. Esto da lugar a la siguiente

Definición 11.6 (Complemento ortogonal de $\mathcal{A} \subset \mathbb{R}^n$). En el caso de que $\mathcal{A} \subset \mathbb{R}^n$ sea un subespacio, el subespacio \mathcal{A}^\perp (conjunto ortogonal de \mathcal{A}) se llama **complemento ortogonal** de \mathcal{A} .

Así podemos reescribir el último apartado del Ejercicio 58 en la página anterior como: ‘Demuestre que la intersección entre un subespacio de \mathbb{R}^n y su complemento ortogonal es $\{\mathbf{0}\}$ ’.

11.2.1. Ortogonalidad de los 4 subespacios fundamentales de \mathbf{A}

Definición 11.7 (Subconjuntos ortogonales). Dos subconjuntos A y B de \mathbb{R}^n son **ortogonales** ($A \perp B$), si todo vector de A es perpendicular a todo vector de B ; es decir, si $\mathbf{a} \cdot \mathbf{b} = 0$ para cualesquiera $\mathbf{a} \in A$ y $\mathbf{b} \in B$.

Fíjese que un subespacio \mathcal{A} y su complemento ortogonal \mathcal{A}^\perp son un caso particular de subconjuntos (en este caso subespacios) ortogonales, pues $\mathcal{A} \perp (\mathcal{A}^\perp)$.

EJERCICIO 59. Demuestre la siguiente

Proposición 11.2.2. Sean A y B subconjuntos de \mathbb{R}^n , entonces $A \perp B$ si y solo si $\mathcal{L}(A) \perp \mathcal{L}(B)$.

La siguiente proposición muestra que los subespacios fundamentales de \mathbf{A} son ortogonales dos a dos.

Proposición 11.2.3. Sea \mathbf{A} de orden m por n , entonces

- El espacio fila y el espacio nulo son ortogonales: $\mathcal{C}(\mathbf{A}^\top) \perp \mathcal{N}(\mathbf{A})$.
- El espacio columna y el espacio nulo por la izquierda son ortogonales: $\mathcal{C}(\mathbf{A}) \perp \mathcal{N}(\mathbf{A}^\top)$.

Demostración. El espacio fila es el conjunto de combinaciones lineales de las filas de \mathbf{A} , es decir, es el conjunto de vectores de la forma $\mathbf{f} = \mathbf{y}\mathbf{A}$ con $\mathbf{y} \in \mathbb{R}^m$; y los vectores del espacio nulo son los vectores \mathbf{x} tales que $\mathbf{A}\mathbf{x} = \mathbf{0}$. Evidentemente los vectores \mathbf{f} y \mathbf{x} son perpendiculares puesto que:

$$\mathbf{f} \cdot \mathbf{x} = \mathbf{y}\mathbf{A}\mathbf{x} = \mathbf{y} \cdot \mathbf{0} = 0.$$

Del mismo modo, como el espacio columna es el conjunto de combinaciones lineales de las columnas de \mathbf{A} , es decir, es el conjunto de vectores de la forma $\mathbf{b} = \mathbf{A}\mathbf{x}$ con $\mathbf{x} \in \mathbb{R}^n$; y como los vectores del espacio nulo por la izquierda son los vectores \mathbf{y} tales que $\mathbf{y}\mathbf{A} = \mathbf{0}$; los vectores \mathbf{y} y \mathbf{b} también son perpendiculares:

$$\mathbf{y} \cdot \mathbf{b} = \mathbf{y}\mathbf{A}\mathbf{x} = \mathbf{0} \cdot \mathbf{x} = 0.$$

\square

²En espacios vectoriales de dimensión infinita este resultado es falso en general (véase el ejemplo de la Página 149 donde \mathcal{A} y \mathcal{A}^\perp no son suplementarios); pero sí es verdadero cuando \mathcal{A} es de dimensión finita (Proposición 11.3.4 en la página 147).

```

a = Vector( [1, 1, 1] )
A = Matrix( [a, -a, a, -a] )

C = SubEspacio( A.sis() ) # Espacio columna de A
F = SubEspacio( (~A).sis() ) # Espacio fila de A
N = SubEspacio( A ) # Espacio nulo de A
NI = SubEspacio( ~A ) # Espacio nulo por la izquierda de A

print( F == ~N ) # ¿Es cierto que F es igual a complemento ortogonal de N?
print( NI == ~C ) # ¿Es cierto que NI es igual a complemento ortogonal de C?

```

True
True

El método de eliminación como generador de bases del complemento ortogonal

En la lección anterior vimos que con el *método de eliminación* podemos encontrar bases de los cuatro espacios fundamentales de una matriz. Consecuentemente, podemos encontrar una base del *complemento ortogonal* de cualquier subespacio de \mathbb{R}^m : si tenemos un sistema de vectores Z de \mathbb{R}^m , basta escribir dichos vectores como filas de una matriz y aplicar el método de eliminación “de izquierda a derecha”; las “soluciones especiales” serán base del complemento ortogonal del espacio $\mathcal{L}(Z)$ generado por Z .

Ejemplo 26. Busquemos una base para el complemento ortogonal del espacio generado por los vectores

$$\left[\begin{pmatrix} 1 \\ -3 \\ 0 \\ -1 \end{pmatrix}; \begin{pmatrix} 0 \\ -1 \\ 1 \\ 1 \end{pmatrix}; \begin{pmatrix} 1 \\ -4 \\ 1 \\ 0 \end{pmatrix}; \right].$$

Aplicando la eliminación sobre una matriz cuyas filas son los vectores dados tenemos:

$$\begin{bmatrix} 1 & -3 & 0 & -1 \\ 0 & -1 & 1 & 1 \\ 1 & -4 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{\begin{matrix} \tau \\ [(3)1+2] \\ [(1)1+4] \\ [(1)2+3] \\ [(1)2+4] \end{matrix}} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 1 & 3 & 3 & 4 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \Rightarrow \text{Base del complemento ortogonal: } \left[\begin{pmatrix} 3 \\ 1 \\ 1 \\ 0 \end{pmatrix}; \begin{pmatrix} 4 \\ 1 \\ 0 \\ 1 \end{pmatrix}; \right].$$

```

S = Sistema( [Vector([1,-3,0,-1]), Vector([0,-1,1,1]), Vector([1,-4,1,0])] )
~Matrix(S) # Matriz cuyas filas son los vectores de S
Homogenea(~Matrix(S),1).sgen # Sistema generador del complemento ortogonal
SubEspacio( Homogenea(~Matrix(S)).sgen ) == ~SubEspacio(S) # Comprobación

```

11.2.2. De las ecuaciones paramétricas a las ecuaciones cartesianas

En la Sección 7.4 de la Lección 7 aparecieron por primera vez tanto las ecuaciones paramétricas, como las ecuaciones cartesianas (véase las notas a pie de página números 3 en la página 97 y 1 en la página 93).

Cada modo de representación tiene un propósito distinto: las ecuaciones paramétricas nos permiten generar ejemplos de vectores que pertenecen al conjunto (basta dar valores arbitrarios a los parámetros); las ecuaciones cartesianas nos permiten comprobar si un vector y pertenece al conjunto (basta comprobar si y es solución del sistema de ecuaciones indicado en las ecuaciones cartesianas).

Veamos ahora cómo pasar de unas a las otras; así dispondremos en todo momento de dos modos de expresar el mismo conjunto de vectores.

El camino recorrido hasta ahora: de las ecuaciones cartesianas a las paramétricas

Dado un sistema de ecuaciones lineales $\mathbf{Ax} = \mathbf{b}$, donde \mathbf{A} es de orden m por n , expresamos el conjunto de soluciones con la siguiente *ecuación cartesiana*:

$$\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{Ax} = \mathbf{b}\}$$

Cuando resolvemos el sistema encontrando una solución particular \mathbf{s} y una base $[\mathbf{n}_1; \dots; \mathbf{n}_k]$ de $\mathcal{N}(\mathbf{A})$, re-expresamos el mismo conjunto de soluciones con la *ecuación paramétrica*:

$$\{\mathbf{x} \in \mathbb{R}^n \mid \exists \mathbf{p} \in \mathbb{R}^k \text{ tal que } \mathbf{x} = \mathbf{s} + [\mathbf{n}_1; \dots; \mathbf{n}_k] \mathbf{p}\}.$$

Así, cada vez que hemos resuelto un sistema de ecuaciones, hemos partido de unas ecuaciones cartesianas (o implícitas) para obtener unas ecuaciones paramétricas (el conjunto de soluciones). Veamos ahora como recorrer el camino inverso (partiendo de un conjunto de soluciones, obtener un sistema de ecuaciones que tiene a dicho conjunto como solución).

Recorriendo el camino inverso: de las ecuaciones paramétricas a las cartesianas

En el camino inverso, dadas unas ecuaciones paramétricas del conjunto de soluciones de “un sistema desconocido, $\mathbf{Dx} = \mathbf{f}$ ”, encontraremos las filas de una matriz de coeficientes \mathbf{A} y un vector del lado derecho \mathbf{b} que definan un sistema de ecuaciones $\mathbf{Ax} = \mathbf{b}$ *equivalente* a $\mathbf{Dx} = \mathbf{f}$, es decir, cuyo conjunto de soluciones coincida con el del sistema de ecuaciones desconocido, $\mathbf{Dx} = \mathbf{f}$.

Procedimiento. Cuando disponemos de unas ecuaciones paramétricas

$$\{\mathbf{x} \in \mathbb{R}^n \mid \exists \mathbf{p} \in \mathbb{R}^k \text{ tal que } \mathbf{x} = \mathbf{s} + [\mathbf{n}_1; \dots; \mathbf{n}_k] \mathbf{p}\},$$

disponemos tanto de una solución \mathbf{s} (un vector del conjunto de soluciones) como de una base $[\mathbf{n}_1; \dots; \mathbf{n}_k]$ del espacio nulo de la matriz \mathbf{A} que buscamos, es decir, $\mathcal{N}(\mathbf{A}) = \mathcal{L}([\mathbf{n}_1; \dots; \mathbf{n}_k])$.

Sabemos que por eliminación podemos encontrar una base del complemento ortogonal de $\mathcal{L}([\mathbf{n}_1; \dots; \mathbf{n}_k])$, es decir, una base de $\mathcal{C}(\mathbf{A}^\top)$. Usaremos los vectores de dicha base para formar las *filas* de \mathbf{A} . Los pasos son:

Primero encontrar una base del complemento ortogonal de $\mathcal{N}(\mathbf{A}) = \mathcal{L}([\mathbf{n}_1; \dots; \mathbf{n}_k])$. Para ello usamos la matriz \mathbf{Z} cuyas k filas son los vectores $\mathbf{n}_1, \dots, \mathbf{n}_k$, y aplicamos la eliminación “de izquierda a derecha” siguiendo el esquema:

$$\begin{bmatrix} \mathbf{Z} \\ \mathbf{I} \end{bmatrix} \xrightarrow{\tau_1 \dots \tau_k} \begin{bmatrix} \mathbf{K} \\ \mathbf{E} \end{bmatrix}; \quad \text{donde } \mathbf{Z} = [\mathbf{n}_1; \dots; \mathbf{n}_k]^\top \quad (\text{nótese el símbolo de transposición}).$$

Como $\mathbf{ZE} = \mathbf{K}$; si ν es la lista de índices de las columnas nulas de \mathbf{K} , entonces $\mathbf{ZE}_{|\nu} = \mathbf{K}_{|\nu} = \mathbf{0}$.

Después. Si usamos las columnas de $\mathbf{E}_{|\nu}$ (i.e., las soluciones especiales) como *filas* para construir una matriz \mathbf{A} , es decir, si $\mathbf{E}_{|\nu} = \mathbf{A}^\top$; entonces $\mathbf{Z}(\mathbf{A}^\top) = \mathbf{0}$, o transponiendo ambos lados, $\mathbf{A}[\mathbf{n}_1; \dots; \mathbf{n}_k] = \mathbf{0}$.

Si ahora multiplicamos por \mathbf{A} la ecuación paramétrica $\mathbf{x} = \mathbf{s} + [\mathbf{n}_1; \dots; \mathbf{n}_k] \mathbf{p}$ tenemos que

$$\mathbf{Ax} = \mathbf{As} + \underbrace{\mathbf{A}[\mathbf{n}_1; \dots; \mathbf{n}_k]}_{\mathbf{0}} \mathbf{p} \Rightarrow \mathbf{Ax} = \mathbf{As}, \quad (11.4)$$

donde $\mathbf{b} = \mathbf{As}$ será el vector del lado derecho del sistema que buscamos. Este sistema es, por construcción, un sistema cuyas soluciones son las ecuaciones paramétricas de las que hemos partido, ya que

$$\begin{cases} \mathbf{Ax} = \mathbf{b} & \text{cuando } \mathbf{x} = \mathbf{s} \\ \mathbf{Ax} = \mathbf{0} & \text{cuando } \mathbf{x} \in \mathcal{L}([\mathbf{n}_1; \dots; \mathbf{n}_k]) = \mathcal{N}(\mathbf{A}) \end{cases}.$$

Procedimiento abreviado para usar con lápiz y papel. No es necesario encontrar primero los vectores perpendiculares a $\mathbf{n}_k, \dots, \mathbf{n}_n$ para formar con ellos las filas de \mathbf{A} y multiplicar la ecuación paramétrica para obtener el sistema $\mathbf{A}\mathbf{x} = \mathbf{b}$. Podemos hacerlo todo a la vez con la *eliminación de izquierda a derecha*.

Sea \mathbf{M} la matriz que resulta tras añadir los vectores \mathbf{x} y \mathbf{s} como nuevas filas de \mathbf{Z} . Entonces, aplicando la secuencia de transformaciones $\tau_1 \cdots \tau_k$ que transforma \mathbf{Z} en una forma pre-escalada \mathbf{K} , tenemos que

$$\mathbf{M} = \begin{bmatrix} \mathbf{Z} \\ \mathbf{x} \\ \mathbf{s} \end{bmatrix} \xrightarrow{\tau_1 \cdots \tau_k} \begin{bmatrix} \mathbf{K} \\ \mathbf{x}\mathbf{E} \\ \mathbf{s}\mathbf{E} \end{bmatrix}, \quad \text{donde } \mathbf{E} = \mathbf{I}_{\tau_1 \cdots \tau_k}.$$

Por la Ecuación 11.4 sabemos que $\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{s}$, es decir $\mathbf{x}(\mathbf{A}^\top) = \mathbf{s}(\mathbf{A}^\top)$, por tanto $\mathbf{x}\mathbf{E}_{|\nu} = \mathbf{s}\mathbf{E}_{|\nu}$; donde ν es la lista de índices de las columnas nulas de \mathbf{K} . Así pues, si $\mathbf{K}_{|j} = \mathbf{0}$ entonces $\mathbf{x}\mathbf{E}_{|j} = \mathbf{s}\mathbf{E}_{|j}$. Dicho de otro modo, *bajo las columnas nulas de \mathbf{K} podemos ver las ecuaciones del sistema buscado*.

Ejemplo 27. Encontremos unas ecuaciones cartesianas para el siguiente conjunto P de vectores de \mathbb{R}^4 :

$$P = \left\{ \mathbf{v} \in \mathbb{R}^4 \mid \exists \mathbf{p} \in \mathbb{R}^2, \mathbf{v} = \begin{pmatrix} 0 \\ 2 \\ 0 \\ 5 \end{pmatrix} + \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 0 & 1 \\ -2 & 3 \end{bmatrix} \mathbf{p} \right\}.$$

Aplicando la eliminación sobre la matriz \mathbf{M} , cuya tercera fila es $\mathbf{x} = (x, y, z, w,)$ y cuya última fila es la solución particular $\mathbf{s} = (0, 2, 0, 5,)$, tenemos:

$$\begin{bmatrix} 1 & 2 & 0 & -2 \\ 0 & 0 & 1 & 3 \\ \hline x & y & z & w \\ 0 & 2 & 0 & 5 \end{bmatrix} \xrightarrow{\begin{matrix} [(-2)\mathbf{1}+\mathbf{2}] \\ [(2)\mathbf{1}+\mathbf{4}] \end{matrix}} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 3 \\ \hline x & -2x+y & z & 2x+w \\ 0 & 5 & 0 & 5 \end{bmatrix} \xrightarrow{[(-3)\mathbf{3}+\mathbf{4}]} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline x & -2x+y & z & 2x-3z+w \\ 0 & 2 & 0 & 5 \end{bmatrix}.$$

Mirando bajo las columnas nulas se vislumbran las ecuaciones que buscamos:

$$\begin{cases} -2x + y & = 2 \\ 2x & - 3z + w = 5 \end{cases}.$$

Así pues, unas *ecuaciones cartesianas* para P son: $\left\{ \mathbf{v} \in \mathbb{R}^4 \mid \begin{bmatrix} -2 & 1 & 0 & 0 \\ 2 & 0 & -3 & 1 \end{bmatrix} \mathbf{v} = \begin{pmatrix} 2 \\ 5 \end{pmatrix} \right\}.$

11.2.3. Puntos, rectas, planos e hiper-planos de \mathbb{R}^n . Espacios afines

Dado un sistema de ecuaciones $\mathbf{A}\mathbf{x} = \mathbf{b}$, su conjunto de soluciones posee una forma geométrica que “no se curva”. En particular, si $\dim \mathcal{N}(\mathbf{A}) = 0$ el conjunto es *un punto*; cuando $\dim \mathcal{N}(\mathbf{A}) = 1$ el conjunto es *una recta*; cuando $\dim \mathcal{N}(\mathbf{A}) = 2$ es *un plano*, y si $\dim \mathcal{N}(\mathbf{A}) = d > 2$ es *un hiper-plano de dimensión d* .

Esto sugiere emplear nuevos nombres para conceptos ya conocidos. Podemos denominar ecuaciones paramétricas (o ecuaciones cartesianas) de un punto, de una recta, de un plano, etc. al conjunto de soluciones del correspondiente sistema $\mathbf{A}\mathbf{x} = \mathbf{b}$. Así por ejemplo, podemos preguntarnos por las ecuaciones cartesianas de una recta en \mathbb{R}^n (al decir *recta* indicamos implícitamente que el espacio nulo de la matriz de coeficientes de las ecuaciones cartesianas tiene dimensión uno).

A los subconjuntos de \mathbb{R}^n que podemos describir como la suma de un vector particular \mathbf{s} más las combinaciones lineales de un sistema de vectores $[\mathbf{n}_1; \dots \mathbf{n}_k;]$ (o que alternativamente podemos describir como el conjunto solución de un sistema de ecuaciones $\mathbf{A}\mathbf{x} = \mathbf{b}$) se denominan *espacios afines*. La dimensión de un espacio afín es la dimensión del subespacio generado por las combinaciones lineales del citado sistema de vectores $[\mathbf{n}_1; \dots \mathbf{n}_k;]$ (o, equivalentemente, la dimensión del espacio nulo de \mathbf{A}).

NAcAL guarda el conjunto de soluciones de $\mathbf{A}\mathbf{x} = \mathbf{b}$ como un espacio afín en el atributo **eafin**. Su representación muestra tanto las ecuaciones cartesianas como las paramétricas.

```
A = Matrix( [ [-1,1,0,0], [-1,0,-1,1] ] )
b = Vector([0,1])
SEL(A,b).eafin # conjunto de soluciones
```

$$\left\{ \mathbf{v} \in \mathbb{R}^4 \mid \exists \mathbf{p} \in \mathbb{R}^2, \mathbf{v} = \begin{pmatrix} -1 \\ -1 \\ 0 \\ 0 \end{pmatrix} + \begin{bmatrix} -1 & 1 \\ -1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{p} \right\} = \left\{ \mathbf{v} \in \mathbb{R}^4 \mid \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & -1 & 1 \end{bmatrix} \mathbf{v} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$$

NAcAL permite generar espacios afines de varias maneras. Con una matriz \mathbf{A} y un punto \mathbf{v} :

```
A = Matrix( [ [-1,1,0,0], [-1,0,-1,1] ] )
v = Vector([0,0,0,1])
EAfin(A,v) # contiene al vector v y es paralelo a N(A)
```

$$\left\{ \mathbf{v} \in \mathbb{R}^4 \mid \exists \mathbf{p} \in \mathbb{R}^2, \mathbf{v} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} + \begin{bmatrix} -1 & 1 \\ -1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{p} \right\} = \left\{ \mathbf{v} \in \mathbb{R}^4 \mid \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & -1 & 1 \end{bmatrix} \mathbf{v} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$$

o con un sistema de vectores \mathbf{S} y un punto \mathbf{v} :

```
S = Sistema([ Vector([-1,-1,1,0]), Vector([0,0,1,1]) ])
v = Vector([0,0,0,1])
EAfin(S,v) # contiene al vector v y es paralelo a L(S)
```

$$\left\{ \mathbf{v} \in \mathbb{R}^4 \mid \exists \mathbf{p} \in \mathbb{R}^2, \mathbf{v} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} + \begin{bmatrix} -1 & 0 \\ -1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \mathbf{p} \right\} = \left\{ \mathbf{v} \in \mathbb{R}^4 \mid \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & -1 & 1 \end{bmatrix} \mathbf{v} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$$

(POR HACER: definir el conjunto de soluciones como un elemento de un espacio cociente.)

11.3. Generalizaciones o visiones alternativas (★)



Aunque hasta este momento nos hemos limitado a trabajar en \mathbb{R}^n , lo interesante es que las ideas de longitud, ortogonalidad, norma, distancia y ángulo entre vectores son “de algún modo” extrapolables a espacios vectoriales genéricos \mathcal{V} cuando en ellos definimos *productos escalares*. De hecho, las aplicaciones más interesantes ocurren en espacios vectoriales de dimensión infinita, por ejemplo en la Probabilidad.

En esta sección definimos el concepto de producto escalar en un espacio vectorial abstracto. Luego generalizamos las nociones de longitud, ortogonalidad, norma, distancia y ángulo entre vectores. Para finalizar, trataremos con subespacios ortogonales.

11.3.1. Productos escalares en \mathcal{V}

Llamamos producto escalar entre dos vectores \vec{x} e \vec{y} de \mathcal{V} a cualquier función $\langle _, _ \rangle: \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}$ que verifique los axiomas de

- **Simetría:** $\langle \vec{x}, \vec{y} \rangle = \langle \vec{y}, \vec{x} \rangle$.
- **Linealidad respecto al primer argumento:**
 1. $\langle (\alpha \vec{x}), \vec{y} \rangle = \alpha \langle \vec{x}, \vec{y} \rangle$.
 2. $\langle (\vec{x} + \vec{y}), \vec{z} \rangle = \langle \vec{x}, \vec{z} \rangle + \langle \vec{y}, \vec{z} \rangle$.
- **Positivo:** $\langle \vec{x}, \vec{x} \rangle \geq 0$.
- **Definido:** $\langle \vec{x}, \vec{x} \rangle = 0 \Leftrightarrow \vec{x} = \vec{0}$.

🔗 Como ilustración veamos algunos ejemplos de productos escalares. Primero en \mathbb{R}^n y luego en otros espacios.

Ejemplos de productos escalares en \mathbb{R}^n

Fíjese que el producto punto es un producto escalar en \mathbb{R}^n

$$\begin{aligned} \langle _, _ \rangle: \mathbb{R}^n \times \mathbb{R}^n &\longrightarrow \mathbb{R} \\ (\mathbf{a}, \mathbf{b}) &\longmapsto \mathbf{a} \cdot \mathbf{b} \end{aligned}$$

de hecho, el *producto punto* también se denomina *producto escalar usual en \mathbb{R}^n* . Pero hay otros ejemplos de producto escalar en \mathbb{R}^n de uso muy expandido.

Ejemplo 28. En estadística se emplea un producto escalar que garantiza que la norma del vector $\mathbf{1}$ (cuyas componentes son todas iguales a uno) es 1 independientemente del número de unos que contenga el vector.

$$\begin{aligned} \langle _, _ \rangle_s: \mathbb{R}^n \times \mathbb{R}^n &\longrightarrow \mathbb{R} \\ (\mathbf{a}, \mathbf{b}) &\longmapsto \frac{1}{n}(\mathbf{a} \cdot \mathbf{b}) \end{aligned}$$

(Véase la Sección 19.1 en la página 211).

En la Lección 19 veremos que, usando este producto escalar, la desviación típica y el valor absoluto de la media aritmética son longitudes; y que la correlación es el coseno del ángulo formado entre vectores perpendiculares a $\mathbf{1}$.

EJERCICIO 60. Demuestre que la función $\langle _, _ \rangle_s: \mathcal{V} \times \mathcal{V} \longrightarrow \mathbb{R}$ es un producto escalar en \mathbb{R}^n .

Ejemplo 29. En estadística también se emplean las *medias ponderadas*, en las que unos datos tienen mayor peso que otros. Pero se mantiene la propiedad de que la media (ponderada) de un vector de n datos todos iguales a uno es uno. Un ejemplo es

$$\begin{aligned} \langle _, _ \rangle_{sp}: \mathbb{R}^n \times \mathbb{R}^n &\longrightarrow \mathbb{R} \\ (\mathbf{a}, \mathbf{b}) &\longmapsto \mathbf{a} \mathbf{D} \mathbf{b} \end{aligned}$$

donde \mathbf{D} es una matriz diagonal de orden n cuyos elementos en la diagonal son positivos y suman 1. Fíjese que el anterior producto escalar $\langle _, _ \rangle_s$ corresponde al caso particular $\mathbf{D} = \frac{1}{n} \mathbf{I}$.

EJERCICIO 61. Considere la matriz diagonal \mathbf{D} de orden n y cuyos elementos en la diagonal son positivos y suman 1. Demuestre que la función $\langle _, _ \rangle_{sp}: \mathcal{V} \times \mathcal{V} \longrightarrow \mathbb{R}$, es un producto escalar en \mathbb{R}^n .

Ejemplo 30. Para toda matriz \mathbf{A} de orden n y rango completo, la función

$$\begin{aligned} \langle _, _ \rangle: \mathcal{V} \times \mathcal{V} &\longrightarrow \mathbb{R} \\ (\mathbf{a}, \mathbf{b}) &\longmapsto \mathbf{a} \mathbf{A}^\top \mathbf{A} \mathbf{b} \end{aligned},$$

es un producto escalar en \mathbb{R}^n . De hecho, para cualquier producto escalar definido en \mathbb{R}^n se puede encontrar una matriz \mathbf{A} tal que dicho producto escalar se puede expresar de esta manera (ha de tenerse en cuenta que dicha matriz \mathbf{A} no es única —véase la diagonalización por congruencia de la Lección 18).

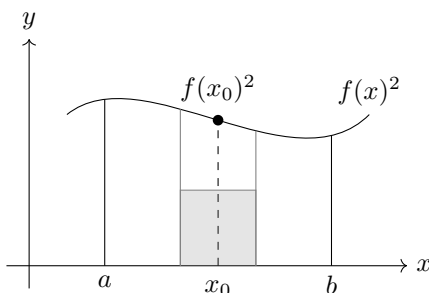
Ejemplos de productos escalares en otros subespacios

Ejemplo 31. El subconjunto $\mathcal{C}[a, b]$ de funciones reales y continuas con dominio en el intervalo $[a, b]$, junto con las operaciones habituales de suma de funciones y producto de funciones por escalares, es un subespacio del espacio $\mathbb{R}^{[a, b]}$ de funciones reales de variable real con dominio en el intervalo $[a, b]$. En $\mathcal{C}[a, b]$, la función

$$\begin{aligned} \langle _, _ \rangle: \mathcal{C}[0, 1] \times \mathcal{C}[0, 1] &\longrightarrow \mathbb{R} \\ (f, g) &\longmapsto \int_a^b f(x)g(x)dx \end{aligned}$$

es un producto escalar, pues verifica los axiomas:

- **Simetría:** $\int_a^b f(x)g(x)dx = \int_a^b g(x)f(x)dx$
- **Linealidad respecto al primer argumento:**
 1. $\int_a^b (f(x) + g(x))h(x)dx = \int_a^b f(x)h(x)dx + \int_a^b g(x)h(x)dx$
 2. $\int_a^b \alpha f(x)g(x)dx = \alpha \int_a^b f(x)g(x)dx$
- **Positivo:** si f no es nula entonces $\int_a^b f(x)^2 dx > 0$, ya que para el punto x_0 en el que la función es distinta de cero, existe un entorno tal que si $z \in (x_0 - \delta, x_0 + \delta)$ entonces $f(z)^2 \geq \frac{1}{2} \int_a^b f(x_0)^2 dx$.



- **Definido:** $\int_a^b f(x)^2 dx = 0$, si y solo si f es la función nula, es decir, $f(x) = 0$ para todo $x \in [a, b]$.

Ejemplo 32. Otro ejemplo que encontrará en estadística y econometría corresponde al conjunto de variables aleatorias de varianza finita definidas en un espacio de probabilidad (Ω, \mathcal{F}, P) . En dicho subespacio, la esperanza $E(XY)$ del producto entre las variables aleatorias X e Y es un producto escalar.

11.3.2. Generalizando los conceptos geométricos a otros espacios vectoriales

🔗 Aquí replicaremos la Sección 11.1.3 sustituyendo la notación de producto punto por la de producto escalar.

Definición 11.8. Decimos que \vec{x} e \vec{y} son ortogonales o perpendiculares ($\vec{x} \perp \vec{y}$) cuando $\langle \vec{x}, \vec{y} \rangle = 0$.

EJERCICIO 62. ¿Qué vector es ortogonal a cualquier otro (incluido él mismo)?

Definición 11.9. La longitud (o norma) de un vector \vec{x} de \mathcal{V} es la raíz cuadrada de $\langle \vec{x}, \vec{x} \rangle$:

$$\text{longitud de } \vec{x} = \|\vec{x}\| = \sqrt{\langle \vec{x}, \vec{x} \rangle}.$$

EJERCICIO 63. Si $\|\vec{x}\| = 0$, ¿Quién es \vec{x} ?

Multiplicando \vec{x} por λ obtenemos un múltiplo de \vec{x} cuya longitud es $|\lambda|$ veces su longitud original.

Definición 11.10 (Vector unitario). Se dice que un vector es unitario si su longitud es uno.

Para obtener un múltiplo unitario de \vec{x} basta dividir el vector \vec{x} por su longitud.

Teorema 11.3.1 (Teorema de Pitágoras en \mathcal{V}). Sean \vec{x} e \vec{y} son dos vectores de \mathcal{V} ; entonces $\langle \vec{x}, \vec{y} \rangle = 0$ (son perpendiculares) si y solo si

$$\|\vec{x} + \vec{y}\|^2 = \|\vec{x}\|^2 + \|\vec{y}\|^2.$$

EJERCICIO 64. Demuestre el teorema de Pitágoras en \mathcal{V} con producto escalar $\langle _, _ \rangle: \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}$.

Proposición 11.3.2 (Desigualdad de Cauchy-Schwarz). Para \vec{x} e \vec{y} de \mathcal{V} y el producto escalar $\langle _, _ \rangle$ definido en \mathcal{V} se verifica que

$$|\langle \vec{x}, \vec{y} \rangle| \leq \|\vec{x}\| \|\vec{y}\|.$$

EJERCICIO 65. Demuestre la desigualdad de Cauchy-Schwarz.

Así, en un extremo $\langle \vec{x}, \vec{y} \rangle = 0$ cuando $\vec{x} \perp \vec{y}$; y en el otro $|\langle \vec{x}, \vec{y} \rangle| = \|\vec{x}\| \|\vec{y}\|$ cuando $\vec{y} = \lambda \vec{x}$.

Definición 11.11. El coseno del ángulo θ formado por los vectores no nulos \vec{x} y \vec{y} de \mathcal{V} es

$$\cos \theta = \frac{\langle \vec{x}, \vec{y} \rangle}{\|\vec{x}\| \|\vec{y}\|}; \quad -1 \leq \cos \theta \leq 1. \quad (11.5)$$

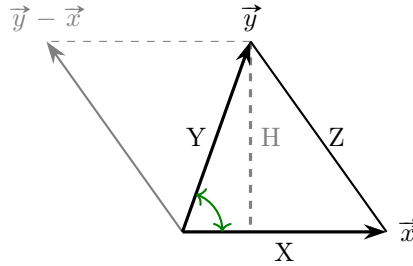


Figura 11.5: Representación esquemática del triángulo formado por $\vec{0}$ y dos vectores \vec{x} y \vec{y} no nulos de \mathcal{V} .

Proposición 11.3.3 (Desigualdad triangular). Sean \vec{x} y \vec{y} vectores de \mathcal{V} , entonces $\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$.

EJERCICIO 66. Demuestre la desigualdad triangular en \mathcal{V} .

La Desigualdad Triangular da lugar a la noción de *distancia entre vectores de \mathcal{V}* :

$$\text{distancia entre } \vec{x} \text{ y } \vec{y}: d(\vec{x}, \vec{y}) = \|\vec{x} - \vec{y}\| = \|\vec{y} - \vec{x}\| = d(\vec{y}, \vec{x}),$$

Puesto que $\|\vec{x}\| = 0$ si y solo si $\vec{x} = \vec{0}$, tenemos que $d(\vec{x}, \vec{y}) = \|\vec{x} - \vec{y}\| = 0$ si y solo si $\vec{x} = \vec{y}$.

Vectores alineados. Cuando $\vec{y} = \lambda \vec{x}$ (cuando \vec{y} es un múltiplo de \vec{x}) el producto punto entre ambos es

$$\langle \vec{y}, \vec{x} \rangle = \lambda \langle \vec{x}, \vec{x} \rangle = \lambda \|\vec{x}\|^2$$

En tal caso (y si $\lambda \neq 0$) el ángulo que forman \vec{x} e \vec{y} es cero o bien 180° (es decir 0 ó π radianes) ya que:

$$\cos \theta = \frac{\langle \vec{x}, \vec{y} \rangle}{\|\vec{x}\| \|\vec{y}\|} = \frac{\lambda \|\vec{x}\|^2}{|\lambda| \cdot \|\vec{x}\| \|\vec{x}\|} = \frac{\lambda}{|\lambda|} = \pm 1.$$

y entonces se dice que los vectores están alineados.

11.3.3. Subespacios ortogonales en espacios vectoriales abstractos

Definición 11.12 (Conjunto ortogonal al subconjunto A de vectores de \mathcal{V}). Llamamos conjunto ortogonal (o perpendicular) a A , que denotamos como A^\perp , al subconjunto de vectores de \mathcal{V} que son perpendiculares a los vectores de A .

$$A^\perp = \{ \vec{x} \in \mathcal{V} \mid \langle \vec{x}, \vec{a} \rangle = 0 \text{ para cada } \vec{a} \in A \}$$

EJERCICIO 67. Demuestre que A^\perp es un subespacio.

¡Fíjese que aunque A no sea un subespacio, A^\perp sí lo es!

EJERCICIO 68.

- (a) Demuestre que si B es un subconjunto de vectores de \mathcal{V} y $A \subset B$ entonces $B^\perp \subset A^\perp$.
- (b) Demuestre que para todo subconjunto A de vectores de \mathcal{V} , se verifica que $A^\perp = \mathcal{L}(A)^\perp$.
(donde $\mathcal{L}(A)$ corresponde a la Definición 9.11 en la página 118)
- (c) Demuestre que $\mathcal{A} \cap \mathcal{A}^\perp = \{ \vec{0} \}$.

11.3.4. Espacios pre-Hilbert

Definición 11.13. Se denomina *espacio pre-Hilbert* al par $(\mathcal{V}, \langle _, _ \rangle)$, donde \mathcal{V} es un espacio vectorial y $\langle _, _ \rangle$ es un producto escalar definido en \mathcal{V} .

Definición 11.14. Cuando \mathcal{V} es de *dimensión finita*, el espacio $(\mathcal{V}, \langle _, _ \rangle)$ se denomina *espacio euclídeo*.

Observación. Cuando nos referimos al *espacio euclídeo* \mathbb{R}^n se sobreentiende que es aquél definido sobre \mathbb{R}^n junto con el *producto punto*, es decir, al par: $(\mathbb{R}^n, _ \cdot _)$.

Complementos ortogonales

Con la Proposición 11.2.1 en la página 139 ya vimos que en el espacio euclídeo $(\mathbb{R}^m, _ \cdot _)$ todo subespacio $\mathcal{A} \subset \mathbb{R}^m$ y su conjunto ortogonal \mathcal{A}^\perp son *suplementarios*.

Ahora vamos a demostrar que en cualquier espacio pre-Hilbert $(\mathcal{V}, \langle _, _ \rangle)$, si el subespacio \mathcal{H} es de *dimensión finita* entonces $\mathcal{V} = \mathcal{H} \oplus \mathcal{H}^\perp$, es decir, que subespacios los \mathcal{H} y \mathcal{H}^\perp son *suplementarios* cuando \mathcal{H} es de *dimensión finita*.

Teorema 11.3.4. Si $(\mathcal{V}, \langle _, _ \rangle)$ es un espacio pre-Hilbert y \mathcal{H} es un subespacio de \mathcal{V} de *dimensión finita*, entonces necesariamente $\mathcal{H} + \mathcal{H}^\perp = \mathcal{V}$. Por tanto si \mathcal{H} es de *dimensión finita* se cumple que $\mathcal{H} \oplus \mathcal{H}^\perp = \mathcal{V}$.

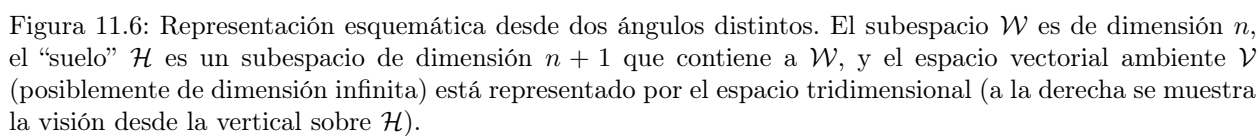
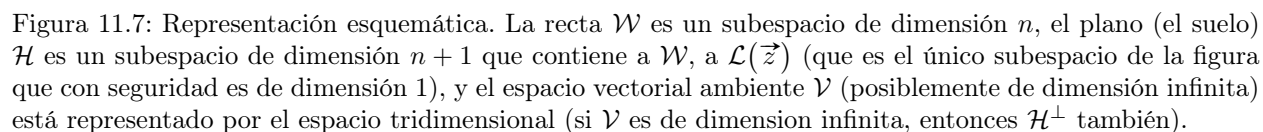
Demostración. Probar que $\mathcal{H} + \mathcal{H}^\perp = \mathcal{V}$ consiste en demostrar que para cualquier $\vec{v} \in \mathcal{V}$ existe $\vec{h} \in \mathcal{H}$ tal que $\vec{v} - \vec{h} \in \mathcal{H}^\perp$. Para ello razonamos por inducción sobre la dimensión de \mathcal{H} .

Si $\dim(\mathcal{H}) = 0$ entonces $\mathcal{H} = \{ \vec{0} \}$ y $\mathcal{H}^\perp = \mathcal{V}$, con lo que $\mathcal{H} + \mathcal{H}^\perp = \mathcal{V}$.

Ahora supongamos que el resultado es cierto para subespacios de dimensión n y que $\dim(\mathcal{H}) = n + 1$. Sea entonces $[\vec{w}_1; \dots; \vec{w}_n]$ un sistema de n vectores linealmente independientes de \mathcal{H} y denotemos con \mathcal{W} al subespacio $\mathcal{L}([\vec{w}_1; \dots; \vec{w}_n])$. Obviamente $\dim(\mathcal{W}) = n$ y $\mathcal{W} \subset \mathcal{H}$; y como $\dim(\mathcal{H}) = n + 1$, existe $\vec{h} \in \mathcal{H} - \mathcal{W}$ tal que $[\vec{w}_1; \dots; \vec{w}_n; \vec{h}]$ es una base de \mathcal{H} . Por hipótesis de inducción sabemos que existe $\vec{h}_w \in \mathcal{W}$ tal que $\vec{h} - \vec{h}_w \in \mathcal{W}^\perp$. Y como $\vec{h}_w \in \mathcal{L}([\vec{w}_1; \dots; \vec{w}_n])$ tendremos que también $[\vec{w}_1; \dots; \vec{w}_n; \vec{h} - \vec{h}_w]$ es una base de \mathcal{H} . Denotemos pues con \vec{z} al vector $\vec{h} - \vec{h}_w$.

Ahora tomemos $\vec{v} \in \mathcal{V}$. Por hipótesis de inducción existe $\vec{v}_w \in \mathcal{W}$ tal que $\vec{v} - \vec{v}_w$ pertenece a \mathcal{W}^\perp . Por tanto $\{ \vec{v} - \vec{v}_w, \vec{z} \} \perp \{ \vec{w}_1, \dots, \vec{w}_n \}$, con lo que para cualquier $\alpha \in \mathbb{R}$ se verifica que

$$\begin{cases} \vec{v} - \vec{v}_w - \alpha \vec{z} & \perp & \{ \vec{w}_1, \dots, \vec{w}_n \} \\ \vec{v}_w + \alpha \vec{z} & \in & \mathcal{H} \end{cases} \quad (11.6)$$


$$\begin{aligned} 0 &= \left\langle (\vec{v} - \vec{v}_w - \alpha_0 \vec{z}), \vec{z} \right\rangle \\ &= \left\langle (\vec{v} - \vec{v}_w), \vec{z} \right\rangle - \alpha_0 \|\vec{z}\|^2 \end{aligned} \Leftrightarrow \alpha_0 = \frac{\langle (\vec{v} - \vec{v}_w), \vec{z} \rangle}{\|\vec{z}\|^2}$$


148

Un subespacio de dimensión infinita y su conjunto ortogonal no siempre son suplementarios



Acabamos de ver que \mathcal{H} y \mathcal{H}^\perp son suplementarios cuando \mathcal{H} es de *dimensión finita*.

Pero \mathcal{H} y \mathcal{H}^\perp pueden *no ser* suplementarios cuando el subespacio \mathcal{H} es de *dimensión infinita*.^a En esta sección veremos un ejemplo.

^aAunque si \mathcal{H} es un subespacio cerrado de un espacio de Hilbert \mathcal{V} , entonces $\mathcal{V} = \mathcal{H} \oplus \mathcal{H}^\perp$ (independientemente de la dimensión de \mathcal{H}). En este caso \mathcal{H}^\perp se llama complemento ortogonal de \mathcal{V} ... pero esto queda fuera del ámbito del curso.

Ejemplo 33. Considere el conjunto $\mathbb{R}[x]$ de sucesiones infinitas de números reales, que se hacen cero a partir de una posición. Este conjunto también es descrito como el conjunto de sucesiones infinitas de números reales con un número finito de componentes son distintas de cero; por lo que dicho conjunto también se llama *conjunto de sucesiones casi nulas*. Algunos elementos de dicho conjunto son:

$$\begin{aligned} a &= (1, 0, 0, 0, 0, 0, 0, 0 \dots) \\ b &= (1, 2, 0, 0, 0, 0, 0, 0 \dots) \\ c &= (3, 2, 1, 0, 0, 0, 0, 0 \dots) \\ d &= (0, 1, 0, 1, 0, 0, 0, 0 \dots) \\ e &= (1, 2, 3, 4, 5, 0, 0, 0 \dots) \end{aligned}$$

Definamos el producto escalar como la suma del producto entre componentes:

$$\begin{aligned} \langle _, _ \rangle: \mathbb{R}[x] \times \mathbb{R}[x] &\longrightarrow \mathbb{R} \\ (p, q) &\longmapsto \sum_{i=0}^{\infty} p_i q_i \end{aligned}$$

(como el producto punto pero con una suma infinita donde a partir de algún índice k los sumandos son cero).

Así, con las sucesiones infinitas anteriores tenemos que: $\langle a, b \rangle = 1$, $\langle c, d \rangle = 2$, y $\langle d, e \rangle = 6$.

Ahora consideremos el subespacio $\mathcal{H} \subset \mathbb{R}[x]$ de sucesiones cuyas componentes suman cero:

$$\mathcal{H} = \left\{ p \in \mathbb{R}[x] \mid \sum_i p_i = 0 \right\}.$$

(fíjese que ninguna de las sucesiones a , b , c , d y e pertenece a \mathcal{H})

El conjunto ortogonal a \mathcal{H} es

$$\mathcal{H}^\perp = \left\{ p \in \mathbb{R}[x] \mid \langle p, q \rangle = 0 \text{ para todo } q \in \mathbb{R}[x] \right\}.$$

Consecuentemente los elementos de \mathcal{H}^\perp deben ser:

- perpendiculares a $(1, -1, 0, 0, 0, 0 \dots)$; así que $p_0 = p_1$,
- perpendiculares a $(1, 1, -2, 0, 0, 0 \dots)$; así que $p_0 + p_1 = 2p_2$ y por tanto $p_0 = p_1 = p_2$,
- perpendiculares a $(1, 1, 1, -3, 0, 0 \dots)$; así que $p_0 + p_1 + p_2 = 3p_3$ y por tanto $p_0 = p_1 = p_2 = p_3$,
- perpendiculares a $(1, 1, 1, 1, -4, 0 \dots)$; así que $p_0 + p_1 + p_2 + p_3 = 4p_4$ y por tanto $p_0 = p_1 = p_2 = p_3 = p_4$,
- ...

Por tanto, los elementos de \mathcal{H}^\perp son sucesiones infinitas cuyas componentes son todas iguales; pero como las sucesiones de $\mathbb{R}[x]$ son cero a partir de alguna posición, el único elemento posible de \mathcal{H}^\perp es la sucesión nula.

Consecuentemente, el subespacio $\mathcal{H} + \mathcal{H}^\perp$ no contiene ninguna de las sucesiones a , b , c , d y e ; es decir, $\mathcal{H} + \mathcal{H}^\perp$ no es todo el espacio $\mathbb{R}[x]$. Así pues, \mathcal{H} y \mathcal{H}^\perp no son suplementarios.

Fíjese que el subespacio de dimensión infinita formado las *sucesiones casi nulas* $\mathbb{R}[x]$ resulta ser el conjunto de polinomios en una variable³, aunque habitualmente se escriben de otra manera... en el caso de los del Ejemplo 33 en la página anterior:

$$a = (1, 0, 0, 0, 0, 0, 0, 0 \dots) = 1$$

$$b = (1, 2, 0, 0, 0, 0, 0, 0 \dots) = 1 + 2x$$

$$c = (3, 2, 1, 0, 0, 0, 0, 0 \dots) = 3 + 2x + x^2$$

$$d = (0, 1, 0, 1, 0, 0, 0, 0 \dots) = x + x^3$$

$$e = (1, 2, 3, 4, 5, 0, 0, 0 \dots) = 1 + 2x + 3x^2 + 4x^3 + 5x^4$$

así, el exponente de cada “ x ” indica la posición de su correspondiente coeficiente dentro de la sucesión; donde se asocia la primera posición al índice 0, es decir, puesto que $y^0 = 1$, entonces $ax^0 = (a, 0, 0 \dots)$.

³Véase la definición de polinomio en la Sección 18 en la página 199

Proyecciones sobre subespacios

12.1. Proyección ortogonal y mínima distancia

Definición 12.1. Llamamos proyección ortogonal sobre el subespacio $\mathcal{V} \subset \mathbb{R}^m$ a la función¹ $f(\mathbf{y})$ tal que la diferencia $\mathbf{y} - f(\mathbf{y})$ es ortogonal a \mathcal{V} ; es decir, tal que $(\mathbf{y} - f(\mathbf{y})) \in \mathcal{V}^\perp$.

Para cada \mathbf{y} de \mathbb{R}^m , el vector $\mathbf{p}_\mathbf{y} = f(\mathbf{y})$ de \mathcal{V} es la *proyección ortogonal de \mathbf{y} sobre el subespacio $\mathcal{V} \subset \mathbb{R}^m$* . Como $\mathbb{R}^m = \mathcal{V} \oplus \mathcal{V}^\perp$, sabemos que el vector $\mathbf{p}_\mathbf{y} \in \mathcal{V}$ existe y es único (Proposición 11.2.1)².

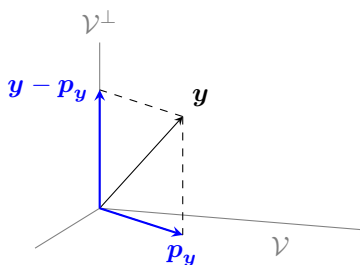


Figura 12.1: Representación esquemática de la proyección ortogonal de \mathbf{y} sobre \mathcal{V}

Fíjese que en la Figura 11.4 en la página 137, el vector $-\alpha\mathbf{a}$ es la proyección ortogonal de \mathbf{b} sobre $\mathcal{L}(\mathbf{a})$. Es decir, aún sin nombrar la proyección ortogonal, la hemos empleado implícitamente en la lección anterior.

Lo más interesante de $\mathbf{p}_\mathbf{y}$ se afirma en la siguiente

Proposición 12.1.1. La proyección ortogonal de $\mathbf{y} \in \mathbb{R}^m$ sobre $\mathcal{V} \subset \mathbb{R}^m$ es el vector de \mathcal{V} más próximo a \mathbf{y}

Demostración. Sea $\mathbf{p}_\mathbf{y}$ la proyección de \mathbf{y} sobre \mathcal{V} y tomemos un vector \mathbf{v} cualquiera de \mathcal{V} . Veamos que \mathbf{y} está más lejos de \mathbf{v} que de su proyección ortogonal $\mathbf{p}_\mathbf{y}$.

Como \mathbf{v} y $\mathbf{p}_\mathbf{y}$ están en \mathcal{V} , su diferencia $(\mathbf{p}_\mathbf{y} - \mathbf{v})$ está en \mathcal{V} (pues \mathcal{V} es subespacio); consecuentemente $(\mathbf{p}_\mathbf{y} - \mathbf{v})$ es ortogonal a $(\mathbf{y} - \mathbf{p}_\mathbf{y}) \in \mathcal{V}^\perp$ (por ser $\mathbf{p}_\mathbf{y}$ la proyección ortogonal sobre \mathcal{V}). Y como la suma de ambos vectores perpendiculares es $(\mathbf{y} - \mathbf{p}_\mathbf{y}) + (\mathbf{p}_\mathbf{y} - \mathbf{v}) = (\mathbf{y} - \mathbf{v})$, por el Tma. de Pitágoras concluimos que

$$\|\mathbf{y} - \mathbf{v}\|^2 = \|\mathbf{y} - \mathbf{p}_\mathbf{y}\|^2 + \|\mathbf{p}_\mathbf{y} - \mathbf{v}\|^2 \geq \|\mathbf{y} - \mathbf{p}_\mathbf{y}\|^2.$$

Por tanto, $\|\mathbf{y} - \mathbf{v}\| \geq \|\mathbf{y} - \mathbf{p}_\mathbf{y}\|$ para todo $\mathbf{v} \in \mathcal{V}$. □

¹Al ser un caso particular del apartado (b) del ejercicio 55 en la página 123, la función proyección es lineal.

²Por otra parte, el vector $\mathbf{y} - \mathbf{p}_\mathbf{y}$ es la *proyección ortogonal de \mathbf{y} sobre \mathcal{V}^\perp* .

☞ Por la Proposición 11.2.1 sabemos que para toda matriz \mathbf{A} de m filas y un vector \mathbf{y} de \mathbb{R}^m , existen $\mathbf{p}_y \in \mathcal{C}(\mathbf{A})$ y $\mathbf{e} \in \mathcal{N}(\mathbf{A}^\top)$ tales que $\mathbf{y} = \mathbf{p}_y + \mathbf{e}$. Pero ¿cómo encontrar dichos vectores \mathbf{p}_y y \mathbf{e} ? Veámoslo...

12.1.1. Sistema de ecuaciones normales

Para encontrar la proyección ortogonal \mathbf{p}_y del vector \mathbf{y} sobre $\mathcal{C}(\mathbf{A})$ combinaremos lo que sabemos:

- \mathbf{p}_y es una combinación lineal de las columnas de \mathbf{A} , es decir, existe $\mathbf{x} \in \mathbb{R}^n$ tal que $\mathbf{Ax} = \mathbf{p}_y$.
- \mathbf{p}_y es la proyección ortogonal de \mathbf{y} sobre $\mathcal{C}(\mathbf{A})$, es decir $(\mathbf{y} - \mathbf{p}_y) \in \mathcal{N}(\mathbf{A}^\top)$.

Así pues, una combinación lineal \mathbf{Ax} resulta ser \mathbf{p}_y si y solo si $\mathbf{A}^\top(\mathbf{y} - \mathbf{Ax}) = \mathbf{0}$:

$$\mathbf{Ax} = \mathbf{p}_y \iff \mathbf{A}^\top(\mathbf{y} - \mathbf{Ax}) = \mathbf{A}^\top\mathbf{y} - (\mathbf{A}^\top\mathbf{A})\mathbf{x} = \mathbf{0} \iff \boxed{(\mathbf{A}^\top\mathbf{A})\mathbf{x} = \mathbf{A}^\top\mathbf{y}}. \quad (12.1)$$

El sistema de ecuaciones $(\mathbf{A}^\top\mathbf{A})\mathbf{x} = \mathbf{A}^\top\mathbf{y}$ se denomina *sistema de ecuaciones normales*; y para obtener la proyección de \mathbf{y} sobre $\mathcal{C}(\mathbf{A})$ basta multiplicar \mathbf{A} por cualquier vector \mathbf{x} que sea solución a dicho sistema.

Lo llamativo es que para resolver $\mathbf{Ax} = \mathbf{p}_y$ (para encontrar la combinación lineal de las columnas de \mathbf{A} requerida) resolvemos un sistema diferente (el *sistema de ecuaciones normales*) donde no aparece \mathbf{p}_y . Este procedimiento indirecto funciona porque el *sistema de ecuaciones normales* y el sistema $\mathbf{Ax} = \mathbf{p}_y$ tienen el mismo conjunto de soluciones (fíjese en las implicaciones “si y solo si” de (12.1)).

Cuando dos sistemas de ecuaciones lineales tienen el mismo conjunto de soluciones decimos que son *sistemas de ecuaciones equivalentes*.

Como el sistema de ecuaciones $\mathbf{Ax} = \mathbf{p}_y$ tiene solución, ya que $\mathbf{p}_y \in \mathcal{C}(\mathbf{A})$, también sabemos que el *sistema de ecuaciones normales* tiene solución, pues ambos son sistemas de ecuaciones equivalentes. Es más, por ser sistemas de ecuaciones equivalentes, *ambos tienen solución única si y solo si las columnas de \mathbf{A} son linealmente independientes*.

12.1.2. Alternativa a un sistema de ecuaciones lineales sin solución

Imagine que tiene el siguiente sistema de ecuaciones

$$\mathbf{Ax} = \mathbf{y} \quad \text{donde } \mathbf{y} \notin \mathcal{C}(\mathbf{A}).$$

Como ninguna combinación de las columnas de \mathbf{A} es igual a \mathbf{y} , no es posible encontrar una solución.

Pero podemos plantearnos el siguiente *problema alternativo*: ¿qué combinación lineal de las columnas de \mathbf{A} está más próxima a \mathbf{y} ? Es decir,

$$\mathbf{Ax} = \mathbf{p}_y,$$

donde \mathbf{p}_y es la proyección ortogonal de \mathbf{y} sobre $\mathcal{C}(\mathbf{A})$. Este segundo problema siempre tiene solución, pues \mathbf{p}_y existe y pertenece al subespacio $\mathcal{C}(\mathbf{A})$.

Ésta es la idea detrás del ajuste mínimo cuadrático en estadística y econometría (Véase la Parte VII)

Observación. Cuando el *sistema de ecuaciones normales* tiene solución única, ésta *suele ser interpretada como la mejor “respuesta”* al sistema de ecuaciones irresoluble

$$\mathbf{Ax} = \mathbf{y}; \quad \text{donde } \mathbf{y} \notin \mathcal{C}(\mathbf{A});$$

pero recuerde que, *en realidad*, es la solución a un sistema de ecuaciones completamente distinto:

$$\mathbf{Ax} = \mathbf{p}_y; \quad \text{donde } \mathbf{p}_y \text{ es la proyección ortogonal de } \mathbf{y} \text{ sobre } \mathcal{C}(\mathbf{A}).$$

12.2. Expresión matricial de la proyección ortogonal



Por último, vamos a encontrar una expresión matricial de la función proyección sobre un subespacio \mathcal{V} . Dicha expresión es la que se emplea en los manuales de estadística y econometría.

Como $\mathbf{A}\mathbf{x} = \mathbf{p}_y$ y $(\mathbf{A}^\top \mathbf{A})\mathbf{x} = \mathbf{A}^\top \mathbf{y}$ son *sistemas de ecuaciones lineales equivalentes*, necesariamente sus matrices de coeficientes tienen el mismo espacio nulo $\mathcal{N}(\mathbf{A}) = \mathcal{N}(\mathbf{A}^\top \mathbf{A})$; y como ambas matrices tienen n columnas, también tienen el mismo rango.

Consecuentemente, si las n columnas de \mathbf{A} son linealmente independientes entonces la matriz cuadrada $\mathbf{A}^\top \mathbf{A}$ es de rango completo (invertible) y la *solución* al sistema de ecuaciones normales $(\mathbf{A}^\top \mathbf{A})\mathbf{x} = \mathbf{A}^\top \mathbf{y}$ es

$$\hat{\mathbf{x}} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{y}.$$

En tal caso el vector *proyección* de \mathbf{y} sobre $\mathcal{C}(\mathbf{A})$ resulta ser

$$\mathbf{p}_y = \mathbf{A}\hat{\mathbf{x}} = \mathbf{A}(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{y}; \quad (12.2)$$

y por tanto el vector *proyección* de \mathbf{y} sobre $\mathcal{N}(\mathbf{A}^\top)$ es

$$\mathbf{y} - \mathbf{p}_y = \mathbf{I}\mathbf{y} - \mathbf{A}\hat{\mathbf{x}} = (\mathbf{I} - \mathbf{A}(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top) \mathbf{y}.$$

12.2.1. Matrices proyección

Si \mathbf{A} es de rango completo por columnas y llamamos \mathbf{P} a la matriz $\mathbf{A}(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top$ de la Ecuación 12.2,

$$\begin{aligned} \mathbf{p}_y &= \mathbf{P}\mathbf{y} && \text{es la proyección de } \mathbf{y} \text{ sobre } \mathcal{C}(\mathbf{A}) \\ \mathbf{e} &= (\mathbf{I} - \mathbf{P})\mathbf{y} && \text{es la proyección de } \mathbf{y} \text{ sobre } \mathcal{N}(\mathbf{A}^\top) \end{aligned}$$

Tanto \mathbf{P} como $(\mathbf{I} - \mathbf{P})$ poseen dos propiedades íntimamente relacionadas con la proyección ortogonal:

- Ambas matrices son *simétricas*. Por una parte

$$\mathbf{P}^\top = (\mathbf{A}(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top)^\top = (\mathbf{A}^\top)^\top ((\mathbf{A}^\top \mathbf{A})^{-1})^\top \mathbf{A}^\top = \mathbf{A}(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top = \mathbf{P},$$

pues la inversa de una matriz simétrica es simétrica: $((\mathbf{A}^\top \mathbf{A})^{-1})^\top = ((\mathbf{A}^\top \mathbf{A})^\top)^{-1} = (\mathbf{A}^\top \mathbf{A})^{-1}$.

EJERCICIO 69. Por otra parte, demuestre que $(\mathbf{I} - \mathbf{P})^\top = \mathbf{I} - \mathbf{P}$.

- Ambas matrices son *idempotentes*, es decir, son iguales a su cuadrado. Por una parte

$$\mathbf{P}^2 = \mathbf{P}\mathbf{P} = \mathbf{A}(\mathbf{A}^\top \mathbf{A})^{-1} \underbrace{\mathbf{A}^\top \mathbf{A}(\mathbf{A}^\top \mathbf{A})^{-1}}_{\mathbf{I}} \mathbf{A}^\top = \mathbf{A}(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top = \mathbf{P}.$$

EJERCICIO 70. Por otra parte, demuestre que $(\mathbf{I} - \mathbf{P})^2 = \mathbf{I} - \mathbf{P}$.

De hecho, en la Sección 12.3 se demuestra que si \mathbf{B} es *simétrica* e *idempotente*, entonces

$$\begin{aligned} f_{\mathbf{B}}: \mathbb{R}^m &\longrightarrow \mathcal{C}(\mathbf{B}) \\ \mathbf{y} &\longmapsto \mathbf{B}\mathbf{y} \end{aligned}$$

es la función proyección ortogonal sobre el subespacio $\mathcal{C}(\mathbf{B})$. Por ello, las matrices simétricas e idempotentes se denominan *matrices proyección*. Así pues, \mathbf{P} y $(\mathbf{I} - \mathbf{P})$ son matrices proyección correspondientes a las respectivas proyecciones ortogonales sobre $\mathcal{C}(\mathbf{A})$ y $\mathcal{N}(\mathbf{A}^\top)$ (véase el siguiente ejercicio).

EJERCICIO 71. Sea $\mathbf{P} = \mathbf{A}(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top$ y sea $\mathbf{M} = \mathbf{I} - \mathbf{P}$.

(a) Demuestre que $\mathcal{C}(\mathbf{P}) = \mathcal{C}(\mathbf{A})$.

(b) Demuestre que $\mathcal{C}(\mathbf{M}) = \mathcal{N}(\mathbf{A}^\top)$.

Resumiendo, si las columnas de \mathbf{A} (de orden m por n) son una base de $\mathcal{C}(\mathbf{A})$ y si $\mathbf{P} = \mathbf{A}(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top$, entonces la función *proyección ortogonal* sobre $\mathcal{C}(\mathbf{A})$ es:

$$\begin{aligned} f_{\mathbf{P}} : \mathbb{R}^m &\longrightarrow \mathcal{C}(\mathbf{A}). \\ \mathbf{y} &\longmapsto \mathbf{P}\mathbf{y} \end{aligned}$$

Fíjese que como la proyección $f_{\mathbf{P}}(\mathbf{y})$ de \mathbf{y} sobre $\mathcal{C}(\mathbf{A})$ pertenece a $\mathcal{C}(\mathbf{A})$, el vector de $\mathcal{C}(\mathbf{A})$ más próximo a $f_{\mathbf{P}}(\mathbf{y})$ resulta ser él mismo; es decir, $f_{\mathbf{P}}(f_{\mathbf{P}}(\mathbf{y})) = \mathbf{P}\mathbf{P}\mathbf{y} = \mathbf{P}\mathbf{y} = f_{\mathbf{P}}(\mathbf{y})$.

12.3. Generalizaciones o visiones alternativas (★)



Hemos definido las proyecciones ortogonales sobre subespacios de \mathbb{R}^m ; también hemos visto que si las columnas de \mathbf{A} son linealmente independientes, podemos definir la matriz *simétrica* e *idempotente* $\mathbf{P} = \mathbf{A}(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top$. Entonces el producto $\mathbf{P}\mathbf{y}$ es la proyección ortogonal de \mathbf{y} sobre $\mathcal{C}(\mathbf{A})$. De manera similar, podemos definir la matriz *simétrica* e *idempotente* $(\mathbf{I} - \mathbf{P})$. En este caso $(\mathbf{I} - \mathbf{P})\mathbf{y}$ es la proyección ortogonal de \mathbf{y} sobre el complemento ortogonal de $\mathcal{C}(\mathbf{A})$.

Aquí vamos a generalizar todo esto a subespacios abstractos. Para lograrlo, primero debemos definir propiedades “equivalentes” a la idempotencia y simetría de matrices, pero para funciones.

Funciones idempotentes

Definición 12.2. Una función $f: V \rightarrow V$ se dice que es *idempotente* si verifica:

$$f \circ f = f,$$

es decir, para todo $x \in V$, $f(f(x)) = f(x)$.

Dicho de otro modo, f es idempotente si

$$f(x) = x \quad \text{para todo } x \in \text{imagen}(f).$$

Fíjese que si $f_{\mathbf{A}}: \mathbb{R}^m \longrightarrow \mathbb{R}^m$ es *idempotente* implica que \mathbf{A} es idempotente, pues

$$\mathbf{x} \longmapsto \mathbf{A}\mathbf{x}$$

$$f(f(\mathbf{x})) = f(\mathbf{x}) \quad \Leftrightarrow \quad \mathbf{A}\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{x}.$$

Funciones autoadjuntas

La propiedad equivalente a la simetría tiene un nombre diferente en el caso de las funciones.

Definición 12.3. Una función $f: \mathcal{V} \rightarrow \mathcal{V}$, donde en \mathcal{V} hay definido un producto escalar $\langle _, _ \rangle$, se dice que es *autoadjunta* si verifica que

$$\text{para todo } \vec{x}, \vec{y} \in \mathcal{V} \quad \langle \vec{x}, f(\vec{y}) \rangle = \langle f(\vec{x}), \vec{y} \rangle.$$

En particular, para \mathbb{R}^n con el producto punto, si $f_{\mathbf{A}}: \mathbb{R}^m \longrightarrow \mathbb{R}^m$ es *autoadjunta*, entonces:

$$\mathbf{x} \longmapsto \mathbf{A}\mathbf{x}$$

$$\mathbf{x}\mathbf{A}\mathbf{y} = \langle \mathbf{x}, \mathbf{A}\mathbf{y} \rangle = \langle \mathbf{A}\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{A}\mathbf{x} \rangle = \mathbf{y}\mathbf{A}\mathbf{x};$$


de donde se deduce que \mathbf{A} es necesariamente simétrica, pues si $\mathbf{x} = {}_i\mathbf{l} = \mathbf{l}_{|i}$ e $\mathbf{y} = \mathbf{l}_{|j} = {}_j\mathbf{l}$,

$${}_i\mathbf{A}_{|j} = {}_i\mathbf{A}\mathbf{l}_{|j} = {}_j\mathbf{A}\mathbf{l}_{|i} = {}_j\mathbf{A}_{|i}.$$

El recíproco también es cierto:

EJERCICIO 72. Demuestre que si \mathbf{A} es simétrica, entonces $f_{\mathbf{A}}: \mathbb{R}^m \rightarrow \mathbb{R}^m$ es autoadjunta.
 $\mathbf{x} \mapsto \mathbf{A}\mathbf{x}$

12.3.1. Proyección ortogonal

 Vayamos ahora con las funciones proyección ortogonal en espacios pre-Hilbert abstractos.

Definición 12.4. Sea $(\mathcal{V}, \langle _, _ \rangle)$ un espacio pre-Hilbert. Llamamos *proyección ortogonal* a toda función $f: \mathcal{V} \rightarrow \mathcal{V}$ que es idempotente y autoadjunta

$$f: \mathcal{V} \rightarrow \mathcal{V} \quad y \quad \begin{cases} f \circ f = f \\ \langle \vec{x}, f(\vec{y}) \rangle = \langle f(\vec{x}), \vec{y} \rangle \end{cases} \quad \text{para todo } \vec{x}, \vec{y} \in \mathcal{V}.$$

(contrariamente a lo que ocurre con la Definición 12.1, aquí no se requiere especificar el subespacio sobre el que se proyecta; dicho espacio es la imagen de la función f . Además, f es lineal como se prueba en el siguiente ejercicio).

EJERCICIO 73. (Opcional) Sea un espacio pre-Hilbert $(\mathcal{V}, \langle _, _ \rangle)$ y sea una función $f: \mathcal{V} \rightarrow \mathcal{V}$ tal que $f \circ f = f$ (idempotente) y $\forall \vec{u}, \vec{v} \in \mathcal{V}, \langle f(\vec{u}), \vec{v} \rangle = \langle \vec{u}, f(\vec{v}) \rangle$ (autoadjunta).

- (a) Demuestre que $\langle f(\vec{u}), f(\vec{w}) \rangle = \langle f(\vec{u}), \vec{w} \rangle$.
- (b) Demuestre que f es lineal.

Proposición 12.3.1. $f: \mathcal{V} \rightarrow \mathcal{V}$ es una proyección ortogonal (i.e., idempotente y autoadjunta) si y solo si para todo $\vec{h} \in \mathcal{V}$ y todo $\vec{y} \in \text{imagen}(f)$ se verifica que $\langle \vec{h} - f(\vec{h}), \vec{y} \rangle = 0$.

Demostración. Empecemos por demostrar que:

$$f \text{ es proyección ortogonal} \Rightarrow \langle \vec{h} - f(\vec{h}), \vec{y} \rangle = 0 \text{ para todo } \vec{h} \in \mathcal{V} \text{ y todo } \vec{y} \in \text{imagen}(f)$$

$$\begin{aligned} \langle \vec{h} - f(\vec{h}), \vec{y} \rangle &= \langle \vec{h}, \vec{y} \rangle - \langle f(\vec{h}), \vec{y} \rangle \\ &= \langle \vec{h}, \vec{y} \rangle - \langle \vec{h}, f(\vec{y}) \rangle && \text{por ser } f \text{ autoadjunta} \\ &= \langle \vec{h}, \vec{y} - f(\vec{y}) \rangle \end{aligned}$$

y como $\vec{y} \in \text{imagen}(f)$ sabemos que $\vec{y} = f(\vec{z})$ para algún $\vec{z} \in \mathcal{V}$

$$\begin{aligned} &= \langle \vec{h}, f(\vec{z}) - f(f(\vec{z})) \rangle && \text{sustituyendo } \vec{y} \text{ por } f(\vec{z}) \\ &= \langle \vec{h}, f(\vec{z}) - f(\vec{z}) \rangle && \text{por ser } f \text{ idempotente} \\ &= \langle \vec{h}, \vec{0} \rangle = 0. \end{aligned}$$

Ahora demostremos que

$$\langle \vec{h} - f(\vec{h}), \vec{y} \rangle = 0 \text{ para todo } \vec{h} \in \mathcal{V} \text{ y todo } \vec{y} \in \text{imagen}(f) \Rightarrow f \text{ es proyección ortogonal}$$

Demostremos que si $\langle \vec{h} - f(\vec{h}), \vec{y} \rangle = 0$ para todo $\vec{h} \in \mathcal{V}$ y todo $\vec{y} \in \text{imagen}(f)$ entonces f es idempotente y autoadjunta.

- Vamos a demostrar que $f(\vec{p}) = \vec{p}$ para todo $p \in \text{imagen}(f)$, y que por tanto f es idempotente.

Reemplazando en la hipótesis la variable \vec{h} por un vector $\vec{p} \in \text{imagen}(f)$, tenemos $\langle \vec{p} - f(\vec{p}), \vec{y} \rangle = 0$ para todo $\vec{y} \in \text{imagen}(f)$, en particular también para $\vec{y} = \vec{p} - f(\vec{p})$:

$$\langle \vec{p} - f(\vec{p}), \vec{p} - f(\vec{p}) \rangle = 0,$$

ya que $(\vec{p} - f(\vec{p})) \in \text{imagen}(f)$, pues $\text{imagen}(f)$ es subespacio por ser f lineal, y $\vec{p}, f(\vec{p}) \in \text{imagen}(f)$.

Por tanto $\vec{p} - f(\vec{p}) = \vec{0}$, es decir, $f(\vec{p}) = \vec{p}$; consecuentemente f es idempotente.

- Para demostrar que es autoadjunta, por un lado tenemos que

$$\begin{aligned} \langle \vec{v} - f(\vec{v}), \vec{w} - f(\vec{w}) \rangle &= \langle \vec{v}, \vec{w} - f(\vec{w}) \rangle - \underbrace{\langle f(\vec{v}), \vec{w} - f(\vec{w}) \rangle}_{=0} \quad \text{pues } f(\vec{v}) \in \text{imagen}(f) \\ &= \langle \vec{v}, \vec{w} \rangle - \langle \vec{v}, f(\vec{w}) \rangle \end{aligned}$$

por otro lado

$$\begin{aligned} \langle \vec{v} - f(\vec{v}), \vec{w} - f(\vec{w}) \rangle &= \langle \vec{v} - f(\vec{v}), \vec{w} \rangle - \underbrace{\langle \vec{v} - f(\vec{v}), f(\vec{w}) \rangle}_{=0} \quad \text{pues } f(\vec{w}) \in \text{imagen}(f) \\ &= \langle \vec{v}, \vec{w} \rangle - \langle f(\vec{v}), \vec{w} \rangle. \end{aligned}$$

Así que necesariamente $\langle \vec{v}, f(\vec{w}) \rangle = \langle f(\vec{v}), \vec{w} \rangle$; es decir, f es autoadjunta.

□

🔗 Conocemos una condición suficiente para asegurar que existe la proyección ortogonal sobre \mathcal{H} , Veámoslo...

Por la Proposición 10.1.4 en la página 123, sabemos que si el subespacio $\mathcal{H} \subset \mathcal{V}$ y su conjunto ortogonal \mathcal{H}^\perp son suplementarios, entonces para todo $\vec{v} \in \mathcal{V}$ existe una descomposición única $\vec{v} = \vec{h} + \vec{c}$ con $\vec{h} \in \mathcal{H}$ y $\vec{c} \in \mathcal{H}^\perp$. Es decir, si $\mathcal{V} = \mathcal{H} \oplus \mathcal{H}^\perp$ entonces existe proyección ortogonal tanto sobre \mathcal{H} como sobre \mathcal{H}^\perp .

Así pues, por el Teorema 11.3.4 sabemos que si $\mathcal{H} \subset (\mathcal{V}, \langle _, _ \rangle)$ es un subespacio de dimensión finita entonces $\mathcal{V} = \mathcal{H} \oplus \mathcal{H}^\perp$ y por tanto existe proyección ortogonal sobre \mathcal{H} y sobre \mathcal{H}^\perp .

🔗 Veamos ahora que, en el caso de que exista la función proyección ortogonal de $\vec{y} \in \mathcal{V}$ sobre $\mathcal{H} \subset \mathcal{V}$, la proyección del vector \vec{y} es el vector de \mathcal{H} “más próximo” a \vec{y} .

Después, y para finalizar la lección, veremos un ejemplo de subespacio de dimensión infinita en el que no existe la función proyección ortogonal, pues no existe el vector “más próximo” a uno dado.

EJERCICIO 74. Demuestre la siguiente proposición:

Proposición 12.3.2. Sea $(\mathcal{V}, \langle _, _ \rangle)$ un espacio pre-Hilbert. En el caso de que exista la proyección ortogonal de $\vec{y} \in \mathcal{V}$ sobre $\mathcal{H} \subset \mathcal{V}$, dicha proyección de \vec{y} es el vector de \mathcal{H} más próximo a \vec{y} .

Sobre algunos subespacios de dimensión infinita no existe la proyección ortogonal

Hay subespacios \mathcal{H} de dimensión infinita en espacios pre-Hilbert \mathcal{V} donde no hay función proyección ortogonal... es decir, donde no existe “el vector más próximo” en \mathcal{H} a un vector $\vec{y} \in \mathcal{V}$ dado.

Un ejemplo son las sucesiones casi nulas $\mathbb{R}[x]$ de la Página 149 y el subespacio $\mathcal{H} \subset \mathbb{R}[x]$ de sucesiones cuyas componentes suman cero:

$$\mathcal{H} = \left\{ p \in \mathbb{R}[x] \mid \sum_i p_i = 0 \right\}.$$

Es fácil ver que aunque podemos aproximarnos tanto como queramos al vector $\vec{a} = (1, 0, 0, 0, 0, 0, 0, \dots)$ con vectores del subespacio \mathcal{H} , no existe un vector en \mathcal{H} de que sea “el más próximo” a \vec{a} .

Para $\vec{b} = (1, -1, 0, 0, 0, 0, 0, \dots)$ tenemos que

$$d(\vec{a}, \vec{b}) = \|\vec{a} - \vec{b}\|^2 = \|(0, -1, 0, 0, 0, 0, 0, \dots)\|^2 = (-1)^2 = 1.$$

Para $\vec{c} = (1, -\frac{1}{2}, -\frac{1}{2}, 0, 0, 0, 0, \dots)$ tenemos que

$$d(\vec{a}, \vec{c}) = \|\vec{a} - \vec{c}\|^2 = \|(0, -\frac{1}{2}, -\frac{1}{2}, 0, 0, 0, 0, \dots)\|^2 = \frac{1}{2^2} + \frac{1}{2^2} = \frac{1}{2}.$$

Para $\vec{d} = (1, -\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}, 0, 0, 0, \dots)$ tenemos que

$$d(\vec{a}, \vec{d}) = \|\vec{a} - \vec{d}\|^2 = \|(0, -\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}, 0, 0, 0, \dots)\|^2 = \frac{1}{3^2} + \frac{1}{3^2} + \frac{1}{3^2} = \frac{1}{3}.$$

Para $\vec{e} = (1, -\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4}, 0, 0, \dots)$ tenemos que

$$d(\vec{a}, \vec{e}) = \|\vec{a} - \vec{e}\|^2 = \|(0, -\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4}, 0, 0, \dots)\|^2 = \frac{1}{4^2} + \frac{1}{4^2} + \frac{1}{4^2} + \frac{1}{4^2} = \frac{1}{4}.$$

y en general para $\vec{z} = (1, \underbrace{-\frac{1}{n}, -\frac{1}{n}, \dots, -\frac{1}{n}}_{n \text{ componentes}}, 0, \dots)$ tenemos que

$$d(\vec{a}, \vec{z}) = \|\vec{a} - \vec{z}\|^2 = \|(0, \underbrace{-\frac{1}{n}, -\frac{1}{n}, \dots, -\frac{1}{n}}_{n \text{ componentes}}, 0, \dots)\|^2 = \frac{1}{n}.$$

Es decir, podemos aproximarnos al vector \vec{a} tanto como queramos. Sin embargo, la distancia nunca puede ser cero, pues $d(\vec{a}, \vec{z}) = 0 \Rightarrow \vec{a} = \vec{z}$, y esto es imposible, pues entonces \vec{a} pertenecería a \mathcal{H} , algo que no puede ocurrir puesto que sus componentes no suman cero.

Como no existe el vector \vec{z} más próximo a \vec{a} en \mathcal{H} , no existe la proyección ortogonal sobre \mathcal{H} .

Ejercicios

EJERCICIO 75. Sean $\mathbf{P} = \mathbf{A}(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top$ y $\mathbf{M} = \mathbf{I} - \mathbf{P}$, con \mathbf{A} de orden m por n y rango n .

- (a) Demuestre que para todo $\mathbf{y} \in \mathbb{R}^m$, la diferencia $\mathbf{y} - \mathbf{P}\mathbf{y}$ es perpendicular a $\mathcal{C}(\mathbf{A})$.
- (b) Demuestre que $f_{\mathbf{P}}$ actúa como la función identidad con todo $\mathbf{y} \in \mathcal{C}(\mathbf{A})$.
- (c) Demuestre que $f_{\mathbf{P}}$ actúa como la función nula con todo $\mathbf{y} \in \mathcal{N}(\mathbf{A}^\top)$.
- (d) Demuestre que $f_{\mathbf{M}}: \mathbb{R}^m \rightarrow \mathcal{N}(\mathbf{A}^\top)$ es la función *proyección ortogonal* sobre $\mathcal{N}(\mathbf{A}^\top)$.

$$\mathbf{y} \mapsto \mathbf{M}\mathbf{y}$$

Pista. Fíjese en que $f_{\mathbf{M}}(\mathbf{y}) = \mathbf{M}\mathbf{y} = (\mathbf{I} - \mathbf{P})\mathbf{y} = \mathbf{y} - \mathbf{P}\mathbf{y} = \mathbf{y} - f_{\mathbf{P}}(\mathbf{y})$ y también en que $\mathcal{N}(\mathbf{A}^\top)^\perp = \mathcal{C}(\mathbf{A})$.

- (e) Demuestre también que $f_{\mathbf{M}}$ actúa como la función identidad con todo $\mathbf{y} \in \mathcal{N}(\mathbf{A}^\top)$ y que actúa como la función nula con todo $\mathbf{y} \in \mathcal{C}(\mathbf{A})$.

Aquí hemos explotado la equivalencia de dos sistemas de ecuaciones y la definición de proyección ortogonal. En otros manuales se sigue un argumento distinto que básicamente consiste en demostrar las siguientes proposiciones (que propongo como ejercicio *opcional*, pues son resultados que ya las hemos visto más arriba).

EJERCICIO 76. (*Opcional*) Demuestre las siguientes proposiciones:

- (a) **Proposición 12.3.3.** $\mathcal{N}(\mathbf{A}) = \mathcal{N}(\mathbf{A}^\top \mathbf{A})$.
- (b) **Corolario 12.3.4.** El rango de $\mathbf{A}^\top \mathbf{A}$ es igual al rango de \mathbf{A} .
- (c) **Proposición 12.3.5.** El sistema de ecuaciones normales $(\mathbf{A}^\top \mathbf{A})\mathbf{x} = \mathbf{A}^\top \mathbf{y}$ tiene solución.

Pista. Compruebe que la matriz de coeficientes y la matriz ampliada tienen el mismo rango.

En econometría se emplea el siguiente resultado

EJERCICIO 77. Si una matriz \mathbf{Q} de orden n es idempotente entonces su rango es igual a su traza. (encontrará el significado de *traza* en la Definición 3.6 en la página 40).

Parte V

Determinantes

Propiedades de los determinantes

13.1. Función determinante y función volumen

13.1.1. Tres propiedades de la función volumen de un paralelogramo

La función volumen¹ (*base \times altura*) de un paralelepípedo de dimensión n , cuyas aristas son las columnas de una matriz cuadrada \mathbf{A} , tiene las siguientes tres propiedades:

- Dado que las columnas de la matriz identidad tienen norma uno ($\|\mathbf{I}_j\| = 1$), el volumen (o área) del hipercubo de dimensión n descrito por las columnas de \mathbf{I} es siempre uno:

$$\text{Volumen}(\mathbf{I}_{n \times n}) = 1.$$

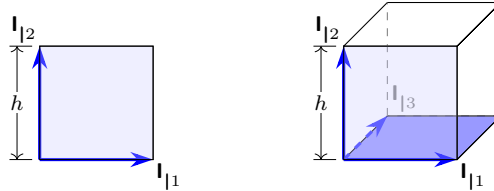


Figura 13.1: $\text{Area}[\mathbf{I}_1; \mathbf{I}_2] = 1$ (izquierda) y $\text{Volumen}[\mathbf{I}_1; \mathbf{I}_2; \mathbf{I}_3] = 1$ (derecha).

- Al sumar a una de las aristas (columnas) un múltiplo de otra arista (columna), el volumen (o el área) del paralelogramo resultante no cambia; pues se mantiene la misma base y la misma altura h . Por tanto, aplicar una transformación elemental de Tipo I no modifica el volumen:

$$\text{Vol}(\mathbf{A}) = \text{Vol}(\mathbf{A}_{[(\alpha)\tau_{k+j}]}) \text{ para } i \neq k.$$

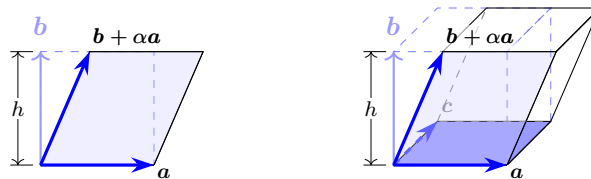


Figura 13.2: $\text{Area}[\mathbf{a}; \mathbf{b}] = \text{Area}[\mathbf{a}; (\alpha\mathbf{a} + \mathbf{b})]$ (izquierda) y $\text{Vol}[\mathbf{a}; \mathbf{b}; \mathbf{c}] = \text{Vol}[\mathbf{a}; (\alpha\mathbf{a} + \mathbf{b}); \mathbf{c}]$ (derecha).

¹en dimensión 2 se denomina área

- Al multiplicar una de las aristas (columnas) por un escalar, el volumen (o área) queda multiplicado por el valor absoluto de dicho escalar (no hay áreas o volúmenes negativos).

$$|\alpha| \cdot \text{Vol}(\mathbf{A}) = |\alpha| \cdot \text{Vol}[\dots \mathbf{A}_{|k}; \dots] = \text{Vol}[\dots \alpha \mathbf{A}_{|k}; \dots]$$

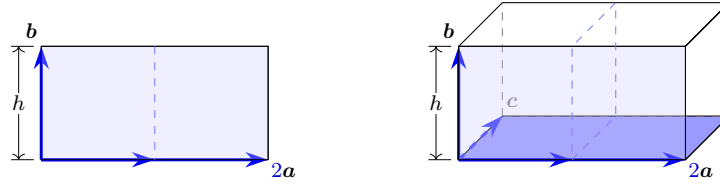


Figura 13.3: $|\alpha| \cdot \text{Area}[\mathbf{a}; \mathbf{b}] = \text{Area}[\alpha \mathbf{a}; \mathbf{b}]$ (izquierda) y $|\alpha| \cdot \text{Vol}[\mathbf{a}; \mathbf{b}; \mathbf{c}] = \text{Vol}[\alpha \cdot \mathbf{a}; \mathbf{b}; \mathbf{c}]$ (derecha).

Por tanto, aplicar la transformación, $\tau_{[(\alpha)\mathbf{j}]}$, donde $\alpha \neq 0$, multiplica el volumen por $|\alpha|$. Y multiplicar una arista por 0 multiplica el volumen por 0.

13.1.2. Las tres primeras propiedades (P-1 to P-3)

Por analogía con la función volumen, definimos el determinante del siguiente modo

Definición 13.1. Denominamos función determinante a toda función que asigna a cada sistema de n vectores de \mathbb{R}^n (a cada matriz cuadrada de orden n) un número real

$$\det : \mathbb{R}^{n \times n} \longrightarrow \mathbb{R}$$

y que verifica las siguientes tres propiedades:

P-1 P-1 **Determinante de las matrices identidad:**

$$\det \mathbf{I} = 1$$

P-2 P-2 **Sumar a una columna un múltiplo de otra no altera el determinante:**

$$\det \mathbf{A} = \det \left(\mathbf{A} \tau_{[(\alpha)\mathbf{k}+\mathbf{j}]} \right) \quad \text{para } i \neq k.$$

P-3 P-3 **Multiplicar una columna multiplica el determinante**

$$\alpha \cdot \det \mathbf{A} = \det [\dots \alpha \mathbf{A}_{|k}; \dots] \quad \text{para cualquier } k \in \{1 : n\} \text{ y } \alpha \in \mathbb{R}$$

Nótese que como caso particular, concluimos que aplicar una transformación elemental de Tipo II, $\tau_{[(\alpha)\mathbf{i}]}$, multiplica el determinante por $\alpha \neq 0$. Consecuentemente, los determinantes pueden tomar valores negativos.

La relación entre la función Volumen y la función Determinante

Nótese que la única diferencia con la función Volumen es que los determinantes pueden tomar valores negativos. De hecho, la relación entre la función Volumen y la función Determinante es tan estrecha que podemos definir la función Volumen del paralelogramo cuyas aristas son las columnas de la matriz cuadrada \mathbf{A} como el valor absoluto del Determinante de \mathbf{A} :

$$\text{Vol}(\mathbf{A}) = \text{Valor absoluto del determinante de } \mathbf{A}.$$

Notación Emplearemos dos notaciones alternativas para el determinante de la matriz \mathbf{A} :

$$\text{determinante de } \mathbf{A} \equiv \det(\mathbf{A}) \equiv |\mathbf{A}|$$

de manera que las tres propiedades anteriores se pueden expresar para matrices de orden 3 como:

$$\mathbf{P-1:} \quad \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1; \quad \det [\mathbf{I}_1; \mathbf{I}_2; \mathbf{I}_3] = 1.$$

$$\mathbf{P-2:} \quad \begin{vmatrix} a_1 & (b_1 + \alpha c_1) & c_1 \\ a_2 & (b_2 + \alpha c_2) & c_2 \\ a_3 & (b_3 + \alpha c_3) & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}; \quad \det [a; (b + \alpha c); c] = \det [a; b; c]$$

$$\mathbf{P-3:} \quad \begin{vmatrix} a_1 & (\beta b_1) & c_1 \\ a_2 & (\beta b_2) & c_2 \\ a_3 & (\beta b_3) & c_3 \end{vmatrix} = \beta \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}; \quad \det [a; \beta b; c] = \beta \det [a; b; c].$$

Advertencia: Una barra vertical a cada lado de una matriz $|\mathbf{A}|$ denota el determinante de \mathbf{A} . Una barra vertical a cada lado de un número $|a|$ significa valor absoluto del número. Es decir, el significado de las barras verticales viene dado por el objeto encerrado entre ellas: si es un número es el *valor absoluto*, y si es una matriz es el *determinante*. Jugando con esto, podemos decir que

$$\text{Vol}(\mathbf{A}) = \text{Valor absoluto de } \det(\mathbf{A}) = |\det \mathbf{A}| = ||\mathbf{A}||.$$

Las propiedades **P-1**, **P-2** y **P-3** definen la función determinante. Ahora queda deducir todo lo demás...

13.2. Seis propiedades del determinante

13.2.1. Determinante de una matriz con una columna de ceros

EJERCICIO 78. Demuestre la siguiente propiedad:

P-4 Determinante de una matriz con una columna nula.

P-4

Si \mathbf{A} tiene una columna nula entonces:

$$\det(\mathbf{A}) = 0$$

Esta propiedad es análoga a considerar el volumen de un paralelogramo con una arista de longitud cero.

13.2.2. Determinantes de matrices elementales

Ya sabemos que

$$\det \left(\mathbf{A}_{[(\alpha)\mathbf{k}+\mathbf{j}]}^{\boldsymbol{\tau}} \right) = |\mathbf{A}|; \quad \det \left(\mathbf{A}_{[(\alpha)\mathbf{k}]}^{\boldsymbol{\tau}} \right) = \alpha |\mathbf{A}|; \quad (13.1)$$

así que en particular, y puesto que $\det(\mathbf{I}) = 1$:

$$\det \left(\mathbf{I}_{[(\alpha)\mathbf{k}+\mathbf{j}]}^{\boldsymbol{\tau}} \right) = 1 \quad \text{y} \quad \det \left(\mathbf{I}_{[(\alpha)\mathbf{j}]}^{\boldsymbol{\tau}} \right) = \alpha.$$

Ahora recordando que $\mathbf{A}_{\boldsymbol{\tau}} = \mathbf{A}(\mathbf{I}_{\boldsymbol{\tau}})$, y analizando tanto la igualdad de la izquierda como la de la derecha de (13.1), concluimos que

$$|\mathbf{A}(\mathbf{I}_{\boldsymbol{\tau}})| = |\mathbf{A}| \cdot |\mathbf{I}_{\boldsymbol{\tau}}| \quad (13.2)$$

donde $\mathbf{I}_{\boldsymbol{\tau}}$ es una matriz elemental (nos da igual de qué tipo sea).

13.2.3. Sucesión de transformaciones elementales por columnas

Veamos más resultados relacionados con las matrices elementales.

EJERCICIO 79. Demuestre las siguientes proposiciones.

- (a) $\det(\mathbf{A}_{\tau_1 \dots \tau_k}) = |\mathbf{A}| \cdot |\mathbf{I}_{\tau_1}| \cdots |\mathbf{I}_{\tau_k}|$.
- (b) Si \mathbf{B} es de rango completo, es decir, si $\mathbf{B} = \mathbf{I}_{\tau_1 \dots \tau_k}$, entonces $|\mathbf{B}| = |\mathbf{I}_{\tau_1}| \cdots |\mathbf{I}_{\tau_k}|$, y por tanto $|\mathbf{B}| \neq 0$.
- (c) Si \mathbf{A} y \mathbf{B} son de orden n , y \mathbf{B} es de rango completo entonces

$$\det(\mathbf{AB}) = |\mathbf{A}| \cdot |\mathbf{B}| \quad (13.3)$$

Ejemplo 34. Una sucesión $\tau_1 \cdots \tau_k$ de transformaciones *Tipo I* de la matriz \mathbf{A} no altera el determinante.

$$|\mathbf{A}_{\tau_1 \dots \tau_k}| = |\mathbf{AI}_{\tau_1 \dots \tau_k}| = |\mathbf{A}| \cdot |\mathbf{I}_{\tau_1 \dots \tau_k}| = |\mathbf{A}| \cdot 1 = |\mathbf{A}|$$

Ejemplo 35. Pero una sucesión de transformaciones *Tipo II* si puede modificar el valor del determinante:

$$\begin{vmatrix} 2a & 3c \\ 2b & 3d \end{vmatrix} = 6 \begin{vmatrix} a & c \\ b & d \end{vmatrix}$$

Intercambio de columnas. Propiedad antisimétrica

EJERCICIO 80. Demuestre la siguiente propiedad:

P-5 **P-5** **[Propiedad antisimétrica]** Intercambiar dos columnas cambia el signo del determinante.

Por ejemplo,

$$\begin{vmatrix} c_1 & b_1 & a_1 \\ c_2 & b_2 & a_2 \\ c_3 & b_3 & a_3 \end{vmatrix} = (-1) \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}.$$

13.2.4. Matriz singular, matriz inversa y producto de matrices

Determinante de matrices singulares y determinante de matrices inversas

Vamos a demostrar conjuntamente dos propiedades más de la función determinante:

P-6 **P-6** Si \mathbf{A} es singular entonces $|\mathbf{A}| = 0$.

P-7 **P-7** $\det(\mathbf{A}^{-1}) = (\det \mathbf{A})^{-1}$.

Demostración. Sea \mathbf{A} y sea $\mathbf{I}_{\tau_1 \dots \tau_k}$ tal que $\mathbf{A}(\mathbf{I}_{\tau_1 \dots \tau_k}) = \mathbf{R}$ es una forma escalonada reducida. Entonces $|\mathbf{A}| \cdot |\mathbf{I}_{\tau_1 \dots \tau_k}| = |\mathbf{R}|$ y solo se pueden dar dos casos:

$$\begin{cases} \mathbf{A} \text{ singular } (\mathbf{R}_{|n} = \mathbf{0}) : & |\mathbf{R}| = |\mathbf{A}| \cdot |\mathbf{I}_{\tau_1 \dots \tau_k}| = 0 \quad \Rightarrow \quad |\mathbf{A}| = 0 \\ \mathbf{A} \text{ invertible } (\mathbf{R} = \mathbf{I}) : & |\mathbf{R}| = |\mathbf{A}| \cdot |\mathbf{I}_{\tau_1 \dots \tau_k}| = 1 \quad \Rightarrow \quad |\mathbf{I}_{\tau_1 \dots \tau_k}| = |\mathbf{A}^{-1}| = \frac{1}{|\mathbf{A}|} \end{cases}.$$

□

Así pues, podemos calcular el determinante de la matriz inversa mediante la eliminación gaussiana.

- Cuando la matriz es singular (rango $< n$) el determinante es cero;
- Cuando la matriz es de rango completo, basta con mirar qué transformaciones elementales *Tipo II* han ido cambiando el valor del determinante de \mathbf{A} hasta llegar a la matriz \mathbf{I} (las de *Tipo I* no importan!...)

Ejemplo 36. Sea $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 2 \end{bmatrix}$, entonces

$$\begin{bmatrix} 1 & 2 \\ 2 & 2 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \xrightarrow[\text{(Tipo I)}]{[(-2)\mathbf{1}+2]} \begin{bmatrix} 1 & 0 \\ 2 & -2 \\ 1 & -2 \\ 0 & 1 \end{bmatrix} \xrightarrow[\text{(Tipo II)}]{[(-1/2)\mathbf{2}]} \begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 1 & 1 \\ 0 & -1/2 \end{bmatrix} \xrightarrow[\text{(Tipo I)}]{[(-2)\mathbf{2}+1]} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 1 \\ 1 & -1/2 \end{bmatrix}$$

Por tanto, $|\mathbf{A}^{-1}| = |\mathbf{I}_{[(-2)\mathbf{1}+2]}| \cdot |\mathbf{I}_{[(-1/2)\mathbf{2}]}| \cdot |\mathbf{I}_{[(-2)\mathbf{2}+1]}| = 1 \cdot \frac{-1}{2} \cdot 1 = \frac{-1}{2}$; es decir $|\mathbf{A}| = -2$.

Determinante del producto.

Puesto que para cualquier \mathbf{B} de orden n , existe una sucesión de transformaciones elementales tal que $\mathbf{B}(\mathbf{I}_{\tau_1 \dots \tau_k}) = \mathbf{L}$ es una forma escalonada de \mathbf{B} , si \mathbf{B} es singular entonces $\mathbf{L}_{|n} = \mathbf{0}$. Aplicando las mismas transformaciones elementales a las columnas de \mathbf{AB} tenemos

$$(\mathbf{AB}(\mathbf{I}_{\tau_1 \dots \tau_k}))_{|n} = (\mathbf{AL})_{|n} = \mathbf{0};$$

y como $(\mathbf{I}_{\tau_1 \dots \tau_k})_{|n} \neq \mathbf{0}$, necesariamente (\mathbf{AB}) es singular (véase Corolario 8.1.2 en la página 104).

Este resultado nos permite enunciar una nueva propiedad de la función determinante.

P-8 **[Determinante del producto de matrices]** P-8

$$\det(\mathbf{AB}) = \det(\mathbf{A}) \cdot \det(\mathbf{B}). \quad (13.4)$$

Demostración. $\begin{cases} \text{Si } \mathbf{B} \text{ es singular, también lo es } \mathbf{AB} \Rightarrow \det(\mathbf{AB}) = 0 = \det(\mathbf{A}) \cdot \det(\mathbf{B}) \\ \text{Si } \mathbf{B} = \mathbf{I}_{\tau_1 \dots \tau_k} \Rightarrow \det(\mathbf{AB}) = \det(\mathbf{A}) \cdot \det(\mathbf{B}) \end{cases}$ (Ecuación (13.3)) □

13.2.5. Determinante de la matriz transpuesta.

EJERCICIO 81.

- (a) ¿Qué relación hay entre el determinante de \mathbf{I}_{τ} y el determinante de su transpuesta ${}_{\tau}\mathbf{I}$?
 (b) Sea \mathbf{B} de rango completo, demuestre que $|\mathbf{B}| = |\mathbf{B}^{\top}|$.

Así llegamos a la última propiedad que veremos en esta lección:

P-9 **Determinante de la transpuesta** P-9

$$|\mathbf{A}| = |\mathbf{A}^{\top}|.$$

Demostración. Por una parte, si \mathbf{A} es singular, \mathbf{A}^{\top} también es singular (véase Proposición 9.5.13 en la página 117); y por tanto ambas matrices tienen determinante nulo. Y por otra, si \mathbf{A} es de rango completo, acabamos de ver en el ejercicio anterior que su determinante es igual al de su transpuesta. □

Fórmulas para el cálculo del determinante y aplicaciones

Definición 14.1. Llamaremos **matriz extendida** de \mathbf{A} a la matriz $\begin{bmatrix} \mathbf{A} \\ 1 \end{bmatrix}$ de orden $n + 1$.

Fíjese que si \mathbf{I}_τ es de orden n , entonces τ transforma una de las n primeras columnas, pues tanto si τ es de *Tipo I*: $\begin{bmatrix} \tau \\ [(\alpha)p+s] \end{bmatrix}$, como si es de *Tipo II*: $\begin{bmatrix} \tau \\ [(\alpha)p] \end{bmatrix}$, los índices p, s son necesariamente menores o iguales a n .

Por tanto, para la transformación τ y las matrices extendidas de \mathbf{B}_τ y \mathbf{B} , se verifica que

$$\begin{bmatrix} \mathbf{B}_\tau \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{B} \\ 1 \end{bmatrix}_\tau. \quad (14.1)$$

Y como $\begin{bmatrix} \mathbf{I} \\ 1 \end{bmatrix}_\tau$ y \mathbf{I}_τ son matrices elementales del mismo tipo, **ambas tienen idéntico determinante**. Así pues, si \mathbf{B} es invertible, repitiendo k veces el paso dado en (14.1) tenemos que

$$\left| \begin{bmatrix} \mathbf{I}_{\tau_1 \dots \tau_k} \\ 1 \end{bmatrix} \right| = \det \left(\begin{bmatrix} \mathbf{I} \\ 1 \end{bmatrix}_{\tau_1 \dots \tau_k} \right) = \det \left(\begin{bmatrix} \mathbf{I} \\ 1 \end{bmatrix}_{\tau_1} \cdots \begin{bmatrix} \mathbf{I} \\ 1 \end{bmatrix}_{\tau_k} \right) = |\mathbf{I}_{\tau_1}| \cdots |\mathbf{I}_{\tau_k}| = |\mathbf{I}_{\tau_1 \dots \tau_k}|. \quad (14.2)$$

Vamos a emplear este resultado para relacionar el determinante de dos matrices de distinto orden:

Proposición 14.0.1. Si \mathbf{A} es la matriz extendida de \mathbf{B} , entonces $\det(\mathbf{A}) = \det(\mathbf{B})$.

Demostración. Solo caben dos posibilidades. Si \mathbf{B} es de rango completo, es decir, si $\mathbf{B} = \mathbf{I}_{\tau_1 \dots \tau_k}$, por (14.2) ya sabemos que

$$|\mathbf{A}| = \left| \begin{bmatrix} \mathbf{B} \\ 1 \end{bmatrix} \right| = \left| \begin{bmatrix} \mathbf{I}_{\tau_1 \dots \tau_k} \\ 1 \end{bmatrix} \right| = |\mathbf{I}_{\tau_1 \dots \tau_k}| = |\mathbf{B}|.$$

Si \mathbf{B} es singular, entonces las columnas de \mathbf{B} son linealmente dependientes; así que también lo son las de \mathbf{A} :

$$|\mathbf{A}| = 0 = |\mathbf{B}|.$$

□

14.1. Determinante de matrices triangulares

EJERCICIO 82.

- (a) ¿A qué es igual el determinante de una matriz triangular inferior \mathbf{L} de rango completo?
- (b) ¿A qué es igual el determinante de una matriz triangular con algún elemento nulo en su diagonal?
- (c) ¿A qué es igual el determinante de una matriz triangular superior \mathbf{U} ?

14.2. Cálculo del determinante por eliminación Gaussiana

Ahora sabemos que el determinante de una matriz cuadrada y triangular es igual al producto de los elementos de la diagonal principal. Este descubrimiento nos permite calcular de manera sencilla el determinante mediante eliminación: basta con encontrar una forma escalonada de la matriz extendida, compensando en su última fila las transformaciones elementales Tipo II aplicadas sobre las n primeras columnas.

Ejemplo 37. Considere la matriz $\begin{bmatrix} 1 & 5 \\ 2 & 3 \end{bmatrix}$ entonces

$$\left[\begin{array}{cc|c} 1 & 5 & 0 \\ 2 & 3 & 0 \\ 0 & 0 & 1 \end{array} \right] \xrightarrow{[(-5)1+2]} \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 2 & -7 & 0 \\ 0 & 0 & 1 \end{array} \right]$$

Multiplicando los elementos de la diagonal obtenemos el valor del determinante de \mathbf{A} . En este caso $\det \mathbf{A} = -7$.

Ejemplo 38. Para la matriz $\begin{bmatrix} 0 & 2 & 1 \\ 9 & 6 & 3 \\ 0 & 1 & 1 \end{bmatrix}$ tenemos

$$\left[\begin{array}{ccc|c} 0 & 2 & 1 & 0 \\ 9 & 6 & 3 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right] \xrightarrow{\begin{matrix} [(\frac{\tau}{2})3] \\ [(-1)2+3] \\ [(\frac{\tau}{2})4] \end{matrix}} \left[\begin{array}{ccc|c} 0 & 2 & 0 & 0 \\ 9 & 6 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & \frac{1}{2} \end{array} \right] \xrightarrow{\begin{matrix} [(\frac{\tau}{3})2] \\ [(-2)1+2] \\ [(\frac{\tau}{3})4] \end{matrix}} \left[\begin{array}{ccc|c} 0 & 6 & 0 & 0 \\ 9 & 0 & 0 & 0 \\ 0 & 3 & 1 & 0 \\ 0 & 0 & 0 & \frac{1}{6} \end{array} \right] \xrightarrow{\begin{matrix} [1 \rightleftharpoons 2] \\ [(-1)4] \end{matrix}} \left[\begin{array}{ccc|c} 6 & 0 & 0 & 0 \\ 0 & 9 & 0 & 0 \\ 3 & 0 & 1 & 0 \\ 0 & 0 & 0 & -\frac{1}{6} \end{array} \right]$$

Y por tanto $\det \mathbf{A} = (6)(9)(1)(-\frac{1}{6}) = -9$.

14.3. Determinante de matrices diagonales por bloques

Puesto que para cualquier matriz cuadrada \mathbf{B} se verifica que $\begin{vmatrix} \mathbf{B} & \\ & 1 \end{vmatrix} = |\mathbf{B}|$; si \mathbf{A} es de orden m y \mathbf{I} de orden n , entonces, aplicando la anterior igualdad repetidamente n veces, deducimos que

$$\begin{vmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{vmatrix}_{\substack{n \times m \\ m \times n}} = \begin{vmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{vmatrix}_{\substack{(n-1) \times m \\ m \times (n-1)}} = \begin{vmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{vmatrix}_{\substack{(n-2) \times m \\ m \times (n-2)}} = \dots = \begin{vmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & 1 \end{vmatrix} = |\mathbf{A}|.$$

Pero todo lo realizado en las anteriores secciones se podía haber desarrollado de manera similar si en lugar de extender la matriz \mathbf{B} con la última fila y columna de la matriz identidad $n+1$, se hubiera hecho anteponiendo la primera fila y columna de la matriz identidad $n+1$. Así, para cualquier matriz cuadrada \mathbf{B} también se verifica que

$$\begin{vmatrix} 1 & \\ & \mathbf{B} \end{vmatrix} = |\mathbf{B}|; \quad \text{por lo que también se deduce que} \quad \begin{vmatrix} \mathbf{I} & \\ & \mathbf{A} \end{vmatrix} = |\mathbf{A}|.$$

Producto de matrices por bloques. En la Sección 3.B se vió cómo calcular el producto de matrices por bloques siguiendo la fórmula de la Ecuación 3.6 en la página 39.

EJERCICIO 83. Demuestre las siguientes propiedades:

- (a) Sean \mathbf{A} de orden m y \mathbf{B} de orden n , entonces $\begin{vmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{vmatrix}_{\substack{n \times m \\ m \times n}} = |\mathbf{A}| \cdot |\mathbf{B}|$.
- (b) Sean \mathbf{A} de orden m , \mathbf{B} de orden n y \mathbf{C} de orden n por m , entonces

$$\begin{vmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{C} & \mathbf{B} \end{vmatrix} = |\mathbf{A}| \cdot |\mathbf{B}|.$$

14.4. Fórmulas sencillas para matrices de orden menor que 4

Para matrices de orden 1, 2 o 3, hay formulas sencillas. ¡Ojo, para matrices de orden 4 o más NO HAY REGLAS SENCILLAS! No extrapole que lo que funciona para una matriz 3 por 3 se puede hacer de manera similar para una matriz de orden mayor o igual a 4.

Para matrices de orden 1: Para $\mathbf{A} = \begin{bmatrix} a \end{bmatrix}$ tenemos

$$\left[\begin{array}{c|c} a & 0 \\ \hline 0 & 1 \end{array} \right] \Rightarrow \boxed{\det \mathbf{A} = a.}$$

Para matrices de orden 2. Deduzcamos la formula general para una matriz de la forma $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

■ Caso $a \neq 0$:

$$\left[\begin{array}{cc|c} a & b & 0 \\ c & d & 0 \\ \hline 0 & 0 & 1 \end{array} \right] \xrightarrow{[(\frac{-b}{a})\tau]1+2} \left[\begin{array}{cc|c} a & 0 & 0 \\ c & d - \frac{bc}{a} & 0 \\ \hline 0 & 0 & 1 \end{array} \right] \Rightarrow \det \mathbf{A} = ad - bc.$$

■ Caso $a = 0$ y $b \neq 0$:

$$\left[\begin{array}{cc|c} a & b & \\ c & d & \\ \hline & & 1 \end{array} \right] \xrightarrow[\tau]{[1 \Rightarrow 2]} \left[\begin{array}{cc|c} b & a & \\ d & c & \\ \hline & & -1 \end{array} \right] \Rightarrow \det \mathbf{A} = (bc - ad)(-1) = ad - bc.$$

Es decir, en todos los casos

$$\boxed{\det \mathbf{A} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.}$$

Para matrices de orden 3. Deduzcamos la formula general para una matriz de la forma $\mathbf{A} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$

■ Caso $a \neq 0$ y $(ae - bd) \neq 0$:

$$\left[\begin{array}{ccc|c} a & b & c & 0 \\ d & e & f & 0 \\ g & h & i & 0 \\ \hline 0 & 0 & 0 & 1 \end{array} \right] \xrightarrow{[(\frac{-b}{a})\tau]1+2, [(\frac{-c}{a})\tau]1+3} \left[\begin{array}{ccc|c} a & 0 & 0 & 0 \\ d & e - \frac{bd}{a} & f - \frac{cd}{a} & 0 \\ g & h - \frac{bg}{a} & i - \frac{cg}{a} & 0 \\ \hline 0 & 0 & 0 & 1 \end{array} \right] \xrightarrow{[(\frac{-af+cd}{ae-bd})\tau]2+3} \left[\begin{array}{ccc|c} a & 0 & 0 & 0 \\ d & e - \frac{bd}{a} & 0 & 0 \\ g & h - \frac{bg}{a} & \frac{aei-afh-bdi+bfh+cdh-ceg}{ae-bd} & 0 \\ \hline 0 & 0 & 0 & 1 \end{array} \right] \Rightarrow \det \mathbf{A} = (bc-ad)(-1) = aei-afh-bdi+bfh+cdh-ceg.$$

■ Caso $a = 0$, $b \neq 0$ y $(bd - ae) \neq 0$:

$$\left[\begin{array}{ccc|c} a & b & c & \\ d & e & f & \\ g & h & i & \\ \hline & & & 1 \end{array} \right] \xrightarrow[\tau]{[1 \Rightarrow 2]} \left[\begin{array}{ccc|c} b & a & c & \\ e & d & f & \\ h & g & i & \\ \hline & & & -1 \end{array} \right] \Rightarrow \det \mathbf{A} = aei - afh - bdi + bfg + cdh - ceg.$$

■ ... y de modo similar para el caso $a = 0$, $b = 0$, $c \neq 0$ y $(ce - bf) \neq 0$.

Reordenando los términos, en todos los casos resulta la misma fórmula que se conoce comúnmente como *Regla de Sarrus*.

$$\det \mathbf{A} = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei + bfg + dhc - ceg - bdi - fha$$

14.5. No hay fórmulas sencillas para matrices de orden mayor a 3

A modo de ejemplo, tratemos de calcular el determinante de una matriz genérica de orden 4 (asumiremos que las componentes de la matriz son tales que todos los denominadores son distintos de cero).

$$\begin{aligned} & \left[\begin{array}{cccc|c} a & b & c & d & 0 \\ e & f & h & i & 0 \\ k & l & m & n & 0 \\ o & p & q & r & 0 \\ \hline 0 & 0 & 0 & 0 & 1 \end{array} \right] \xrightarrow{\begin{matrix} \tau \\ [(-\frac{b}{a})1+2] \\ [(-\frac{c}{a})1+3] \\ [(-\frac{d}{a})1+4] \end{matrix}} \left[\begin{array}{cccc|c} a & 0 & 0 & 0 & 0 \\ e & f - \frac{be}{a} & h - \frac{ce}{a} & i - \frac{de}{a} & 0 \\ k & l - \frac{bk}{a} & m - \frac{ck}{a} & n - \frac{dk}{a} & 0 \\ o & p - \frac{bo}{a} & q - \frac{co}{a} & r - \frac{do}{a} & 0 \\ \hline 0 & 0 & 0 & 0 & 1 \end{array} \right] \xrightarrow{\begin{matrix} \tau \\ [(-\frac{ah+ce}{af-be})2+3] \\ [(-\frac{ai+de}{af-be})2+4] \end{matrix}} \\ & \left[\begin{array}{cccc|c} a & 0 & 0 & 0 & 0 \\ e & f - \frac{be}{a} & 0 & 0 & 0 \\ k & l - \frac{bk}{a} & \frac{afm-ahl-bem+bhk+cel-cfk}{af-be} & \frac{afn-ail-ben+bik+del-dfk}{af-be} & 0 \\ o & p - \frac{bo}{a} & \frac{afq-ahp-beq+bho+cep-cfo}{af-be} & \frac{afr-aip-ber+bio+dep-dfo}{af-be} & 0 \\ \hline 0 & 0 & 0 & 0 & 1 \end{array} \right] \xrightarrow{\begin{matrix} \tau \\ [(-\frac{afn+ail+ben-bik-del+dfk}{afm-ahl-bem+bhk+cel-cfk})3+4] \end{matrix}} \\ & \left[\begin{array}{cccc|c} a & 0 & 0 & 0 & 0 \\ e & f - \frac{be}{a} & 0 & 0 & 0 \\ k & l - \frac{bk}{a} & \frac{afm-ahl-bem+bhk+cel-cfk}{af-be} & 0 & 0 \\ o & p - \frac{bo}{a} & \frac{afq-ahp-beq+bho+cep-cfo}{af-be} & r + \frac{\left(-n + \frac{(i-\frac{de}{a})(l-\frac{bk}{a}) + dk}{f-\frac{be}{a}}\right)\left(-q + \frac{(h-\frac{ce}{a})(p-\frac{bo}{a}) + co}{f-\frac{be}{a}}\right) - \frac{(i-\frac{de}{a})(p-\frac{bo}{a})}{f-\frac{be}{a}} - \frac{do}{a}}{-m + \frac{(h-\frac{ce}{a})(l-\frac{bk}{a}) + ck}{f-\frac{be}{a}}} & 0 \\ \hline 0 & 0 & 0 & 0 & 1 \end{array} \right] \end{aligned}$$

Calculando el producto de la diagonal y simplificando llegamos a la expresión final: $\det \mathbf{A} = \begin{vmatrix} a & b & c & d \\ e & f & h & i \\ k & l & m & n \\ o & p & q & r \end{vmatrix} =$

$$afmr - afnq - ahlr + ahnp + ailq - aimp - bemr + benq + bhkr - bhno - bikq + bimo + celr - cenp - cfkr + cfno + cikap - cilo - delq + demp + dfkq - dfmo - dhkp + dhlo.$$

Está claro que la fórmula no es ni sencilla ni fácil de recordar... ¡No puedo imaginar lo espantosa que tiene que ser la expresión para una matriz de orden 5 o mayor! Es importante que sepa que:

No hay expresiones sencillas para calcular determinantes de orden mayor que 3

14.6. Expansión de Laplace (desarrollo por cofactores).

Hay otra forma de calcular el determinante de una matriz de orden n empleando los determinantes de submatrices de orden $n-1$. Es una descripción recursiva de la función determinante que no emplea la eliminación gaussiana y se llama *Expansión de Laplace*. Pero tenga en cuenta que, con cualquier procedimiento, el cálculo del determinante de una matriz genérica de orden elevado es computacionalmente intenso.

Para deducir la citada *Expansión de Laplace* necesitamos enunciar una última propiedad de la función determinante, y definir los *menores* y los *cofactores*.

14.6.1. Propiedad multilineal

P-10

Propiedad multilineal

P-10

$$\det [\dots (\beta \mathbf{a} + \psi \mathbf{b}); \dots] = \beta \det [\dots \mathbf{a}; \dots] + \psi \det [\dots \mathbf{b}; \dots]$$

Demostración. Puesto que se verifica la propiedad del producto **P-3**, nos basta con demostrar que

$$\det [\dots (\mathbf{a} + \mathbf{b}); \dots] = \det [\dots \mathbf{a}; \dots] + \det [\dots \mathbf{b}; \dots].$$

Haremos la demostración para la primera columna. La demostración para el resto de columnas es similar.

Hay dos casos: si $\mathbf{y}_1, \dots, \mathbf{y}_{n-1}$ son linealmente dependientes, entonces la demostración es inmediata pues

$$\det [(\mathbf{a} + \mathbf{b}); \mathbf{y}_1; \dots \mathbf{y}_{n-1}] = 0 = 0 + 0 = \det [\mathbf{a}; \mathbf{y}_1; \dots \mathbf{y}_{n-1}] + \det [\mathbf{b}; \mathbf{y}_1; \dots \mathbf{y}_{n-1}].$$

Si por el contrario existe \mathbf{x} tal que $[\mathbf{x}; \mathbf{y}_1; \dots \mathbf{y}_{n-1}]$ es una base de \mathbb{R}^n , entonces

$$\det [(\alpha \mathbf{x} + \psi_1 \mathbf{y}_1 + \dots + \psi_{n-1} \mathbf{y}_{n-1}); \mathbf{y}_1; \dots \mathbf{y}_{n-1}] \stackrel{*}{=} \det [\alpha \mathbf{x}; \mathbf{y}_1; \dots \mathbf{y}_{n-1}] \quad (14.3)$$

$$= \alpha \det [\mathbf{x}; \mathbf{y}_1; \dots \mathbf{y}_{n-1}] \quad (14.4)$$

(*) ya que con una secuencia de transformaciones elementales se puede “simplificar” la primera columna. Ahora, si

$$\begin{aligned} \mathbf{a} &= \alpha \mathbf{x} + \psi_1 \mathbf{y}_1 + \dots + \psi_{n-1} \mathbf{y}_{n-1} \\ \mathbf{b} &= \beta \mathbf{x} + \gamma_1 \mathbf{y}_1 + \dots + \gamma_{n-1} \mathbf{y}_{n-1} \end{aligned}$$

tendremos que

$$\mathbf{a} + \mathbf{b} = (\alpha + \beta) \mathbf{x} + (\psi_1 + \gamma_1) \mathbf{y}_1 + \dots + (\psi_{n-1} + \gamma_{n-1}) \mathbf{y}_{n-1}$$

y consecuentente

$$\begin{aligned} \det [\mathbf{a} + \mathbf{b}; \mathbf{y}_1; \dots; \mathbf{y}_{n-1}] &= (\alpha + \beta) \det [\mathbf{x}; \mathbf{y}_1; \dots; \mathbf{y}_{n-1}] && \text{por (14.4)} \\ &= \alpha \det [\mathbf{x}; \mathbf{y}_1; \dots; \mathbf{y}_{n-1}] + \beta \det [\mathbf{x}; \mathbf{y}_1; \dots; \mathbf{y}_{n-1}] && \text{pues } \det \mathbf{A} \text{ es un número} \\ &= \det [\alpha \mathbf{x}; \mathbf{y}_1; \dots; \mathbf{y}_{n-1}] + \det [\beta \mathbf{x}; \mathbf{y}_1; \dots; \mathbf{y}_{n-1}] && \text{por P-3} \\ &= \det [\mathbf{a}; \mathbf{y}_1; \dots; \mathbf{y}_{n-1}] + \det [\mathbf{b}; \mathbf{y}_1; \dots; \mathbf{y}_{n-1}] && \text{por (14.3)} \end{aligned}$$

□

Ejemplo 39. Para \mathbb{R}^2

$$\begin{vmatrix} a + \alpha & c \\ b + \beta & d \end{vmatrix} = \begin{vmatrix} a & c \\ b & d \end{vmatrix} + \begin{vmatrix} \alpha & c \\ \beta & d \end{vmatrix};$$

por tanto

$$\begin{vmatrix} a & c \\ b & d \end{vmatrix} = \begin{vmatrix} a & c \\ 0 & d \end{vmatrix} + \begin{vmatrix} 0 & c \\ b & d \end{vmatrix}.$$

14.6.2. Menores y cofactores



En esta sección veremos cómo calcular los determinantes mediante formulas más sencillas. Fórmulas que expresan determinantes de matrices de orden n como sumas de determinantes de matrices de orden $n - 1$.

Proposición 14.6.1. Si \mathbf{A} es de la forma $_{n \times n}$ $\begin{bmatrix} \mathbf{B} & \begin{smallmatrix} 0 \\ \vdots \\ 0 \end{smallmatrix} \\ a_{11} & \cdots & a_{1n-1} & 1 \end{bmatrix}$ entonces $\boxed{\det(\mathbf{A}) = \det(\mathbf{B})}.$

Demostración. Mediante operaciones elementales por columnas de *Tipo I*, podemos reducir \mathbf{A} a

$$\mathbf{A} = \left[\begin{array}{ccc|c} & & & 0 \\ & \mathbf{B} & & \vdots \\ & & & 0 \\ \hline a_{11} & \cdots & a_{1n-1} & 1 \end{array} \right] \xrightarrow{\text{transformaciones Tipo I}} \left[\begin{array}{ccc|c} & \mathbf{B} & & 0 \\ & & & \vdots \\ & & & 0 \\ \hline 0 & \cdots & 0 & 1 \end{array} \right] \equiv \begin{bmatrix} \mathbf{B} & \\ & 1 \end{bmatrix}$$

Por emplear tan solo transformaciones *Tipo I*, tenemos que $|\mathbf{A}| = \begin{vmatrix} \mathbf{B} & \\ & 1 \end{vmatrix} = |\mathbf{B}|$ (Proposición 14.0.1). \square

Nueva notación y definición de menores y cofactores

Sea $\mathbf{q} = (q_1, \dots, q_n)$ de \mathbb{R}^n , quitando la j -ésima componente, q_j , obtenemos un vector de $\mathbb{R}^{(n-1)}$ que denotamos con $\mathbf{q}^{\uparrow j}$:

$$\mathbf{q}^{\uparrow j} = (q_1, \dots, q_{j-1}, q_{j+1}, \dots, q_n) \in \mathbb{R}^{n-1}.$$

Análogamente, si \mathbf{A} es de orden m por n , denotamos a la submatriz que resulta de quitar la i -ésima fila con ${}^i\mathbf{A}$, y a la submatriz que resulta de quitar la j -ésima columna con $\mathbf{A}^{\uparrow j}$. Así, ${}^i\mathbf{A}^{\uparrow j}$, es la submatriz de orden $m-1$ por $n-1$ que resulta de quitar la fila i -ésima y la columna j -ésima.

Definición 14.2 (menores y cofactores). Sea ${}^i\mathbf{A}^{\uparrow j}$ la submatriz de \mathbf{A} resultante de eliminar la fila i y la columna j . Llamamos menor de ${}_{ij}\mathbf{A}$ al determinante

$$\det({}^i\mathbf{A}^{\uparrow j}).$$

Los menores con los signos alternados en función de si $(i+j)$ es par (en cuyo caso el signo no cambia) o impar (en cuyo caso se invierte el signo) se denominan cofactores. Así pues, el cofactor de ${}_{ij}\mathbf{A}$ es

$$\text{cof}_{ij}(\mathbf{A}) = (-1)^{i+j} \det({}^i\mathbf{A}^{\uparrow j}).$$

Ejemplo 40. Para $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$, tenemos

$${}^1\mathbf{A}^{\uparrow 2} = \begin{bmatrix} 4 & 6 \\ 7 & 9 \end{bmatrix}, \quad {}^3\mathbf{A}^{\uparrow 3} = \begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix},$$

y por tanto

$$\text{cof}_{12}(\mathbf{A}) = (-1)^{1+2} \det({}^1\mathbf{A}^{\uparrow 2}) = (-1) \begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix}.$$

y

$$\text{cof}_{33}(\mathbf{A}) = (-1)^{3+3} \det({}^3\mathbf{A}^{\uparrow 3}) = \begin{vmatrix} 1 & 2 \\ 4 & 5 \end{vmatrix}.$$

14.6.3. Desarrollo del determinante por cofactores (Expansión de Laplace).

Teorema 14.6.2 ([Expansión de Laplace]). Para \mathbf{A} de orden n , $\det(\mathbf{A})$ se puede expresar como suma de los productos de los elementos de cualquier columna (fila) de \mathbf{A} por sus correspondientes cofactores:

$$\det(\mathbf{A}) = \sum_{i=1}^n a_{ij} \text{cof}_{ij}(\mathbf{A}), \quad \text{expansión por la columna } j\text{-ésima};$$

o bien: $\det(\mathbf{A}) = \sum_{j=1}^n a_{ij} \text{cof}_{ij}(\mathbf{A})$, expansión por la fila i -ésima.

Demostración. Probaremos la expansión por la columna j -ésima. La expansión por filas es similar. Primero movemos la columna j -ésima a la última posición con una secuencia de intercambios:

$$\mathbf{A}_{[j=(j+1)] \ [(j+1) \rightleftharpoons (j+2)] \ [(j+2) \rightleftharpoons (j+3)] \ \cdots \ [(n-1) \rightleftharpoons n]} = \left[\mathbf{A}^{r_j} \mid \mathbf{A}_{|j} \right].$$

Puesto que hay $n - j$ permutaciones en la sucesión, tenemos que: $\det \mathbf{A} = (-1)^{n-j} \cdot \det \left[\mathbf{A}^{r_j} \mid \mathbf{A}_{|j} \right]$.

Escribiendo $\mathbf{A}_{|j}$ como $\begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{pmatrix} = \begin{pmatrix} a_{1j} \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ a_{2j} \\ \vdots \\ 0 \end{pmatrix} + \cdots + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ a_{nj} \end{pmatrix}$, y empleando la propiedad multilineal (**P-10**):

$$\det \mathbf{A} = (-1)^{n-j} \left(\det \left[\mathbf{A}^{r_j} \mid \begin{pmatrix} a_{1j} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \right] + \det \left[\mathbf{A}^{r_j} \mid \begin{pmatrix} 0 \\ a_{2j} \\ \vdots \\ 0 \end{pmatrix} \right] + \cdots + \det \left[\mathbf{A}^{r_j} \mid \begin{pmatrix} 0 \\ \vdots \\ 0 \\ a_{nj} \end{pmatrix} \right] \right).$$

Si “hundimos” la primera fila de la matriz del primer determinante hasta la última posición mediante una sucesión de $(n - 1)$ intercambios: $\begin{matrix} \tau & \tau & \cdots & \tau \\ [1 \rightleftharpoons 2] & [2 \rightleftharpoons 3] & & [(n-1) \rightleftharpoons n] \end{matrix}$; y la segunda fila de la matriz del segundo determinante mediante $(n - 2)$ intercambios: $\begin{matrix} \tau & \tau & \cdots & \tau \\ [2 \rightleftharpoons 3] & [3 \rightleftharpoons 4] & & [(n-1) \rightleftharpoons n] \end{matrix}$; —y en general, “hundimos” la fila i -ésima de la matriz i -ésima con $(n - i)$ intercambios: $\begin{matrix} \tau & \tau & \cdots & \tau \\ [i \rightleftharpoons (i+1)] & [(i+1) \rightleftharpoons (i+2)] & \cdots & [(n-1) \rightleftharpoons n] \end{matrix}$, obtenemos:

$$|\mathbf{A}| = (-1)^{n-j} \left((-1)^{n-1} \cdot \det \left[\begin{array}{c|c} {}^1\mathbf{A}^{r_j} & \mathbf{0} \\ \hline {}_1\mathbf{A}^{r_j} & a_{1j} \end{array} \right] + (-1)^{n-2} \cdot \det \left[\begin{array}{c|c} {}^2\mathbf{A}^{r_j} & \mathbf{0} \\ \hline {}_2\mathbf{A}^{r_j} & a_{2j} \end{array} \right] + \cdots + \det \left[\begin{array}{c|c} {}^n\mathbf{A}^{r_j} & \mathbf{0} \\ \hline {}_n\mathbf{A}^{r_j} & a_{nj} \end{array} \right] \right),$$

donde ${}_i\mathbf{A}^{r_j}$ es la fila i -ésima de la submatriz \mathbf{A}^{r_j} (y que hemos “hundido” hasta la última posición junto con el elemento a_{ij}). Así, por la Proposición 14.6.1

$$\det(\mathbf{A}) = (-1)^{n-j} \sum_{i=1}^n a_{ij} (-1)^{n-i} \det({}^i\mathbf{A}^{r_j}) = \sum_{i=1}^n a_{ij} \operatorname{cof}_{ij}(\mathbf{A});$$

puesto que $(-1)^{n-j} \cdot (-1)^{n-i} = (-1)^{2n-(i+j)} = (-1)^{-(i+j)} = (-1)^{i+j}$. □

EJERCICIO 84. Calcule el siguiente determinante: $\det \mathbf{A} = \begin{vmatrix} 2 & 0 & 3 & 2 \\ 5 & 1 & 2 & 4 \\ 3 & 0 & 1 & 2 \\ 5 & 3 & 2 & 1 \end{vmatrix}$.

EJERCICIO 85. Calcule $\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix}$ desarrollado por los cofactores de la segunda columna.

14.7. Aplicación de los determinantes

14.7.1. Regla de Cramer para la resolución de un sistema de ecuaciones

Suponga el sistema de ecuaciones $\mathbf{A}\mathbf{x} = \mathbf{b}$, donde \mathbf{A} es cuadrada y $|\mathbf{A}| \neq 0$. Sabemos que dicho sistema siempre tiene solución única, sea cual sea el vector \mathbf{b} . Dicha solución es

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

que verifica:

$$\mathbf{b} = (\mathbf{A}_{|1})x_1 + \cdots + (\mathbf{A}_{|j})x_j + \cdots + (\mathbf{A}_{|n})x_n, \quad (14.5)$$

donde los coeficiente x_i son los componentes¹ del vector solución \mathbf{x} .

Calculemos el determinante de una nueva matriz, idéntica a \mathbf{A} excepto por su j -ésima columna $\mathbf{A}_{|j}$, que ha sido sustituida por el vector \mathbf{b} :

$$\det[\mathbf{A}_{|1}; \mathbf{A}_{|2}; \dots \overbrace{\mathbf{b}}^{\text{pos. } j}; \dots \mathbf{A}_{|n};].$$

Si nos fijamos en la Ecuación (14.5) es fácil constatar que para pasar de \mathbf{A} a la nueva matriz, hemos multiplicado $\mathbf{A}_{|j}$ por x_j y le hemos sumado una combinación lineal del resto de columnas; por tanto:

$$\det[\mathbf{A}_{|1}; \dots \overbrace{\mathbf{b}}^{\text{pos. } j}; \dots \mathbf{A}_{|n};] = x_j \cdot \det(\mathbf{A}).$$

Despejando x_j encontramos la regla de Cramer para la resolución de sistemas de ecuaciones determinados: *cada componente x_j de la solución \mathbf{x} del sistema $\mathbf{A}\mathbf{x} = \mathbf{b}$ se puede calcular del siguiente modo*

$$x_j = \frac{\det[\mathbf{A}_{|1}; \dots \overbrace{\mathbf{b}}^{\text{pos. } j}; \dots \mathbf{A}_{|n};]}{\det(\mathbf{A})}.$$

14.7.2. Cálculo de la inversa de una matriz

Definición 14.3. Para \mathbf{A} , de orden n , la matriz $\text{Adj}(\mathbf{A})$ (la matriz adjunta de \mathbf{A}) es la transpuesta de la matriz resultante de sustituir cada elemento a_{ij} por su correspondiente cofactor: $\text{cof}_{ij}(\mathbf{A})$,

$$\text{Adj}(\mathbf{A}) = \begin{bmatrix} \text{cof}_{11}(\mathbf{A}) & \text{cof}_{12}(\mathbf{A}) & \cdots & \text{cof}_{1n}(\mathbf{A}) \\ \text{cof}_{21}(\mathbf{A}) & \text{cof}_{22}(\mathbf{A}) & \cdots & \text{cof}_{2n}(\mathbf{A}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cof}_{n1}(\mathbf{A}) & \text{cof}_{n2}(\mathbf{A}) & \cdots & \text{cof}_{nn}(\mathbf{A}) \end{bmatrix}^T.$$

Si \mathbf{A} es invertible ¿qué obtenemos si multiplicamos la matriz adjunta de \mathbf{A} por la matriz \mathbf{A} ?

$$\underbrace{\begin{bmatrix} \text{cof}_{11}(\mathbf{A}) & \text{cof}_{21}(\mathbf{A}) & \cdots & \text{cof}_{n1}(\mathbf{A}) \\ \text{cof}_{12}(\mathbf{A}) & \text{cof}_{22}(\mathbf{A}) & \cdots & \text{cof}_{n2}(\mathbf{A}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cof}_{1n}(\mathbf{A}) & \text{cof}_{2n}(\mathbf{A}) & \cdots & \text{cof}_{nn}(\mathbf{A}) \end{bmatrix}}_{\text{Adj}(\mathbf{A})} \underbrace{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}}_{\mathbf{A}}$$

El primer elemento de la diagonal de la matriz resultante es el producto punto entre la primera fila de $\text{Adj}(\mathbf{A})$ y la primera columna de \mathbf{A} ; y resulta ser el desarrollo de la Expansión de Laplace de $|\mathbf{A}|$ por la primera columna de \mathbf{A} . El segundo elemento de la diagonal es el desarrollo de $|\mathbf{A}|$ por la segunda columna, etc. Y en general el componente j -ésimo de la diagonal es la expansión por la columna j -ésima:

$$j|\text{Adj}(\mathbf{A})\mathbf{A}_{|j} = \text{cof}_{1j}(\mathbf{A})a_{1j} + \text{cof}_{2j}(\mathbf{A})a_{2j} + \text{cof}_{3j}(\mathbf{A})a_{3j} + \cdots + \text{cof}_{nj}(\mathbf{A})a_{nj} = \sum_{i=1}^n a_{ij} \text{cof}_{ij}(\mathbf{A}) = \det(\mathbf{A}).$$

Así pues, la diagonal de la matriz resultante está compuesta por el valor del determinante de \mathbf{A} .

¹las coordenadas de \mathbf{b} en la base formada por las columnas de \mathbf{A}

Los elementos fuera de la diagonal ${}_i(\mathbf{Adj}(\mathbf{A})\mathbf{A})_{|j}$ (con $i \neq j$) son determinantes de matrices singulares por la particularidad de que las columnas i y j son iguales, es decir, que $\mathbf{A}_{|i} = \mathbf{A}_{|j}$. Veámoslo.

Por ejemplo, ${}_1(\mathbf{Adj}(\mathbf{A})\mathbf{A})_{|2}$ es el desarrollo por la primera columna del determinante de una matriz cuya primera columna es igual que la segunda, y por lo que el desarrollo por la primera columnas es igual a cero:

$$\begin{vmatrix} a_{12} & a_{12} & \dots & a_{1n} \\ a_{22} & a_{22} & \dots & a_{2n} \\ a_{32} & a_{32} & \dots & a_{3n} \\ \vdots & \vdots & & \vdots \\ a_{n2} & a_{n2} & \dots & a_{nn} \end{vmatrix} = a_{12} \text{cof}_{11}(\mathbf{A}) + a_{22} \text{cof}_{21}(\mathbf{A}) + a_{32} \text{cof}_{31}(\mathbf{A}) + \dots + a_{n2} \text{cof}_{n1}(\mathbf{A}) = \sum_{i=1}^n a_{i2} \text{cof}_{i1}(\mathbf{A}) = 0.$$

Y el elemento k -ésimo ($k \neq 1$) de la primera fila, ${}_1(\mathbf{Adj}(\mathbf{A})\mathbf{A})_{|k}$, es el desarrollo por la primera columna del determinante de una matriz cuya primera y k -ésima columnas son iguales:

$$\begin{vmatrix} a_{1k} & a_{12} & \dots & a_{1n} \\ a_{2k} & a_{22} & \dots & a_{2n} \\ a_{3k} & a_{32} & \dots & a_{3n} \\ \vdots & \vdots & & \vdots \\ a_{nk} & a_{n2} & \dots & a_{nn} \end{vmatrix} = a_{1k} \text{cof}_{11}(\mathbf{A}) + a_{2k} \text{cof}_{21}(\mathbf{A}) + a_{3k} \text{cof}_{31}(\mathbf{A}) + \dots + a_{nk} \text{cof}_{n1}(\mathbf{A}) = \sum_{i=1}^n a_{ik} \text{cof}_{i1}(\mathbf{A}) = 0.$$

Es decir, ${}_i(\mathbf{Adj}(\mathbf{A})\mathbf{A})_{|j} = \begin{cases} |\mathbf{A}| & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$. Así, como $\mathbf{Adj}(\mathbf{A})\mathbf{A} = \begin{bmatrix} |\mathbf{A}| & 0 & \dots & 0 \\ 0 & |\mathbf{A}| & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & |\mathbf{A}| \end{bmatrix} = |\mathbf{A}| \cdot \mathbf{I}$,

concluimos que

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}(\mathbf{Adj}(\mathbf{A})).$$

Parte VI

Autovalores y autovectores. Diagonalización y formas cuadráticas

Autovalores y autovectores

Considere la siguiente ecuación donde \mathbf{A} es cuadrada y de orden n :

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \quad (15.1)$$

Definición 15.1 (Autovalor, autovector y espectro). Un **autovalor** de una matriz cuadrada \mathbf{A} es cualquier número λ tal que (15.1) tiene soluciones no nulas ($\mathbf{x} \neq \mathbf{0}$). En tal caso, los vectores \mathbf{x} se llaman **autovectores** de \mathbf{A} correspondientes al autovalor λ . El conjunto de autovalores de \mathbf{A} se denomina **espectro** de \mathbf{A} .

Por ejemplo, es fácil ver que

$$\mathbf{x}_1 = \begin{pmatrix} 4 \\ 1 \end{pmatrix} \text{ y } \mathbf{x}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \text{ son autovectores de } \mathbf{A} = \begin{bmatrix} 5 & 4 \\ 1 & 2 \end{bmatrix}$$

correspondientes a los autovalores 6 y 1 respectivamente (basta multiplicar $\mathbf{A}(\mathbf{x}_i)$ para comprobarlo). El espectro de \mathbf{A} es $\{6, 1\}$.

EJERCICIO 86. Demuestre la siguiente proposición:

Proposición 15.0.1. Una combinación lineal no nula de autovectores correspondientes al autovalor λ es otro autovector correspondiente al mismo autovalor λ .

El anterior resultado justifica la siguiente definición:

Definición 15.2 (Autoespacio). Sea \mathbf{A} de orden n . Denominamos **autoespacio** correspondiente al autovalor λ al subespacio formado por los autovectores correspondientes al autovalor λ de \mathbf{A} junto con el vector nulo:

$$\mathcal{E}_\lambda(\mathbf{A}) = \left\{ \mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} = \lambda\mathbf{x} \right\}.$$

15.1. Cálculo de los autovalores y los autovectores

¿Cómo podemos encontrar los autovalores de una matriz? ¿Y qué podemos decir acerca de la existencia de autovalores de una matriz en general? Para responder, escribamos la Ecuación (15.1) de manera diferente

$$\begin{aligned} \mathbf{A}\mathbf{x} &= \lambda\mathbf{x} \\ \mathbf{A}\mathbf{x} &= \lambda\mathbf{I}\mathbf{x} \\ \mathbf{A}\mathbf{x} - \lambda\mathbf{I}\mathbf{x} &= \mathbf{0} \\ (\mathbf{A} - \lambda\mathbf{I})\mathbf{x} &= \mathbf{0}, \end{aligned} \quad (15.2)$$

es decir, el autoespacio correspondiente a λ es el espacio nulo de $(\mathbf{A} - \lambda\mathbf{I})$:

$$\mathcal{E}_\lambda(\mathbf{A}) = \mathcal{N}(\mathbf{A} - \lambda\mathbf{I}).$$

Como $\mathcal{E}_\lambda(\mathbf{A})$ contiene vectores no nulos (los autovectores), $\mathcal{N}(\mathbf{A} - \lambda\mathbf{I})$ también los contiene, es decir, la matriz cuadrada $(\mathbf{A} - \lambda\mathbf{I})$ es singular, y por tanto:

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0. \quad (15.3)$$

Así pues, el problema reside en encontrar los valores de λ para los que $(\mathbf{A} - \lambda\mathbf{I})$ es singular (y por tanto que $|\mathbf{A} - \lambda\mathbf{I}| = 0$). Como veremos a continuación, al desarrollar el determinante se comprueba que existen $n + 1$ coeficientes p_0, \dots, p_n de \mathbb{R} , tales que para todo λ

$$\det(\mathbf{A} - \lambda\mathbf{I}) = p_0 + p_1\lambda + \dots + p_{n-1}\lambda^{n-1} + p_n\lambda^n.$$

$|\mathbf{A} - \lambda\mathbf{I}|$ es un polinomio en λ que denominamos *polinomio característico* de \mathbf{A} y que denotamos con $P_{\mathbf{A}}(\lambda)$.

En el ejemplo de más arriba:

$$P_{\mathbf{A}}(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}) = \begin{vmatrix} 5 - \lambda & 4 \\ 1 & 2 - \lambda \end{vmatrix} = 6 - 7\lambda + \lambda^2 = 0.$$

Vamos a demostrar que efectivamente para cualquier matriz \mathbf{A} cuadrada, $P_{\mathbf{A}}(\lambda)$ siempre es un polinomio.

Demostración. Vamos a demostrarlo por inducción sobre el orden n .

El resultado es evidente para cualquier matriz de orden uno, pues: $\det(a_{11} - \lambda) = a_{11} - \lambda$. Supuesto que es cierto para cualquier matriz de orden $(n - 1)$, veamos que también es cierto para cualquier matriz de orden n :

$$\begin{aligned} \det(\mathbf{A} - \lambda\mathbf{I}) &= \det \left[\begin{array}{c|c} {}^n\mathbf{A}^{\text{r}n} - \lambda\mathbf{I} & \begin{matrix} a_{1n} \\ \vdots \\ a_{(n-1)n} \end{matrix} \\ \hline * & \dots & * & a_{nn} - \lambda \end{array} \right] \\ &= \det \left[\begin{array}{c|c} {}^n\mathbf{A}^{\text{r}n} - \lambda\mathbf{I} & \begin{matrix} a_{1n} \\ \vdots \\ a_{(n-1)n} \end{matrix} \\ \hline * & \dots & * & a_{nn} \end{array} \right] + \det \left[\begin{array}{c|c} {}^n\mathbf{A}^{\text{r}n} - \lambda\mathbf{I} & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ \hline * & \dots & * & -\lambda \end{array} \right] \\ &= \det \left[\begin{array}{c|c} {}^n\mathbf{A}^{\text{r}n} - \lambda\mathbf{I} & \begin{matrix} a_{1n} \\ \vdots \\ a_{(n-1)n} \end{matrix} \\ \hline * & \dots & * & a_{nn} \end{array} \right] - \lambda \det({}^n\mathbf{A}^{\text{r}n} - \lambda\mathbf{I}), \end{aligned} \quad (15.4)$$

donde ${}^n\mathbf{A}^{\text{r}n}$ es la submatriz de \mathbf{A} que resulta tras quitar la fila y columna n -ésimas. Por hipótesis de inducción, $\det({}^n\mathbf{A}^{\text{r}n} - \lambda\mathbf{I})$ es un polinomio de grado $(n - 1)$; por tanto, el término que resta en (15.4) es un polinomio de grado n .

Ahora se presentan dos casos. Si $a_{nn} \neq 0$, mediante transformaciones elementales de *Tipo I* se pueden anular todos los componentes de la última fila que están a la izquierda de a_{nn} . Entonces

$$\det \left[\begin{array}{c|c} {}^n\mathbf{A}^{\text{r}n} - \lambda\mathbf{I} & \begin{matrix} a_{1n} \\ \vdots \\ a_{(n-1)n} \end{matrix} \\ \hline * & \dots & * & a_{nn} \end{array} \right] = \det \left[\begin{array}{c|c} \mathbf{B} - \lambda\mathbf{I} & \begin{matrix} a_{1n} \\ \vdots \\ a_{(n-1)n} \end{matrix} \\ \hline 0 & \dots & 0 & a_{nn} \end{array} \right] = a_{nn} \cdot \det(\mathbf{B} - \lambda\mathbf{I}),$$

que por hipótesis de inducción es un polinomio de grado $(n-1)$, por lo que la diferencia de polinomios en (15.4) es un polinomio de grado n . En el segundo caso $a_{nn} = 0$, por tanto

$$\det \left[\begin{array}{ccc|c} n^{\text{th}} \mathbf{A}^n - \lambda \mathbf{I} & & & a_{1n} \\ & & & \vdots \\ & & & a_{(n-1)n} \\ \hline * & \dots & * & 0 \end{array} \right] = \det \left[\begin{array}{ccc|c} n^{\text{th}} \mathbf{A}^n - \lambda \mathbf{I} & & & a_{1n} \\ & & & \vdots \\ & & & a_{(n-1)n} \\ \hline * & \dots & * & -1 \end{array} \right] + \det \left[\begin{array}{ccc|c} n^{\text{th}} \mathbf{A}^n - \lambda \mathbf{I} & & & 0 \\ & & & \vdots \\ & & & 0 \\ \hline * & \dots & * & 1 \end{array} \right].$$

Repitiendo el argumento de más arriba, constatamos que ambos determinantes son polinomios de grado $(n-1)$, por lo que su diferencia es un polinomio de grado menor o igual que $(n-1)$. Así que de nuevo la diferencia de polinomios en (15.4) es un polinomio de grado n . \square

Corolario 15.1.1. *Todo autovalor de \mathbf{A} es raíz del polinomio característico $P_{\mathbf{A}}(\lambda)$.*

Para el ejemplo de más arriba, puesto que $\det(\mathbf{A} - \lambda \mathbf{I}) = \lambda^2 - 7\lambda + 6 = 0$; tenemos que

$$\lambda = \frac{7 \pm \sqrt{49 - 24}}{2} = \frac{7 \pm 5}{2} = \begin{cases} 6 \\ 1 \end{cases} \implies \text{El espectro de } \mathbf{A} \text{ es } \{6, 1\}.$$

El *Teorema Fundamental del Álgebra*¹ establece que un polinomio $P(\lambda)$ con coeficientes complejos² y de grado $n > 0$ se puede factorizar como

$P_{\mathbf{A}}(\lambda) = p_0 + p_1\lambda + \dots + p_{n-1}\lambda^{n-1} + p_n\lambda^n = \alpha(\lambda_1 - \lambda)(\lambda_2 - \lambda) \dots (\lambda_n - \lambda)$, donde $\alpha, \lambda_1, \dots, \lambda_n \in \mathbb{C}$.
y por tanto, *tiene como mínimo una raíz y como máximo n raíces complejas distintas.*



A partir de ahora, y para poder hacer uso del Teorema Fundamental del Álgebra, asumiremos que tanto los vectores como las matrices están formadas por números complejos: \mathbb{C}^n y $\mathbb{C}^{n \times m}$ respectivamente.

Como consecuencia tenemos el siguiente resultado:

Teorema 15.1.2. *Los autovalores de una matriz \mathbf{A} de orden n son las raíces del Polinomio Característico $P_{\mathbf{A}}(\lambda)$. Por tanto, \mathbf{A} tiene como mucho n autovalores distintos.*

Se denomina *multiplicidad* de una raíz al número de veces que aparece en la factorización del polinomio. Se extiende esta nomenclatura a los autovalores añadiendo la “coletilla” *algebraica*:

Definición 15.3 (Multiplicidad algebraica de un autovalor). *Si λ es una raíz de $P_{\mathbf{A}}$ de multiplicidad k diremos que λ es un autovalor de \mathbf{A} de multiplicidad algebraica k ; que denotamos con $\mu(\lambda) = k$.*

Siguiendo con el ejemplo de más arriba, la multiplicidad algebraica tanto del autovalor $\lambda_1 = 6$ como de $\lambda_2 = 1$ es uno; y los autoespacios de \mathbf{A} correspondientes a 6 y a 1 son respectivamente los espacios nulos

$$\mathcal{N}(\mathbf{A} - 6\mathbf{I}); \quad \text{y} \quad \mathcal{N}(\mathbf{A} - \mathbf{I});$$

por tanto los autovectores de \mathbf{A} correspondientes a 6 y 1 son las soluciones no nulas de los respectivos sistemas

$$\begin{bmatrix} -1 & 4 \\ 1 & -4 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{y} \quad \begin{bmatrix} 4 & 4 \\ 1 & 1 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Definición 15.4 (Multiplicidad geométrica de un autovalor). *La dimensión del autoespacio $\mathcal{N}(\mathbf{A} - \lambda \mathbf{I})$ se denomina multiplicidad geométrica del autovalor λ ; que denotamos con $\gamma(\lambda)$.*

¹Cuya demostración está fuera del alcance de este curso.

²El conjunto de números complejos se denota con \mathbb{C} .



El problema de encontrar los autovalores y autovectores de una matriz requiere los siguientes pasos:

1. Encontrar el polinomio característico $P_{\mathbf{A}}(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I})$.
2. Encontrar las raíces λ_i de la ecuación característica $P_{\mathbf{A}}(\lambda) = 0$.
3. Resolver los sistemas homogéneos $(\mathbf{A} - \lambda_i \mathbf{I})\mathbf{x} = \mathbf{0}$ para encontrar los autovectores.

La multiplicidad *algebraica* de λ es el número de veces que se repite la raíz λ en el polinomio característico $P_{\mathbf{A}}(\lambda)$.

La multiplicidad *geométrica* de λ es la dimensión del correspondiente autoespacio $\mathcal{E}_{\lambda}(\mathbf{A})$.

Diagonalización en bloques triangulares por semejanza



En esta lección veremos que toda matriz cuadrada se puede transformar en una matriz que es diagonal **por bloques triangulares**, es decir, tal que cada submatriz (cada bloque) en la diagonal principal es una matriz triangular; además cada uno de dichos bloques tiene repetido un mismo autovalor en su correspondiente diagonal principal. También veremos la condición necesaria para que los bloques sean de tamaño 1, es decir, para que la matriz sea *diagonalizable*.

Pero primero es *fundamental* que lea la Sección 4.B en la página 57 sobre las transformaciones “espejo”.

16.1. Matrices semejantes

Definición 16.1. Decimos que \mathbf{A} y \mathbf{C} (del mismo orden) son *semejantes* si existe \mathbf{S} , invertible, tal que

$$\mathbf{A} = \mathbf{S}^{-1}\mathbf{C}\mathbf{S}.$$

Las matrices semejantes comparten muchas características; por ejemplo, *tienen el mismo determinante*:

$$\det \mathbf{A} = \det (\mathbf{S}^{-1}\mathbf{C}\mathbf{S}) = \det \mathbf{S}^{-1} \cdot \det \mathbf{C} \cdot \det \mathbf{S} = \det \mathbf{C}; \quad \text{ya que } \det(\mathbf{S}^{-1}) = (\det \mathbf{S})^{-1}.$$

Las matrices semejantes también tienen idéntico polinomio característico:

Proposición 16.1.1. Si \mathbf{A} y \mathbf{C} son semejantes, entonces tienen el mismo polinomio característico.

Demostración. Puesto que son similares, existe una matriz invertible \mathbf{S} tal que $\mathbf{C} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S}$, entonces

$$\begin{aligned} P_{\mathbf{C}}(\lambda) &= \det(\mathbf{C} - \lambda\mathbf{I}) = \det(\mathbf{S}^{-1}\mathbf{A}\mathbf{S} - \lambda\mathbf{S}^{-1}\mathbf{S}) && \text{pues } \mathbf{C} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S} \text{ y } \mathbf{S}^{-1}\mathbf{S} = \mathbf{I} \\ &= \det(\mathbf{S}^{-1}(\mathbf{A} - \lambda\mathbf{I})\mathbf{S}) && \text{sacando factor común} \\ &= \det \mathbf{S}^{-1} \cdot \det(\mathbf{A} - \lambda\mathbf{I}) \cdot \det \mathbf{S} && \text{determinante de un producto de matrices} \\ &= \det(\mathbf{A} - \lambda\mathbf{I}) = P_{\mathbf{A}}(\lambda) && \text{ya que } \det(\mathbf{S}^{-1}) = (\det \mathbf{S})^{-1}. \end{aligned}$$

□

Y puesto que cada raíz del polinomio característico es un autovalor de la matriz, y que la multiplicidad de cada raíz es la multiplicidad *algebraica* de cada autovalor; concluimos que *dos matrices semejantes tienen los mismos autovalores y con la misma multiplicidad algebraica*.

Pero además *los autovalores de dos matrices semejantes también tienen la misma multiplicidad geométrica*. Es fácil deducirlo. Sea \mathbf{A} de orden n ; sabemos que si \mathbf{S} es invertible y del mismo orden, entonces $\mathcal{C}(\mathbf{A}\mathbf{S}) = \mathcal{C}(\mathbf{A})$, y por tanto las formas escalonadas de las matrices $\mathbf{A}\mathbf{S}$ y \mathbf{A} tienen el mismo número de

columnas y de pivotes; por tanto $\dim \mathcal{N}(\mathbf{A}\mathbf{S}) = \dim \mathcal{N}(\mathbf{A})$. Por otra parte sabemos que si \mathbf{S} es invertible, $\mathcal{N}(\mathbf{S}\mathbf{A}) = \mathcal{N}(\mathbf{A})$.

Combinando ambos resultados y teniendo en cuenta que

$$\mathbf{S}^{-1}\mathbf{A}\mathbf{S} - \lambda_i\mathbf{I} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S} - \lambda_i\mathbf{S}^{-1}\mathbf{S} = \mathbf{S}^{-1}(\mathbf{A} - \lambda_i\mathbf{I})\mathbf{S},$$

tenemos que para cada autovalor λ_i de \mathbf{A} ,

$$\dim \mathcal{N}(\mathbf{S}^{-1}\mathbf{A}\mathbf{S} - \lambda_i\mathbf{I}) = \dim \mathcal{N}(\mathbf{S}^{-1}(\mathbf{A} - \lambda_i\mathbf{I})\mathbf{S}) = \dim \mathcal{N}(\mathbf{A} - \lambda_i\mathbf{I}).$$

Recordemos que la *traza* de \mathbf{A} es la suma de los elementos de su diagonal principal, es decir:

$$\text{tr}(\mathbf{A}) = a_{11} + a_{22} + \cdots + a_{kk}, \quad \text{donde } k = \min(m, n).$$

Pues bien, *dos matrices semejantes también tienen la misma traza*. Para demostrarlo comenzamos multiplicando \mathbf{A} por la matriz elemental \mathbf{I}_τ por la derecha y por su inversa por la izquierda. Evidentemente la matriz resultante es similar a \mathbf{A} . Veamos que la traza no cambia ¹...

Proposición 16.1.2. Si $\mathbf{A} = (\mathbf{I}_{\tau^{-1}})\mathbf{B}(\mathbf{I}_\tau)$, es decir, si $\mathbf{A} =_{\text{esp}(\tau^{-1})} \mathbf{B}_\tau$, entonces $\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{B})$.

Demostración. Veamos que la traza se mantiene tanto con transformaciones *Tipo I* como *Tipo II*:

Si τ es de Tipo II y τ multiplica por α a la columna i -ésima, entonces $\text{esp}(\tau^{-1})$ dividirá por α la fila i -ésima. Por tanto, el i -ésimo componente de la diagonal no cambiará. Veámoslo:

$$\mathbf{A} \left(\mathbf{I}_{\tau} \right)_{[(\alpha)j]} \Big|_j = \left(\mathbf{A} \right)_{[(\alpha)j]} \Big|_j = \alpha(\mathbf{A})_{jj};$$

pero al aplicar la inversa también tenemos

$$\left(\mathbf{I}_{\tau} \right)_{[(\frac{1}{\alpha})j]} \Big|_j \mathbf{A} = \left(\mathbf{I}_{\tau} \right)_{[(\frac{1}{\alpha})j]} \Big|_j \mathbf{A} = \frac{1}{\alpha}(\mathbf{A})_{jj}.$$

Como la primera operación multiplica la componente $j|\mathbf{A}|_j$ por α y la segunda la divide por α , la diagonal no cambia; y por tanto tampoco cambia la traza.

Si τ es de Tipo I y τ suma α veces la columna i -ésima a la j -ésima, entonces la inversa de la correspondiente matriz elemental $\mathbf{I}_{[(\alpha)i+j]}$ es la matriz elemental $\mathbf{I}_{[(-\alpha)j+i]}$; y por tanto

$$\left(\mathbf{A} \mathbf{I}_{[(\alpha)i+j]} \right)_{jj} = \alpha(\mathbf{A})_{jj} + \mathbf{A}_{jj} \Rightarrow \left(\mathbf{A} \mathbf{I}_{[(\alpha)i+j]} \right)_{jj} = \alpha(j|\mathbf{A}|_i) + j|\mathbf{A}|_j,$$

pero por otra parte

$$\left(\mathbf{I}_{[(-\alpha)j+i]} \mathbf{A} \right)_{ii} = -\alpha(j|\mathbf{A}) + i|\mathbf{A} \Rightarrow \left(\mathbf{I}_{[(-\alpha)j+i]} \mathbf{A} \right)_{ii} = -\alpha(j|\mathbf{A}|_i) + i|\mathbf{A}|_i.$$

Así pues, estas transformaciones cambian los elementos j -ésimo e i -ésimo de la diagonal, a uno se le suma $\alpha_{j|}\mathbf{A}_{|i}$ y al otro se le resta $\alpha_{j|}\mathbf{A}_{|i}$. Por tanto (aunque cambia la diagonal) la traza no cambia. \square

... y puesto que las matrices invertibles son producto de matrices elementales, tenemos el siguiente

Corolario 16.1.3. Si \mathbf{A} y \mathbf{B} son semejantes, entonces tienen la misma traza.

¹repase las transformaciones *espejo* en la Sección 4.B en la página 57.


```
A = Matrix([[1,1,1],[1,1,1],[1,1,1]])
Tr = T((5,2)) & T((-2,3,1))
(Tr**-1).espejo() & Matrix(A) & Tr           # No cambia la traza
```

Y ahora veamos el resultado más importante de la lección: que dada \mathbf{A} de orden n , y *conocidos sus autovalores*, siempre es posible transformar \mathbf{A} en una matriz diagonal por bloques triangulares similar a \mathbf{A} y cuya diagonal contiene todos los autovalores de \mathbf{A} .

De propina deduciremos que la suma de los autovalores es la traza y que su producto es el determinante.

Nota 4. ¡Recuerde que para obtener matrices semejantes hay que operar tanto con las filas como con las columnas! (... si multiplicamos por la derecha con \mathbf{S} debemos multiplicar por la izquierda con \mathbf{S}^{-1}).

16.2. Diagonalización por bloques triangulares

Teorema 16.2.1. Para toda matriz cuadrada $\mathbf{A} \in \mathbb{C}^{n \times n}$ existe una matriz invertible \mathbf{S} tal que

$$\mathbf{S}^{-1}\mathbf{A}\mathbf{S} = \begin{bmatrix} \boxed{\mathbf{T}_{\lambda_k}} & & \\ & \ddots & \\ & & \boxed{\mathbf{T}_{\lambda_2}} & \\ & & & \boxed{\mathbf{T}_{\lambda_1}} \end{bmatrix} \quad \text{donde } \mathbf{T}_{\lambda_i} = \begin{bmatrix} \lambda_i & & \\ * & \lambda_i & \\ * & * & \ddots \\ * & * & \cdots & \lambda_i \end{bmatrix} \quad \text{es triangular y de orden igual a la multiplicidad}$$

algebraica de λ_i , y donde $\{\lambda_1, \dots, \lambda_k\}$ es el espectro de \mathbf{A} , es decir, el conjunto² de autovalores de \mathbf{A} .

Antes de demostrar este importante teorema, veamos algunos resultados previos que nos ayudarán a entender los pasos del algoritmo de diagonalización por bloques.

Lema 16.2.2 (De paso inicial). Si $\mathbf{A} \in \mathbb{C}^{n \times n}$ es de la forma

$$\mathbf{A} = \left[\begin{array}{c|c} \mathbf{C} & \\ \hline * & \mathbf{L} \end{array} \right],$$

donde \mathbf{C} (de orden m) es singular y \mathbf{L} es triangular inferior sin ceros en la diagonal principal, entonces existe una matriz invertible \mathbf{S} tal que

$$\mathbf{S}^{-1}\mathbf{A}\mathbf{S} = \left[\begin{array}{c|c|c} \mathbf{C}' & & \\ \hline & 0 & \\ * & & \mathbf{L} \end{array} \right], \quad \text{donde } \mathbf{C}' \text{ es de orden } (m-1).$$

Y si \mathbf{A} es simplemente singular, entonces $\mathbf{S}^{-1}\mathbf{A}\mathbf{S} = \left[\begin{array}{c|c} \mathbf{C}' & \\ \hline * & 0 \end{array} \right]$, donde \mathbf{C}' es de orden $(m-1)$.

Demostración. Etapa 1 [Anulando la última columna de \mathbf{C} (su columna m -ésima)]. Como \mathbf{C} (de orden m) es singular, podemos anular su última columna por eliminación Gaussiana usando una sucesión de

²Asumimos que en el conjunto $\{\lambda_1, \dots, \lambda_k\}$ no hay elementos repetidos.

transformaciones elementales, $\mathbf{I}_{\tau_1 \dots \tau_k} = \mathbf{R}$, que involucran únicamente a las m primeras columnas de \mathbf{A} . Por tanto, al aplicar las correspondientes transformaciones inversas “espejo” a las filas, $_{\text{esp}}(\tau_1^{-1} \dots \tau_k^{-1}) \mathbf{I} = \mathbf{R}^{-1}$, únicamente modificaremos las primeras m filas de \mathbf{A} (todas ellas con un cero en la posición m -ésima). Así obtenemos una matriz de la forma

$$_{\text{esp}}(\tau_1^{-1} \dots \tau_k^{-1}) \mathbf{A}_{(\tau_1 \dots \tau_k)} = \mathbf{R}^{-1} \mathbf{A} \mathbf{R} = \left[\begin{array}{c|cc} * & 0 & \\ \vdots & \vdots & \\ * & 0 & \\ \hline & d_{m+1} & \beta_{m+1} \\ & d_{m+2} & * \quad \beta_{m+2} \\ & \vdots & * \quad * \quad \ddots \\ & d_n & * \quad * \quad \dots \quad \beta_n \end{array} \right].$$

Etapas 2 [Anulando el resto de coeficientes de la columna m -ésima de \mathbf{A} (aquellos a la izquierda de \mathbf{L})]. Gracias a que las componentes β_j de la diagonal principal de \mathbf{L} son pivotes, mediante una sucesión de transformaciones elementales $\mathbf{I}_{\tau_{(k+1)} \dots \tau_p}$, (del tipo $\mathbf{I}_{[(\alpha_j)j+m]} = \mathbf{P}$, con $j > m$), se pueden anular las componentes d_{m+1}, \dots, d_n de la columna m -ésima. Al aplicar la sucesión de las correspondientes transformaciones inversas “espejo”, $_{\text{esp}}(\tau_{k+1}^{-1} \dots \tau_p^{-1}) \mathbf{I} = \mathbf{P}^{-1}$, (del tipo $\mathbf{I}_{[(-\alpha_j)m+j]}$, con $j > m$), solo varían las columnas correspondientes a los asteriscos “*”, ya que la fila m -ésima contiene únicamente ceros a partir de la posición m ; resultando la matriz

$$_{\text{esp}}(\tau_1^{-1} \dots \tau_k^{-1}, \tau_{(k+1)}^{-1} \dots \tau_p^{-1}) \mathbf{A}_{(\tau_1 \dots \tau_k, \tau_{k+1} \dots \tau_p)} = ((\mathbf{P}^{-1})(\mathbf{R}^{-1})) \mathbf{A} (\mathbf{R} \mathbf{P}) = \left[\begin{array}{c|cc} \mathbf{C}' & & \\ \hline * & 0 & \\ & 0 & \mathbf{L} \end{array} \right], \quad \text{donde } \mathbf{P} = \mathbf{I}_{\tau_{(k+1)} \dots \tau_p}.$$

Llamando \mathbf{S} a la matriz $\mathbf{R} \mathbf{P} = \mathbf{I}_{\tau_1 \dots \tau_p}$ hemos terminado la demostración del primer caso.

Demostrar el caso en que \mathbf{A} es simplemente singular es más sencillo... basta con aplicar la *Etapas 1*. \square

Lema 16.2.3 (Paso de continuación). Si $\mathbf{A} \in \mathbb{C}^{n \times n}$ es de la forma

$$\mathbf{A} = \left[\begin{array}{c|cc} \mathbf{C} & & \\ \hline * & \mathbf{T} & \\ * & & \mathbf{L} \end{array} \right].$$

donde \mathbf{C} (de orden k) es singular, donde \mathbf{L} (de orden $n - m$) es triangular inferior sin ceros en la diagonal principal y donde \mathbf{T} es triangular inferior con la diagonal principal llena de ceros, entonces existe \mathbf{S} (invertible) tal que

$$\mathbf{S}^{-1} \mathbf{A} \mathbf{S} = \left[\begin{array}{c|cc} \mathbf{C}' & & \\ \hline * & \mathbf{T}' & \\ * & & \mathbf{L} \end{array} \right],$$

donde \mathbf{C}' es de orden $(k - 1)$ y \mathbf{T}' es triangular inferior con la diagonal principal llena de ceros.

Y si \mathbf{A} es simplemente de la forma $\left[\begin{array}{c|c} \mathbf{C} & \\ \hline * & \mathbf{T} \end{array} \right]$, entonces $\mathbf{S}^{-1} \mathbf{A} \mathbf{S} = \left[\begin{array}{c|c} \mathbf{C}' & \\ \hline * & \mathbf{T}' \end{array} \right]$, donde \mathbf{C}' es de orden $k - 1$ y \mathbf{T}' es triangular inferior con la diagonal principal llena de ceros.

Demostración. Etapa 1 [Anulando la última columna de \mathbf{C} (su columna k ésima)]. Aplicando la eliminación Gaussiana como en la *Etapa 1* del lema anterior obtenemos una matriz de la forma

$$\left[\begin{array}{c|c|c|c} * & 0 & & \\ k \times (k-1) & \vdots & & \\ & 0 & & \\ \hline * & * & \mathbf{T} & \\ \hline & d_{m+1} & & \beta_{m+1} \\ * & d_{m+2} & & * \beta_{m+2} \\ & \vdots & & * \quad * \quad \ddots \\ & d_n & & * \quad * \quad \cdots \beta_n \end{array} \right], \quad \text{donde } \mathbf{T}' = \left[\begin{array}{c|c} 0 & \\ * & \mathbf{T} \end{array} \right].$$

Etapa 2 [Anulando coeficientes de la columna k ésima que están a la izquierda de \mathbf{L}]. Hacemos lo mismo que en la *Etapa 2* del lema anterior quedando una matriz de la forma

$$\left[\begin{array}{c|c|c|c} \mathbf{C}' & & & \\ \hline & 0 & & \\ * & * & \mathbf{T} & \\ \hline & 0 & & \mathbf{L} \end{array} \right], \quad \text{donde } \mathbf{T}' = \left[\begin{array}{c|c} 0 & \\ * & \mathbf{T} \end{array} \right].$$


Y si \mathbf{A} es simplemente de la forma $\left[\begin{array}{c|c} \mathbf{C}' & \\ * & \mathbf{L} \end{array} \right]$, entonces basta aplicar el *Etapa 1*. □

Antes de pasar al siguiente corolario, recuérdese que para una *matriz triangular por bloques* (véase el EJERCICIO 83 en la página 168)

$$\mathbf{A} = \left[\begin{array}{c|c} \mathbf{B} & \mathbf{0} \\ \hline \mathbf{C} & \mathbf{D} \end{array} \right] \quad \text{se verifica que} \quad |\mathbf{A}| = |\mathbf{B}| \cdot |\mathbf{D}|;$$

por tanto, el polinomio característico de la matriz triangular por bloques \mathbf{A} es igual al producto de los polinomios característicos de \mathbf{B} y \mathbf{D} ; es decir $P_{\mathbf{A}}(\lambda) = P_{\mathbf{B}}(\lambda) \cdot P_{\mathbf{D}}(\lambda)$, o expresado con determinantes:

$$\det(\mathbf{A} - \lambda \mathbf{I}) = \det(\mathbf{B} - \lambda \mathbf{I}) \cdot \det(\mathbf{D} - \lambda \mathbf{I}).$$

 Usaremos este resultado sobre polinomios característicos en la última parte de la demostración del siguiente corolario, que nos indica como iniciar el algoritmo de diagonalización para generar un primer bloque:

Corolario 16.2.4. Si $\mathbf{A} \in \mathbb{C}^{n \times n}$ y λ_1 es un autovalor de \mathbf{A} , entonces existe \mathbf{S} invertible tal que

$$\text{o bien} \quad \mathbf{S}^{-1} \mathbf{A} \mathbf{S} = \left[\begin{array}{c|c} \mathbf{C}' & \\ * & \mathbf{T}_{\lambda_1} \end{array} \right]; \quad \text{o bien} \quad \mathbf{S}^{-1} \mathbf{A} \mathbf{S} = \mathbf{T}_{\lambda_1};$$

donde $\mathbf{T}_{\lambda_1} = \left[\begin{array}{c} \lambda_1 \\ * \lambda_1 \\ * \quad * \quad \ddots \\ * \quad * \quad \cdots \quad \lambda_1 \end{array} \right]$ es de orden igual a la multiplicidad algebraica de λ_1 .

Demostración. Como $(\mathbf{A} - \lambda_1 \mathbf{I})$ es singular, aplicando el Lema *de paso inicial* e iterando el Lema *de paso de continuación* mientras sea posible (mientras la submatriz de la esquina superior izquierda sea singular), llegamos a

$$(\mathbf{S}_k^{-1} \cdots \mathbf{S}_1^{-1})(\mathbf{A} - \lambda_1 \mathbf{I})(\mathbf{S}_1 \cdots \mathbf{S}_k) = \begin{cases} \begin{bmatrix} \mathbf{T} \end{bmatrix} & \text{si pudimos continuar hasta el final} \\ \left[\begin{array}{c|c} \mathbf{C}' & \\ \hline * & \mathbf{T} \end{array} \right] & \text{si topamos con una submatriz } \mathbf{C}' \text{ invertible} \end{cases}, \quad (16.1)$$

donde en ambos casos \mathbf{T} es triangular inferior de orden k , con la diagonal llena de ceros (donde k es el número de *pasos* que hemos dado). Como para cualquier \mathbf{S} invertible se verifica que $\mathbf{S}^{-1}(\mathbf{A} - \lambda_1 \mathbf{I})\mathbf{S} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S} - \lambda_1 \mathbf{I}$, sumando $\lambda_1 \mathbf{I}$ en (16.1) tenemos

$$(\mathbf{S}_k^{-1} \cdots \mathbf{S}_1^{-1})\mathbf{A}(\mathbf{S}_1 \cdots \mathbf{S}_k) = \begin{cases} \begin{bmatrix} \mathbf{T} + \lambda_1 \mathbf{I} \end{bmatrix} \\ \left[\begin{array}{c|c} \mathbf{C}' + \lambda_1 \mathbf{I} & \\ \hline * & \mathbf{T} + \lambda_1 \mathbf{I} \end{array} \right] \end{cases}; \quad \text{donde } \mathbf{T}_{\lambda_1} = \begin{bmatrix} \mathbf{T} + \lambda_1 \mathbf{I} \end{bmatrix}.$$

En el primer caso $k = n$ y el polinomio característico de la matriz \mathbf{A} coincide con el polinomio característico de \mathbf{T}_{λ_1} , que es $(\lambda_1 - \lambda)^n$, donde n es el orden tanto de la matriz \mathbf{T} como de \mathbf{A} .

En el segundo caso, el polinomio característico de \mathbf{A} es igual al producto de los polinomios característicos de las dos submatrices, es decir, $(P_{(\mathbf{C}' + \lambda_1 \mathbf{I})}(\lambda)) \cdot (\lambda_1 - \lambda)^k$. Ahora bien, como λ_1 no es un autovalor³ de $(\mathbf{C}' + \lambda_1 \mathbf{I})$, entonces k es la multiplicidad del autovalor λ_1 . \square



... el último corolario nos dice cómo continuar el algoritmo para seguir generando el resto de bloques en la diagonal (nótese que es casi idéntico al anterior)...

Corolario 16.2.5. Si $\mathbf{A} \in \mathbb{C}^{n \times n}$ de la forma $\left[\begin{array}{c|c} \mathbf{C} & \\ \hline * & \begin{matrix} \boxed{\mathbf{T}_{\lambda_r}} & & \\ & \ddots & \\ & & \boxed{\mathbf{T}_{\lambda_1}} \end{matrix} \end{array} \right]$, donde $\mathbf{T}_{\lambda_i} = \begin{bmatrix} \lambda_i & & \\ * & \lambda_i & \\ & * & \ddots \\ * & * & \cdots & \lambda_i \end{bmatrix}$ es

de orden igual a la multiplicidad algebraica del autovalor λ_i y donde λ_{r+1} es un autovalor⁴ de \mathbf{C} , entonces existe \mathbf{S} invertible tal que

$$\text{o bien } \mathbf{S}^{-1}\mathbf{A}\mathbf{S} = \left[\begin{array}{c|c} \mathbf{C}' & \\ \hline * & \begin{matrix} \boxed{\mathbf{T}_{\lambda_{r+1}}} & & \\ & \ddots & \\ & & \boxed{\mathbf{T}_{\lambda_1}} \end{matrix} \end{array} \right], \quad \text{o bien } \mathbf{S}^{-1}\mathbf{A}\mathbf{S} = \left[\begin{array}{c|c} \mathbf{T}_{\lambda_{r+1}} & \\ \hline * & \begin{matrix} \boxed{\mathbf{T}_{\lambda_r}} & & \\ & \ddots & \\ & & \boxed{\mathbf{T}_{\lambda_1}} \end{matrix} \end{array} \right],$$

$$\text{donde } \mathbf{T}_{\lambda_{r+1}} = \begin{bmatrix} \lambda_{r+1} & & \\ * & \lambda_{r+1} & \\ * & * & \ddots \\ * & * & \cdots & \lambda_{r+1} \end{bmatrix} \quad \text{es de orden igual a la multiplicidad algebraica de } \lambda_{r+1}.$$

³si fuera autovalor, entonces \mathbf{C}' sería singular.

⁴por tanto también es un autovalor de \mathbf{A} distinto de $\lambda_1, \dots, \lambda_r$

Demostración. Como $(\mathbf{C} - \lambda_{r+1}\mathbf{I})$ es singular, repitiendo los pasos de la anterior demostración:

$$(\mathbf{S}_k^{-1} \cdots \mathbf{S}_1^{-1})\mathbf{A}(\mathbf{S}_1 \cdots \mathbf{S}_k) = \begin{cases} \left[\begin{array}{c|c} \mathbf{T}_{\lambda_{r+1}} & \\ \hline * & \mathbf{T}_{\lambda_r} \\ & \ddots \\ & \mathbf{T}_{\lambda_1} \end{array} \right] & \text{si pudimos continuar hasta el final} \\ \left[\begin{array}{c|c} \mathbf{C}' + \lambda_{r+1}\mathbf{I} & \\ \hline * & \mathbf{T}_{\lambda_{r+1}} \\ & \ddots \\ & \mathbf{T}_{\lambda_1} \end{array} \right] & \text{si topamos con una submatriz } \mathbf{C}' \text{ invertible} \end{cases},$$

donde en ambos casos $\mathbf{T}_{\lambda_{r+1}}$ es triangular inferior de orden k , con λ_{r+1} en la diagonal principal y donde k es el número de *pasos* que hemos dado.

En el primer caso el polinomio característico de la matriz \mathbf{A} es el producto de los polinomios característicos

$$P_{\mathbf{A}}(\lambda) = P_{\mathbf{T}_{\lambda_1}}(\lambda) \cdots P_{\mathbf{T}_{\lambda_r}}(\lambda) \cdot P_{\mathbf{T}_{\lambda_{r+1}}}(\lambda) = (\lambda_1 - \lambda)^{\mu(\lambda_1)} \cdots (\lambda_r - \lambda)^{\mu(\lambda_r)} \cdot (\lambda_{r+1} - \lambda)^k,$$

donde $\mu(\lambda_i)$ es la multiplicidad *algebraica* del autovalor λ_i . En el segundo caso

$$P_{\mathbf{A}}(\lambda) = \left(P_{(\mathbf{C}' + \lambda_{r+1}\mathbf{I})}(\lambda) \right) \cdot (\lambda_1 - \lambda)^{\mu(\lambda_1)} \cdots (\lambda_r - \lambda)^{\mu(\lambda_r)} \cdot (\lambda_{r+1} - \lambda)^k.$$

Y como λ_{r+1} no es un autovalor de $(\mathbf{C}' + \lambda_{r+1}\mathbf{I})$, entonces k es la multiplicidad del autovalor λ_{r+1} . \square

Ya solo resta demostrar el Teorema 16.2.1, pero...

Demostración del Teorema 16.2.1. Ya no hay nada que demostrar. Basta aplicar el primer corolario e iterar el segundo hasta finalizar la generación de bloques. \square

Ejemplo 41. Sea $\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & -2 & 1 \end{bmatrix}$ con autovalores 1, 1 y 0. Vamos a diagonalizar por bloques triangulares:

$$\left[\begin{array}{c} \mathbf{A} \\ \mathbf{I} \end{array} \right] \xrightarrow[0\mathbf{I}]{(-)} \left[\begin{array}{ccc|ccc} 1 & -1 & 0 & & & \\ 0 & 0 & 0 & & & \\ 0 & -2 & 1 & & & \\ \hline 1 & 0 & 0 & & & \\ 0 & 1 & 0 & & & \\ 0 & 0 & 1 & & & \end{array} \right] \xrightarrow[\begin{smallmatrix} \tau \\ [(1)1+2] \\ [(2)3+2] \\ [2 \rightleftharpoons 3] \end{smallmatrix}]{\tau} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & & & \\ 0 & 0 & 0 & & & \\ 0 & 1 & 0 & & & \\ \hline 1 & 0 & 1 & & & \\ 0 & 0 & 1 & & & \\ 0 & 1 & 2 & & & \end{array} \right] \xrightarrow[\begin{smallmatrix} \tau \\ [(-1)2+1] \\ [(-2)2+3] \\ [2 \rightleftharpoons 3] \end{smallmatrix}]{\tau} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & & & \\ 0 & 1 & 0 & & & \\ 0 & 0 & 0 & & & \\ \hline 1 & 0 & 1 & & & \\ 0 & 0 & 1 & & & \\ 0 & 1 & 2 & & & \end{array} \right] \xrightarrow[0\mathbf{I}]{(+)} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & & & \\ 0 & 1 & 0 & & & \\ 0 & 0 & 0 & & & \\ \hline 1 & 0 & 1 & & & \\ 0 & 0 & 1 & & & \\ 0 & 1 & 2 & & & \end{array} \right] = \left[\begin{array}{c} \mathbf{D} \\ \mathbf{S} \end{array} \right]$$

Así tenemos la matriz diagonal $\mathbf{D} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S}.$

Librería NAcAL para Python

```
A = Matrix([[1,-1,0],[0,0,0],[0,-2,1]]); L=[1,1,0];
DiagonalizaS(A, L, 1)
```

Ejemplo 42. Sea la matriz $\mathbf{A} = \begin{bmatrix} -2 & 0 & 3 \\ 3 & -2 & -9 \\ -1 & 2 & 6 \end{bmatrix}$ con autovalores 1, 1 y 0. Vamos a diagonalizar por bloques.

$$\begin{aligned}
 \left[\begin{array}{c|c} \mathbf{A} & \\ \hline \mathbf{I} & \end{array} \right] & \xrightarrow[1\mathbf{I}]{(-)} \left[\begin{array}{ccc|ccc} -3 & 0 & 3 & 1 & 0 & 0 \\ 3 & -3 & -9 & 0 & 1 & 0 \\ -1 & 2 & 5 & 0 & 0 & 1 \end{array} \right] \xrightarrow[\tau]{\begin{smallmatrix} [(1)1+3] \\ [(-2)2+3] \end{smallmatrix}} \left[\begin{array}{ccc|ccc} -3 & 0 & 0 & 1 & 0 & 1 \\ 3 & -3 & 0 & 0 & 1 & -2 \\ -1 & 2 & 0 & 0 & 0 & 1 \end{array} \right] \xrightarrow[\tau]{\begin{smallmatrix} [(-1)3+1] \\ [(2)3+2] \end{smallmatrix}} \left[\begin{array}{ccc|ccc} -2 & -2 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & -2 \\ -1 & 2 & 0 & 0 & 0 & 1 \end{array} \right] \xrightarrow[1\mathbf{I}]{(+)} \left[\begin{array}{ccc|ccc} -1 & -2 & 0 & 1 & 0 & 1 \\ 1 & 2 & 0 & 0 & 1 & -2 \\ -1 & 2 & 1 & 0 & 0 & 1 \end{array} \right] \\
 & \xrightarrow[1\mathbf{I}]{(-)} \left[\begin{array}{ccc|ccc} -2 & -2 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & -2 \\ -1 & 2 & 0 & 0 & 0 & 1 \end{array} \right] \xrightarrow[\tau]{[(1)1+2]} \left[\begin{array}{ccc|ccc} -2 & 0 & 0 & 1 & -1 & 1 \\ 1 & 0 & 0 & 0 & 1 & -2 \\ -1 & 3 & 0 & 0 & 0 & 1 \end{array} \right] \xrightarrow[\tau]{[(1)2+1]} \left[\begin{array}{ccc|ccc} -1 & 0 & 0 & 1 & -1 & 1 \\ 1 & 0 & 0 & 0 & 1 & -2 \\ -1 & 3 & 0 & 0 & 0 & 1 \end{array} \right] \xrightarrow[1\mathbf{I}]{(+)} \left[\begin{array}{ccc|ccc} 0 & 0 & 0 & 1 & -1 & 1 \\ 1 & 1 & 0 & 0 & 1 & -2 \\ -1 & 3 & 1 & 0 & 0 & 1 \end{array} \right] \\
 & \xrightarrow[0\mathbf{I}]{(-)} \left[\begin{array}{ccc|ccc} 0 & 0 & 0 & 1 & -1 & 1 \\ 1 & 1 & 0 & 0 & 1 & -2 \\ -1 & 3 & 1 & 0 & 0 & 1 \end{array} \right] \xrightarrow[\tau]{\begin{smallmatrix} [(-1)2+1] \\ [(4)3+1] \end{smallmatrix}} \left[\begin{array}{ccc|ccc} 0 & 0 & 0 & 6 & -1 & 1 \\ 0 & 1 & 0 & -9 & 1 & -2 \\ 0 & 3 & 1 & 4 & 0 & 1 \end{array} \right] \xrightarrow[\tau]{\begin{smallmatrix} [(1)1+2] \\ [(-4)1+3] \end{smallmatrix}} \left[\begin{array}{ccc|ccc} 0 & 0 & 0 & 6 & -1 & 1 \\ 0 & 1 & 0 & -9 & 1 & -2 \\ 0 & 3 & 1 & 4 & 0 & 1 \end{array} \right] \xrightarrow[0\mathbf{I}]{(+)} \left[\begin{array}{ccc|ccc} 0 & 0 & 0 & 6 & -1 & 1 \\ 0 & 1 & 0 & -9 & 1 & -2 \\ 0 & 3 & 1 & 4 & 0 & 1 \end{array} \right] = \begin{bmatrix} \mathbf{C} \\ \mathbf{S} \end{bmatrix}
 \end{aligned}$$

Así pues,

$$\mathbf{C} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 6 & -1 & 1 \\ -9 & 1 & -2 \\ 4 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} -2 & 0 & 3 \\ 3 & -2 & -9 \\ -1 & 2 & 6 \end{bmatrix} \begin{bmatrix} 6 & -1 & 1 \\ -9 & 1 & -2 \\ 4 & 0 & 1 \end{bmatrix} = \mathbf{S}^{-1} \mathbf{A} \mathbf{S}.$$

Por lo que hemos llegado a una matriz diagonal por bloques triangulares.

Librería NAcAL para Python

```

A = Matrix([[-2,0,3],[3,-2,-9],[-1,2,6]]); L=[1,1,0];
C = DiagonalizaS (A, L, 1) # Matriz diagonal por bloques triangulares
S = C.S                  # C = (S**-1) * A * S (La matriz S es un atributo de C)
(S**-1) * A * S          # Comprobación

```

Como se ha visto, el algoritmo es bastante pesado para ser calculado con “papel y lápiz”. . . ¡pero para eso se inventaron los ordenadores!

Ejemplo 43. $\mathbf{A} = \begin{bmatrix} 3 & 0 & -1 & 1 \\ 3 & 2 & -2 & 2 \\ 1 & -2 & 2 & 0 \\ -3 & -2 & 3 & -1 \end{bmatrix}$ con autovalores 2 (doble) y 1 (doble). Diagonalicemos por bloques:

Librería NAcAL para Python

```

A = Matrix([[3,0,-1,1],[3,2,-2,2],[1,-2,2,0],[-3,-2,3,-1]]); L=[1,1,2,2];
C = DiagonalizaS (A, L, 1) # Quitando el tercer argumento no se visualizan los pasos
C.S                        # La matriz S se guarda como atributo S

```

Por tanto,
$$\mathbf{C} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & -2 \\ 1 & 0 & 2 & -2 \\ 1 & -1 & 2 & -2 \\ 1 & -1 & 0 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 3 & 0 & -1 & 1 \\ 3 & 2 & -2 & 2 \\ 1 & -2 & 2 & 0 \\ -3 & -2 & 3 & -1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & -2 \\ 1 & 0 & 2 & -2 \\ 1 & -1 & 2 & -2 \\ 1 & -1 & 0 & 2 \end{bmatrix} = \mathbf{S}^{-1} \mathbf{A} \mathbf{S}.$$

16.2.1. Autovalores, determinante y traza de matrices semejantes

Como las matrices semejantes tienen el mismo determinante, la misma traza y los mismos autovalores (con la misma multiplicidad aritmética y geométrica); y puesto que por el Teorema 16.2.1 sabemos que para toda matriz cuadrada podemos encontrar otra similar a ella, que es diagonal por bloques *con sus autovalores en la diagonal principal*, concluimos que:

Corolario 16.2.6. *La suma de los autovalores de \mathbf{A} es igual a la traza de \mathbf{A} .*

Corolario 16.2.7. *El producto de los autovalores de \mathbf{A} es igual al determinante de \mathbf{A} .*

16.2.2. Autovectores

Hemos visto que para toda matriz \mathbf{A} cuadrada, existe una matriz \mathbf{S} invertible tal que

$$\mathbf{C} = \mathbf{S}^{-1} \mathbf{A} \mathbf{S}$$

es diagonal por bloques triangulares, y donde cada bloque triangular tiene repetido el mismo autovalor λ_i en su diagonal principal. Fijémonos ahora en aquellas columnas de \mathbf{C} en la que aparezca un autovalor λ_i con únicamente ceros por debajo, es decir, $\mathbf{C}_{|j} = \lambda_i \mathbf{I}_{|j}$ (por ejemplo, las columnas $\mathbf{C}_{|2}$ y $\mathbf{C}_{|4}$ del último ejemplo). Puesto que $\mathbf{C} = \mathbf{S}^{-1} \mathbf{A} \mathbf{S}$, tenemos que

$$\mathbf{A} \mathbf{S} = \mathbf{S} \mathbf{C} \implies \mathbf{A} \mathbf{S}_{|j} = \mathbf{S} \mathbf{C}_{|j},$$

y como para dichas columnas $\mathbf{C}_{|j} = \lambda_i \mathbf{I}_{|j}$, tenemos que $\mathbf{S} \mathbf{C}_{|j} = \lambda_i \mathbf{S} \mathbf{I}_{|j} = \lambda_i \mathbf{S}_{|j}$ y por tanto

$$\mathbf{A}(\mathbf{S}_{|j}) = \lambda_i(\mathbf{S}_{|j}).$$

Corolario 16.2.8. *Si $\mathbf{C} = \mathbf{S}^{-1} \mathbf{A} \mathbf{S}$ es diagonal por bloques triangulares y $\mathbf{C}_{|j} = \lambda_i \mathbf{I}_{|j}$, entonces $\mathbf{S}_{|j}$ es un autovector asociado al autovalor λ_i .*

Así, usando nuestro módulo de Python con el último ejemplo, llegamos a
$$\begin{bmatrix} \mathbf{C} \\ \mathbf{S} \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & -2 \\ 1 & 0 & 2 & -2 \\ 1 & -1 & 2 & -2 \\ 1 & -1 & 0 & 2 \end{bmatrix},$$

y deducimos que $(0, 0, -1, -1)$ es un autovector para $\lambda = 2$; y $(-2, -2, -2, 2)$ lo es para $\lambda = 1$.

Nótese que tras diagonalizar por bloques triangulares $\begin{bmatrix} \mathbf{A} \\ \mathbf{I} \end{bmatrix} \xrightarrow[\text{esp}(\tau_1^{-1} \dots \tau_p^{-1})]{\tau_1 \dots \tau_p} \begin{bmatrix} \mathbf{C} \\ \mathbf{S} \end{bmatrix}$ donde $\mathbf{S} = \mathbf{I}_{\tau_1 \dots \tau_p}$, como \mathbf{S} es invertible, necesariamente *los autovectores correspondientes a autovalores distintos son linealmente independientes*.

16.3. Matrices diagonalizables

Definición 16.2 (Matriz diagonalizable). Se dice que \mathbf{A} (de orden n) es **diagonalizable** si existe una matriz \mathbf{S} tal que $\mathbf{S}^{-1}\mathbf{A}\mathbf{S}$ es diagonal.

Ejemplo 44.

$$\begin{aligned}
 & \left[\begin{array}{c} \mathbf{A} \\ \mathbf{I} \end{array} \right] \xrightarrow{(-)} \begin{array}{c} \text{OI} \\ \left[\begin{array}{ccc|ccc} 3 & 4 & 2 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array} \right] \end{array} \xrightarrow{\begin{array}{c} \tau \\ [(-2)2+3] \\ [(2)1+3] \end{array}} \begin{array}{c} \left[\begin{array}{ccc|ccc} 3 & 4 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \hline 1 & 0 & 2 & 0 & 1 & 0 \\ 0 & 1 & -2 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array} \right] \end{array} \xrightarrow{\begin{array}{c} \tau \\ [(-2)3+1] \\ [(2)3+2] \end{array}} \begin{array}{c} \left[\begin{array}{ccc|ccc} 3 & 4 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \hline 1 & 0 & 2 & 0 & 1 & 0 \\ 0 & 1 & -2 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array} \right] \end{array} \xrightarrow{(+)} \begin{array}{c} \text{OI} \\ \left[\begin{array}{ccc|ccc} 3 & 4 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \hline 1 & 0 & 2 & 0 & 1 & 0 \\ 0 & 1 & -2 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array} \right] \end{array} \\
 & \xrightarrow{(-)} \begin{array}{c} \text{OI} \\ \left[\begin{array}{ccc|ccc} 2 & 4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ \hline 1 & 0 & 2 & 0 & 1 & 0 \\ 0 & 1 & -2 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array} \right] \end{array} \xrightarrow{\begin{array}{c} \tau \\ [(-2)1+2] \end{array}} \begin{array}{c} \left[\begin{array}{ccc|ccc} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ \hline 1 & -2 & 2 & 0 & 1 & 0 \\ 0 & 1 & -2 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array} \right] \end{array} \xrightarrow{\begin{array}{c} \tau \\ [(2)2+1] \end{array}} \begin{array}{c} \left[\begin{array}{ccc|ccc} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ \hline 1 & -2 & 2 & 0 & 1 & 0 \\ 0 & 1 & -2 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array} \right] \end{array} \xrightarrow{(+)} \begin{array}{c} \text{OI} \\ \left[\begin{array}{ccc|ccc} 3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 1 & -2 & 2 & 0 & 1 & 0 \\ 0 & 1 & -2 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array} \right] \end{array} = \left[\begin{array}{c} \mathbf{D} \\ \mathbf{S} \end{array} \right]
 \end{aligned}$$

Fíjese que

$$\mathbf{A}\mathbf{S}_{|1} = \begin{bmatrix} 3 & 4 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = 3\mathbf{S}_{|1}; \quad \mathbf{A}\mathbf{S}_{|2} = \begin{bmatrix} 3 & 4 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix} \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} = 1\mathbf{S}_{|2}; \quad \mathbf{A}\mathbf{S}_{|3} = \begin{bmatrix} 3 & 4 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix} \begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix} = 0\mathbf{S}_{|3}.$$

Cuando la matriz es diagonalizable, todas las columnas de \mathbf{S} son autovectores (linealmente independientes), y por tanto, la multiplicidad *algebraica* de cada autovalor (el número de veces que aparece cada λ_i en la diagonal) necesariamente coincide con la multiplicidad *geométrica* (el número de autovectores linealmente independientes asociados a dicho autovalor).

Corolario 16.3.1. Si para cada autovalor λ_i de $\mathbf{A} \in \mathbb{C}^{n \times n}$ las multiplicidades algebraica y geométrica coinciden, la matriz es diagonalizable (es decir, los bloques de la diagonalización por bloques son diagonales).

Corolario 16.3.2. Si todos los autovalores \mathbf{A} son distintos entre sí, entonces \mathbf{A} es diagonalizable.

Demostración. Si la multiplicidad algebraica de cada autovalor es uno, entonces al diagonalizar por bloques, cada bloque resultará de orden 1. \square



Si \mathbf{A} es diagonalizable e invertible tenemos que

$$\mathbf{A}^{-1} = (\mathbf{S}\mathbf{D}(\mathbf{S}^{-1}))^{-1} = \mathbf{S}\mathbf{D}^{-1}(\mathbf{S}^{-1}).$$

Además, si seguimos el siguiente convenio

$$\mathbf{A}^0 = \mathbf{A}^{(1-1)} = \mathbf{A}^1(\mathbf{A}^{-1}) = \mathbf{I};$$

entonces, para toda \mathbf{A} diagonalizable e invertible y todo número entero n : $\mathbf{A}^n = \mathbf{S}\mathbf{D}^n(\mathbf{S}^{-1})$.

Diagonalización ortogonal de matrices simétricas

17.1. Método de Gram-Schmidt

Definición 17.1. Un *sistema de vectores es ortogonal* si cada uno de los vectores es ortogonal al resto.

Proposición 17.1.1. Un sistema ortogonal sin vectores nulos es linealmente independiente.

Demostración. Sea $\mathbf{Z} = [\mathbf{z}_1; \dots; \mathbf{z}_n;]$ un sistema ortogonal sin vectores nulos. Si $\mathbf{0} = a_1\mathbf{z}_1 + \dots + a_n\mathbf{z}_n$ entonces para cada $j = 1 : n$

$$0 = \mathbf{0} \cdot \mathbf{z}_j = (a_1\mathbf{z}_1 + \dots + a_n\mathbf{z}_n) \cdot \mathbf{z}_j = a_j\mathbf{z}_j \cdot \mathbf{z}_j,$$

y como $\mathbf{z}_j \cdot \mathbf{z}_j \neq 0$ (pues $\mathbf{z}_j \neq \mathbf{0}$), necesariamente $a_j = 0$. □

Teorema 17.1.2 (Método de Gram-Schmidt). *Dado un sistema de vectores, existe una sucesión de transformaciones elementales “de izquierda a derecha” que transforman el sistema en otro equivalente¹ ortogonal.*

Demostración. Lo demostraremos por inducción sobre el número de vectores del sistema.

Si el sistema contiene un único vector no hay nada que hacer. Veamos que si el resultado es cierto para sistemas de n vectores, entonces también es cierto para sistemas de $n + 1$ vectores.

Sea el sistema $[\mathbf{y}_1; \dots; \mathbf{y}_n; \mathbf{y}_{(n+1)}]$, aplicando la hipótesis de inducción, existe una sucesión, τ_k, \dots, τ_n , de transformaciones elementales de “izquierda a derecha” que transforman el sub-sistema formado por los n primeros vectores en otro sistema $[\mathbf{z}_1; \dots; \mathbf{z}_n;]$ que es equivalente, pero además es ortogonal; por tanto, aplicando las transformaciones elementales sobre el sistema completo tenemos

$$[\mathbf{y}_1; \dots; \mathbf{y}_n; \mathbf{y}_{(n+1)};]_{\tau_1 \dots \tau_k} = [\mathbf{z}_1; \dots; \mathbf{z}_n; \mathbf{y}_{(n+1)};].$$

Ahora, con una segunda sucesión de transformaciones elementales “de izquierda a derecha”, $\tau_{(k+1)}, \dots, \tau_p$, podemos transformar el último vector en $\mathbf{z}_{(n+1)} = \mathbf{y}_{(n+1)} - a_1\mathbf{z}_1 - a_2\mathbf{z}_2 - \dots - a_n\mathbf{z}_n$. Pues bien, hagámoslo de manera que el nuevo vector $\mathbf{z}_{(n+1)}$ sea ortogonal a todos los vectores \mathbf{z}_j que le anteceden, es decir, de manera que

$$\mathbf{z}_{(n+1)} \cdot \mathbf{z}_j = 0, \quad \text{para } j = 1, \dots, n.$$

Para lograrlo fijémonos en lo siguiente. Para cada $j = 1, \dots, n$ tenemos dos casos posibles:

- Si \mathbf{z}_j es cero, entonces \mathbf{z}_j y $\mathbf{z}_{(n+1)}$ ya son ortogonales (y no hay nada que hacer).
- Si $\mathbf{z}_j \neq \mathbf{0}$, entonces

$$0 = \mathbf{z}_{(n+1)} \cdot \mathbf{z}_j = (\mathbf{y}_{(n+1)} - \widehat{a_1}\mathbf{z}_1 - \dots - \widehat{a_n}\mathbf{z}_n) \cdot \mathbf{z}_j = (\mathbf{y}_{(n+1)} \cdot \mathbf{z}_j) - \widehat{a_j}(\mathbf{z}_j \cdot \mathbf{z}_j) \Leftrightarrow \widehat{a_j} = \frac{\mathbf{y}_{(n+1)} \cdot \mathbf{z}_j}{\mathbf{z}_j \cdot \mathbf{z}_j}.$$

¹decimos que dos sistemas son equivalentes si generan el mismo espacio (véase la Definición 9.4 en la página 110)

Por tanto, la segunda sucesión de transformaciones elementales realiza la siguiente transformación.

$$\mathbf{z}_{(n+1)} = \mathbf{y}_{(n+1)} - \sum_{\mathbf{z}_j \neq \mathbf{0}} \widehat{a_j} \mathbf{z}_j, \quad \text{donde} \quad \widehat{a_j} = \frac{\mathbf{y}_{(n+1)} \cdot \mathbf{z}_j}{\mathbf{z}_j \cdot \mathbf{z}_j}, \quad (17.1)$$

de manera que tras las dos sucesiones de transformaciones elementales tenemos

$$[\mathbf{y}_1; \dots; \mathbf{y}_n; \mathbf{y}_{(n+1)};] \xrightarrow{\tau_1, \dots, \tau_k} [\mathbf{z}_1; \dots; \mathbf{z}_n; \mathbf{y}_{(n+1)};] \xrightarrow{\tau_{(k+1)}, \dots, \tau_p} [\mathbf{z}_1; \dots; \mathbf{z}_n; \mathbf{z}_{(n+1)};],$$

que es un sistema *ortogonal* equivalente al de partida. \square

Compare el Teorema 17.1.2 en la página anterior con el Teorema 11.3.4 en la página 147. Fíjese además que $\sum_{\mathbf{z}_j \neq \mathbf{0}} \widehat{a_j} \mathbf{z}_j$ en la Ecuación 17.1 es la proyección ortogonal de $\mathbf{y}_{(n+1)}$ sobre $\mathcal{L}([\mathbf{z}_1; \dots; \mathbf{z}_n;])$. Fíjese en la Figura 11.4.

Corolario 17.1.3. *Cualquier sistema ortogonal de vectores de \mathbb{R}^n no nulos se puede extender hasta formar una base ortogonal de \mathbb{R}^n .*

Demostración. Sea $[\mathbf{z}_1; \dots; \mathbf{z}_r;]$ un sistema ortogonal de vectores de \mathbb{R}^n y sea $[\mathbf{y}_1; \dots; \mathbf{y}_m;]$ un sistema generador de \mathbb{R}^n . Entonces aplicando Gram-Schmidt sobre el sistema ampliado

$$[\mathbf{z}_1; \dots; \mathbf{z}_r; \mathbf{y}_1; \dots; \mathbf{y}_m;]$$

comenzando sobre el vector $(r+1)$ ésimo, obtenemos un nuevo sistema equivalente al ampliado y ortogonal²

$$[\mathbf{z}_1; \dots; \mathbf{z}_r; \mathbf{z}_{r+1}; \dots; \mathbf{z}_{r+m};].$$

Por ser un sistema equivalente al anterior, es un sistema generador de \mathbb{R}^n . Si de este sistema quitamos los vectores nulos, seguirá siendo generador y ortogonal, pero además será linealmente independiente. \square

17.2. Matrices ortogonales

Definición 17.2. Un *sistema de vectores es ortonormal* si es ortogonal y cada vector es de norma uno.

En la demostración del Teorema espectral que enunciaremos más adelante, usaremos el siguiente resultado.

EJERCICIO 87. Demuestre el siguiente corolario:

Corolario 17.2.1. Dado \mathbf{q} de \mathbb{R}^n y *unitario*, existe una base ortonormal de \mathbb{R}^n cuyo último vector es \mathbf{q} .

Habitualmente denotaremos con \mathbf{Q} a las matrices cuyas columnas forman un sistema *ortonormal*.

EJERCICIO 88. Demuestre la siguiente proposición:

Proposición 17.2.2. Las columnas de \mathbf{Q} son ortonormales si y solo si $(\mathbf{Q}^\top)\mathbf{Q} = \mathbf{I}$.

Nótese que cuando $m > n = r$ entonces $(\mathbf{Q}^\top)\mathbf{Q} = \mathbf{I}_{n \times n} \neq \mathbf{Q}(\mathbf{Q}^\top)$. Por ejemplo, si $\mathbf{Q} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$ entonces

$$(\mathbf{Q}^\top)\mathbf{Q} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}_{2 \times 2}; \quad \text{pero} \quad \mathbf{Q}(\mathbf{Q}^\top) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \neq \mathbf{I}_{3 \times 3}.$$

²En realidad se puede empezar desde la posición 1, pues los vectores \mathbf{y}_i ya son ortogonales entre si, y por tanto el método no los modifica.

Definición 17.3. Decimos que \mathbf{Q} es **matriz ortogonal** si es cuadrada y sus columnas son ortonormales.

Corolario 17.2.3. \mathbf{Q} es ortogonal si y solo si $\mathbf{Q}^\top = \mathbf{Q}^{-1}$.

Nótese por tanto que las columnas de una matriz ortogonal \mathbf{Q} de orden n forman una base de \mathbb{R}^n .

EJERCICIO 89. Demuestre las siguientes proposiciones:

(a) **Proposición 17.2.4.** El producto de matrices ortogonales es ortogonal.

(b) **Proposición 17.2.5.** Si \mathbf{A} es simétrica y \mathbf{Q} es ortogonal, entonces $(\mathbf{Q}^{-1})\mathbf{A}\mathbf{Q}$ es simétrica.

17.3. Nota sobre la conjugación de números complejos

Un número complejo se expresa de la forma $a + bi$, donde a y b son números reales, y donde i es la solución de la ecuación $x^2 = -1$ (es decir, $i^2 = -1$).

Puesto que ningún número real satisface dicha ecuación, a i se le denomina *número imaginario*. Para el número complejo $a + bi$, denominamos *parte real* a “ a ” y *parte imaginaria* a “ b ”.

- Para un número complejo $a + bi$, su conjugado es: $\overline{a + bi} = a - bi$ (es decir, se cambia el signo de la parte imaginaria).
- Un número complejo es real si y solo si es igual a su conjugado: $\bar{x} = x$.
- El producto del número $a + bi$ por su conjugado es $(a + bi) \cdot \overline{(a + bi)} = a^2 + b^2$.
- El conjugado de una suma es la suma de los conjugados: $\overline{x + y} = \bar{x} + \bar{y}$.
- El conjugado de un producto es el producto de los conjugados: $\overline{x \cdot y} = \bar{x} \cdot \bar{y}$.

17.4. Diagonalización de matrices simétricas

Proposición 17.4.1. Los autovalores λ de una matriz \mathbf{A} real y simétrica son reales.

Demostración. Suponga $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ (con $\mathbf{x} \neq \mathbf{0}$); y donde tanto λ como \mathbf{x} son complejos. Entonces multiplicando $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ por el conjugado \bar{x} , tenemos

$$\bar{x}\mathbf{A}\mathbf{x} = \bar{x} \cdot (\lambda\mathbf{x}) = \lambda(\bar{x} \cdot \mathbf{x}), \quad (17.2)$$

donde $\bar{x} \cdot \mathbf{x} \neq 0$ por ser $\mathbf{x} \neq \mathbf{0}$. Puesto que el conjugado de \mathbf{A} es \mathbf{A} , por ser una matriz real; tomando el conjugado de $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$, tenemos $\mathbf{A}\bar{x} = \bar{\lambda}\bar{x}$. Como \mathbf{A} es simétrica, para cualquier \mathbf{y} tenemos que $\mathbf{A}\mathbf{y} = \mathbf{y}\mathbf{A}$, en particular $\mathbf{A}\bar{x} = \bar{x}\mathbf{A} = \bar{\lambda}\bar{x}$. Y multiplicando por \mathbf{x} a ambos lados tenemos

$$\bar{x}\mathbf{A}\mathbf{x} = (\bar{\lambda}\bar{x}) \cdot \mathbf{x} = \bar{\lambda}(\bar{x} \cdot \mathbf{x}). \quad (17.3)$$

Como en (17.2) y (17.3) los lados izquierdos son idénticos, necesariamente los lados derechos son iguales; por tanto

$$\bar{\lambda} = \lambda,$$

algo que sólo es posible si la parte imaginaria es cero. Por tanto los autovalores son reales. \square

Además, como los autovectores provienen de resolver la ecuación real $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$, necesariamente son reales; por tanto, obtenemos el siguiente corolario:

Corolario 17.4.2. Los autovectores \mathbf{x} de una matriz real y simétrica son reales.

Teorema 17.4.3. Si \mathbf{A} es real y simétrica, entonces existe \mathbf{Q} ortogonal y tal que $\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}$ es diagonal.

$n \times n$

Demostración. Lo demostraremos describiendo un algoritmo que construye dicha base en dos etapas.

Paso inicial. Sea \mathbf{q}_n un autovector unitario de \mathbf{A} correspondiente al autovalor λ_n . Aplicando el Corolario 17.2.1 extendemos el sistema $[\mathbf{q}_n;]$ hasta obtener una base ortonormal de \mathbb{R}^n

$$\mathbf{Q}_n = [\mathbf{q}_1; \dots; \mathbf{q}_n].$$

Por tanto \mathbf{Q}_n es ortogonal, y $\mathbf{Q}_n^{-1} \mathbf{A} \mathbf{Q}_n$ es simétrica y de la forma $\mathbf{Q}_n^{-1} \mathbf{A} \mathbf{Q}_n = \begin{bmatrix} \mathbf{A}' & 0 \\ (n-1) \times (n-1) & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$, ya que

$$(\mathbf{Q}_n^{-1} \mathbf{A} \mathbf{Q}_n)_{|n} = (\mathbf{Q}_n^T \mathbf{A} \mathbf{Q}_n)_{|n} = \mathbf{Q}_n^T \mathbf{A} \mathbf{q}_n = \lambda_n \cdot \mathbf{Q}_n^T \mathbf{q}_n = \lambda_n \begin{pmatrix} \mathbf{q}_1 \cdot \mathbf{q}_n \\ \mathbf{q}_2 \cdot \mathbf{q}_n \\ \vdots \\ \mathbf{q}_n \cdot \mathbf{q}_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \lambda_n \end{pmatrix} = \lambda_n \cdot \mathbf{I}_{|n}.$$

Paso de continuación. Supongamos que tras $n - k$ pasos tenemos

$$\mathbf{Q}_{k+1}^{-1} \dots \mathbf{Q}_n^{-1} \cdot \mathbf{A} \cdot \mathbf{Q}_n \dots \mathbf{Q}_{k+1} = \left[\begin{array}{c|ccc} \mathbf{A}' & & & \\ \hline k \times k & & & \\ \hline & \lambda_{k+1} & & \\ & & \ddots & \\ & & & \lambda_n \end{array} \right], \quad (17.4)$$

como \mathbf{A}' es simétrica, tomamos un autovector $\mathbf{x}_k \in \mathbb{R}^k$ de \mathbf{A}' de norma 1 y correspondiente al autovalor λ_k . Aplicando el Corolario 17.2.1 formamos una base ortonormal $\mathbf{R} = [\mathbf{x}_1; \dots; \mathbf{x}_k;]$ de \mathbb{R}^k . Ahora, si consideramos la matriz por bloques

$$\mathbf{Q}_k = \left[\begin{array}{c|c} \mathbf{R} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{I} \\ \hline & (n-k) \times (n-k) \end{array} \right] = [\mathbf{y}_1; \dots; \mathbf{y}_n],$$

entonces \mathbf{Q}_k es ortogonal y sus últimas columnas $\mathbf{y}_k, \dots, \mathbf{y}_n$ son autovectores de la matriz de la Ecuación (17.4); es decir, llamando \mathbf{B} a la matriz de dicha ecuación tenemos que

$$\mathbf{B} \mathbf{y}_j = \lambda_j \mathbf{y}_j; \quad \text{para } j = k, \dots, n.$$

Consecuentemente, multiplicando (17.4) por \mathbf{Q}_k^{-1} y por \mathbf{Q}_k tenemos

$$\mathbf{Q}_k^{-1} \cdot (\mathbf{Q}_{k+1}^{-1} \dots \mathbf{Q}_n^{-1} \cdot \mathbf{A} \cdot \mathbf{Q}_n \dots \mathbf{Q}_{k+1}) \cdot \mathbf{Q}_k = \mathbf{Q}_k^{-1} \mathbf{B} \mathbf{Q}_k = \left[\begin{array}{c|ccc} \mathbf{A}'' & & & \\ \hline k-1 \times k-1 & & & \\ \hline & \lambda_k & & \\ & & \ddots & \\ & & & \lambda_n \end{array} \right],$$

ya que, $(\mathbf{Q}_k^{-1} \mathbf{B} \mathbf{Q}_k)_{|j} = \mathbf{Q}_k^T (\mathbf{B} \mathbf{Q}_k)_{|j} = \mathbf{Q}_k^T (\mathbf{B} \mathbf{y}_j) = \mathbf{Q}_k^T (\lambda_j \mathbf{y}_j) = \lambda_j \begin{pmatrix} \mathbf{y}_1 \cdot \mathbf{y}_j \\ \mathbf{y}_2 \cdot \mathbf{y}_j \\ \vdots \\ \mathbf{y}_n \cdot \mathbf{y}_j \end{pmatrix} = \lambda_j (\mathbf{I}_{|j}).$

Así, el producto de matrices $\mathbf{Q} = \mathbf{Q}_n \dots \mathbf{Q}_1$ es ortogonal y $\mathbf{D} = \mathbf{Q}^{-1} \mathbf{A} \mathbf{Q}$ es diagonal con los autovalores $\lambda_1, \dots, \lambda_n$ de \mathbf{A} correspondientes a las columnas de \mathbf{Q} en su diagonal principal. \square

Definición 17.4. \mathbf{A} es diagonalizable ortogonalmente si existe \mathbf{Q} ortogonal tal que $\mathbf{Q}^{-1} \mathbf{A} \mathbf{Q}$ es diagonal.

El anterior teorema tiene un importante corolario que se conoce por [Teorema Espectral](#):

Corolario 17.4.4 (Teorema espectral). Si \mathbf{A} (de orden n) es [real y simétrica](#), entonces existe una base ortonormal de \mathbb{R}^n formada por autovectores de \mathbf{A} .

Para finalizar, demostramos un último resultado acerca de la diagonalización de matrices simétricas:

Proposición 17.4.5. Los autovectores correspondientes a autovalores distintos, [de una matriz simétrica](#), son ortogonales entre si.

Demostración. Considere \mathbf{x} e \mathbf{y} , autovectores correspondientes a dos autovalores distintos: $\mathbf{Ax} = \lambda_1\mathbf{x}$ y $\mathbf{Ay} = \lambda_2\mathbf{y}$. Entonces

$$\lambda_1 \mathbf{x} \cdot \mathbf{y} = \mathbf{Ax} \cdot \mathbf{y} = \mathbf{x}(\mathbf{A}^\top)\mathbf{y} = \mathbf{xAy} = \lambda_2(\mathbf{x} \cdot \mathbf{y}).$$

Como $\lambda_1 \neq \lambda_2$; necesariamente:

$$\lambda_1(\mathbf{x} \cdot \mathbf{y}) - \lambda_2(\mathbf{x} \cdot \mathbf{y}) = 0 \implies (\lambda_1 - \lambda_2)\mathbf{x} \cdot \mathbf{y} = 0 \implies \mathbf{x} \cdot \mathbf{y} = 0.$$

□

EJERCICIO 90. Demuestre la siguiente

Proposición 17.4.6. Si una matriz es diagonalizable ortogonalmente, entonces es simétrica.

Para finalizar, la demostración del Teorema 17.4.3 nos describe un algoritmo para diagonalizar \mathbf{A} ortogonalmente, obteniendo una matriz ortogonal \mathbf{Q} cuyas columnas son autovectores de \mathbf{A} :

```
A = Matrix([[1,1,1],[1,0,0],[1,0,0]]); # Matriz simétrica
L = [-1,0,2];                          # Espectro de A
A
```

Librería NAcAL para Python

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

```
D=Diagonaliza0(A,L); # Diagonaliza Ortogonalmente
D
```

Librería NAcAL para Python

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

El atributo Q de la matriz diagonalizada es la matriz ortogonal de autovectores:

```
Q=D.Q # Matriz de autovectores
```

Librería NAcAL para Python

$$\begin{bmatrix} \frac{\sqrt{6}}{3} & 0 & -\frac{\sqrt{3}}{3} \\ \frac{\sqrt{6}}{6} & \frac{\sqrt{2}}{2} & \frac{\sqrt{3}}{3} \\ \frac{\sqrt{6}}{6} & -\frac{\sqrt{2}}{2} & \frac{\sqrt{3}}{3} \end{bmatrix}$$

Comprobemos que efectivamente \mathbf{Q} es ortogonal:

```
~Q*Q      # Matriz de autovectores es ortogonal
```

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

y que diagonaliza a **A**:

```
~Q*A*Q      # Matriz de autovectores diagonaliza A
```

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

No obstante, como toda matriz simétrica es diagonalizable, también podemos obtener una matriz **Q** aplicando el algoritmo descrito en la demostración del Teorema 16.2.1 de diagonalización por bloques triangulares: primero aplicamos la diagonalización por semejanza

$$\begin{bmatrix} \mathbf{A} \\ \mathbf{I} \end{bmatrix} \xrightarrow[\text{esp}(\tau_1^{-1} \dots \tau_p^{-1})]{\tau_1 \dots \tau_p} \begin{bmatrix} \mathbf{C} \\ \mathbf{S} \end{bmatrix} \quad \text{donde } \mathbf{S} = \mathbf{I}_{\tau_1 \dots \tau_p},$$

```
D = DiagonalizaS(A, L) # D matriz diagonalizada por semejanza
S = D.S                # S matriz de autovectores
S
```

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

y a continuación aplicamos Gram-Schmidt sobre las columnas de **S** para asegurarnos que las columnas de **S** que corresponden a autovalores repetidos son perpendiculares, y normalizamos todas las columnas para obtener una matriz **Q**.

```
Q = S.GS().normalizada() # Q matriz ortogonal de autovectores
Q
```

$$\begin{bmatrix} \frac{\sqrt{6}}{3} & 0 & -\frac{\sqrt{3}}{3} \\ \frac{\sqrt{6}}{6} & \frac{\sqrt{2}}{2} & \frac{\sqrt{3}}{3} \\ \frac{\sqrt{6}}{6} & -\frac{\sqrt{2}}{2} & \frac{\sqrt{3}}{3} \end{bmatrix}$$

Diagonalización por congruencia de matrices simétricas. Formas cuadráticas

18.1. Formas cuadráticas y matrices definidas positivas

Un polinomio es una expresión que contiene variables y coeficientes, y en la que únicamente están involucradas las operaciones de suma, resta de *términos* que contienen números, productos y exponentes no negativos de las variables. Un ejemplo de polinomio en x con tres términos es: $x^2 - 4x + 7$. Y un ejemplo de polinomio en tres variables y cuatro términos es: $x^3 + 2xyz^2 - yz + 1$.

El grado de cada término es la suma de los exponentes de las variables que aparecen en él, y por tanto nunca es negativo. El grado del polinomio es el mayor de los grados de sus términos con coeficientes no nulos. Así, para los ejemplos anteriores, el grado del polinomio $x^2 - 4x^1 + 7x^0$ es dos; y el grado del polinomio $x^3 + 2x^1y^1z^2 - y^1z^1 + 1$ es cuatro: $4 = 1 + 1 + 2$ por el término $2x^1y^1z^2$.

Definición 18.1. Una forma cuadrática es un polinomio en el que todos sus términos son de grado dos.

Por ejemplo, $4x^2 + 2xy - 3y^2$ es una forma cuadrática en las variables x e y .

18.1.1. Formas cuadráticas reales

Toda matriz \mathbf{A} simétrica de orden n define una forma cuadrática $q_{\mathbf{A}}$ en n variables mediante la formula

$$q_{\mathbf{A}}(\mathbf{x}) = \mathbf{x}\mathbf{A}\mathbf{x},$$

donde $\mathbf{x} = (x, y, \dots)$ es el vector que contiene las variables de la forma cuadrática.

Recíprocamente, dada una forma cuadrática en n variables, sus coeficientes se pueden arreglar en una matriz simétrica de orden n . Volviendo al ejemplo de más arriba

$$4x^2 + 2xy - 3y^2 = (x, \quad y) \begin{bmatrix} 4 & 1 \\ 1 & -3 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix},$$

y con la matriz $\begin{bmatrix} 1 & -2 \\ -2 & 3 \\ 5 \end{bmatrix}$ formamos la forma cuadrática

$$(x, \quad y, \quad z) \begin{bmatrix} 1 & -2 & \\ -2 & 3 & \\ & & 5 \end{bmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = x^2 - 4xy + 3y^2 + 5z^2.$$

18.1.2. Matrices definidas positivas

Definición 18.2. Una matriz simétrica $\mathbf{A} \in \mathbb{R}^{n \times n}$ es **definida positiva** si para todo $\mathbf{x} \neq \mathbf{0}$ de \mathbb{R}^n , $\mathbf{x}\mathbf{A}\mathbf{x} > 0$.

Entonces también se dice que la forma cuadrática $\mathbf{x}\mathbf{A}\mathbf{x}$ es **definida positiva**.

Como consecuencia de la definición tenemos que

Proposición 18.1.1. Si \mathbf{A} $m \times n$ tiene rango n , entonces $\mathbf{A}^\top \mathbf{A}$ es definida positiva.

Demostración. Por una parte,

$$\mathbf{x} \mathbf{A}^\top \mathbf{A} \mathbf{x} = (\mathbf{A} \mathbf{x}) \cdot (\mathbf{A} \mathbf{x}) = \|\mathbf{A} \mathbf{x}\|^2 \geq 0.$$

Como \mathbf{A} es de rango completo por columnas, si $\mathbf{x} \neq \mathbf{0}$, entonces $\mathbf{A} \mathbf{x} \neq \mathbf{0}$; y por tanto $\mathbf{x} \mathbf{A} \mathbf{x} = \|\mathbf{A} \mathbf{x}\|^2 > 0$. \square

EJERCICIO 91. Demuestre las siguientes proposiciones:

- (a) si \mathbf{A} es definida positiva los elementos de la diagonal son positivos.
- (b) si \mathbf{A} y \mathbf{B} son definidas positivas, entonces la suma $(\mathbf{A} + \mathbf{B})$ también es definida positiva.

La primera de las anteriores proposiciones se puede generalizar a todas las **submatrices principales** de \mathbf{A} (véase el Corolario 18.1.5 en el Ejercicio 93 más abajo).

Hay una estrecha relación entre los signos de los autovalores de una matriz \mathbf{A} y el signo de la forma cuadrática asociada: $q_{\mathbf{A}}(\mathbf{x}) = \mathbf{x} \mathbf{A} \mathbf{x}$.

Proposición 18.1.2. Una matriz real y simétrica de orden n es definida positiva si y sólo si, son positivos todos sus autovalores.

Demostración. La demostración se basa en que toda matriz real y simétrica es diagonalizable ortogonalmente. Puesto que \mathbf{A} es real y simétrica, tenemos que $\mathbf{A} = \mathbf{Q} \mathbf{D} (\mathbf{Q}^{-1})$ con $\mathbf{Q}^{-1} = \mathbf{Q}^\top$; y por tanto

$$\mathbf{x} \mathbf{A} \mathbf{x} = \mathbf{x} \mathbf{Q} \mathbf{D} (\mathbf{Q}^\top) \mathbf{x} = (\mathbf{Q}^\top \mathbf{x}) \mathbf{D} (\mathbf{Q}^\top \mathbf{x}) = \mathbf{y} \mathbf{D} \mathbf{y}, \quad (\text{donde } \mathbf{y} = \mathbf{Q}^\top \mathbf{x}).$$

De esta expresión es evidente que $\mathbf{x} \mathbf{A} \mathbf{x}$ es una suma ponderada de cuadrados:

$$\mathbf{x} \mathbf{A} \mathbf{x} = \mathbf{y} \mathbf{D} \mathbf{y} = (y_1, \dots, y_n) \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \sum_{j=1}^n \lambda_j (y_j^2),$$

donde las ponderaciones λ_j son los autovalores de \mathbf{A} . Puesto que \mathbf{Q} es invertible, $\mathbf{y} = \mathbf{Q}^\top \mathbf{x}$ es distinto de cero siempre que $\mathbf{x} \neq \mathbf{0}$, la suma $\sum_{j=1}^n \lambda_j (\mathbf{y}_j)^2$ es positiva cuando los autovalores son positivos.

Por otra parte, la forma cuadrática no puede ser positiva si algún autovalor λ_j es negativo o cero; para verlo basta elegir $\mathbf{y} = \mathbf{I}_{|j}$ pues $\mathbf{y} \mathbf{D} \mathbf{y} = \mathbf{I}_{|j} \mathbf{D} \mathbf{I}_{|j} = \mathbf{I}_{|j} \mathbf{D}_{|j} = \lambda_j$. \square

Así, puesto que el determinante de \mathbf{A} es igual al producto de sus autovalores, llegamos al siguiente

Corolario 18.1.3. Si una matriz es definida positiva su determinante es positivo.

EJERCICIO 92. Demuestre que si \mathbf{A} es simétrica y definida positiva, entonces \mathbf{A}^{-1} también es definida positiva.

Para finalizar esta sección:

EJERCICIO 93. Revise la Definición 1.21 en la página 23 y demuestre las siguientes proposiciones:

- (a) **Proposición 18.1.4.** Si \mathbf{A} $n \times n$ es definida positiva y \mathbf{X} $n \times k$ de rango k entonces $(\mathbf{X}^\top) \mathbf{A} \mathbf{X}$ es definida positiva.
- (b) **Corolario 18.1.5.** Si \mathbf{A} es definida positiva, todas sus submatrices principales son definidas positivas.
- (c) **Proposición 18.1.6.** Si \mathbf{A} es definida positiva entonces la mayor componente de \mathbf{A} está en la diagonal.

Pista. Demuestre antes que $a_{ij}^2 < a_{ii} a_{jj}$ usando el determinante de $_{(i,j)} \mathbf{A}_{(i,j)}$.

18.1.3. Algunos tipos de formas cuadráticas

Decimos que una forma cuadrática (y su correspondiente matriz de orden n) es

- Definida positiva si para cualquier $\mathbf{x} \in \mathbb{R}^n$ no nulo se verifica que $\mathbf{x}\mathbf{A}\mathbf{x} > 0$.
- Semi-definida positiva si para cualquier $\mathbf{x} \in \mathbb{R}^n$ no nulo se verifica que $\mathbf{x}\mathbf{A}\mathbf{x} \geq 0$.
- Definida negativa si para cualquier $\mathbf{x} \in \mathbb{R}^n$ no nulo se verifica que $\mathbf{x}\mathbf{A}\mathbf{x} < 0$.
- Semi-definida negativa si para cualquier $\mathbf{x} \in \mathbb{R}^n$ no nulo se verifica que $\mathbf{x}\mathbf{A}\mathbf{x} \leq 0$.
- Indefinida si no es ni semi-definida positiva ni semi-definida negativa.

Nótese que si \mathbf{A} es (*semi*-) definida positiva, entonces $(-\mathbf{A})$ es (*semi*-) definida negativa.

EJERCICIO 94. Demuestre que si \mathbf{A} es semi-definida positiva los elementos de la diagonal son no negativos.

18.2. Diagonalización de matrices simétricas por congruencia



Vamos a ver que podemos diagonalizar cualquier matriz por congruencia (mediante eliminación Gaussiana de filas y columnas). En general, la matriz obtenida con este procedimiento no contiene los autovalores de la matriz original en su diagonal, pero permite expresar cualquier forma cuadrática como sumas y/o restas de términos al cuadrado. Así podremos comprobar si una matriz es definida positiva sin necesidad de calcular sus autovalores (basta comprobar que la expresión solo contiene sumas de cuadrados).

Este resultado tiene importancia práctica. No es posible encontrar las raíces de un polinomio cualquiera (tan solo está asegurado encontrar las raíces de polinomios de grado menor o igual a 4). Afortunadamente la diagonalización por congruencia nos revelará *los signos* de los autovalores de “cualquier” matriz real y simétrica. Este resultado se llama *Ley de inercia*.

Definición 18.3. Dos matrices \mathbf{A} y \mathbf{C} son congruentes si existe una matriz \mathbf{B} invertible tal que

$$\mathbf{C} = (\mathbf{B}^T)\mathbf{A}\mathbf{B}.$$

En esta sección veremos que siempre es posible encontrar una matriz diagonal que sea congruente con una matriz simétrica \mathbf{A} . La demostración describirá los pasos a seguir para diagonalizar por congruencia (paso de inicio y paso de continuación). Pero antes de exponer la demostración, veamos tres ejemplos de diagonalización por congruencia. El método consiste en tratar de escalar una matriz, pero aplicando a las filas todas las operaciones que hayamos aplicado a las columnas.

Ejemplo 45. Vamos a diagonalizar la matriz $\mathbf{A} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$ mediante eliminación, pero cada operación sobre las columnas, la repetiremos también sobre las filas:¹

$$\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \xrightarrow[\begin{smallmatrix} [(2)2] \\ [(1)1+2] \end{smallmatrix}]{\begin{smallmatrix} \tau \\ \tau \end{smallmatrix}}} \begin{bmatrix} 2 & 0 & 0 \\ -1 & 3 & -1 \\ 0 & -2 & 2 \end{bmatrix} \xrightarrow[\begin{smallmatrix} [(1)1+2] \\ [(2)2] \end{smallmatrix}]{\tau} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 6 & -2 \\ 0 & -2 & 2 \end{bmatrix} \xrightarrow[\begin{smallmatrix} [(1)2+3] \\ [(3)3] \end{smallmatrix}]{\begin{smallmatrix} \tau \\ \tau \end{smallmatrix}}} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & -2 & 4 \end{bmatrix} \xrightarrow[\begin{smallmatrix} [(1)2+3] \\ [(3)3] \end{smallmatrix}]{\tau} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 12 \end{bmatrix}.$$

Aunque la matriz obtenida es diagonal, ni 6 ni 12 son autovalores de \mathbf{A} (*este método NO encuentra los autovalores*)... Pero como los componentes de la diagonal son positivos, \mathbf{A} es definida positiva.

¹Recuerde que este esquema, indica que el orden en el que se aplican las transformaciones *de las filas* es de abajo a arriba, es decir, primero las más “alejadas” de las flechas y luego las más “cercanas” a las flechas (véase el Apéndice 4.A en la página 54).

Y cuando en la diagonal tenemos un cero ¿qué podemos hacer?... Podemos sumar a la primera columna otra columna (en un ejemplo posterior veremos que hay que hacerlo con cuidado, pues no siempre funciona).

Ejemplo 46. Considere $\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$. La primera componente de la primera fila es un cero, así que no tenemos un pivote con el que anular todo lo que queda a la derecha de la diagonal. Como la segunda columna si tiene un pivote en la primera fila, podemos sumar la segunda columna a la primera columna y lograr tener un pivote donde nos interesa (pero *recuerde que cada operación sobre las columnas se repite sobre las filas*):

$$\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \xrightarrow{[(1)\mathbf{2}+1]} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 2 & 1 & 0 \end{bmatrix} \xrightarrow{[(1)\mathbf{2}+1]} \begin{bmatrix} 2 & 1 & 2 \\ 1 & 0 & 1 \\ 2 & 1 & 0 \end{bmatrix} \xrightarrow{\begin{matrix} \tau \\ [(2)\mathbf{2}] \\ [(-1)\mathbf{1}+2] \\ [(-1)\mathbf{1}+3] \end{matrix}} \begin{bmatrix} 2 & 0 & 0 \\ 1 & -1 & 0 \\ 2 & 0 & -2 \end{bmatrix} \xrightarrow{\begin{matrix} \tau \\ [(-1)\mathbf{1}+3] \\ [(-1)\mathbf{1}+2] \\ [(2)\mathbf{2}] \end{matrix}} \begin{bmatrix} 2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix}.$$

Un tercer ejemplo nos indica que debemos andar con cuidado cuando encontramos un cero en la diagonal.

Ejemplo 47. Considere la matriz $\begin{bmatrix} 0 & 1 \\ 1 & a \end{bmatrix}$, que tiene un cero en donde querríamos tener un pivote con el que eliminar todo lo que queda a su derecha (y por debajo). ¿Qué podemos hacer?

Primer intento: podemos intentar la estrategia del ejemplo anterior, sumando la segunda columna a la primera (y repetir la operación con las filas):

$$\begin{bmatrix} 0 & 1 \\ 1 & a \end{bmatrix} \xrightarrow{[(1)\mathbf{2}+1]} \begin{bmatrix} 1 & 1 \\ 1+a & a \end{bmatrix} \xrightarrow{[(1)\mathbf{2}+1]} \begin{bmatrix} 2+a & 1+a \\ 1+a & a \end{bmatrix},$$

y ahora con el pivote $(2+a)$ podemos anular todo lo que está a su derecha (y por debajo), pero... ¿estamos seguros de que $(2+a)$ es un pivote? Si $a = -2$ entonces estaríamos como al principio. En el anterior ejemplo la estrategia funcionó porque sumamos una columna que tenía un cero en la diagonal... pero en este ejemplo la columna que sumamos tiene el número a en la diagonal, y esto podría hacer fallar el algoritmo.

Segundo intento: para que la estrategia anterior no falle debemos asegurarnos de que la columna que sumamos tiene un cero en la diagonal. Como en este caso no es así, la solución pasa por intercambiar dos columnas (y filas) para “mover” alguna componente no nula de la diagonal (que esté situada a la derecha y por debajo del cero que queremos sustituir) y colocarla en la posición del cero que ocasiona el problema, y luego continuar como siempre:

$$\begin{bmatrix} 0 & 1 \\ 1 & -2 \end{bmatrix} \xrightarrow{[1\mathbf{2}]} \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \xrightarrow{[1\mathbf{2}]} \begin{bmatrix} -2 & 1 \\ 1 & 0 \end{bmatrix} \xrightarrow{\begin{matrix} \tau \\ [(2)\mathbf{2}] \\ [(1)\mathbf{1}+2] \end{matrix}} \begin{bmatrix} -2 & 0 \\ 1 & 1 \end{bmatrix} \xrightarrow{\begin{matrix} \tau \\ [(1)\mathbf{1}+2] \\ [(2)\mathbf{2}] \end{matrix}} \begin{bmatrix} -2 & 0 \\ 0 & 2 \end{bmatrix}.$$

Vayamos con la demostración de los pasos que nos permiten diagonalizar por congruencia:

Proposición 18.2.1 (Paso de inicio). *Dada una matriz \mathbf{A} , simétrica y de orden n , existe una matriz no singular \mathbf{B} tal que $\mathbf{B}^T \mathbf{A} \mathbf{B}$ es una matriz simétrica de la forma*

$$\left[\begin{array}{c|c} * & \\ \hline & \mathbf{A}' \\ & (n-1) \times (n-1) \end{array} \right].$$

Demostración. Tenemos tres casos:

Caso trivial. Si \mathbf{A} ya tiene la forma de más arriba, es decir, si a la derecha y por debajo de a_{11} son todo ceros, entonces basta que \mathbf{B} sea la matriz identidad.

En caso contrario hay dos posibilidades: que a_{11} sea cero o que no sea cero.

Caso 1 ($a_{11} \neq 0$). Cuando a_{11} es un pivote se pueden anular los componentes situados a su derecha por eliminación de izquierda a derecha, y aplicando las mismas operaciones sobre las filas (de arriba a abajo) llegamos a la siguiente matriz congruente con \mathbf{A} :

$$\tau_k \cdots \tau_1 \mathbf{A} \tau_1 \cdots \tau_k = \mathbf{B}^T \mathbf{A} \mathbf{B} = \left[\begin{array}{c|c} a_{11} & \\ \hline & \mathbf{A}' \\ & (n-1) \times (n-1) \end{array} \right]; \quad \text{donde } \mathbf{I}_{\tau_1 \cdots \tau_k} = \mathbf{B}.$$

Caso 2 ($a_{11} = 0$). En este caso hay dos posibilidades: que algún otro elemento de la diagonal principal sea distinto de cero, o que todos los elementos de la diagonal sean cero.

- Si el elemento a_{jj} (con $j > 1$) es distinto de cero; intercambiamos la primera columna con la j -ésima, de manera que ahora el elemento no nulo se encuentra en la posición $(j, 1)$ y a continuación se intercambia la fila j -ésima por la primera, con lo que el componente *no nulo* termina por situarse en la posición $(1, 1)$. Así hemos llegado a una matriz del Caso 1.
- Si todos los elementos de la diagonal principal son nulos —y puesto que algún elemento a_{1j} es distinto de cero (pues no estamos en el Caso Trivial)— sumamos \mathbf{A}_{1j} a la primera columna, y por tanto, también sumamos ${}_j \mathbf{A}$ a la primera fila. Entonces de nuevo habremos llegado al Caso 1 ya que

$${}_{[(1)j+1]} \tau \mathbf{A} {}_{[(1)j+1]} \tau = \left[\begin{array}{c|ccc} 2a_{1j} & * & \dots & * \\ * & & & \\ \vdots & & & \\ * & & & \end{array} \right]; \quad \text{donde } 2a_{1j} \neq 0.$$

□

Proposición 18.2.2 (Paso de continuación). *Dada una matriz simétrica de orden n de la forma*

$$\left[\begin{array}{c|c} * & \\ \hline & \mathbf{A}' \\ & k \times k \end{array} \right],$$

existe una matriz no singular \mathbf{B} tal que $\mathbf{B}^T \mathbf{A} \mathbf{B}$ es una matriz simétrica de la forma

$$\left[\begin{array}{c|c} * & \\ \hline & \mathbf{A}'' \\ & (k-1) \times (k-1) \end{array} \right].$$

Demostración. Basta hacer como en el *Paso de Inicio*, pero trabajando sobre las filas y columnas de \mathbf{A}' . □

Combinando las dos proposiciones anteriores llegamos al siguiente

Corolario 18.2.3. *Para toda \mathbf{A} simétrica de orden n , existe \mathbf{B} invertible tal que $\mathbf{B}^T \mathbf{A} \mathbf{B}$ es diagonal.*

Ejemplo 48. Diagonalice por congruencia la matriz

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

Piense en qué pasos debe sumar una columna (y una fila) para generar un pivote en la diagonal, y cuando debe intercambiar columnas (y filas).

Librería NAcAL para Python

```
A = Matrix([[0,0,1],[0,0,0],[1,0,0]])
D = DiagonalizaC(A,1)
B = D.B
~B*A*B                                # D es congruente con A
```



Nota 5. Fíjese que el Teorema Espectral muestra una forma muy especial de diagonalización pues

$$\mathbf{D} = \mathbf{Q}^{-1}\mathbf{A}\mathbf{Q} = \mathbf{Q}^T\mathbf{A}\mathbf{Q},$$

es simultáneamente una diagonalización por semejanza y por congruencia.

18.2.1. Ley de inercia

Teorema 18.2.4 (Ley de inercia). Sea \mathbf{A} real y simétrica; y sean \mathbf{B} y \mathbf{C} matrices invertibles tales que

$$\mathbf{D} = (\mathbf{B}^T)\mathbf{A}\mathbf{B} \quad y \quad \mathbf{D}^* = (\mathbf{C}^T)\mathbf{A}\mathbf{C}$$

son diagonales. Entonces la diagonal de \mathbf{D} tiene tantas componentes positivas, tantas componentes nulas y tantas componentes negativas como la diagonal de \mathbf{D}^* .

Demostración. Comencemos con el número de componentes nulas. Puesto que \mathbf{B} y \mathbf{C} son de rango completo, el rango de \mathbf{D} y el rango de \mathbf{D}^* deben ser iguales al rango de \mathbf{A} . Así, puesto que \mathbf{D} y \mathbf{D}^* son diagonales y con el mismo rango, necesariamente tienen el mismo número de ceros. Pasemos a las componentes positivas:

Sean p y p^* el número de componentes *positivas* en la diagonal de \mathbf{D} y de \mathbf{D}^* respectivamente. Por tanto $q = n - p$ y $q^* = n - p^*$ son el número de componentes menores o iguales a cero en las respectivas diagonales. Veamos que $p = p^*$.

Definamos el vector $\mathbf{p} = (i \mid d_{ii} > 0)$ con la lista de índices de las p componentes positivas en \mathbf{D} ; y el vector $\mathbf{p}^* = (i \mid d_{ii}^* > 0)$ con la correspondiente lista de las p^* componentes positivas en \mathbf{D}^* . Entonces

$$\mathbf{x}_{(\mathbf{p} \mid \mathbf{D}_{|\mathbf{p}})}\mathbf{x} = \sum d_{ii}(\mathbf{x}_{|i})^2 > 0; \quad \text{para todo } \mathbf{x} \in \mathbb{R}^p \text{ tal que } \mathbf{x} \neq \mathbf{0}.$$

De manera similar, definamos $\mathbf{q} = (i \mid d_{ii} \leq 0)$ con la lista de índices de las q componentes negativas o nulas en la diagonal de \mathbf{D} ; y $\mathbf{q}^* = (i \mid d_{ii}^* \leq 0)$ con la correspondiente lista para las q^* componentes negativas o nulas en la diagonal de \mathbf{D}^* . Entonces

$$\mathbf{y}_{(\mathbf{q}^* \mid \mathbf{D}_{|\mathbf{q}^*}^*)}\mathbf{y} = \sum d_{ii}^*(\mathbf{y}_{|i})^2 \leq 0; \quad \text{para todo } \mathbf{y} \in \mathbb{R}^{q^*} \text{ tal que } \mathbf{y} \neq \mathbf{0}.$$

Ahora veamos que necesariamente

$$\mathcal{C}(\mathbf{B}_{|\mathbf{p}}) \cap \mathcal{C}(\mathbf{C}_{|\mathbf{q}^*}) = \{\mathbf{0}\}.$$

Para probarlo supondremos que $\mathbf{v} \neq \mathbf{0}$ pertenece a dicha intersección de subespacios y llegaremos a la contradicción de que $\mathbf{v}\mathbf{A}\mathbf{v}$ es simultáneamente mayor que cero y también menor o igual a cero. Por pertenecer a $\mathcal{C}(\mathbf{B}_{|p})$ y ser \mathbf{B} invertible, el vector \mathbf{v} es de la forma $\mathbf{v} = (\mathbf{B}_{|p})\mathbf{x}$ para algún $\mathbf{x} \in \mathbb{R}^p$ no nulo. Así pues,

$$\mathbf{v}\mathbf{A}\mathbf{v} = \mathbf{x}_{(p|}(\mathbf{B}^\top))\mathbf{A}(\mathbf{B}_{|p})\mathbf{x} = \mathbf{x}_{(p|}(\mathbf{B}^\top)\mathbf{A}\mathbf{B}_{|p})\mathbf{x} = \mathbf{x}_{(p|}\mathbf{D}_{|p})\mathbf{x} > 0.$$

Análogamente, por pertenecer a $\mathcal{C}(\mathbf{C}_{|q^*})$ y ser \mathbf{C} invertible, \mathbf{v} es de la forma $\mathbf{v} = (\mathbf{C}_{|q^*})\mathbf{y}$ para algún $\mathbf{y} \in \mathbb{R}^{q^*}$ no nulo. Por tanto

$$\mathbf{v}\mathbf{A}\mathbf{v} = \mathbf{y}_{(q^*|}(\mathbf{C}^\top))\mathbf{A}(\mathbf{C}_{|q^*})\mathbf{y} = \mathbf{y}_{(q^*|}(\mathbf{C}^\top)\mathbf{A}\mathbf{C}_{|q^*})\mathbf{y} = \mathbf{y}_{(q^*|}\mathbf{D}_{|q^*}^*)\mathbf{y} \leq 0.$$

Así, como ambos son subespacios de \mathbb{R}^n con intersección $\{\mathbf{0}\}$, tenemos que $\mathcal{C}(\mathbf{B}_{|p}) + \mathcal{C}(\mathbf{C}_{|q^*}) \subset \mathbb{R}^n$; consecuentemente (Corolario 10.1.3 en la página 122)


$$n = \dim \mathbb{R}^n \geq \dim \left(\mathcal{C}(\mathbf{B}_{|p}) + \mathcal{C}(\mathbf{C}_{|q^*}) \right) = \dim \mathcal{C}(\mathbf{B}_{|p}) + \dim \mathcal{C}(\mathbf{C}_{|q^*}) - \dim \left(\mathcal{C}(\mathbf{B}_{|p}) \cap \mathcal{C}(\mathbf{C}_{|q^*}) \right) = p + (n - p^*) + 0,$$

es decir, $n \geq n + (p - p^*)$, y por tanto $\boxed{p \leq p^*}$.

De igual manera, si repetimos el razonamiento pero con las submatrices principales $\mathbf{p}^*| \mathbf{D}_{|p}^*$ y $q| \mathbf{D}_{|q}$ y los subespacios $\mathcal{C}(\mathbf{C}_{|p^*})$ y $\mathcal{C}(\mathbf{B}_{|q})$ llegaremos a la conclusión de que $\boxed{p^* \leq p}$.

Así pues, $p = p^*$, es decir, ambas matrices tienen igual número de componentes positivas y, como tienen el mismo número de ceros en la diagonal, también tienen el mismo número de componentes negativas. \square

18.3. Completar el cuadrado para clasificar las formas cuadráticas

 Ahora usaremos el signo de los pivotes de la matriz diagonalizada por congruencia (y su rango) para decidir el signo de las formas cuadráticas.

Se denomina “completar el cuadrado” a expresar una forma cuadrática $\mathbf{x}\mathbf{A}\mathbf{x}$ como sumas (y/o restas) de términos al cuadrado. Pues bien, la diagonalización por congruencia nos permite encontrar muchas formas distintas de completar el cuadrado. Basta darse cuenta de que dada \mathbf{A} (simétrica), para toda \mathbf{B} invertible tal que $\mathbf{B}^\top \mathbf{A} \mathbf{B}$ es diagonal, y puesto que $(\mathbf{B}^{-1})^\top = (\mathbf{B}^\top)^{-1}$, tenemos que

$$\mathbf{D} = (\mathbf{B}^\top)\mathbf{A}\mathbf{B} \quad \Rightarrow \quad \mathbf{A} = ((\mathbf{B}^{-1})^\top)\mathbf{D}(\mathbf{B}^{-1})$$

y por tanto, denotando con \mathbf{y} al vector $(\mathbf{B}^{-1})\mathbf{x}$, la forma cuadrática se puede expresar como

$$\mathbf{x}\mathbf{A}\mathbf{x} = \mathbf{y}\mathbf{D}\mathbf{y} = \sum d_{jj}(y_j)^2;$$

es decir, como una suma de los cuadrados de los elementos de \mathbf{y} ponderados por los elementos de la diagonal de \mathbf{D} . Así, si todos los pivotes de \mathbf{D} son positivos, la correspondiente forma cuadrática es definida positiva. Por tanto, hemos demostrado la siguiente proposición:

Proposición 18.3.1. *Una matriz real y simétrica de orden n es definida positiva si y sólo si, es congruente con una matriz diagonal cuyos pivotes son todos mayores que cero.*

Corolario 18.3.2. *Si una matriz real y simétrica de orden n es congruente con una matriz diagonal cuyos pivotes son todos mayores que cero, entonces sus autovalores también son mayores que cero.*

Es lo que ocurría con la matriz del Ejemplo 45 en la página 201. Vamos repetir su diagonalización, pero ahora realizando solo transformaciones *Tipo I* (así minimizamos el número de operaciones y además no modificamos su determinante, por lo que el producto de los sucesivos pivotes de \mathbf{D} calcula los sucesivos *menores* de la matriz):

$$\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \xrightarrow{[(\frac{1}{2})_{1+2}]} \begin{bmatrix} 2 & 0 & 0 \\ -1 & \frac{3}{2} & -1 \\ 0 & -1 & 2 \end{bmatrix} \xrightarrow{[(\frac{1}{2})_{1+2}]} \begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{3}{2} & -1 \\ 0 & -1 & 2 \end{bmatrix} \xrightarrow{[(\frac{2}{3})_{2+3}]} \begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{3}{2} & 0 \\ 0 & -1 & \frac{4}{3} \end{bmatrix} \xrightarrow{[(\frac{2}{3})_{2+3}]} \begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{3}{2} & 0 \\ 0 & 0 & \frac{4}{3} \end{bmatrix}.$$

Como

$$\mathbf{B} = \mathbf{I} \xrightarrow{[(\frac{1}{2})_{1+2}]} \xrightarrow{[(\frac{2}{3})_{2+3}]} \mathbf{B}^{-1} = \mathbf{I} \xrightarrow{[(\frac{-2}{3})_{2+3}]} \xrightarrow{[(\frac{-1}{2})_{1+2}]} = \begin{bmatrix} 1 & -1/2 & 0 \\ 0 & 1 & -2/3 \\ 0 & 0 & 1 \end{bmatrix}.$$

Así, la forma cuadrática \mathbf{xAx} se puede re-escribir como suma de cuadrados:

$$\begin{aligned} \mathbf{xAx} &= (x, y, z) \begin{bmatrix} 1 & 0 & 0 \\ -1/2 & 1 & 0 \\ 0 & -2/3 & 1 \end{bmatrix} \begin{bmatrix} 2 & & \\ & 3/2 & \\ & & 4/3 \end{bmatrix} \begin{bmatrix} 1 & -1/2 & 0 \\ 0 & 1 & -2/3 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \mathbf{x}((\mathbf{B}^{-1})^T) \mathbf{D}(\mathbf{B}^{-1}) \mathbf{x} \\ &= \left((x - \frac{1}{2}y), (y - \frac{2}{3}z), z \right) \begin{bmatrix} 2 & & \\ & 3/2 & \\ & & 4/3 \end{bmatrix} \begin{pmatrix} (x - \frac{1}{2}y) \\ (y - \frac{2}{3}z) \\ z \end{pmatrix} = (\mathbf{B}^{-1}\mathbf{x}) \mathbf{D}(\mathbf{B}^{-1}\mathbf{x}) \\ &= 2(x - \frac{1}{2}y)^2 + \frac{3}{2}(y - \frac{2}{3}z)^2 + \frac{4}{3}z^2 = \sum ({}_j\mathbf{D}_{|j})(w_j)^2. \end{aligned}$$

donde $\mathbf{w} = (\mathbf{B}^{-1})\mathbf{x}$. Así, $\mathbf{xAx} = \mathbf{wDw}$ es suma de tres términos al cuadrado (los componentes de \mathbf{w}) multiplicados por los correspondientes pivotes de \mathbf{D} . Como dichos pivotes son positivos, la forma cuadrática es definida positiva; algo que podemos verificar calculando los autovalores (las raíces de $|\mathbf{A} - \lambda\mathbf{I}| = 0$):

$$\begin{vmatrix} 2 - \lambda & -1 & 0 \\ -1 & 2 - \lambda & -1 \\ 0 & -1 & 2 - \lambda \end{vmatrix} = (2 - \lambda)^3 - 2(2 - \lambda) = 0 \implies \begin{cases} \lambda = 2 \\ (2 - \lambda)^2 - 2 = \lambda^2 - 4\lambda + 2 \end{cases} \begin{cases} \lambda = 2 + \sqrt{2} \\ \lambda = 2 - \sqrt{2} \end{cases}.$$

Por el contrario, la matriz diagonal del Ejemplo 46 en la página 202 tiene dos pivotes negativos y uno positivo, así que la correspondiente forma cuadrática es indefinida (sus autovalores son -1 (doble) y 2).

Apéndice a la lección

18.A. Factorización de Cholesky de una matriz definida positiva

De los Teoremas 4.2.1 y 5.1.16 y del Corolario 5.1.17 se deduce el siguiente

Corolario 18.A.1. *Toda matriz \mathbf{A} se puede factorizar en $\mathbf{A} = \mathbf{KU}$ donde \mathbf{K} es pre-escalónada y \mathbf{U} es triangular superior unitaria.*

Demostración. Sea $\mathbf{K} = \mathbf{A}(\mathbf{I}_{\tau_1 \dots \tau_k})$ una forma pre-escalónada obtenida mediante eliminación “de izquierda a derecha”. Por el Teorema 4.2.1 en la página 51 necesariamente $\mathbf{I}_{\tau_1 \dots \tau_k}$ es triangular superior unitaria. Si denotamos por \mathbf{U} a $(\mathbf{I}_{\tau_1 \dots \tau_k})^{-1}$, que también es triangular superior unitaria, entonces $\mathbf{A} = \mathbf{KU}$. \square

Más interesante es el caso de las matrices cuyas submatrices principales superiores son de rango completo.

Proposición 18.A.2 (Factorización LU). *Sea \mathbf{A} de orden n . Si para $p = 1 : (n - 1)$ las submatrices ${}_{(1:p)}\mathbf{A}_{|(1:p)}$ son invertibles, entonces \mathbf{A} se puede factorizar en la forma $\mathbf{A} = \mathbf{LU}$, donde \mathbf{L} es triangular inferior y \mathbf{U} es triangular superior unitaria. Si \mathbf{A} es invertible, esta factorización es única.*

Demostración. Por el anterior corolario sabemos que podemos factorizar \mathbf{A} en \mathbf{KU} donde $\mathbf{K} = \mathbf{A}(\mathbf{I}_{\tau_1 \dots \tau_h})$ es una forma pre-escalada obtenida mediante eliminación “de izquierda a derecha” y \mathbf{U} (triangular superior unitaria) es la inversa de $\mathbf{I}_{\tau_1 \dots \tau_h}$. Ahora necesitamos demostrar que *si las submatrices ${}_{(1:p)}\mathbf{A}_{|(1:p)}$ son invertibles entonces los p primeros componentes de la diagonal de \mathbf{K} son pivotes* (i.e., no son nulos); y por tanto \mathbf{K} es escalonada (i.e., triangular inferior).

Para la matriz principal superior de orden 1 es trivial: como ${}_{(1,)}\mathbf{A}_{|(1,)}$ ya está escalonada, entonces ${}_{(1,)}\mathbf{K}_{|(1,)} = {}_{(1,)}\mathbf{A}_{|(1,)}$; y como además es invertible, su único componente, ${}_{11}\mathbf{A}_{|1}$, es distinto de cero.

Vamos a demostrar que si la forma pre-escalada de la matriz principal superior de orden p , obtenida por eliminación “de izquierda a derecha”, es triangular inferior y sin ceros en la diagonal, entonces la forma pre-escalada obtenida por eliminación “de izquierda a derecha” de la matriz principal superior de orden $p+1$ también es triangular inferior, y tampoco tiene ceros en la diagonal si es de orden menor o igual a $n-1$.

Sea ${}_{(1:p)}\mathbf{A}_{|(1:p)}^{\tau_1 \dots \tau_k}$ la forma pre-escalada de ${}_{(1:p)}\mathbf{A}_{|(1:p)}$ obtenida por eliminación de “izquierda a derecha” y que, por hipótesis, tiene p pivotes en la diagonal. Si aplicamos la misma secuencia de transformaciones elementales sobre la siguiente submatriz principal superior, la de orden $(p+1)$, entonces ${}_{(1:p+1)}\mathbf{A}_{|(1:p+1)}^{\tau_1 \dots \tau_k}$ tendrá exactamente los mismos p pivotes en la diagonal (pues las p primeras componentes de las p primeras filas de ${}_{(1:p+1)}\mathbf{A}_{|(1:p+1)}$ coinciden con las de ${}_{(1:p)}\mathbf{A}_{|(1:p)}$). Por tanto, podemos usar esos p pivotes para, mediante eliminación de “izquierda a derecha”, anular las p primeras componentes de la última columna, para obtener una matriz escalonada (i.e., triangular inferior).

Para finalizar la primera parte de la demostración: tras el proceso de eliminación de “izquierda a derecha” por el que hemos obtenido una matriz escalonada ¿puede ser cero la última componente de la última columna? Nunca en el caso de las $(n-1)$ primeras submatrices principales superiores, puesto que son de rango completo (y tampoco al aplicar la eliminación sobre toda la matriz \mathbf{A} si ésta es invertible).

En cuanto a la unicidad. Si $\mathbf{A} = \mathbf{L}_1\mathbf{U}_1$ y $\mathbf{A} = \mathbf{L}_2\mathbf{U}_2$ son dos factorizaciones LU de una matriz \mathbf{A} de rango completo, entonces, multiplicando por \mathbf{L}_2^{-1} por la izquierda y \mathbf{U}_1^{-1} por la derecha tenemos que $(\mathbf{L}_2^{-1})\mathbf{L}_1 = \mathbf{U}_2(\mathbf{U}_1^{-1})$. Puesto que el producto de dos matrices triangulares superiores unitarias también es triangular superior unitario (Corolario 3.A.4) y el producto de dos matrices triangulares inferiores también es triangular inferior, necesariamente las matrices a ambos lados de la igualdad anterior son la matriz identidad (es la única matriz cuadrada que simultáneamente es triangular inferior y superior unitaria). Por tanto $\mathbf{L}_1 = \mathbf{L}_2$ y $\mathbf{U}_1 = \mathbf{U}_2$. \square

EJERCICIO 95. Demuestre el siguiente

Corolario 18.A.3 (Factorización LDU). *Sea \mathbf{A} de orden n . Tal que todas sus submatrices principales superiores son invertibles, entonces existe una única matriz triangular inferior unitaria \mathbf{L} , una única matriz triangular superior unitaria \mathbf{U} y una única matriz diagonal \mathbf{D} tales que $\mathbf{A} = \mathbf{LDU}$.*

La factorización LDU es incluso más simple cuando, además, la matriz es simétrica.

EJERCICIO 96. Demuestre el siguiente

Corolario 18.A.4. $\mathbf{L} = \mathbf{U}^\top$ cuando $\mathbf{A} = \mathbf{LDU}$ es la factorización LDU de una matriz simétrica e invertible.

Ahora ya estamos en disposición de presentar la Factorización de Cholesky.

Teorema 18.A.5 (Factorización de Cholesky). *Si \mathbf{A} de orden n es simétrica y definida positiva, entonces existe una única matriz triangular superior \mathbf{C} con componentes positivas en la diagonal tal que $\mathbf{A} = \mathbf{C}(\mathbf{C}^\top)$.*

Demostración. Por el corolario anterior, existe una matriz triangular inferior unitaria \mathbf{L} y una matriz diagonal \mathbf{D} tal que $\mathbf{A} = \mathbf{LD}(\mathbf{L}^\top)$. Puesto que las componentes de la diagonal de \mathbf{D} son positivas, podemos definir la matriz diagonal \mathbf{D}^* cuyas componentes son la raíz cuadrada de las componentes de \mathbf{D} , es decir, ${}_{ii}\mathbf{D}^*_{|i} =$

$\sqrt{d_i}$. Entonces $\mathbf{C} = \mathbf{L}\mathbf{D}^*$ es una matriz real y triangular inferior con componentes positivas en la diagonal. Además satisface $\mathbf{A} = \mathbf{C}(\mathbf{C}^\top)$. La unicidad proviene de la unicidad de la factorización LDU. \square

18.B. Productos escalares en \mathbb{R}^n

Considere la siguiente función $\langle _, _ \rangle_{\mathbf{A}} : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}$, donde \mathbf{A} es definida positiva.

$$(\mathbf{x}, \mathbf{y}) \longmapsto \mathbf{x}\mathbf{A}\mathbf{y}$$

EJERCICIO 97. Demuestre la anterior función es un producto escalar en \mathbb{R}^n .

Pista. Debe verificar los cuatro axiomas indicados en la Sección 11.3.1 en la página 144, que son los mismos el EJERCICIO 11 en la página 25.

La relación entre productos escalares en \mathbb{R}^n y matrices definidas positivas es mucho más estrecha de lo que se pueda pensar...

Sea \mathcal{V} un subespacio de \mathbb{R}^m de dimensión n y sea \mathbf{B} (m por n) tal que sus columnas son una base de \mathcal{V} . Considere dos vectores \mathbf{x} e \mathbf{y} de \mathcal{V} , que por tanto son de la forma $\mathbf{x} = \mathbf{B}\mathbf{a}$ e $\mathbf{y} = \mathbf{B}\mathbf{c}$. Entonces su producto escalar se puede expresar como

$$\mathbf{x} \cdot \mathbf{z} = \mathbf{B}\mathbf{a} \cdot \mathbf{B}\mathbf{c} = \mathbf{a}(\mathbf{B}^\top)\mathbf{B}\mathbf{c} = \mathbf{a}\mathbf{A}\mathbf{c},$$

donde $\mathbf{A} = (\mathbf{B}^\top)\mathbf{B}$ es definida positiva (Proposición 18.1.1). Por tanto el producto punto entre \mathbf{x} e \mathbf{y} se puede expresar con un producto escalar entre las coordenadas de \mathbf{x} e \mathbf{y} respecto a la base formada por las columnas de \mathbf{B} .

Parte VII

Estadística y econometría

“por hacer” (en estado muy preliminar)

Visión geométrica de la Estadística

Una característica esencial de la estadística y la probabilidad es que los subespacios en los que se opera siempre incluyen el vector constante “uno”¹, y que los productos escalares que se emplean (Sección 11.3.1) son tales que *la norma de dicho vector constante “uno” es 1*.²

Fíjese que esto no ocurre con el producto punto de vectores de \mathbb{R}^m , pues $\mathbf{1}_{|i} = 1$ para $i = 1 : m$, y entonces:

$$\|\mathbf{1}\|^2 = \langle \mathbf{1}, \mathbf{1} \rangle = \mathbf{1} \cdot \mathbf{1} = \sum_{i=1}^m (\mathbf{1}_{|i})(\mathbf{1}_{|i}) = \sum_{i=1}^m 1 = m \quad \Rightarrow \quad \|\mathbf{1}\| = \sqrt{m}.$$

19.1. El producto escalar más habitual en estadística descriptiva

Cuando en estadística descriptiva se trabaja con listas de m datos numéricos se está operando en \mathbb{R}^m . Consecuentemente es necesario un *producto escalar* tal que *el vector constante $\mathbf{1} \in \mathbb{R}^m$ tenga longitud 1*. El producto escalar más empleado en estadística cuando se trabaja con muestras de tamaño m es³

$$\langle \mathbf{x}, \mathbf{y} \rangle_s = \frac{1}{m} (\mathbf{x} \cdot \mathbf{y}) \quad (19.1)$$

(Fíjese que pongo un subíndice “s” para indicar que nos referimos al producto escalar usado en estadística).

Con este producto escalar el cuadrado de la norma de un vector \mathbf{x} es

$$\|\mathbf{x}\|^2 = \frac{1}{m} (\mathbf{x} \cdot \mathbf{x});$$

que arroja el resultado deseado, pues independientemente del número m de componentes del vector $\mathbf{1}$ tenemos:

$$\|\mathbf{1}\|^2 = \frac{1}{m} (\mathbf{1} \cdot \mathbf{1}) = \frac{1}{m} \sum_{i=1}^m 1 = \frac{m}{m} = 1; \quad \Rightarrow \quad \|\mathbf{1}\| = \sqrt{1} = 1. \quad (19.2)$$

En esta lección, si no se indica lo contrario, al referirnos a “la norma de un vector \mathbf{x} ”, siempre será la norma correspondiente al producto escalar $\langle _, _ \rangle_s$ de la Ecuación 19.1.

Fíjese además en que si dos vectores de \mathbb{R}^m son perpendiculares con el producto punto, entonces también son perpendiculares con el producto escalar, $\langle _, _ \rangle_s$

$$\mathbf{x} \cdot \mathbf{y} = 0 \quad \Leftrightarrow \quad \frac{1}{m} (\mathbf{x} \cdot \mathbf{y}) = 0.$$

¹en \mathbb{R}^m es el vector $\mathbf{1}$, es decir, aquel cuyas componentes son todas iguales a 1; en el espacio de las variables aleatorias \mathbb{R}^Ω , es la variable aleatoria constante 1, es decir, aquella que para todo suceso elemental $w \in \Omega$ toma el valor 1.

²En probabilidad el producto escalar es la esperanza del producto de variables aleatorias, que asegura que cualquier variable aleatoria igual a la variable aleatoria constante uno (salvo en un conjunto de sucesos de probabilidad cero) tiene un valor esperado igual a uno.

³(véase el *Ejemplo 28 en la página 144*)

19.2. La media aritmética

Hay dos formas de ver la media aritmética.

La media aritmética, μ_y , es el producto escalar de y con el vector constante $\mathbf{1}$

La *media aritmética*, μ_y , del vector y de \mathbb{R}^m es el producto escalar, $\langle y, \mathbf{1} \rangle_s$:

$$\mu_y = \langle y, \mathbf{1} \rangle_s = \frac{1}{m}(\mathbf{1} \cdot y), \quad \text{es decir,} \quad \mu_y = \frac{1}{m} \sum_i y_i. \quad (19.3)$$

La media aritmética, μ_y , es el valor por el que hay que multiplicar el vector $\mathbf{1}$ para lograr *el vector constante más próximo a y*

Denotaremos con μ_y a la *proyección ortogonal de y sobre la recta de vectores constantes $\mathcal{L}([\mathbf{1};])$* ; que, consecuentemente, satisface dos condiciones:

- μ_y es un múltiplo de $\mathbf{1}$, es decir, es de forma $\mu_y = \hat{a}\mathbf{1}$,
- la diferencia $(y - \mu_y)$ es perpendicular a $\mathbf{1}$, es decir, $\frac{1}{m}(y - \mu_y) \cdot \mathbf{1} = 0$.

Combinando ambas condiciones tenemos que

$$\mu_y = \hat{a}\mathbf{1} \iff \frac{1}{m}(y - \hat{a}\mathbf{1}) \cdot \mathbf{1} = 0 \iff \frac{1}{m}(y \cdot \mathbf{1}) - \frac{1}{m}(\mathbf{1} \cdot \mathbf{1})\hat{a} = 0; \iff \frac{1}{m}(y \cdot \mathbf{1}) - \hat{a} = 0,$$

pues $\frac{1}{m}(\mathbf{1} \cdot \mathbf{1}) = \|\mathbf{1}\| = 1$. Concluimos que $\hat{a} = \frac{1}{m}(y \cdot \mathbf{1}) = \mu_y$ (por 19.3). Es decir, la media aritmética μ_y es el valor por el que hay que multiplicar el vector $\mathbf{1}$ para lograr *el vector constante más próximo a y* .

Llamaremos *vector de medias* de y a dicha proyección: $\mu_y = \mathbf{1}(\mu_y) = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \mu_y = \begin{pmatrix} \mu_y \\ \mu_y \\ \vdots \\ \mu_y \end{pmatrix}$.

Como el vector de medias μ_y es el resultado de multiplicar el *vector unitario $\mathbf{1}$* por el escalar μ_y , concluimos que *el valor absoluto de la media aritmética, μ_y , es la longitud de la proyección μ_y* .

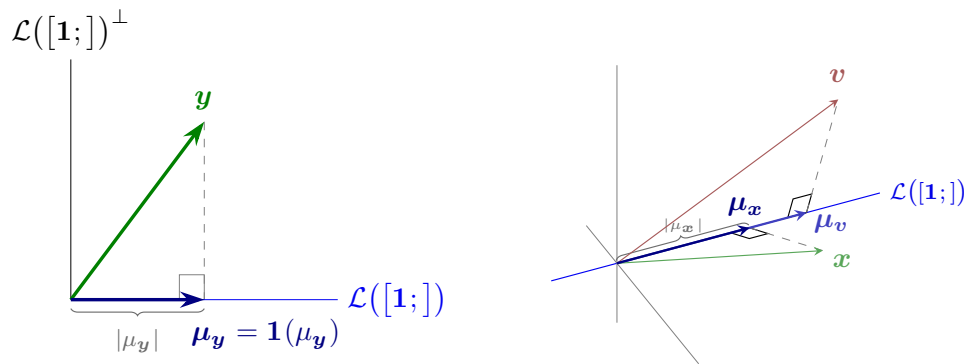


Figura 19.1: Representaciones esquemáticas de la proyección de y sobre el subespacio de vectores constantes (izquierda); y de la proyección de los vectores v y x sobre el subespacio de vectores constantes (derecha).

Además, como las proyecciones ortogonales sobre el subespacio de vectores constantes $\mathcal{L}([\mathbf{1};])$ tanto del vector y como del vector de medias μ_y son en ambos casos $\mu_y \mathbf{1}$, concluimos que:

Proposición 19.2.1. Las medias del vector \mathbf{y} y del vector $\mu_{\mathbf{y}}$ son iguales.

Consecuentemente $\frac{1}{m}(\mathbf{1} \cdot \mathbf{y}) = \mu_{\mathbf{y}} = \frac{1}{m}(\mathbf{1} \cdot \mu_{\mathbf{y}})$, así que necesariamente

Proposición 19.2.2. Los productos punto $(\mathbf{1} \cdot \mu_{\mathbf{y}})$ y $(\mathbf{1} \cdot \mathbf{y})$ son iguales.

Y como $(\mathbf{y} - \mu_{\mathbf{y}})$ es perpendicular a $\mathbf{1}$, por ser $\mu_{\mathbf{y}}$ la proyección ortogonal de \mathbf{y} sobre $\mathcal{L}([\mathbf{1};])$, tenemos que

$$\frac{1}{m}(\mathbf{1} \cdot (\mathbf{y} - \mu_{\mathbf{y}})) = \frac{1}{m} \sum_i (y_i - \mu_{\mathbf{y}}) = 0;$$

así, denominando a $(\mathbf{y} - \mu_{\mathbf{y}})$ como el vector de *desviaciones de \mathbf{y} respecto a su valor medio*, hemos llegado al conocido resultado de que “la suma de las desviaciones de \mathbf{y} respecto a su valor medio es cero”.

19.3. La desviación típica

La desviación típica $\sigma_{\mathbf{y}}$ es la longitud de la componente de \mathbf{y} ortogonal a $\mu_{\mathbf{y}}$

Sabemos que la *longitud* de $\mu_{\mathbf{y}}$ (la proyección de \mathbf{y} sobre los múltiplos de $\mathbf{1}$) es el *valor absoluto de la media aritmética* de \mathbf{y}

$$\|\mu_{\mathbf{y}}\| = |\mu_{\mathbf{y}}|.$$

Pues bien, la *longitud* de la componente de \mathbf{y} ortogonal a $\mathbf{1}$ (el vector $(\mathbf{y} - \mu_{\mathbf{y}})$ de la Figura 19.2) se denomina *desviación típica* de \mathbf{y} ; que denotaremos con $\sigma_{\mathbf{y}}$:

$$\|\mathbf{y} - \mu_{\mathbf{y}}\| = \sigma_{\mathbf{y}}.$$

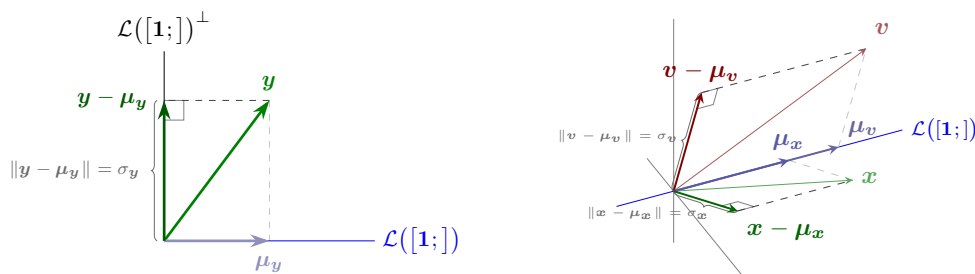


Figura 19.2: Representaciones esquemáticas de la proyección de \mathbf{y} sobre el subespacio ortogonal a los vectores constantes (izquierda); y de la proyección de los vectores \mathbf{v} y \mathbf{x} sobre el subespacio ortogonal a los vectores constantes (derecha).

EJERCICIO 98. Demuestre la siguiente

Proposición 19.3.1. Sumar a \mathbf{y} un vector constante (no nulo) cambia la media pero no la desviación típica.

Pista. Véase la Figura 19.3.

EJERCICIO 99.

- (a) Demuestre que \mathbf{y} tiene media cero si y solo si es perpendicular al vector constante $\mathbf{1}$.
- (b) Demuestre la siguiente

Proposición 19.3.2. Sumar a \mathbf{y} un vector (no nulo) de media cero no cambia la media, aunque si puede cambiar la desviación típica.

Pista. Véase la Figura 19.4.

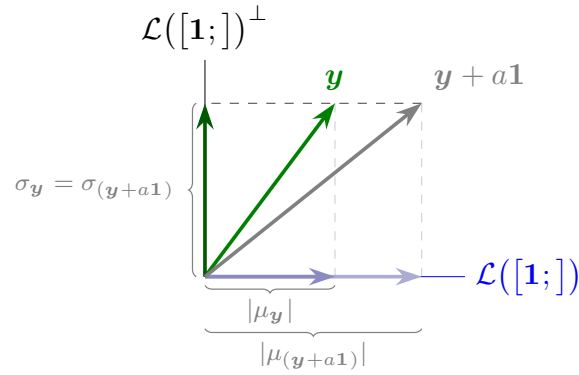


Figura 19.3: Sumar un vector constante no nulo, $a\mathbf{1}$, cambia la media pero no la desviación típica.

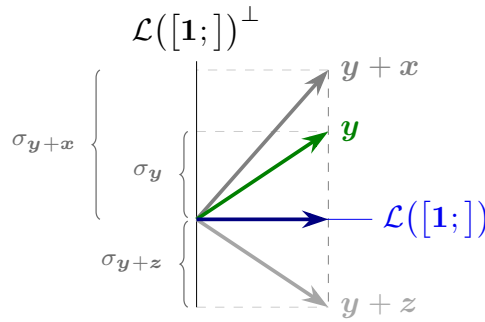


Figura 19.4: Sumar un vector de media nula (perpendicular a $\mathbf{1}$) no cambia la media pero puede cambiar la desviación típica. En la figura $\sigma_{\mathbf{y}} \neq \sigma_{\mathbf{y}+\mathbf{x}}$, pero sin embargo $\sigma_{\mathbf{y}} = \sigma_{\mathbf{y}+\mathbf{z}}$.

19.4. La varianza $\sigma_{\mathbf{y}}^2$ y el Teorema de Pitágoras

El cuadrado de la *desviación típica*, $\sigma_{\mathbf{y}}^2$ (el cuadrado de la longitud del vector $\mathbf{y} - \mu_{\mathbf{y}}$) se llama *varianza*.

$$\sigma_{\mathbf{y}}^2 = \|\mathbf{y} - \mu_{\mathbf{y}}\|^2 = \frac{1}{m}(\mathbf{y} - \mu_{\mathbf{y}}) \cdot (\mathbf{y} - \mu_{\mathbf{y}}), \quad \text{es decir,} \quad \sigma_{\mathbf{y}}^2 = \frac{1}{m} \sum_i (y_i - \mu_{\mathbf{y}})^2.$$

Como consecuencia de la [Proposición 19.3.1](#), la varianza no cambia al sumar un vector constante.

Puesto que $\mu_{\mathbf{y}}$ es perpendicular a $(\mathbf{y} - \mu_{\mathbf{y}})$, por el Teorema de Pitágoras, sabemos que $\|\mathbf{y}\|^2 = \|\mu_{\mathbf{y}}\|^2 + \|\mathbf{y} - \mu_{\mathbf{y}}\|^2$; así que despejando $\|\mathbf{y} - \mu_{\mathbf{y}}\|^2$ obtenemos otra expresión para la varianza de \mathbf{y}

$$\sigma_{\mathbf{y}}^2 = \|\mathbf{y} - \mu_{\mathbf{y}}\|^2 = \|\mathbf{y}\|^2 - \|\mu_{\mathbf{y}}\|^2 = \frac{1}{m}(\mathbf{y} \cdot \mathbf{y}) - \mu_{\mathbf{y}}^2, \quad \text{es decir,} \quad \sigma_{\mathbf{y}}^2 = \frac{\sum_i y_i^2}{m} - \mu_{\mathbf{y}}^2.$$

EJERCICIO 100. Demuestre la siguiente

Proposición 19.4.1. Sea \mathbf{z} un vector perpendicular a $\mathbf{1}$, y sea \mathbf{x} otro vector de \mathbb{R}^m , entonces, para todo $a \in \mathbb{R}$ se verifica que $\mathbf{z} \cdot (\mathbf{x} - a\mathbf{1}) = \mathbf{z} \cdot \mathbf{x}$.

La anterior propiedad de los vectores ortogonales a $\mathbf{1}$ nos permite obtener otra expresión más para la varianza. Como $(\mathbf{y} - \mu_{\mathbf{y}})$ es perpendicular a $\mathbf{1}$:

$$\sigma_{\mathbf{y}}^2 = \frac{1}{m}(\mathbf{y} - \mu_{\mathbf{y}}) \cdot (\mathbf{y} - \mu_{\mathbf{y}}) = \frac{1}{m}(\mathbf{y} - \mu_{\mathbf{y}}) \cdot \mathbf{y}.$$

19.5. Covarianza y correlación

La *covarianza* entre \mathbf{x} e \mathbf{y} es el producto escalar entre las proyecciones $(\mathbf{x} - \mu_{\mathbf{x}})$ y $(\mathbf{y} - \mu_{\mathbf{y}})$:

$$\sigma_{\mathbf{x}\mathbf{y}} = \frac{1}{m}(\mathbf{x} - \mu_{\mathbf{x}}) \cdot (\mathbf{y} - \mu_{\mathbf{y}});$$

que por la Proposición 19.4.1, podemos escribir como $\sigma_{\mathbf{x}\mathbf{y}} = \frac{1}{m}\mathbf{x} \cdot (\mathbf{y} - \mu_{\mathbf{y}})$ ó como $\sigma_{\mathbf{x}\mathbf{y}} = \frac{1}{m}\mathbf{y} \cdot (\mathbf{x} - \mu_{\mathbf{x}})$.

EJERCICIO 101. Demuestre la siguiente proposición:

Proposición 19.5.1. Sean \mathbf{x} e \mathbf{y} vectores de \mathbb{R}^m , entonces $\sigma_{\mathbf{x}\mathbf{y}} = \frac{1}{m}\mathbf{x} \cdot \mathbf{y} - \mu_{\mathbf{x}}\mu_{\mathbf{y}}$.

Fíjese que la varianza es un caso particular en el que $\mathbf{x} = \mathbf{y}$, es decir, $\sigma_{\mathbf{y}\mathbf{y}} = \sigma_{\mathbf{y}}^2$.

Definición 19.1. Se denomina *coeficiente de correlación lineal* entre \mathbf{x} e \mathbf{y} al coseno del ángulo⁴ formado por las proyecciones $(\mathbf{x} - \mu_{\mathbf{x}})$ y $(\mathbf{y} - \mu_{\mathbf{y}})$:

$$\rho_{\mathbf{x}\mathbf{y}} = \frac{\langle (\mathbf{x} - \mu_{\mathbf{x}}), (\mathbf{y} - \mu_{\mathbf{y}}) \rangle_s}{\|(\mathbf{x} - \mu_{\mathbf{x}})\| \cdot \|(\mathbf{y} - \mu_{\mathbf{y}})\|} = \frac{\frac{1}{m}(\mathbf{x} - \mu_{\mathbf{x}}) \cdot (\mathbf{y} - \mu_{\mathbf{y}})}{\|(\mathbf{x} - \mu_{\mathbf{x}})\| \cdot \|(\mathbf{y} - \mu_{\mathbf{y}})\|} = \frac{\sigma_{\mathbf{x}\mathbf{y}}}{\sigma_{\mathbf{x}}\sigma_{\mathbf{y}}}. \quad (19.4)$$

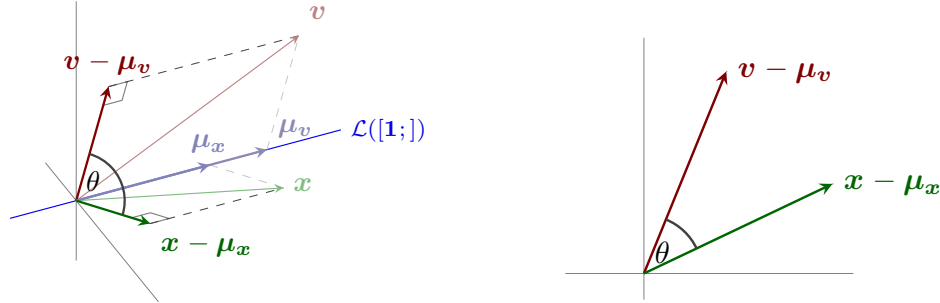


Figura 19.5: Visión desde dos perspectivas distintas de la parte derecha de la Figura 19.2 con la representación del ángulo θ cuyo coseno es la correlación entre los vectores \mathbf{v} y \mathbf{x} .

Como la correlación es el coseno de un ángulo, su valor siempre toma valores entre -1 y 1 (véase las secciones 11.1.3 y 11.3.2). Por tanto, cuando dos vectores tienen correlación cero sus correspondientes componentes perpendiculares a $\mathbf{1}$ son ortogonales entre si. Es más, si dos vectores \mathbf{x} e \mathbf{y} tienen correlación cero ($\rho_{\mathbf{x}\mathbf{y}} = 0$) y uno de ellos tiene media cero, entonces ambos vectores son perpendiculares: $\mathbf{x} \perp \mathbf{y}$.

19.6. Regresión lineal o ajuste MCO

Consideremos ahora la proyección de un vector \mathbf{y} de m datos (m componentes) sobre el espacio columna de una matriz \mathbf{X} de orden m por k , tal que $\mathcal{C}(\mathbf{X})$ contiene el subespacio de vectores constantes $\mathcal{L}([1;])$. En este contexto las k columnas de \mathbf{X} se denominan *regresores* y la matriz \mathbf{X} se llama *matriz de regresores*.

Denotaremos con $\hat{\mathbf{y}}$ a la proyección de $\mathbf{y} \in \mathbb{R}^m$ sobre el espacio columna $\mathcal{C}(\mathbf{X})$ de la matriz de regresores. Por ser la proyección ortogonal de \mathbf{y} sobre $\mathcal{C}(\mathbf{X})$ satisface dos condiciones:

⁴véase la Definición 11.11 en la página 146 de coseno del ángulo entre dos vectores.

- $\hat{\mathbf{y}}$ es combinación lineal de las columnas de la matriz de regresores \mathbf{X} , es decir, $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ con $\hat{\boldsymbol{\beta}} \in \mathbb{R}^k$;
- $(\mathbf{y} - \hat{\mathbf{y}})$ es perpendicular a $\mathcal{C}(\mathbf{X})$, es decir, $\frac{1}{m}(\mathbf{X}^\top)(\mathbf{y} - \hat{\mathbf{y}}) = \mathbf{0}$.

Combinando ambas condiciones tenemos

$$\frac{1}{m}(\mathbf{X}^\top)(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{0} \iff \frac{1}{m}(\mathbf{X}^\top)\mathbf{y} - \frac{1}{m}(\mathbf{X}^\top)\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{0}.$$

Así llegamos a las ecuaciones normales ⁵

$$\frac{1}{m}(\mathbf{X}^\top)\mathbf{X}\hat{\boldsymbol{\beta}} = \frac{1}{m}(\mathbf{X}^\top)\mathbf{y}. \quad (19.5)$$

Asumiremos que las k columnas de \mathbf{X} son linealmente independientes; consecuentemente el vector $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ es “la” combinación lineal de los regresores más próxima a \mathbf{y} ; es decir, es el vector $\hat{\mathbf{y}}$ de $\mathcal{C}(\mathbf{X})$ que minimiza

$$\|\mathbf{y} - \hat{\mathbf{y}}\|^2 = \frac{1}{m}(\mathbf{y} - \hat{\mathbf{y}}) \cdot (\mathbf{y} - \hat{\mathbf{y}}) = \frac{1}{m} \sum_i (y_i - \hat{y}_i)^2; \quad (19.6)$$

donde cada sumando $(y_i - \hat{y}_i)$ se denomina *error de ajuste* de la correspondiente componente i -ésima. Es decir, $\hat{\mathbf{y}}$ minimiza la suma del cuadrado de los errores de ajuste. Por ello, la proyección de \mathbf{y} sobre $\mathcal{C}(\mathbf{X})$ se llama *ajuste por mínimos cuadrados ordinarios* o por sus siglas, *ajuste MCO*.

19.6.1. Tres casos sencillos

El vector constante 1 como único regresor

Este caso ya lo hemos visto: la proyección $\hat{\mathbf{y}}$ del vector \mathbf{y} sobre $\mathcal{L}([1;])$ es el vector de medias $\boldsymbol{\mu}_y$. El siguiente ejercicio le pide que reproduzca este resultado, pero ahora usando las ecuaciones normales.

EJERCICIO 102. Emplee las ecuaciones normales (19.5) para encontrar el vector $\hat{\boldsymbol{\beta}}$ y con él, el vector $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ cuando la matriz de regresores \mathbf{X} únicamente tiene la columna 1.

Regresión lineal simple y la recta de regresión

Veamos el caso particular de la proyección ortogonal de \mathbf{y} sobre el espacio columna de la matriz de regresores $\mathbf{X} = [1; \mathbf{x}]$, es decir, busquemos la combinación lineal de los vectores $\mathbf{1}$ y \mathbf{x} más próxima a \mathbf{y} (donde \mathbf{x} no es un vector constante).

Recordando que $_{il}(\mathbf{X}^\top)\mathbf{X}_{lj} = \mathbf{X}_{li} \cdot \mathbf{X}_{lj}$ y que $_{il}(\mathbf{X}^\top)\mathbf{y} = (\mathbf{X}_{li}) \cdot \mathbf{y}$, y que en este caso $\mathbf{X}_{11} = \mathbf{1}$ y $\mathbf{X}_{12} = \mathbf{x}$, podemos reescribir el sistema de ecuaciones normales (19.5) como

$$\begin{bmatrix} \frac{1}{m}(\mathbf{1} \cdot \mathbf{1}) & \frac{1}{m}(\mathbf{1} \cdot \mathbf{x}) \\ \frac{1}{m}(\mathbf{x} \cdot \mathbf{1}) & \frac{1}{m}(\mathbf{x} \cdot \mathbf{x}) \end{bmatrix} \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} \frac{1}{m}(\mathbf{1} \cdot \mathbf{y}) \\ \frac{1}{m}(\mathbf{x} \cdot \mathbf{y}) \end{pmatrix}; \quad \text{es decir} \quad \begin{bmatrix} 1 & \mu_x \\ \mu_x & \frac{1}{m}(\mathbf{x} \cdot \mathbf{x}) \end{bmatrix} \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} \mu_y \\ \frac{1}{m}(\mathbf{x} \cdot \mathbf{y}) \end{pmatrix};$$

que resolvemos por eliminación (recordando que $\sigma_x^2 = \frac{1}{m}(\mathbf{x} \cdot \mathbf{x}) - \mu_x^2$ y que $\sigma_{xy} = \frac{1}{m}(\mathbf{x} \cdot \mathbf{y}) - \mu_x \mu_y$):

$$\left[\begin{array}{cc|c} 1 & \mu_x & -\mu_y \\ \mu_x & \frac{1}{m}(\mathbf{x} \cdot \mathbf{x}) & -\frac{1}{m}(\mathbf{x} \cdot \mathbf{y}) \\ \hline 1 & 0 & 0 \\ 0 & 1 & 0 \end{array} \right] \xrightarrow{\begin{bmatrix} (-\mu_x)1+2 \\ (-\mu_y)1+3 \end{bmatrix}} \left[\begin{array}{cc|c} 1 & 0 & 0 \\ \mu_x & \sigma_x^2 & -\sigma_{xy} \\ \hline 1 & -\mu_x & \mu_y \\ 0 & 1 & 0 \end{array} \right] \xrightarrow{\left[\left(\frac{\sigma_{xy}}{\sigma_x^2} \right) 2+3 \right]} \left[\begin{array}{cc|c} 1 & 0 & 0 \\ \mu_x & \sigma_x^2 & 0 \\ \hline 1 & -\mu_x & \mu_y - \mu_x \left(\frac{\sigma_{xy}}{\sigma_x^2} \right) \\ 0 & 1 & \left(\frac{\sigma_{xy}}{\sigma_x^2} \right) \end{array} \right].$$

⁵Normalmente se escribe el sistema $(\mathbf{X}^\top)\mathbf{X}\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top)\mathbf{y}$, pero aquí he preferido incluir los factores $\frac{1}{m}$ para recordar que vamos a operar con el producto escalar $\langle _, _ \rangle_s$ habitual en estadística.

Es decir, la combinación lineal de $\mathbf{1}$ y \mathbf{x} más próxima a \mathbf{y} es:

$$\hat{\mathbf{y}} = \underbrace{\left(\mu_{\mathbf{y}} - \mu_{\mathbf{x}} \frac{\sigma_{\mathbf{x}\mathbf{y}}}{\sigma_{\mathbf{x}}^2}\right)}_{\hat{a}} \mathbf{1} + \underbrace{\left(\frac{\sigma_{\mathbf{x}\mathbf{y}}}{\sigma_{\mathbf{x}}^2}\right)}_{\hat{b}} \mathbf{x} = \underbrace{\left(\mu_{\mathbf{y}} - \mu_{\mathbf{x}} \hat{b}\right)}_{\hat{a}} \mathbf{1} + \underbrace{\left(\frac{\sigma_{\mathbf{x}\mathbf{y}}}{\sigma_{\mathbf{x}}^2}\right)}_{\hat{b}} \mathbf{x} \quad (\text{véase la Figura 19.6}).$$

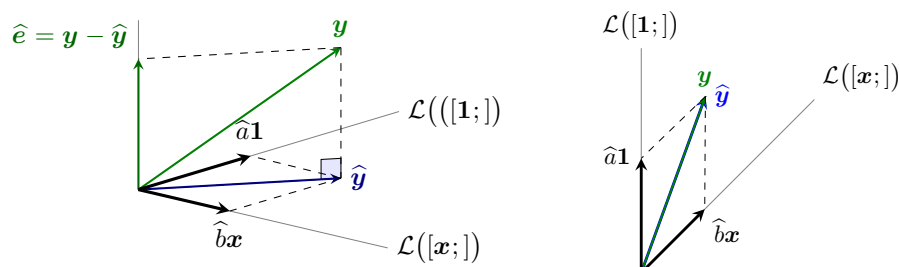


Figura 19.6: Representación esquemática de la proyección sobre el espacio columna de $\mathbf{X} = [\mathbf{1}; \mathbf{x}]$; con $\mathbf{x} \notin \mathcal{L}([\mathbf{1};])$. El “suelo” de la figura representa el subespacio $\mathcal{C}(\mathbf{X}) \subset \mathbb{R}^m$. La figura de la derecha es una visión desde “arriba” (desde esta perspectiva $\hat{\mathbf{y}}$ está detrás de \mathbf{y}).

Recta de regresión. Los textos de estadística y econometría suelen mostrar otra manera alternativa de visualizar el juste de *regresión lineal simple* (aunque carece de la interpretación geométrica que hemos usado).

El *diagrama de dispersión* de \mathbf{x} e \mathbf{y} es la representación del conjunto de pares (x_i, y_i) en el plano XY ; por ejemplo en la Figura 19.7 aparece el diagrama de dispersión de los vectores $\mathbf{x} = (1, 2, 3,)$ e $\mathbf{y} = (1, 2, 2,)$, es decir, la representación de los pares $(1, 1)$, $(2, 2)$ y $(3, 2)$; que son las filas de la matriz $[\mathbf{x}; \mathbf{y}]$.

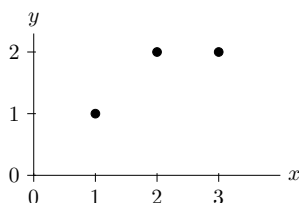


Figura 19.7: Diagrama de dispersión de los vectores $\mathbf{x} = (1, 2, 3,)$ e $\mathbf{y} = (1, 2, 2,)$.

Si \mathbf{z} es una *combinación lineal* de $\mathbf{1}$ y \mathbf{x} (si es de la forma $\alpha \mathbf{1} + \beta \mathbf{x}$) el diagrama de dispersión de \mathbf{x} con \mathbf{z} es la representación de una *función lineal* cuyo dominio son los componentes de \mathbf{x} . Por ello los puntos aparecen alineados sobre la *recta* cuya ordenada en el origen es α y cuya pendiente es β (véase la Figura 19.8).

Así, como $\hat{\mathbf{y}}$ es la *combinación lineal* $\hat{a} \mathbf{1} + \hat{b} \mathbf{x}$, los puntos del correspondiente diagrama de dispersión con \mathbf{x} están alineados sobre la recta cuya ordenada en el origen es \hat{a} y cuya pendiente es \hat{b} . La *recta* a lo largo de la que se alinean los puntos del diagrama de dispersión de \mathbf{x} y $\hat{\mathbf{y}}$ se denomina *recta de regresión*.

Al representar juntos los diagramas de dispersión de \mathbf{x} con \mathbf{y} y de \mathbf{x} con $\hat{\mathbf{y}}$, la diferencia vertical entre los puntos del diagrama de \mathbf{y} y los del diagrama de $\hat{\mathbf{y}}$ sobre cada valor x_i corresponden a cada uno de los errores de ajuste (véase la Figura 19.9). Por ese motivo se dice que la *recta de regresión* es la recta que minimiza la suma del cuadrado de los *errores de ajuste* (véase la Ecuación 19.6 en la página anterior).

EJERCICIO 103. Encuentre la ordenada en el origen y la pendiente de la recta de regresión de la Figura 19.9. Encuentre también $\hat{\mathbf{y}}$ y el vector $\hat{\mathbf{e}}$ de errores del ajuste MCO ($\mathbf{y} - \hat{\mathbf{y}}$).

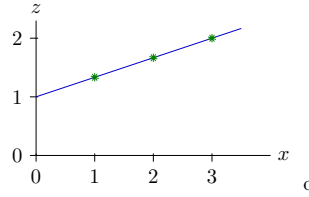


Figura 19.8: Diagrama de dispersión de $\mathbf{x} = (1, 2, 3,)$ y la combinación lineal $\mathbf{z} = \mathbf{1} + \frac{1}{3}\mathbf{x}$. Los puntos se sitúan sobre la recta cuya ordenada en el origen es 1 y cuya pendiente es $\frac{1}{3}$.

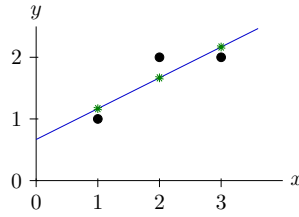


Figura 19.9: Ajuste de la recta de regresión $\hat{\mathbf{y}} = \hat{a}\mathbf{1} + \hat{b}\mathbf{x}$ donde $\mathbf{x} = (1, 2, 3,)$ e $\mathbf{y} = (1, 2, 2,)$.

Regresión lineal con tres regresores (de ellos uno constante). El plano de regresión

El tercer caso sencillo (aunque ya no tanto) corresponde a la proyección ortogonal de \mathbf{y} sobre el espacio columna de la matriz de regresores $\mathbf{X} = [\mathbf{1}; \mathbf{x}; \mathbf{z}]$, es decir, la búsqueda de la combinación lineal de los vectores $\mathbf{1}$, \mathbf{x} y \mathbf{z} más próxima a \mathbf{y} (y donde \mathbf{X} es de rango 3). Como en el caso del modelo lineal simple, podemos reescribir el sistema de ecuaciones normales (19.5) como

$$\begin{bmatrix} 1 & \mu_{\mathbf{x}} & \mu_{\mathbf{z}} \\ \mu_{\mathbf{x}} & \frac{1}{n}(\mathbf{x} \cdot \mathbf{x}) & \frac{1}{n}(\mathbf{x} \cdot \mathbf{z}) \\ \mu_{\mathbf{z}} & \frac{1}{n}(\mathbf{x} \cdot \mathbf{z}) & \frac{1}{n}(\mathbf{z} \cdot \mathbf{z}) \end{bmatrix} \begin{pmatrix} \hat{a} \\ \hat{b} \\ \hat{c} \end{pmatrix} = \begin{pmatrix} \mu_{\mathbf{y}} \\ \frac{1}{n}(\mathbf{x} \cdot \mathbf{y}) \\ \frac{1}{n}(\mathbf{z} \cdot \mathbf{y}) \end{pmatrix}$$

y como antes, lo resolvemos por eliminación simplificando las expresiones de las varianzas y covarianzas a medida que van apareciendo, y obtenemos el vector solución $\hat{\beta} = (\hat{a}, \hat{b}, \hat{c})$:

$$\begin{bmatrix} 1 & \mu_x & \mu_z & -\mu_y \\ \mu_x & \frac{1}{n}(\mathbf{x} \cdot \mathbf{x}) & \frac{1}{n}(\mathbf{x} \cdot \mathbf{z}) & -\frac{1}{n}(\mathbf{x} \cdot \mathbf{y}) \\ \mu_z & \frac{1}{n}(\mathbf{x} \cdot \mathbf{z}) & \frac{1}{n}(\mathbf{z} \cdot \mathbf{z}) & -\frac{1}{n}(\mathbf{z} \cdot \mathbf{y}) \\ 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{\begin{matrix} \tau \\ [(-\mu_x)1+2] \\ [(-\mu_z)1+3] \\ [(\mu_y)1+4] \\ \left[-\frac{\sigma_{xz}}{\sigma_z^2}\right]2+3 \\ \left[\left(\frac{\sigma_{xy}}{\sigma_x^2}\right)2+4\right] \\ \left[\left(\frac{\sigma_x^2\sigma_{zy}-\sigma_{xy}\sigma_{xz}}{\sigma_x^2\sigma_z^2-\sigma_{xz}^2}\right)3+4 \right] \end{matrix}} \begin{bmatrix} 1 & 0 & 0 & 0 \\ \mu_x & \sigma_x^2 & 0 & 0 \\ \mu_z & \sigma_{xz} & \sigma_z^2 - \frac{\sigma_{xz}^2}{\sigma_x^2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{\begin{matrix} -\mu_x\sigma_z^2\sigma_{xy} + \mu_x\sigma_{xz}\sigma_{zy} + \mu_y\sigma_x^2\sigma_z^2 - \mu_y\sigma_{xz}^2 - \mu_z\sigma_x^2\sigma_{zy} + \mu_z\sigma_{xz}\sigma_{xy} \\ \sigma_x^2\sigma_z^2 - \sigma_{xz}^2 \\ \sigma_z^2\sigma_{xy} - \sigma_{xz}\sigma_{zy} \\ \sigma_x^2\sigma_z^2 - \sigma_{xz}^2 \\ \sigma_x^2\sigma_{zy} - \sigma_{xz}\sigma_{xy} \\ \sigma_x^2\sigma_z^2 - \sigma_{xz}^2 \\ 1 \end{matrix}} \begin{bmatrix} 0 \\ 0 \\ 0 \\ \frac{-\mu_x\sigma_z^2\sigma_{xy} + \mu_x\sigma_{xz}\sigma_{zy} + \mu_y\sigma_x^2\sigma_z^2 - \mu_y\sigma_{xz}^2 - \mu_z\sigma_x^2\sigma_{zy} + \mu_z\sigma_{xz}\sigma_{xy}}{\sigma_x^2\sigma_z^2 - \sigma_{xz}^2} \\ \frac{\sigma_z^2\sigma_{xy} - \sigma_{xz}\sigma_{zy}}{\sigma_x^2\sigma_z^2 - \sigma_{xz}^2} \\ \frac{\sigma_x^2\sigma_{zy} - \sigma_{xz}\sigma_{xy}}{\sigma_x^2\sigma_z^2 - \sigma_{xz}^2} \\ 1 \end{bmatrix}.$$

Es decir, la combinación lineal de $\mathbf{1}$, \mathbf{x} y \mathbf{z} más próxima a \mathbf{y} es:

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\beta} = \underbrace{(\mu_y - \mu_x \hat{b} - \mu_z \hat{c})}_{\hat{a}} \mathbf{1} + \underbrace{\left(\frac{\sigma_z^2\sigma_{xy} - \sigma_{xz}\sigma_{zy}}{\sigma_x^2\sigma_z^2 - \sigma_{xz}^2}\right)}_{\hat{b}} \mathbf{x} + \underbrace{\left(\frac{\sigma_x^2\sigma_{zy} - \sigma_{xy}\sigma_{xz}}{\sigma_x^2\sigma_z^2 - \sigma_{xz}^2}\right)}_{\hat{c}} \mathbf{z}.$$

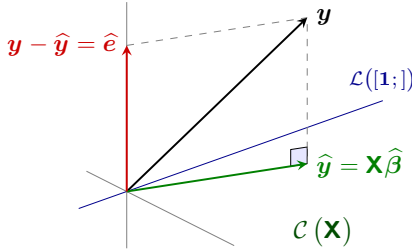


Figura 19.10: Representación esquemática de la proyección de $\mathbf{y} \in \mathbb{R}^m$ sobre el subespacio $\mathcal{C}(\mathbf{X})$ que contienen el subespacio de vectores constantes $\mathcal{L}([1;])$. El “suelo” de la figura representa el subespacio $\mathcal{C}(\mathbf{X})$.

La Figura 19.10 es una representación esquemática de la proyección ortogonal (o ajuste MCO) sobre el subespacio engendrado por las columnas de \mathbf{X} (los regresores). En el caso de tres regresores (cuando uno de ellos es constante) disponemos de una representación alternativa mediante diagramas de dispersión. Pero ahora el diagrama de dispersión de (\mathbf{x}, \mathbf{z}) e \mathbf{y} requiere de una representación tridimensional de los pares $((x_i, z_i), y_i)$, donde sobre cada punto (x_i, z_i) del plano XZ (el “suelo” de la Figura 19.11) se dibuja un punto, cuya altura indica el valor de y_i .

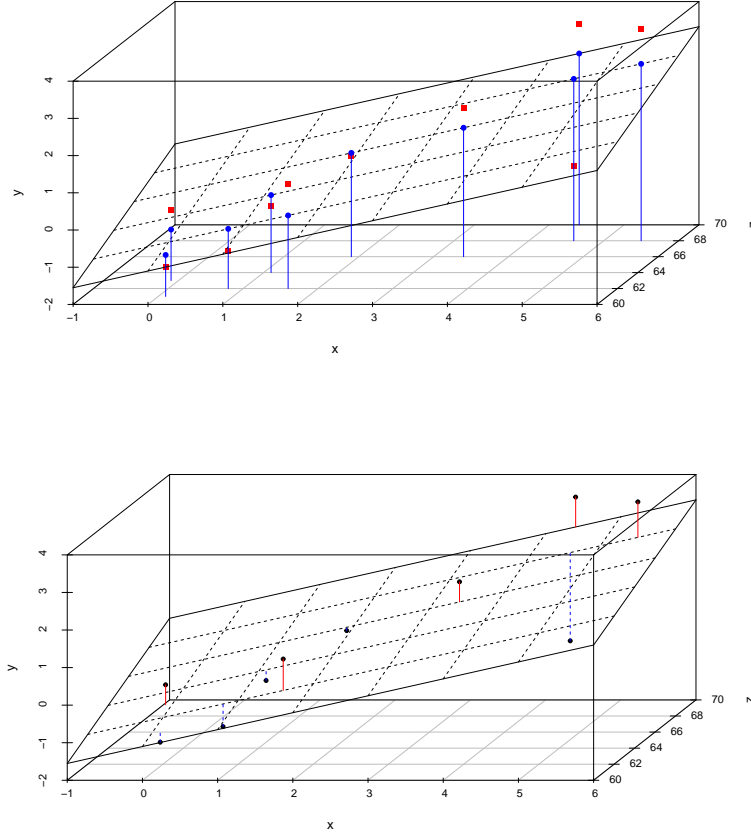


Figura 19.11: Ajuste del plano de regresión $\hat{\mathbf{y}} = \hat{a}\mathbf{1} + \hat{b}\mathbf{x} + \hat{c}\mathbf{z}$ que minimiza los errores de ajuste. **Arriba:** las barras verticales arrancan de los puntos (x_i, z_i) . La altura de los cuadrados rojos corresponde a los correspondientes valores y_i y la altura de los círculos azules (en el extremo superior de las barras) a los valores ajustados \hat{y}_i (dichos círculos están contenidos en el plano de regresión). Las diferencias entre ambas alturas, $y_i - \hat{y}_i$, son los errores de ajuste (los componentes de $\hat{\mathbf{e}} = \mathbf{y} - \hat{\mathbf{y}}$). **Abajo:** Representación de dichos errores de ajuste mediante barras continuas si son positivos, y discontinuas si son negativos.

Mientras que la Figura 19.10 es una representación esquemática del ajuste MCO, la Figura 19.11 es una representación descriptiva, pues muestra la magnitud y el signo de cada uno de los errores cometidos por el *plano de regresión* (exactamente igual que cuando se representa la recta de regresión en la Figura 19.9). Dada la naturaleza del ajuste MCO, cualquier otro plano genera unos errores de ajuste tales que la suma de sus cuadrados será mayor que la suma de los cuadrados de los errores cometidos con este plano, es decir, será mayor que $\sum_i (y_i - \hat{y}_i)^2 = \|\hat{\mathbf{e}}\|^2$.

19.6.2. Caso general

Como se indicó al comienzo de esta Sección 19.6, la regresión lineal (o ajuste MCO) consiste en proyectar un vector \mathbf{y} de \mathbb{R}^m sobre el espacio columna de una matriz de regresores \mathbf{X} tal que $\mathcal{C}(\mathbf{X})$ contiene el conjunto de vectores constantes $\mathcal{L}([\mathbf{1};])$, de hecho se suele asumir que la primera columna de \mathbf{X} es el vector $\mathbf{1}$ (tal como ocurre en los tres casos vistos más arriba). También se asume que los regresores (las columnas de \mathbf{X}) son *linealmente independientes* y por tanto existe un único vector $\hat{\boldsymbol{\beta}}$ tal que $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$, donde $\hat{\mathbf{y}}$ es la proyección


ortogonal de \mathbf{y} sobre $\mathcal{C}(\mathbf{X})$. Dicho vector es *la solución* a las ecuaciones normales.

A medida que crece el número de regresores las expresiones que aparecen al resolver las ecuaciones normales se complican, pero todas ellas comparten una característica común: las expresiones están formadas por las medias, varianzas y covarianzas de \mathbf{y} y de las columnas de \mathbf{X} .

Lamentablemente cuando hay más de tres regresores no disponemos de una representación similar a la recta o el plano de regresión de las figuras 19.9 y 19.11; pero sigue siendo válida la representación esquemática de la Figura 19.10, donde el “suelo” representa el subespacio $\mathcal{C}(\mathbf{X})$ de dimensión k que contiene el subespacio de vectores constantes $\mathcal{L}([1;])$. Esta representación indica la relación correspondiente al Teorema de Pitágoras:

$$\|\mathbf{y}\|^2 = \|\hat{\mathbf{y}}\|^2 + \|\hat{\mathbf{e}}\|^2.$$

19.7. Interpretación estadística

 Hemos visto la media aritmética, la desviación típica y la regresión lineal desde un punto de vista geométrico. Pero lo habitual es dar una interpretación estadística que omita cualquier referencia explícita a la geometría de la proyección ortogonal. . . Veamos una interpretación estadística habitual en las aplicaciones.

19.7.1. El vector de medias como predictor

Imagine un banco con m cajas de seguridad numeradas de 1 a m . En el interior de cada caja hay una pepita de oro. Usted dispone de una lista, \mathbf{y} , con el peso de las m pepitas, pero desconoce en qué caja está contenida cada una de las pepita. Además, dentro de un bombo como los del sorteo de Navidad hay m bolas numeradas de 1 a m .

Se le propone el siguiente juego. Usted tiene que indicar un peso en gramos para cada una de las pepitas; la restricción es que dicho peso ha de ser *el mismo para todas* las pepitas. Después se extrae del bombo una bola al azar, se abre la caja cuyo número corresponde con el de la bola y usted recibe la pepita guardada en ella; pero debe pagar 5 céntimos de euro⁶ por cada miligramo de discrepancia entre el peso de la pepita y la cantidad que usted indicó. ¿Cuál es la estrategia óptima en dicho juego?

Evidentemente nuestra ganancia disminuye con la distancia entre el peso indicado y el peso de la pepita recibida. Y, salvo que todas las pepitas pesen lo mismo, no hay manera de saber con antelación el peso de la pepita que se le va a entregar. Así que debemos considerar la discrepancia entre el peso indicado y el peso de *cada una de las pepitas de la lista*, pues es igualmente esperable que cualquiera de ellas sea entregada.

La estrategia óptima es indicar la media aritmética, μ_y , de los pesos de las pepitas en la lista. Así, el vector de medias, $\boldsymbol{\mu}_y = \mu_y \mathbf{1}$, es interpretado como *predictor óptimo*; pues minimiza el error de predicción (es decir, es el vector constante que minimiza la distancia al vector de verdaderos pesos)

$$\|\mathbf{y} - \boldsymbol{\mu}_y\|.$$

Por ello, la media aritmética μ_y también se interpreta como el valor que *mejor representa* a la muestra.

Extrapolación. Bajo ciertas circunstancias esta interpretación puede ser válida incluso fuera de la muestra. Imagine que todas las pepitas se han extraído de una misma zona de un único yacimiento. Ahora se le propone un nuevo juego: hay otro banco con n cajas de seguridad; cada una de ellas contiene una pepita de oro extraída exactamente del mismo lugar que las m pepitas del otro banco. Además sabe que no ha habido ningún proceso de selección por el que unas pepitas hayan acabado en el primer banco y otras en el segundo.

Se le propone repetir el juego con las n cajas de este segundo banco y un bombo con n bolas numeradas de 1 a n . Pero ahora no se le entrega una nueva lista con los pesos de estas n pepitas; solo dispone de la anterior lista de pesos de las m pepitas del primer banco. ¿Qué estrategia puede seguir?

⁶50 euros por gramo

Si usted considera que, dado que no ha habido ningún proceso de selección, las m pepitas del primer banco representan adecuadamente el tipo de pepitas extraídas en el yacimiento, podría usar como vector constante $\mathbf{z} = \mu_y \mathbf{1} \in \mathbb{R}^n$, donde μ_y es la media del peso de las m pepitas del primer banco. Si es así, está interpretando que la media aritmética de la muestra de las m pepitas del *primer banco* es el valor que permite realizar *buenas predicciones* para los pesos de las pepitas de *ambos bancos*.

Pero las circunstancias podrían ser diferentes. Imagine que recibimos un soplo por el que somos informados de que las pepitas del segundo banco sí han sido seleccionadas, a resultas de lo cual son aproximadamente el doble de grandes que las del primero... entonces podríamos indicar como peso, para cada pepita del nuevo banco, el doble del peso medio calculado con las pepitas del primer banco.

19.7.2. La desviación típica como medida del error de predicción de la media.

Hemos visto que emplear el vector de medias es la estrategia óptima si queremos dar una predicción puntual (un único valor) para el peso de las pepitas. Pero no implica que logremos una “buena” predicción (solo que lo haremos aún peor empleando como predictor otro vector constante distinto del vector de medias).

La desviación típica es por definición la norma del error de predicción cometido con el vector de medias:

$$\sigma_y = \|\mathbf{y} - \mu_y\|.$$

Hay un único caso en el que las predicciones de μ_y son perfectas: cuando todas las pepitas pesan lo mismo, sea cual fuere la pepita entregada, siempre sabremos con antelación su peso exacto. En este caso, como el vector de pesos es constante, coincide con su proyección sobre $\mathcal{L}([\mathbf{1};])$, es decir, $\mathbf{y} = \mu_y$. Consecuentemente la desviación típica es cero.

En cualquier otro caso el error de predicción deja de ser nulo. De hecho, el tamaño (o norma) del error de predicción crece cuando aumenta el número de pepitas cuyo peso es distinto del peso medio y también si aumentan las discrepancias entre los pesos de las pepitas y el peso medio, pues:

$$\sigma_y = \|\mathbf{y} - \mu_y\| = \sqrt{\frac{1}{m} \sum_i (y_i - \mu_y)^2};$$

por eso también se dice que la desviación típica es *una medida de la “variabilidad” de los datos*.

Nótese que la desviación típica crece conforme crece la magnitud de las discrepancias $y_i - \mu_y$; y que además está expresada en las mismas unidades que los datos⁷. Consecuentemente las unidades de medida de los datos afectan a la magnitud de la desviación típica. En el ejemplo de las pepitas, el valor de la desviación típica depende de si el peso está expresado en kilos, en gramos o en miligramos. Así, salvo cuando $\sigma_y = 0$, un valor concreto de la desviación típica no nos da una idea clara de la variabilidad de los datos; pero sí nos sirve para comparar la variabilidad respecto a la de otros datos que estén expresados en las mismas unidades.

19.7.3. Regresión lineal como predictor

19.8. Medida de ajuste en la regresión lineal

Apendices a la lección

19.A. Otros productos escalares usados en estadística descriptiva

En estadística también se emplean las *medias ponderadas* en las que unos datos tienen mayor peso que otros. La única restricción es que las ponderaciones sean positivas y que se cumpla que la media ponderada de un vector de n datos todos iguales a uno es uno. Por ello, los pesos (todos positivos) deben sumar uno. Por ejemplo podemos emplear el siguiente producto escalar⁸

$$\begin{aligned} \langle _, _ \rangle_{sp} : \mathbb{R}^n \times \mathbb{R}^n &\longrightarrow \mathbb{R} \quad , \\ (\mathbf{a}, \mathbf{b}) &\longmapsto \mathbf{a} \mathbf{D} \mathbf{b} \end{aligned}$$

⁷ puesto que las discrepancias se elevan al cuadrado pero luego se aplica la raíz cuadrada a la suma ponderada.

⁸ Véase el EJERCICIO 61 en la página 144

donde \mathbf{D} es una matriz diagonal de orden n cuyos elementos en la diagonal son positivos y suman 1. Fíjese que el producto escalar $\langle _, _ \rangle_s$ corresponde al caso particular $\mathbf{D} = \frac{1}{n} \mathbf{I}$.

Medias ponderadas. Imagine la siguiente variante del primer juego: en la lista de pesos se nos informa sobre el contenido de cada una de las cajas. Además las bolas del bombo con numeración par son más pesadas que las de numeración impar; y por ese motivo la frecuencia con la que se extrae una bola de numeración par es el doble que una bola impar.

En tal caso podemos cambiar de producto escalar y en lugar de ponderar el peso de cada pepita por $\frac{1}{m}$, podemos ponderar el doble los pesos de las pepitas guardadas en cajas pares, pues aparecerán con el doble de frecuencia (véase el Ejemplo 29 en la página 144).

No obstante, una vez hemos establecido un nuevo producto escalar $\langle _, _ \rangle_{sp}$, la media (en este caso ponderada) sigue siendo el producto escalar del vector de pesos \mathbf{y} por el vector constante: $\langle \mathbf{y}, \mathbf{1} \rangle_{sp}$. Ahora al medir los errores de predicción con el nuevo producto escalar ponderaremos el doble los errores producidos en las cajas pares (las que se abren con el doble de frecuencia). Así pues, la media (ponderada) minimiza los errores de predicción (ponderados), por lo que, como antes, la estrategia es indicar la media (ponderada) del peso de las m pepitas.

Consecuentemente, también la media ponderada, $\langle \mathbf{y}, \mathbf{1} \rangle_{sp}$, puede interpretarse como el *mejor predictor* para los elementos de la muestra bajo ciertas circunstancias (o incluso para elementos fuera de la muestra, si la muestra es representativa de la población sobre la que realizamos la previsión).

Resumen de los temas por lecciones

A.1. Resumen del Tema 1

El tema consta de cinco lecciones que resumimos a continuación.

Lección 1.

En la primera lección solo se establece la notación para vectores, matrices, y operadores selectores junto con algunas reglas de reescritura. Por ejemplo, estas dos reglas de reescritura

$$\boxed{(a + b)_{|i} = a_{|i} + b_{|i}} \quad \text{y} \quad \boxed{(\lambda b)_{|i} = \lambda(b_{|i})}$$

definen la suma de vectores y su producto por escalares. Jugando con estas reglas se deducen las propiedades de la suma de vectores y producto por escalares que permitirán definir el *espacio vectorial* \mathbb{R}^n en el Tema 2. Las reglas

$$\boxed{(\mathbf{A} + \mathbf{B})_{|j} = \mathbf{A}_{|j} + \mathbf{B}_{|j}} \quad \text{y} \quad \boxed{(\lambda \mathbf{B})_{|j} = \lambda(\mathbf{B}_{|j})}$$

definen la suma de matrices y su producto por escalares (operando con las columnas). Como estas reglas son esencialmente iguales a las anteriores, se obtienen propiedades análogas, lo que permitirá definir el espacio vectorial $\mathbb{R}^{n \times m}$. ¡El juego con las reglas de reescritura, independientemente del significado que tengan los símbolos \mathbf{a} o \mathbf{A} da lugar a las mismas propiedades! Con la transposición de una matriz

$$\boxed{(\mathbf{A}^\top)_{|j} = {}_j|\mathbf{A}}$$

mostraremos que las operaciones con matrices también se pueden definir como operaciones entre filas ${}_i|\mathbf{A}$ (y también entre componentes ${}_i|\mathbf{A}_{|j}$). Teniendo en cuenta que $\boxed{\lambda \mathbf{a} = \mathbf{a} \lambda}$ y que $\boxed{\lambda \mathbf{A} = \mathbf{A} \lambda}$, y jugando con la notación se deducen una serie de reglas de reescritura que se usarán en las siguientes lecciones: ¹

Reglas distributivas

$$\begin{aligned} (a + b)_{|i} &= a_{|i} + b_{|i} & {}_i|(a + b) &= {}_i|a + {}_i|b \\ (\mathbf{A} + \mathbf{B})_{|j} &= \mathbf{A}_{|j} + \mathbf{B}_{|j} & {}_i|(\mathbf{A} + \mathbf{B}) &= {}_i|\mathbf{A} + {}_i|\mathbf{B} \end{aligned}$$

Reglas asociativas (desplazando el paréntesis)

$$\begin{aligned} (\lambda b)_{|i} &= \lambda(b_{|i}) & {}_i|(b \lambda) &= ({}_i|b) \lambda \\ (\lambda \mathbf{A})_{|j} &= \lambda(\mathbf{A}_{|j}) & {}_i|(\mathbf{A} \lambda) &= ({}_i|\mathbf{A}) \lambda \end{aligned}$$

¹De hecho, la librería de Python solo se implementan las reglas recuadradas... y entonces ¡todo lo demás funciona automáticamente!

Intercambio entre el escalar y el operador

$$\begin{aligned}(b\lambda)_{|i} &= (b_{|i})\lambda & {}_{|i}(\lambda b) &= \lambda({}_{|i}b) \\ (\mathbf{A}\lambda)_{|j} &= (\mathbf{A}_{|j})\lambda & {}_{|i}(\lambda \mathbf{A}) &= \lambda({}_{|i}\mathbf{A})\end{aligned}$$

Lección 2

Comienza con el *producto punto* (o producto escalar usual) de dos vectores de \mathbb{R}^n y sus propiedades.²

$$\mathbf{a} \cdot \mathbf{b} = \sum_i a_i b_i$$

El uso del *producto punto* nos dota de una notación muy compacta (sin sumatorios).

La Lección 2 trata de *combinaciones lineales* de vectores de \mathbb{R}^n y su notación matricial $\mathbf{A}\mathbf{b}$ (combinación de las columnas de \mathbf{A}) y $\mathbf{a}\mathbf{B}$ (combinación de las filas de \mathbf{B}).

$$\mathbf{A}\mathbf{b} = \sum_j (\mathbf{A}_{|j})b_j \quad \text{y} \quad \mathbf{a}\mathbf{B} = (\mathbf{B}^\top)\mathbf{a}$$

Usando las reglas de reescritura de la lección anterior, se deducen nuevas reglas para el producto de una matriz por un vector a su derecha $\mathbf{A}\mathbf{b}$:

Propiedades de linealidad

- $\mathbf{A}(\mathbf{b} + \mathbf{c}) = \mathbf{A}\mathbf{b} + \mathbf{A}\mathbf{c}$
- $\mathbf{A}(\lambda\mathbf{b}) = \lambda(\mathbf{A}\mathbf{b})$

Otras propiedades

- $\mathbf{A}(\lambda\mathbf{b}) = (\lambda\mathbf{A})\mathbf{b}$
- $(\mathbf{A} + \mathbf{B})\mathbf{c} = \mathbf{A}\mathbf{c} + \mathbf{B}\mathbf{c}$
- $\mathbf{A}(\mathbf{B}\mathbf{c}) = [\mathbf{A}(\mathbf{B}_{|1}); \dots \mathbf{A}(\mathbf{B}_{|n});] \mathbf{c}$

y propiedades análogas para el producto de un vector por una matriz $\mathbf{a}\mathbf{B}$.

Adicionalmente, y solo en las transparencias de la clase correspondiente a esta lección, se adelanta la interpretación geométrica de un sistema de ecuaciones $\mathbf{A}\mathbf{x} = \mathbf{b}$... ¿qué combinaciones lineales de las columnas ($\mathbf{A}\mathbf{x}$) son iguales al vector del lado derecho \mathbf{b} ?

Lección 3.

Trata sobre el *producto de matrices*.

$$(\mathbf{AB})_{|j} = \mathbf{A}(\mathbf{B}_{|j})$$

(cada columna de \mathbf{AB} es una combinación de las columnas de \mathbf{A}). Jugando con la definición y con las reglas de reescritura de las lecciones anteriores se deducen las siguientes propiedades

- | | |
|---|---|
| ▪ $\mathbf{A}(\mathbf{B}\mathbf{c}) = (\mathbf{AB})\mathbf{c}$ | ▪ $\mathbf{A}(\lambda\mathbf{B}) = \lambda(\mathbf{AB})$ |
| ▪ $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$ | ▪ $\mathbf{A}(\lambda\mathbf{B}) = (\lambda\mathbf{A})\mathbf{B}$ |
| ▪ $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$ | ▪ $\mathbf{IA} = \mathbf{A}$ |
| ▪ $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$ | ▪ $\mathbf{AI} = \mathbf{A}$ |

²Como en este material se hace una marcada distinción entre vectores y matrices, aquí carece de sentido “transponer un vector”. Así evitamos el frecuente abuso de notación del que hacen uso otros textos.

Y continuado con el mismo juego de manipulación de símbolos, se deducen nuevas interpretaciones del producto

$${}_i(\mathbf{AB}) = ({}_i\mathbf{A})\mathbf{B} \quad \text{y} \quad {}_i(\mathbf{AB})_{|j} = ({}_i\mathbf{A}) \cdot (\mathbf{B}_{|j})$$

es decir, las filas de \mathbf{AB} son combinaciones lineales de las filas de \mathbf{B} , y los elementos de \mathbf{AB} son productos punto de las filas de \mathbf{A} con las columnas de \mathbf{B} . Aquí se evidencian las bondades de la notación. Una expresión como

$${}_i\mathbf{AB}_{|j}$$

se puede interpretar como el elemento de la fila i , y columna j de \mathbf{AB} , como el elemento i -ésimo de la combinación de las columnas $\mathbf{A}(\mathbf{B}_{|j})$, como el elemento j -ésimo de una combinación de las filas $({}_i\mathbf{A})\mathbf{B}$, o como el producto escalar de la fila ${}_i\mathbf{A}$ con la columna $\mathbf{B}_{|j}$. Todas estas interpretaciones son correctas, y todas ellas están sugeridas en la expresión, ${}_i\mathbf{AB}_{|j}$. Es destacable la potencia computacional de la notación (véase la demostración de $(\mathbf{AB})^\top = \mathbf{B}^\top(\mathbf{A}^\top)$ así como su implementación en la librería de Python).

Lección 4

Trata sobre las *transformaciones elementales*, y su uso en el *método de eliminación*.

Aquí solo consideramos dos tipos de transformaciones elementales:

- $\boxed{\begin{smallmatrix} \tau \\ [(\lambda)i+j] \end{smallmatrix}}$ suma λ veces el vector i -ésimo al j -ésimo.
- $\boxed{\begin{smallmatrix} \tau \\ [(\alpha)i] \end{smallmatrix}}$ multiplica por α el vector i -ésimo.

Llamamos *matriz elemental* a la matriz resultante de aplicar una única transformación elemental sobre las columnas (o bien sobre las filas) de una matriz identidad

$$\mathbf{I}_\tau \quad \text{o} \quad {}_\tau\mathbf{I}$$

Dada la transformación elemental, τ , las dos matrices elementales de más arriba son una la transpuesta de la otra. Por tanto, la transpuesta de una matriz elemental es otra matriz elemental. Aplicar una transformación elemental a una matriz es equivalente a multiplicarla por la correspondiente matriz elemental, es decir

$$\mathbf{A}_\tau = \mathbf{A}(\mathbf{I}_\tau) \quad \text{y} \quad {}_\tau\mathbf{A} = ({}_\tau\mathbf{I})\mathbf{A}.$$

Para describir la aplicación de la secuencia $\tau_1 \cdots \tau_k$ de k transformaciones elementales de las *columnas* de \mathbf{A} usamos el esquema.³

$$\mathbf{A} \xrightarrow{\tau_1} (\mathbf{A}_{\tau_1}) \xrightarrow{\tau_2} (\mathbf{A}_{\tau_1\tau_2}) \cdots \xrightarrow{\tau_k} (\mathbf{A}_{\tau_1\cdots\tau_k}) \quad \text{o agregando varios pasos} \quad \mathbf{A} \xrightarrow[\tau_1]{\tau_1 \cdots \tau_p} (\mathbf{A}_{\tau_1\cdots\tau_p}) \xrightarrow[\tau_{p+1}]{\tau_{p+1} \cdots \tau_k} (\mathbf{A}_{\tau_1\cdots\tau_k}).$$

Cuando se aplica la sucesión $\tau_1 \cdots \tau_k$ de k transformaciones elementales sobre las *columnas*, o bien la sucesión $\tau_k \cdots \tau_1$ de k transformaciones elementales sobre las *filas* (nótese el distinto orden en las sucesiones) se obtienen relaciones similares a las que encontramos al aplicar una única transformación:

$$\mathbf{A}_{\tau_1\cdots\tau_k} = \mathbf{A}(\mathbf{I}_{\tau_1\cdots\tau_k}) \quad \text{y} \quad {}_{\tau_1\cdots\tau_k}\mathbf{A} = ({}_{\tau_1\cdots\tau_k}\mathbf{I})\mathbf{A}.$$

³De manera análoga, para describir una secuencia de transformaciones elementales de las *filas* de \mathbf{A} (donde τ_1 es la primera que se aplica, luego τ_1, \dots y por último τ_k) se debe usar el siguiente esquema

$$\mathbf{A} \xrightarrow[\tau_1]{\tau_1} ({}_{\tau_1}\mathbf{A}) \xrightarrow[\tau_2]{\tau_2} ({}_{\tau_2\tau_1}\mathbf{A}) \cdots \xrightarrow[\tau_k]{\tau_k} ({}_{\tau_k\cdots\tau_1}\mathbf{A}) \quad \text{o agregando varios pasos} \quad \mathbf{A} \xrightarrow[\tau_1]{\tau_p \cdots \tau_1} ({}_{\tau_p\cdots\tau_1}\mathbf{A}) \xrightarrow[\tau_{p+1}]{\tau_k \cdots \tau_{p+1}} ({}_{\tau_k\cdots\tau_1}\mathbf{A});$$

donde la secuencia $\tau_k \cdots \tau_1$ es la transpuesta de la secuencia $\tau_1 \cdots \tau_k$, es decir, $(\tau_1 \cdots \tau_k)^\top = \tau_k \cdots \tau_1$, pues al actuar por la izquierda, las primeras transformaciones que se aplican son la que están más a la derecha de la secuencia. Por tanto ${}_{\tau_k\cdots\tau_1}\mathbf{I}$ es la transpuesta de $\mathbf{I}_{\tau_1\cdots\tau_k}$ (cuando ambas tienen el mismo orden).

Es posible realizar un *intercambio* de posición entre dos vectores mediante una sucesión de transformaciones elementales:

- $\begin{bmatrix} \tau \\ [i \rightleftharpoons j] \end{bmatrix}$ intercambia de posición los vectores i -ésimo y j -ésimo.

Llamamos matriz de intercambio a la matriz resultante de aplicar un único intercambio entre dos columnas (o bien dos filas) de una matriz identidad.

$$\mathbf{I}_{\begin{bmatrix} \tau \\ [i \rightleftharpoons j] \end{bmatrix}} \quad \text{o} \quad \begin{bmatrix} \tau \\ [i \rightleftharpoons j] \end{bmatrix} \mathbf{I}$$

Las matrices intercambio son simétricas (por lo que estas dos de arriba son iguales si tienen el mismo orden). Aplicar un intercambio a una matriz es equivalente a multiplicarla por la correspondiente matriz de intercambio, es decir

$$\mathbf{A}_{\begin{bmatrix} \tau \\ [i \rightleftharpoons j] \end{bmatrix}} = \mathbf{A} \begin{bmatrix} \mathbf{I} & \tau \\ & [i \rightleftharpoons j] \end{bmatrix} \quad \text{y} \quad \begin{bmatrix} \tau \\ [i \rightleftharpoons j] \end{bmatrix} \mathbf{A} = \begin{bmatrix} \tau & \mathbf{I} \\ [i \rightleftharpoons j] \end{bmatrix} \mathbf{A}.$$

La aplicación de una sucesión de intercambios da lugar a un reordenamiento de los vectores. Denominamos *matriz permutación* a la matriz que resulta tras una sucesión de intercambios en las columnas (o en las filas) de la matriz identidad. De nuevo tenemos que

$$\mathbf{A}_{\begin{bmatrix} \tau \\ [\mathfrak{S}] \end{bmatrix}} = \mathbf{A} \begin{bmatrix} I_{\tau} \\ [\mathfrak{S}] \end{bmatrix} \quad \text{y} \quad \begin{bmatrix} \tau \\ [\mathfrak{S}] \end{bmatrix} \mathbf{A} = \begin{bmatrix} \tau & I \\ [\mathfrak{S}] \end{bmatrix} \mathbf{A}$$

(la flecha circular denota un reordenamiento de las columnas (o de las filas) de la matriz).

Mediante una sucesión de transformaciones elementales es posible pre-escalonar cualquier matriz. La demostración de este importante teorema describe la implementación del método en Python. A este procedimiento se le llama *Método de eliminación*. Hay dos extensiones más: la *eliminación Gaussiana* que escalona la matriz reordenando las columnas y la eliminación *Gauss-Jordan*, que reduce la matriz escalonada.

Lección 5.

Trata sobre la *inversión de las transformaciones elementales* y la *inversión de matrices* (cuando es posible) aplicando el Método de eliminación Gauss-Jordan.

Definimos la inversa de \mathbf{A} de orden n (por tanto *cuadrada*) como aquella matriz \mathbf{B} tal que $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$. Y demostramos que, si existe \mathbf{B} , es única y la denotamos por \mathbf{A}^{-1} . A continuación se demuestra que:

- Si \mathbf{A} y \mathbf{B} tienen inversa, entonces $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}(\mathbf{A}^{-1})$.
- Si \mathbf{B} es invertible, entonces \mathbf{AB} es invertible si y solo si \mathbf{A} es invertible.
- Si \mathbf{A} es invertible, entonces \mathbf{AB} es invertible si y solo si \mathbf{B} es invertible.
- Si \mathbf{A} es invertible, entonces $(\mathbf{A}^{\top})^{-1} = (\mathbf{A}^{-1})^{\top}$.
- Si \mathbf{A} tiene alguna columna (o fila) nula entonces no tiene inversa.

Todas *las transformaciones elementales* se pueden deshacer (todas *son invertibles*), lo que implica que *todas las matrices elementales son invertibles*. En la segunda parte de lección relacionamos la invertibilidad de una matriz con las transformaciones elementales, pues cualquier matriz de la forma $\mathbf{I}_{\tau_1 \dots \tau_k}$ es invertible por ser producto de matrices elementales:

$$(\mathbf{I}_{\tau_1 \dots \tau_k})(\mathbf{I}_{\tau_k^{-1} \dots \tau_1^{-1}}) = \mathbf{E}_1 \dots \mathbf{E}_k \cdot \mathbf{E}_k^{-1} \dots \mathbf{E}_1^{-1} = \mathbf{I}_{\tau_1 \dots \tau_k \cdot \tau_k^{-1} \dots \tau_1^{-1}} = \mathbf{I}.$$

Siguiendo esta idea, se demuestra que

- Si \mathbf{A} es producto de matrices elementales, entonces es invertible.

- Si \mathbf{A} tiene columnas que son combinación lineal del resto, entonces no es invertible.
- \mathbf{A} es invertible si y solo si sus formas escalonadas \mathbf{L} son invertibles.

A continuación se demuestra que toda matriz cuadrada \mathbf{L} escalonada y sin columnas nulas se puede transformar en la matriz identidad mediante transformaciones elementales (de nuevo la demostración indica los pasos para implementar el algoritmo en Python). Esto da lugar a un corolario que establece que las siguientes propiedades son equivalentes

1. El resultado de escalar \mathbf{A} no tiene columnas nulas.
2. \mathbf{A} es producto de matrices elementales.
3. \mathbf{A} tiene inversa.

A continuación se demuestra que si \mathbf{A} y \mathbf{B} son cuadradas y del mismo orden, y $\mathbf{AB} = \mathbf{I}$, entonces \mathbf{A} y \mathbf{B} son una la inversa de la otra.

En la lección anterior vimos que el [Método de eliminación](#) permite encontrar una forma pre-escalonada de toda matriz (Teorema 4.2.2 en la página 53)

$$\mathbf{A}_{\tau_1 \dots \tau_p} = \mathbf{K},$$

y, si \mathbf{K} no tiene columnas nulas, se puede continuar con las transformaciones elementales hasta transformar \mathbf{K} en \mathbf{I} (Teorema 5.1.10 en la página 64)

$$\mathbf{K}_{\tau_{(p+1)} \dots \tau_k} = \mathbf{I}.$$

Por tanto, la sucesión de k transformaciones elementales $\tau_1 \dots \tau_p$, $\tau_{(p+1)} \dots \tau_k$ transforma \mathbf{A} en \mathbf{I} , es decir,

$$\mathbf{A}_{\tau_1 \dots \tau_k} = \mathbf{A}(\mathbf{I}_{\tau_1 \dots \tau_k}) = \mathbf{I};$$

así, si \mathbf{L} no tiene columnas nulas, entonces $\boxed{\mathbf{A}^{-1} = \mathbf{I}_{\tau_1 \dots \tau_k}}$. Por otra parte, \mathbf{A}^{-1} no existe cuando \mathbf{L} tiene columnas nulas.

La lección finaliza proponiendo un método para encontrar la inversa a la vez que se transforma \mathbf{A} en \mathbf{I} , y dando una definición de *rango* de una matriz (el número de columnas no nulas de cualquiera de sus formas escalonadas).

Aquí finaliza el primer tema del curso.

Soluciones a los Ejercicios

Ejercicio 1. $(3*b)|2$, primero multiplica el vector por 3. Después selecciona la segunda componente. En $3*(b|2)$, primero se selecciona la segunda componente. Luego multiplica dicha componente por 3. $3*b|2$ hace exactamente lo mismo que con la primera expresión (pues $*$ tiene precedencia sobre $|$).

□

Ejercicio 2. En $b+b*2$ lo primero es el producto. En $b+b|2$ lo primero es la suma.

□

Ejercicio 3. Como en Python el operador $+$ tiene precedencia sobre $|$, en la expresión $b+b|2$ primero se suman los vectores, y luego se selecciona la segunda componente del resultado.

Pero para nosotros la expresión es confusa ya que entendemos $b + b|_2$ como la suma del vector b con la segunda componente de b ; y la operación suma entre un vector y un número *no está definida*.

Para nosotros el modo correcto de escribir la operación es $(b + b)|_2$. Así que, por claridad, es recomendable escribir $(b+b)|2$ también en Python.

□

Ejercicio 4. Mostraremos ambas estrategias en cada caso. En el caso de la segunda, en azul aparecen las operaciones entre números reales.

1. Estrategia 1:

$$a + b = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} + \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ \vdots \\ a_m + b_m \end{pmatrix} = \begin{pmatrix} b_1 + a_1 \\ \vdots \\ b_m + a_m \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} + \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} = b + a.$$

$$\text{Estrategia 2: } (a + b)|_i = a_{|i} + b_{|i} = b_{|i} + a_{|i} = (b + a)|_i.$$

2. Estrategia 1:

$$\begin{aligned} a + (b + c) &= \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} + \left(\begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} + \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix} \right) = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} + \begin{pmatrix} b_1 + c_1 \\ \vdots \\ b_m + c_m \end{pmatrix} = \begin{pmatrix} a_1 + b_1 + c_1 \\ \vdots \\ a_m + b_m + c_m \end{pmatrix} = \\ &= \begin{pmatrix} a_1 + b_1 \\ \vdots \\ a_m + b_m \end{pmatrix} + \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix} = \left(\begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} + \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \right) + \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix} = (a + b) + c. \end{aligned}$$

Estrategia 2:

$$\begin{aligned} (a + (b + c))|_i &= a_{|i} + (b + c)|_i = a_{|i} + (b_{|i} + c_{|i}) = (a_{|i} + b_{|i}) + c_{|i} \\ &= (a + b)|_i + c_{|i} = ((a + b) + c)|_i. \end{aligned}$$

3. Estrategia 1:

$$0 + a = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} 0 + a_1 \\ \vdots \\ 0 + a_m \end{pmatrix} = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} = a.$$

$$\text{Estrategia 2: } (0 + a)|_i = 0_{|i} + a_{|i} = 0 + a_{|i} = a_{|i}.$$

4. Estrategia 1:

$$\mathbf{a} + (-\mathbf{a}) = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} + \begin{pmatrix} -a_1 \\ \vdots \\ -a_m \end{pmatrix} = \begin{pmatrix} a_1 - a_1 \\ \vdots \\ a_m - a_m \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{0}.$$

Estrategia 2: $(\mathbf{a} + (-\mathbf{a}))_{|i} = \mathbf{a}_{|i} + (-\mathbf{a})_{|i} = \mathbf{a}_{|i} - \mathbf{a}_{|i} = \mathbf{0} = \mathbf{0}_{|i}.$

5. Estrategia 1:

$$\lambda(\mathbf{a} + \mathbf{b}) = \begin{pmatrix} \lambda(a_1 + b_1) \\ \vdots \\ \lambda(a_m + b_m) \end{pmatrix} = \begin{pmatrix} \lambda a_1 + \lambda b_1 \\ \vdots \\ \lambda a_m + \lambda b_m \end{pmatrix} = \begin{pmatrix} \lambda a_1 \\ \vdots \\ \lambda a_m \end{pmatrix} + \begin{pmatrix} \lambda b_1 \\ \vdots \\ \lambda b_m \end{pmatrix} = \lambda \mathbf{a} + \lambda \mathbf{b}.$$

Estrategia 2: $(\lambda(\mathbf{a} + \mathbf{b}))_{|i} = \lambda((\mathbf{a} + \mathbf{b})_{|i}) = \lambda(\mathbf{a}_{|i} + \mathbf{b}_{|i}) = \lambda(\mathbf{a}_{|i}) + \lambda(\mathbf{b}_{|i}) = (\lambda \mathbf{a})_{|i} + (\lambda \mathbf{b})_{|i} = (\lambda \mathbf{a} + \lambda \mathbf{b})_{|i}.$

6. Estrategia 1:

$$(\lambda + \varphi)\mathbf{a} = \begin{pmatrix} (\lambda + \varphi)a_1 \\ \vdots \\ (\lambda + \varphi)a_m \end{pmatrix} = \begin{pmatrix} \lambda a_1 + \varphi a_1 \\ \vdots \\ \lambda a_m + \varphi a_m \end{pmatrix} = \begin{pmatrix} \lambda a_1 \\ \vdots \\ \lambda a_m \end{pmatrix} + \begin{pmatrix} \varphi a_1 \\ \vdots \\ \varphi a_m \end{pmatrix} = \lambda \mathbf{a} + \varphi \mathbf{a}.$$

Estrategia 2: $((\lambda + \varphi)\mathbf{a})_{|i} = (\lambda + \varphi)(\mathbf{a}_{|i}) = \lambda(\mathbf{a}_{|i}) + \varphi(\mathbf{a}_{|i}) = (\lambda \mathbf{a})_{|i} + (\varphi \mathbf{a})_{|i} = (\lambda \mathbf{a} + \varphi \mathbf{a})_{|i}.$

7. Estrategia 1:

$$\lambda(\varphi \mathbf{a}) = \lambda \begin{pmatrix} \varphi a_1 \\ \vdots \\ \varphi a_m \end{pmatrix} = \begin{pmatrix} \lambda \varphi a_1 \\ \vdots \\ \lambda \varphi a_m \end{pmatrix} = \lambda \varphi \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} = (\lambda \varphi) \mathbf{a}.$$

Estrategia 2: $(\lambda(\varphi \mathbf{a}))_{|i} = \lambda((\varphi \mathbf{a})_{|i}) = \lambda(\varphi(\mathbf{a}_{|i})) = (\lambda \varphi)(\mathbf{a}_{|i}) = ((\lambda \varphi) \mathbf{a})_{|i}.$

8. Estrategia 1:

$$1\mathbf{a} = \begin{pmatrix} 1 \cdot a_1 \\ \vdots \\ 1 \cdot a_m \end{pmatrix} = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} = \mathbf{a}.$$

Estrategia 2: $(1\mathbf{a})_{|i} = 1(\mathbf{a}_{|i}) = \mathbf{a}_{|i}.$

□

Ejercicio 5(a) Primero multiplica el vector por 3 y luego selecciona la primera componente.

□

Ejercicio 5(b) Primero multiplica 3 por 1; así que selecciona la tercera componente del vector.

□

Ejercicio 5(c) Primero multiplica 1 por 3; así que selecciona la tercera componente del vector.

□

Ejercicio 5(d) Como en la primera expresión, primero multiplica el vector por 3 y luego selecciona la primera componente.

□

Ejercicio 6(a) ${}_k|(\mathbf{A}^\top)_{|j} = {}_k|((\mathbf{A}^\top)_{|j}) = {}_k|({}_j|\mathbf{A}) = ({}_j|\mathbf{A})_{|k} = {}_j|\mathbf{A}_{|k}.$

□

Ejercicio 6(b) ${}_j|((\mathbf{A}^\top)^\top)_{|k} = {}_k|(\mathbf{A}^\top)_{|j} = {}_j|\mathbf{A}_{|k}.$

□

Ejercicio 6(c) $\mathbf{A}_{|i} = ((\mathbf{A}^\top)^\top)_{|i} = {}_i|(\mathbf{A}^\top).$

□

Ejercicio 7. Las demostraciones son prácticamente idénticas a las vistas en el caso de los vectores. En azul aparecen las operaciones entre vectores.

1. $(\mathbf{A} + \mathbf{B})_{|j} = \mathbf{A}_{|j} + \mathbf{B}_{|j} = \mathbf{B}_{|j} + \mathbf{A}_{|j} = (\mathbf{B} + \mathbf{A})_{|j}.$
2. $\left((\mathbf{A} + \mathbf{B}) + \mathbf{C}\right)_{|j} = (\mathbf{A} + \mathbf{B})_{|j} + \mathbf{C}_{|j} = (\mathbf{A}_{|j} + \mathbf{B}_{|j}) + \mathbf{C}_{|j} = \mathbf{A}_{|j} + (\mathbf{B}_{|j} + \mathbf{C}_{|j}) = \mathbf{A}_{|j} + (\mathbf{B} + \mathbf{C})_{|j} = \left(\mathbf{A} + (\mathbf{B} + \mathbf{C})\right)_{|j}.$
3. $(\mathbf{0} + \mathbf{A})_{|j} = \mathbf{0}_{|j} + \mathbf{A}_{|j} = \mathbf{0} + \mathbf{A}_{|j} = \mathbf{A}_{|j}.$
4. $(\mathbf{A} + (-\mathbf{A}))_{|j} = \mathbf{A}_{|j} + (-\mathbf{A})_{|j} = \mathbf{A}_{|j} - (\mathbf{A}_{|j}) = \mathbf{0} = \mathbf{0}_{|j}.$
5. $\left(\lambda(\mathbf{A} + \mathbf{B})\right)_{|j} = \lambda\left((\mathbf{A} + \mathbf{B})_{|j}\right) = \lambda(\mathbf{A}_{|j} + \mathbf{B}_{|j}) = \lambda(\mathbf{A}_{|j}) + \lambda(\mathbf{B}_{|j}) = (\lambda\mathbf{A})_{|j} + (\lambda\mathbf{B})_{|j} = (\lambda\mathbf{A} + \lambda\mathbf{B})_{|j}.$
6. $\left((\lambda + \varphi)\mathbf{A}\right)_{|j} = (\lambda + \varphi)(\mathbf{A}_{|j}) = \lambda(\mathbf{A}_{|j}) + \varphi(\mathbf{A}_{|j}) = (\lambda\mathbf{A})_{|j} + (\varphi\mathbf{A})_{|j} = (\lambda\mathbf{A} + \varphi\mathbf{A})_{|j}.$
7. $\left(\lambda(\varphi\mathbf{A})\right)_{|j} = \lambda\left((\varphi\mathbf{A})_{|j}\right) = \lambda\left(\varphi(\mathbf{A}_{|j})\right) = (\lambda\varphi)(\mathbf{A}_{|j}) = ((\lambda\varphi)\mathbf{A})_{|j}.$
8. $(1\mathbf{A})_{|j} = 1(\mathbf{A}_{|j}) = \mathbf{A}_{|j}.$

□

Ejercicio 8. Comenzamos por la suma.

$${}_i|(\mathbf{A} + \mathbf{B})_{|j} = {}_i|\left((\mathbf{A} + \mathbf{B})_{|j}\right) = {}_i|\left(\mathbf{A}_{|j} + \mathbf{B}_{|j}\right) = {}_i|\mathbf{A}_{|j} + {}_i|\mathbf{B}_{|j}.$$

Y ahora el producto.

$${}_i|(\lambda\mathbf{A})_{|j} = {}_i|\left((\lambda\mathbf{A})_{|j}\right) = {}_i|\left(\lambda(\mathbf{A}_{|j})\right) = \lambda\left({}_i|(\mathbf{A}_{|j})\right) = \lambda({}_i|\mathbf{A}_{|j}).$$

□

Ejercicio 9(a) $\left((\lambda\mathbf{A})^\top\right)_{|j} = {}_j|(\lambda\mathbf{A}) = \lambda({}_j|\mathbf{A}) = \lambda\left((\mathbf{A}^\top)_{|j}\right) = \left(\lambda(\mathbf{A}^\top)\right)_{|j}.$

□

Ejercicio 9(b) $((\mathbf{A} + \mathbf{B})^\top)_{|j} = {}_{j|}(\mathbf{A} + \mathbf{B}) = {}_{j|}\mathbf{A} + {}_{j|}\mathbf{B} = (\mathbf{A}^\top)_{|j} + (\mathbf{B}^\top)_{|j} = (\mathbf{A}^\top + \mathbf{B}^\top)_{|j}.$

□

Ejercicio 10. Comenzamos con la suma

$${}_{i|}(\mathbf{A} + \mathbf{B}) = ((\mathbf{A} + \mathbf{B})^\top)_{|i} = (\mathbf{A}^\top + \mathbf{B}^\top)_{|i} = (\mathbf{A}^\top)_{|i} + (\mathbf{B}^\top)_{|i} = {}_{i|}\mathbf{A} + {}_{i|}\mathbf{B}$$

Y continuamos con el producto:

$${}_{i|}(\mathbf{A}\lambda) = ((\mathbf{A}\lambda)^\top)_{|i} = ((\mathbf{A}^\top)\lambda)_{|i} = ((\mathbf{A}^\top)_{|i})\lambda = ({}_{i|}\mathbf{A})\lambda.$$

□

Ejercicio 11(a) $\mathbf{x} \cdot \mathbf{y} = x_1y_1 + \cdots + x_ny_n = y_1x_1 + \cdots + y_nx_n = \mathbf{y} \cdot \mathbf{x}$

□

Ejercicio 11(b)

$$\blacksquare (\lambda\mathbf{x}) \cdot \mathbf{y} = \lambda x_1y_1 + \cdots + \lambda x_ny_n = \lambda(x_1y_1 + \cdots + x_ny_n) = \lambda(\mathbf{x} \cdot \mathbf{y}).$$

$$\blacksquare (\mathbf{x} + \mathbf{y}) \cdot \mathbf{z} = (x_1 + y_1)z_1 + \cdots + (x_n + y_n)z_n = x_1z_1 + \cdots + x_nz_n + y_1z_1 + \cdots + y_nz_n = \mathbf{x} \cdot \mathbf{z} + \mathbf{y} \cdot \mathbf{z}.$$

□

Ejercicio 11(c) $\mathbf{x} \cdot \mathbf{x} = x_1^2 + \cdots + x_n^2 \geq 0.$

□

Ejercicio 11(d) $\mathbf{x} \cdot \mathbf{x} = x_1^2 + \cdots + x_n^2 = 0 \iff x_i = 0 \text{ para } i = 1 : n.$

□

Ejercicio 12.

$$2 \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} + 2 \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} + 2 \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ -4 \end{pmatrix} + \begin{pmatrix} 2 \\ -4 \\ 2 \end{pmatrix} + \begin{pmatrix} -4 \\ 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 2+2-4 \\ 2-4+2 \\ -4+2+2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \mathbf{0}.$$

□

Ejercicio 13. $\mathbf{A}\mathbf{b} = \begin{bmatrix} 1 & 1 & -2 \\ 1 & -2 & 1 \\ -2 & 1 & 1 \end{bmatrix} \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} + 2 \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} + 2 \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$

□

Ejercicio 14. $\mathbf{A}(\mathbf{l}_{|j}) = \mathbf{A} \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = (\mathbf{A}_{|1})0 + \cdots + (\mathbf{A}_{|j})1 + \cdots + (\mathbf{A}_{|n})0 = \mathbf{A}_{|j}.$

□

Ejercicio 15. Sea \mathbf{A} , de n columnas y \mathbf{b} de \mathbb{R}^n , entonces

$$\begin{aligned} {}_i\mathbf{A}(\mathbf{A}\mathbf{b}) &= {}_i\mathbf{A}\left((\mathbf{A}_{|1})b_1 + \cdots + (\mathbf{A}_{|n})b_n\right) \\ &= {}_i\mathbf{A}(\mathbf{A}_{|1})b_1 + \cdots + {}_i\mathbf{A}(\mathbf{A}_{|n})b_n \\ &= ({}_i\mathbf{A})_{|1}b_1 + \cdots + ({}_i\mathbf{A})_{|n}b_n = {}_i\mathbf{A} \cdot \mathbf{b}. \end{aligned}$$

□

Ejercicio 16(a) ${}_i\mathbf{A}(\mathbf{A}(\mathbf{b} + \mathbf{c})) = ({}_i\mathbf{A}) \cdot (\mathbf{b} + \mathbf{c}) = ({}_i\mathbf{A}) \cdot \mathbf{b} + ({}_i\mathbf{A}) \cdot \mathbf{c} = {}_i(\mathbf{A}\mathbf{b}) + {}_i(\mathbf{A}\mathbf{c}) = {}_i(\mathbf{A}\mathbf{b} + \mathbf{A}\mathbf{c}).$

□

Ejercicio 16(b) ${}_i\mathbf{A}(\lambda\mathbf{b}) = ({}_i\mathbf{A}) \cdot (\lambda\mathbf{b}) = \lambda\left(({}_i\mathbf{A}) \cdot \mathbf{b}\right) = \lambda\left({}_i(\mathbf{A}\mathbf{b})\right) = {}_i(\lambda(\mathbf{A}\mathbf{b})).$

□

Ejercicio 17.

$$\begin{aligned} \mathbf{A}(\mathbf{B}\mathbf{c}) &= \mathbf{A}\left((\mathbf{B}_{|1})c_1 + \cdots + (\mathbf{B}_{|n})c_n\right) && \text{por la Definición 2.3 de producto} \\ &= \mathbf{A}(\mathbf{B}_{|1})c_1 + \cdots + \mathbf{A}(\mathbf{B}_{|n})c_n && \text{proposiciones 2.2.1 y 2.2.2} \\ &= \left[\mathbf{A}(\mathbf{B}_{|1}); \quad \dots \quad \mathbf{A}(\mathbf{B}_{|n});\right] \mathbf{c} && \text{por la Definición 2.3 de producto.} \end{aligned}$$

□

Ejercicio 18(a) ${}_i\mathbf{A}(\lambda\mathbf{b}) = ({}_i\mathbf{A}) \cdot (\lambda\mathbf{b}) = \lambda\left(({}_i\mathbf{A}) \cdot \mathbf{b}\right) = \left(\lambda({}_i\mathbf{A})\right) \cdot \mathbf{b} = ({}_i(\lambda\mathbf{A})) \cdot \mathbf{b} = {}_i((\lambda\mathbf{A})\mathbf{b})$

□

Ejercicio 18(b)

$${}_i((\mathbf{A} + \mathbf{B})\mathbf{c}) = ({}_i(\mathbf{A} + \mathbf{B})) \cdot \mathbf{c} = ({}_i\mathbf{A} + {}_i\mathbf{B}) \cdot \mathbf{c} = ({}_i\mathbf{A}) \cdot \mathbf{c} + ({}_i\mathbf{B}) \cdot \mathbf{c} = {}_i(\mathbf{A}\mathbf{c}) + {}_i(\mathbf{B}\mathbf{c}) = {}_i(\mathbf{A}\mathbf{c} + \mathbf{B}\mathbf{c}).$$

□

Ejercicio 19(a) $\mathbf{a}\mathbf{I} = (\mathbf{I}^\top)\mathbf{a} = \mathbf{I}\mathbf{a} = \mathbf{a}.$

□

Ejercicio 19(b) $({}_i\mathbf{I})\mathbf{A} = (\mathbf{A}^\top)({}_i\mathbf{I}) = (\mathbf{A}^\top)(\mathbf{I}_{|i}) = (\mathbf{A}^\top)_{|i} = {}_i\mathbf{A}.$

□

Ejercicio 19(c) $(\mathbf{a} + \mathbf{b})\mathbf{C} = (\mathbf{C}^\top)(\mathbf{a} + \mathbf{b}) \underset{(*)}{=} (\mathbf{C}^\top)\mathbf{a} + (\mathbf{C}^\top)\mathbf{b} = \mathbf{a}\mathbf{C} + \mathbf{b}\mathbf{C}.$

□

Ejercicio 19(d) $(\lambda\mathbf{a})\mathbf{B} = (\mathbf{B}^\top)(\lambda\mathbf{a}) \underset{(*)}{=} \lambda\left((\mathbf{B}^\top)\mathbf{a}\right) = \lambda(\mathbf{a}\mathbf{B})$

□

Ejercicio 19(e) $\mathbf{a}(\lambda\mathbf{B}) = ((\lambda\mathbf{B})^\top)\mathbf{a} = \left(\lambda(\mathbf{B}^\top)\right)\mathbf{a} \underset{(*)}{=} (\mathbf{B}^\top)(\lambda\mathbf{a}) = (\lambda\mathbf{a})\mathbf{B}.$

□

Ejercicio 19(f) $\mathbf{a}(\mathbf{B} + \mathbf{C}) = (\mathbf{B} + \mathbf{C})^\top \mathbf{a} = \left((\mathbf{B}^\top) + (\mathbf{C}^\top) \right) \mathbf{a} \underset{(*)}{=} (\mathbf{B}^\top) \mathbf{a} + (\mathbf{C}^\top) \mathbf{a} = \mathbf{aB} + \mathbf{aC}.$

□

Ejercicio 20(a) Recordando la Proposición 2.2.3 en la página 28 (*) y usando la Definición 3.1 de producto:

$$\mathbf{A}(\mathbf{Bc}) \underset{(*)}{=} \left[\mathbf{A}(\mathbf{B}_{|1}), \quad \dots, \quad \mathbf{A}(\mathbf{B}_{|n}) \right] \mathbf{c} = (\mathbf{AB})\mathbf{c}.$$

□

Ejercicio 20(b) Aplicando repetidamente la definición de producto y una vez (*) la proposición anterior tenemos:

$$\left(\mathbf{A}(\mathbf{BC}) \right)_{|j} = \mathbf{A} \left((\mathbf{BC})_{|j} \right) = \mathbf{A} \left(\mathbf{B}(\mathbf{C}_{|j}) \right) \underset{(*)}{=} \left(\mathbf{AB} \right) (\mathbf{C}_{|j}) = \left((\mathbf{AB})\mathbf{C} \right)_{|j}, \quad j = 1 : n.$$

□

Ejercicio 21(a) Usando la Definición 3.1 de producto matricial y recordando que $(\mathbf{A} + \mathbf{B})\mathbf{c} = \mathbf{Ac} + \mathbf{Bc}$; para $j = 1 : n$ tenemos

$$\left((\mathbf{A} + \mathbf{B})\mathbf{C} \right)_{|j} = (\mathbf{A} + \mathbf{B})(\mathbf{C}_{|j}) = \mathbf{A}(\mathbf{C}_{|j}) + \mathbf{B}(\mathbf{C}_{|j}) = (\mathbf{AC})_{|j} + (\mathbf{BC})_{|j} = (\mathbf{AC} + \mathbf{BC})_{|j}.$$

□

Ejercicio 21(b) Usando la Definición 3.1 de producto matricial y la propiedad de linealidad: $\mathbf{A}(\mathbf{u} + \mathbf{v}) = \mathbf{Au} + \mathbf{Av}$; para $j = 1 : n$ tenemos

$$\left(\mathbf{A}(\mathbf{B} + \mathbf{C}) \right)_{|j} = \mathbf{A} \left((\mathbf{B} + \mathbf{C})_{|j} \right) = \mathbf{A} \left((\mathbf{B}_{|j}) + (\mathbf{C}_{|j}) \right) = \mathbf{A}(\mathbf{B}_{|j}) + \mathbf{A}(\mathbf{C}_{|j}) = (\mathbf{AB})_{|j} + (\mathbf{AC})_{|j} = (\mathbf{AB} + \mathbf{AC})_{|j}.$$

□

Ejercicio 21(c)

$$\left(\mathbf{A}(\lambda \mathbf{B}) \right)_{|j} = \mathbf{A} \left((\lambda \mathbf{B})_{|j} \right) = \mathbf{A} \left(\lambda (\mathbf{B}_{|j}) \right) \begin{cases} = (\lambda \mathbf{A})(\mathbf{B}_{|j}) = \left((\lambda \mathbf{A})\mathbf{B} \right)_{|j} \\ = \lambda \left(\mathbf{A}(\mathbf{B}_{|j}) \right) = \lambda \left((\mathbf{AB})_{|j} \right) = \left(\lambda (\mathbf{AB}) \right)_{|j} \end{cases}, \quad j = 1 : n.$$

□

Ejercicio 21(d) $(\mathbf{IA})_{|j} = \mathbf{I}(\mathbf{A}_{|j}) = \mathbf{A}_{|j}, \quad j = 1 : n.$

□

Ejercicio 21(e) $(\mathbf{AI})_{|j} = \mathbf{A}(\mathbf{I}_{|j}) = \mathbf{A}_{|j}, \quad j = 1 : n.$

□

Ejercicio 22. ${}_i(\mathbf{AB})_{|j} = {}_i((\mathbf{AB})_{|j}) = {}_i(\mathbf{A}(\mathbf{B}_{|j})) = ({}_i\mathbf{A}) \cdot (\mathbf{B}_{|j})$

□

Ejercicio 23. Veamos que las componentes j -ésimas son iguales:

$$\left({}_i(\mathbf{AB})\right)_{|j} = {}_i(\mathbf{AB})_{|j} = ({}_i\mathbf{A}) \cdot (\mathbf{B}_{|j}) = \left({}_i(\mathbf{A})\mathbf{B}\right)_{|j}; \quad \text{por tanto } {}_i(\mathbf{AB}) = ({}_i\mathbf{A})\mathbf{B}.$$

□

Ejercicio 24. $(\mathbf{AB})^\top_{|j} = {}_j\mathbf{AB} = (\mathbf{B}^\top)({}_j\mathbf{A}) = (\mathbf{B}^\top)(\mathbf{A}^\top)_{|j}; \quad \text{por tanto } (\mathbf{AB})^\top = (\mathbf{B}^\top)(\mathbf{A}^\top).$

□

Ejercicio 25(a) Sean las matrices \mathbf{A} y \mathbf{B} tales que ${}_i\mathbf{A}_{|j} = 0$ cuando $i > j$ y ${}_i\mathbf{B}_{|j} = 0$ cuando $i > j$. Tenemos que demostrar que ${}_i\mathbf{AB}_{|j} = 0$ cuando $\boxed{i > j}$:

$${}_i\mathbf{AB}_{|j} = \sum_{k=1}^p ({}_i\mathbf{A}_{|k}) ({}_k\mathbf{B}_{|j}) = \sum_{k=1}^j \underbrace{({}_i\mathbf{A}_{|k})} _0 ({}_k\mathbf{B}_{|j}) + \sum_{k=j+1}^p ({}_i\mathbf{A}_{|k}) \underbrace{({}_k\mathbf{B}_{|j})} _0 = 0.$$

□

Ejercicio 25(b) Sean las matrices \mathbf{A} y \mathbf{B} tales que por una parte ${}_i\mathbf{A}_{|j} = 0$ cuando $i > j$, y ${}_j\mathbf{A}_{|j} = 1$, y por otra ${}_i\mathbf{B}_{|j} = 0$ cuando $i > j$. Por la proposición anterior ya sabemos que \mathbf{AB} es triangular superior, así que nos basta demostrar que ${}_i\mathbf{AB}_{|j} = {}_j\mathbf{B}_{|j}$ cuando $\boxed{i = j}$:

$${}_i\mathbf{AB}_{|j} = \sum_{k=1}^p ({}_i\mathbf{A}_{|k}) ({}_k\mathbf{B}_{|j}) = \sum_{k=1}^{j-1} \underbrace{({}_i\mathbf{A}_{|k})} _0 ({}_k\mathbf{B}_{|j}) + \underbrace{({}_j\mathbf{A}_{|j})} _1 ({}_j\mathbf{B}_{|j}) + \sum_{k=j+1}^p ({}_i\mathbf{A}_{|k}) \underbrace{({}_k\mathbf{B}_{|j})} _0 = {}_j\mathbf{B}_{|j}.$$

□

Ejercicio 25(c) Sean las matrices \mathbf{A} y \mathbf{B} tales que por una parte ${}_i\mathbf{A}_{|j} = 0$ cuando $i > j$, y por otra ${}_i\mathbf{B}_{|j} = 0$ cuando $i > j$ y ${}_j\mathbf{B}_{|j} = 1$. Como antes, nos basta demostrar que ${}_i\mathbf{AB}_{|j} = {}_j\mathbf{A}_{|j}$ cuando $\boxed{i = j}$:

$${}_i\mathbf{AB}_{|j} = \sum_{k=1}^p ({}_i\mathbf{A}_{|k}) ({}_k\mathbf{B}_{|j}) = \sum_{k=1}^{j-1} \underbrace{({}_i\mathbf{A}_{|k})} _0 ({}_k\mathbf{B}_{|j}) + \underbrace{({}_j\mathbf{A}_{|j})} _1 ({}_j\mathbf{B}_{|j}) + \sum_{k=j+1}^p ({}_i\mathbf{A}_{|k}) \underbrace{({}_k\mathbf{B}_{|j})} _0 = {}_j\mathbf{A}_{|j}.$$

□

Ejercicio 26(a)

- Sea a un escalar, entonces $\text{tr}(\mathbf{aA}) = \sum_{i=1}^k {}_i(\mathbf{aA})_{|i} = \sum_{j=1}^k a({}_j\mathbf{A}_{|j}) = a \sum_{j=1}^k {}_j\mathbf{A}_{|j} = a \text{tr}(\mathbf{A}).$
- Sean \mathbf{A} y \mathbf{B} del mismo orden: $\text{tr}(\mathbf{A} + \mathbf{B}) = \sum_{j=1}^k {}_j(\mathbf{A} + \mathbf{B})_{|j} = \sum_{j=1}^k {}_j\mathbf{A}_{|j} + \sum_{j=1}^k {}_j\mathbf{B}_{|j} = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B}),$

□

Ejercicio 26(b) $\text{tr}(\mathbf{A}) = \sum_{j=1}^k {}_j|\mathbf{A}|_j = \sum_{j=1}^k {}_j|(\mathbf{A}^\top)|_j = \text{tr}(\mathbf{A}^\top).$

□

Ejercicio 26(c) Recordando la Ecuación 3.5 en la página 38 respecto al producto de matrices; que para cualesquiera vectores \mathbf{a} y \mathbf{b} , se verifica que $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$ y que son iguales los vectores ${}_{(p,)}|\mathbf{A}|_j = {}_p|\mathbf{A}|_{(j,)}:$

$$\begin{aligned} \text{tr}(\mathbf{A}^\top \mathbf{A}) &= \sum_{j=1}^n {}_j|(\mathbf{A}^\top \mathbf{A})|_j = \sum_{j=1}^n {}_j| \left(\sum_{p=1}^m ({}_{(p,)}|(\mathbf{A}^\top)|) ({}_{(p,)}|\mathbf{A}) \right) |_{|j} = \sum_{j=1}^n \sum_{p=1}^m ({}_j|(\mathbf{A}^\top)|_{(p,)}) \cdot ({}_{(p,)}|\mathbf{A}|_j) \\ &= \sum_{j=1}^n \sum_{p=1}^m ({}_{(p,)}|\mathbf{A}|_j) \cdot ({}_j|(\mathbf{A}^\top)|_{(p,)}) = \sum_{p=1}^m \sum_{j=1}^n {}_p|(\mathbf{A}|_{(j,)}) ({}_{(j,)}|(\mathbf{A}^\top)|)_{|p} = \sum_{p=1}^m {}_p|(\mathbf{A} \mathbf{A}^\top)|_p = \text{tr}(\mathbf{A} \mathbf{A}^\top). \end{aligned}$$

□

Ejercicio 26(d) Dado que $\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{A}^\top)$, basta con demostrar que $\text{tr}(\mathbf{A}^\top \mathbf{B}) = \text{tr}(\mathbf{B} \mathbf{A}^\top)$. Repitiendo los pasos del ejercicio anterior:

$$\begin{aligned} \text{tr}(\mathbf{A}^\top \mathbf{B}) &= \sum_{j=1}^n {}_j|(\mathbf{A}^\top \mathbf{B})|_j = \sum_{j=1}^n \sum_{p=1}^m ({}_j|(\mathbf{A}^\top)|_{(p,)}) \cdot ({}_{(p,)}|\mathbf{B}|_j) = \sum_{j=1}^n \sum_{p=1}^m ({}_{(p,)}|\mathbf{B}|_j) \cdot ({}_j|(\mathbf{A}^\top)|_{(p,)}) \\ &= \sum_{p=1}^m \sum_{j=1}^n {}_p|(\mathbf{B}|_{(j,)}) ({}_{(j,)}|(\mathbf{A}^\top)|)_{|p} = \sum_{p=1}^m \left(\sum_{j=1}^n ({}_{(j,)}|\mathbf{B}|_{(j,)}) ({}_{(j,)}|(\mathbf{A}^\top)|) \right)_{|p} = \sum_{p=1}^m {}_p|(\mathbf{B} \mathbf{A}^\top)|_p = \text{tr}(\mathbf{B} \mathbf{A}^\top). \end{aligned}$$

□

Ejercicio 27(a) $\begin{bmatrix} 1 & 0 & -7 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$

□

Ejercicio 27(b) $\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$

□

Ejercicio 27(c) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$

□

Ejercicio 27(d) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -10 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$

□

Ejercicio 28. a) $\tau_{[(-7)1+3]}$ transforma $\mathbf{B}_{|3}$, restándole siete veces $\mathbf{B}_{|1}$, es decir: $\left(\mathbf{B}_{[(-7)1+3]}\right)_{|3} = -7\mathbf{B}_{|1} + \mathbf{B}_{|3}$.

b) $\tau_{[(1)1+4]}$ transforma la cuarta columna de \mathbf{B} , sumándole la primera columna: $\left(\mathbf{B}_{[(1)1+4]}\right)_{|4} = \mathbf{B}_{|1} + \mathbf{B}_{|4}$.

c) $\tau_{[(3)2+1]}$ transforma la primera columna de \mathbf{B} , sumándole tres veces la segunda: $\left(\mathbf{B}_{[(3)2+1]}\right)_{|1} = 3\mathbf{B}_{|2} + \mathbf{B}_{|1}$.

d) $\tau_{[(-10)3]}$ transforma la tercera columna de \mathbf{B} , multiplicándola por -10 , es decir: $\left(\mathbf{B}_{[(-10)3]}\right)_{|3} = -10\mathbf{B}_{|3}$.

□

Ejercicio 29(a) $\mathbf{A}_{\tau_{[(3)1]}} = \mathbf{A}\left(\mathbf{I}_{\tau_{[(3)1]}}\right) = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 3 & 2 & 3 & 4 \end{bmatrix}.$

□

Ejercicio 29(b) $\mathbf{A}_{\tau_{[(-1)1+2]}} = \mathbf{A}\left(\mathbf{I}_{\tau_{[(-1)1+2]}}\right) = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 3 & 4 \end{bmatrix}.$

□

Ejercicio 30. Por una parte,

$$(\lambda\mathbf{A})_{\tau_1 \dots \tau_k} = (\lambda\mathbf{A})\left(\mathbf{I}_{\tau_1 \dots \tau_k}\right) = \lambda(\mathbf{A}\left(\mathbf{I}_{\tau_1 \dots \tau_k}\right)) = \lambda(\mathbf{A}_{\tau_1 \dots \tau_k}),$$

y por otra

$$(\mathbf{A} + \mathbf{B})_{\tau_1 \dots \tau_k} = (\mathbf{A} + \mathbf{B})\left(\mathbf{I}_{\tau_1 \dots \tau_k}\right) = \mathbf{A}\left(\mathbf{I}_{\tau_1 \dots \tau_k}\right) + \mathbf{B}\left(\mathbf{I}_{\tau_1 \dots \tau_k}\right) = \mathbf{A}_{\tau_1 \dots \tau_k} + \mathbf{B}_{\tau_1 \dots \tau_k}.$$

□

Ejercicio 31(a) Una posible sucesión de transformaciones elementales por columnas es:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow[\text{Tipo I}]{[(1)\tau_1+2]} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow[\text{Tipo I}]{[(-1)\tau_2+1]} \begin{bmatrix} 0 & 1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow[\text{Tipo I}]{[(1)\tau_1+2]} \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow[\text{Tipo II}]{[(-1)\tau_1]} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

es decir, $\mathbf{I}_{\tau_{[1 \rightleftharpoons 2]}} = \mathbf{I}_{[(1)\tau_1+2][(-1)\tau_2+1][(1)\tau_1+2][(-1)\tau_1]}.$

□

Ejercicio 31(b) Hay muchas combinaciones posibles. Por ejemplo: $\mathbf{I}_{\tau_{[i \rightleftharpoons j]}} = \mathbf{I}_{[(-1)\tau_j][(-1)\tau_j+i][(1)\tau_i+j][(-1)\tau_j+i]}.$

□

Ejercicio 32(a) Por una parte $\mathbf{ABB}^{-1}\mathbf{A}^{-1} = \mathbf{AIA}^{-1} = \mathbf{I}$, y por otra $\mathbf{B}^{-1}\mathbf{A}^{-1}\mathbf{AB} = \mathbf{B}^{-1}\mathbf{IB} = \mathbf{I}$.

□

Ejercicio 32(b) Basta aplicar iterativamente la proposición anterior:

$$(\mathbf{A}_1 \cdots \mathbf{A}_k)^{-1} = (\mathbf{A}_2 \cdots \mathbf{A}_k)^{-1} \mathbf{A}_1^{-1} = (\mathbf{A}_3 \cdots \mathbf{A}_k)^{-1} \mathbf{A}_2^{-1} \mathbf{A}_1^{-1} = \cdots = (\mathbf{A}_{k-1} \mathbf{A}_k)^{-1} \mathbf{A}_{k-2}^{-1} \cdots \mathbf{A}_1^{-1} = \mathbf{A}_k^{-1} \cdots \mathbf{A}_1^{-1}.$$

□

Ejercicio 32(c) Si \mathbf{A} es invertible, entonces \mathbf{AB} es invertible (Proposición 5.1.2 en la página 61).

Supongamos que \mathbf{AB} es invertible. Entonces sabemos que existe \mathbf{C} tal que

$$(\mathbf{AB})\mathbf{C} = \mathbf{I}, \quad \text{y} \quad \mathbf{C}(\mathbf{AB}) = \mathbf{I}.$$

De la primera igualdad tenemos que \mathbf{A} es invertible por la derecha, pues

$$\mathbf{A}(\mathbf{BC}) = \mathbf{I}.$$

Solo nos falta comprobar que también es invertible por la izquierda, es decir, que $(\mathbf{BC})\mathbf{A} = \mathbf{I}$. Para ello nos fijamos en la segunda igualdad, y multiplicando por \mathbf{B} por la izquierda y por \mathbf{B}^{-1} por la derecha (a ambos lados de la ecuación) tenemos

$$\mathbf{BC}(\mathbf{AB})\mathbf{B}^{-1} = \mathbf{BIB}^{-1} \quad \Rightarrow \quad (\mathbf{BC})\mathbf{A} = \mathbf{I}.$$

□

Ejercicio 32(d) Si \mathbf{B} es invertible, entonces \mathbf{AB} es invertible (Proposición 5.1.2 en la página 61).

Supongamos que \mathbf{AB} es invertible. Entonces sabemos que existe \mathbf{C} tal que

$$(\mathbf{AB})\mathbf{C} = \mathbf{I}, \quad \text{y} \quad \mathbf{C}(\mathbf{AB}) = \mathbf{I}.$$

De la segunda igualdad tenemos que \mathbf{B} es invertible por la izquierda, pues

$$(\mathbf{CA})\mathbf{B} = \mathbf{I}.$$

Solo nos falta comprobar que también es invertible por la derecha, es decir, que $\mathbf{B}(\mathbf{CA}) = \mathbf{I}$. Para ello nos fijamos en la primera igualdad y multiplicando por \mathbf{A}^{-1} por la izquierda y por \mathbf{A} por la derecha (a ambos lados de la ecuación) tenemos

$$\mathbf{A}^{-1}(\mathbf{AB})\mathbf{CA} = \mathbf{A}^{-1}\mathbf{IA}, \quad \Rightarrow \quad (\mathbf{BC})\mathbf{A} = \mathbf{I}.$$

□

Ejercicio 32(e) Aplicando la pista tenemos por un lado: $\mathbf{A}^{\top}(\mathbf{A}^{-1})^{\top} = (\mathbf{A}^{-1}\mathbf{A})^{\top} = \mathbf{I}^{\top} = \mathbf{I}$ y por el otro $(\mathbf{A}^{-1})^{\top}\mathbf{A}^{\top} = (\mathbf{AA}^{-1})^{\top} = \mathbf{I}^{\top} = \mathbf{I}$. Es decir, $(\mathbf{A}^{-1})^{\top} = (\mathbf{A}^{\top})^{-1}$.

□

Ejercicio 32(f) Sea \mathbf{A} de orden n una matriz cuya i -ésima columna es nula: $\mathbf{A}_{|j} = \mathbf{0}$. Si asumimos que existe \mathbf{A}^{-1} llegamos a una contradicción. Para verlo multipliquemos la ecuación por \mathbf{A}^{-1} :

$$\mathbf{A}^{-1}(\mathbf{A}_{|j}) = \mathbf{A}^{-1}(\mathbf{0}) \quad \Rightarrow \quad (\mathbf{A}^{-1}\mathbf{A})_{|j} = \mathbf{0} \quad \Rightarrow \quad \mathbf{I}_{|j} = \mathbf{0}.$$

Pero sabemos que $\mathbf{I}_{|i}$ (la columna i -ésima de \mathbf{I}) no es $\mathbf{0}$, así que hemos llegado a una contradicción. Por tanto \mathbf{A}^{-1} no puede existir: así, \mathbf{A} es singular (no tiene inversa) si tiene alguna columna nula.

Para las filas, basta repetir lo anterior con \mathbf{A}^T pues, si ${}_i\mathbf{A} = (\mathbf{A}^T)_{|i} = \mathbf{0}$, llegamos a la misma contradicción. □

Ejercicio 33. Si para alguna columna $\mathbf{A}_{|j}$ se verifica que $\mathbf{A}_{|j} = \mathbf{A}\mathbf{x}$ (con $x_j = 0$) es decir, $\mathbf{A}_{|j}$ es una combinación lineal del resto de columnas, entonces mediante $(n-1)$ transformaciones elementales ${}^{[(-x_k)\mathbf{k}+j]}$ (con $k = 1 : n$ excepto $k = j$) es posible transformar la columna $\mathbf{A}_{|j}$ en un vector de ceros, es decir, hay transformaciones elementales tales que $(\mathbf{A}_{\tau_1 \dots \tau_{(n-1)}})_{|j} = \mathbf{0}$, pues a la columna $\mathbf{A}_{|j}$ se le resta $\mathbf{A}\mathbf{x}$:

$$\mathbf{A}(\mathbf{I}_{\tau_1 \dots \tau_{(n-1)}}) = \mathbf{A} \begin{matrix} \tau \\ [(-x_1)\mathbf{1}+j] \\ [(-x_2)\mathbf{2}+j] \\ \vdots \\ [(-x_n)\mathbf{n}+j] \end{matrix} = \mathbf{A} \begin{bmatrix} 1 & & -x_1 & & \\ & 1 & -x_2 & & \\ & & \ddots & \ddots & \\ & & & 1 & \\ & & & \vdots & \ddots \\ & & & -x_n & 1 \end{bmatrix} = [\mathbf{A}_{|1}; \mathbf{A}_{|2}; \dots \mathbf{0}; \dots \mathbf{A}_{|n}],$$

Como $\mathbf{I}_{\tau_1 \dots \tau_{(n-1)}}$ es invertible y el resultado es una matriz singular (por ser la columna j -ésima nula), \mathbf{A} es singular (Proposición 5.1.4 en la página 62).

Para las filas, basta transponer \mathbf{A} y repetir el argumento con las columnas de \mathbf{A}^T . □

Ejercicio 34(a) En particular, también es cierto para $\vec{x} = \vec{0}$; así tenemos que

$$\begin{aligned} \vec{z} + \vec{0} &= \vec{z} & \text{por la propiedad que hemos pedido a } \vec{z} \\ \vec{z} &= \vec{0} & \text{por la propiedad que tiene } \vec{0}. \end{aligned}$$

□

Ejercicio 34(b) $\vec{z} = \vec{z} + \vec{0} = \vec{z} + (\vec{x} + (-\vec{x})) = (\vec{z} + \vec{x}) + (-\vec{x}) = \vec{0} + (-\vec{x}) = -\vec{x}$.

□

Ejercicio 35.

1. $\vec{x} + \vec{y} = xy = yx = \vec{y} + \vec{x}$.
2. $\vec{x} + (\vec{y} + \vec{z}) = x(yz) = (xy)z = (\vec{x} + \vec{y}) + \vec{z}$.
3. Si $\vec{0} = 1$; entonces $\vec{x} + \vec{0} = x1 = x = \vec{x}$; por tanto $\vec{0} = 1$.

4. Si $-\vec{x} = 1/x$; entonces $\vec{x} + (-\vec{x}) = x/x = 1 = \vec{0}$.

5. $1\vec{x} = x^1 = x = \vec{x}$.

6. $(ab)\vec{x} = x^{(ab)} = (x^b)^a = (b\vec{x})^a = a(b\vec{x})$.

7. $a(\vec{x} + \vec{y}) = (xy)^a = x^a \cdot y^a = a\vec{x} + a\vec{y}$.

8. $(a+b)\vec{x} = (x)^{a+b} = x^a \cdot x^b = a\vec{x} + b\vec{x}$.

Por tanto, en este espacio vectorial el vector nulo es $\vec{0} = 1$ y el vector opuesto de \vec{x} es $-\vec{x} = 1/x$.

□

Ejercicio 36(a) Esquema de la función lineal:
$$\begin{aligned} -|_j : \mathbb{R}^{m \times n} &\longrightarrow \mathbb{R}^m \\ \mathbf{A} &\longmapsto (a_{1j}, \dots, a_{mj}) \end{aligned}$$

Es lineal puesto que

$$\begin{aligned} (\mathbf{A} + \mathbf{B})|_j &= \mathbf{A}|_j + \mathbf{B}|_j \\ (\lambda \mathbf{A})|_j &= \lambda(\mathbf{A}|_j). \end{aligned}$$

□

Ejercicio 36(b) Esquema de la función lineal:
$$\begin{aligned} {}_i|-|_j : \mathbb{R}^{m \times n} &\longrightarrow \mathbb{R} \\ \mathbf{A} &\longmapsto a_{ij} \end{aligned}$$

Es lineal puesto que

$$\begin{aligned} {}_i|(\mathbf{A} + \mathbf{B})|_j &= {}_i|\mathbf{A}|_j + {}_i|\mathbf{B}|_j \\ {}_i|(\lambda \mathbf{A})|_j &= \lambda({}_i|\mathbf{A}|_j). \end{aligned}$$

□

Ejercicio 36(c) Esquema de la función lineal:
$$\begin{aligned} -^\top : \mathbb{R}^{m \times n} &\longrightarrow \mathbb{R}^{n \times m} \\ \mathbf{A} &\longmapsto [{}_1|\mathbf{A}; \dots {}_m|\mathbf{A}] \end{aligned}$$

Es lineal puesto que

$$\begin{aligned} (\mathbf{A} + \mathbf{B})^\top &= \mathbf{A}^\top + \mathbf{B}^\top \\ (\lambda \mathbf{A})^\top &= \lambda(\mathbf{A}^\top). \end{aligned}$$

□

Ejercicio 36(d) Esquema de la función lineal:
$$\begin{aligned} {}_i|_- : \mathbb{R}^{m \times n} &\longrightarrow \mathbb{R}^n \\ \mathbf{A} &\longmapsto (a_{m1}, \dots, a_{mn}) \end{aligned}$$

Es lineal puesto que

$$\begin{aligned} {}_i|(\mathbf{A} + \mathbf{B}) &= {}_j|\mathbf{A} + {}_i|\mathbf{B} \\ {}_i|(\lambda \mathbf{A}) &= \lambda({}_i|\mathbf{A}). \end{aligned}$$

□

Ejercicio 36(e) Esquema de la función lineal: $\begin{array}{l} \mathbf{A} _ : \mathbb{R}^{m \times n} \longrightarrow \mathbb{R}^n \\ \mathbf{A} \longmapsto (\mathbf{A}_{|1})b_1 + (\mathbf{A}_{|2})b_2 + \cdots + (\mathbf{A}_{|n})b_n \end{array}.$

Es lineal puesto que

$$\begin{aligned} (\mathbf{A} + \mathbf{C})\mathbf{b} &= \mathbf{A}\mathbf{b} + \mathbf{C}\mathbf{b} \\ (\lambda\mathbf{A})\mathbf{b} &= \lambda(\mathbf{A}\mathbf{b}). \end{aligned}$$

□

Ejercicio 36(f) Esquema de la función lineal: $\begin{array}{l} \mathbf{A} _ : \mathbb{R}^n \longrightarrow \mathbb{R}^n \\ \mathbf{b} \longmapsto (\mathbf{A}_{|1})b_1 + (\mathbf{A}_{|2})b_2 + \cdots + (\mathbf{A}_{|n})b_n \end{array}.$

Es lineal puesto que

$$\begin{aligned} \mathbf{A}(\mathbf{b} + \mathbf{c}) &= \mathbf{A}\mathbf{b} + \mathbf{A}\mathbf{c} \\ \mathbf{A}(\lambda\mathbf{b}) &= \lambda(\mathbf{A}\mathbf{b}). \end{aligned}$$

□

Ejercicio 36(g) Esquema de la función lineal: $\begin{array}{l} \mathbf{a} _ : \mathbb{R}^{m \times n} \longrightarrow \mathbb{R}^m \\ \mathbf{B} \longmapsto a_1(\mathbf{B}_{|1}) + \cdots + a_m(\mathbf{B}_{|m}) \end{array}.$

Es lineal puesto que

$$\begin{aligned} \mathbf{a}(\mathbf{B} + \mathbf{C}) &= \mathbf{a}\mathbf{B} + \mathbf{a}\mathbf{C} \\ \mathbf{a}(\lambda\mathbf{B}) &= \lambda(\mathbf{a}\mathbf{B}). \end{aligned}$$

□

Ejercicio 36(h) Esquema de la función lineal: $\begin{array}{l} \mathbf{a} _ : \mathbb{R}^m \longrightarrow \mathbb{R}^m \\ \mathbf{a} \longmapsto a_1(\mathbf{B}_{|1}) + \cdots + a_m(\mathbf{B}_{|m}) \end{array}.$

Es lineal puesto que

$$\begin{aligned} (\mathbf{a} + \mathbf{c})\mathbf{B} &= \mathbf{a}\mathbf{B} + \mathbf{c}\mathbf{B} \\ (\lambda\mathbf{a})\mathbf{B} &= \lambda(\mathbf{a}\mathbf{B}). \end{aligned}$$

□

Ejercicio 36(i) Esquema de la función lineal: $\begin{array}{l} \mathbf{A} _ : \mathbb{R}^{m \times n} \longrightarrow \mathbb{R}^{m \times p} \\ \mathbf{A} \longmapsto [\mathbf{A}\mathbf{B}_{|1}; \dots \mathbf{A}\mathbf{B}_{|p}] \end{array}.$

Es lineal puesto que

$$\begin{aligned} (\mathbf{A} + \mathbf{C})\mathbf{B} &= \mathbf{A}\mathbf{B} + \mathbf{C}\mathbf{B} \\ (\lambda\mathbf{A})\mathbf{B} &= \lambda(\mathbf{A}\mathbf{B}). \end{aligned}$$

□

Ejercicio 36(j) Esquema de la función lineal: $\begin{array}{l} \mathbf{B} : \mathbb{R}^{m \times n} \longrightarrow \mathbb{R}^{k \times n} \\ \mathbf{B} \longmapsto [\mathbf{AB}; \dots_k \mathbf{AB}]^T \end{array}$.

Es lineal puesto que

$$\begin{aligned} \mathbf{A}(\mathbf{B} + \mathbf{C}) &= \mathbf{AB} + \mathbf{AC} \\ \mathbf{A}(\lambda \mathbf{B}) &= \lambda(\mathbf{AB}). \end{aligned}$$

□

Ejercicio 36(k) Esquema de la función lineal: $\mathbf{A} \longmapsto \mathbf{A}(\mathbf{I}_\tau)$ donde \mathbf{I}_τ es una matriz elemental del

$$\mathbf{A} \longmapsto \mathbf{A}(\mathbf{I}_\tau)$$

orden n . Es lineal puesto que

$$\begin{aligned} (\mathbf{A} + \mathbf{B})_\tau &= \mathbf{A}_\tau + \mathbf{B}_\tau \\ (\lambda \mathbf{A})_\tau &= \lambda(\mathbf{A}_\tau). \end{aligned}$$

□

Ejercicio 36(l) Esquema de la función lineal: $\mathbf{A} \longmapsto (\mathbf{I}_\tau \mathbf{A})$ donde \mathbf{I}_τ es una matriz elemental del

$$\mathbf{A} \longmapsto (\mathbf{I}_\tau \mathbf{A})$$

orden m . Es lineal puesto que

$$\begin{aligned} \mathbf{I}_\tau(\mathbf{A} + \mathbf{B}) &= \mathbf{I}_\tau \mathbf{A} + \mathbf{I}_\tau \mathbf{B} \\ \mathbf{I}_\tau(\lambda \mathbf{A}) &= \lambda(\mathbf{I}_\tau \mathbf{A}). \end{aligned}$$

□

Ejercicio 37(a) Tenemos que comprobar que la función $[g \circ f]: \mathcal{D} \rightarrow \mathcal{W}$ es lineal.

Dados \vec{x} e \vec{y} de \mathcal{D} se verifica que

$$[g \circ f](\vec{x} + \vec{y}) = g(f(\vec{x} + \vec{y})) = g(f(\vec{x}) + f(\vec{y})) = g(f(\vec{x})) + g(f(\vec{y})) = [g \circ f](\vec{x}) + [g \circ f](\vec{y}).$$

Dados \vec{x} de \mathcal{D} y $\alpha \in \mathbb{R}$ se verifica que

$$[g \circ f](\alpha \vec{x}) = g(f(\alpha \vec{x})) = g(\alpha f(\vec{x})) = \alpha g(f(\vec{x})) = \alpha [g \circ f](\vec{x}).$$

□

Ejercicio 37(b) Tomemos \vec{x} e \vec{y} de \mathcal{V} , entonces

$$\begin{aligned} f^{-1}(\vec{x} + \vec{y}) &= f^{-1}(f(f^{-1}(\vec{x})) + f(f^{-1}(\vec{y}))) && \text{puesto que } \vec{x} = f(f^{-1}(\vec{x})) \text{ e } \vec{y} = f(f^{-1}(\vec{y})) \\ &= f^{-1}(f(f^{-1}(\vec{x}) + f^{-1}(\vec{y}))) && \text{puesto que } f \text{ es lineal} \\ &= f^{-1}(\vec{x}) + f^{-1}(\vec{y}) && \text{puesto que en general } \vec{z} = f^{-1}(f(\vec{z})). \end{aligned}$$

Tomemos $\vec{x} \in \mathcal{V}$ y $\alpha \in \mathbb{R}$ entonces

$$\begin{aligned} f^{-1}(\alpha \vec{x}) &= f^{-1}(\alpha f(f^{-1}(\vec{x}))) && \text{puesto que } \vec{x} = f(f^{-1}(\vec{x})) \\ &= f^{-1}(f(\alpha f^{-1}(\vec{x}))) && \text{puesto que } f \text{ es lineal} \\ &= \alpha f^{-1}(\vec{x}) && \text{puesto que en general } \vec{z} = f^{-1}(f(\vec{z})). \end{aligned}$$

□

Ejercicio 38. Si, lo es; pues la suma de dos combinaciones lineales es una combinación lineal; y cualquier múltiplo de una combinación lineal también es una combinación lineal. Por tanto el conjunto es cerrado para la suma y el producto por escalares.

□

Ejercicio 39. Hay que verificar que se cumplen las ocho propiedades de la Definición 6.1:

Sean los pares (\vec{a}, \vec{x}) , (\vec{b}, \vec{y}) y (\vec{c}, \vec{z}) , donde las primeras componentes son vectores de \mathcal{A} y las segundas componentes son vectores de \mathcal{X} ; y sean los escalares λ y μ . Entonces

1.

$$\begin{aligned} (\vec{a}, \vec{x}) + (\vec{b}, \vec{y}) &= (\vec{a} + \vec{b}, \vec{x} + \vec{y}) && \text{aplicando la definición de suma de pares} \\ &= (\vec{b} + \vec{a}, \vec{y} + \vec{x}) && \text{puesto que } \mathcal{A} \text{ y } \mathcal{X} \text{ son espacios vectoriales} \\ &= (\vec{b}, \vec{y}) + (\vec{a}, \vec{x}) && \text{aplicando la definición de suma de pares.} \end{aligned}$$

2.

$$\begin{aligned} (\vec{a}, \vec{x}) + ((\vec{b}, \vec{y}) + (\vec{c}, \vec{z})) &= (\vec{a}, \vec{x}) + ((\vec{b} + \vec{c}, \vec{y} + \vec{z})) && \text{aplicando la definición de suma de pares} \\ &= (\vec{a} + (\vec{b} + \vec{c}), \vec{x} + (\vec{y} + \vec{z})) && \text{aplicando la definición de suma de pares} \\ &= ((\vec{a} + \vec{b}) + \vec{c}, (\vec{x} + \vec{y}) + \vec{z}) && \text{puesto que } \mathcal{A} \text{ y } \mathcal{X} \text{ son espacios vectoriales} \\ &= ((\vec{a} + \vec{b}), (\vec{x} + \vec{y})) + (\vec{c}, \vec{z}) && \text{aplicando la definición de suma de pares} \\ &= ((\vec{a}, \vec{x}) + (\vec{b}, \vec{y})) + (\vec{c}, \vec{z}). \end{aligned}$$

3.

$$\begin{aligned} (\vec{a}, \vec{x}) + (\vec{0}_{\mathcal{A}}, \vec{0}_{\mathcal{X}}) &= (\vec{a} + \vec{0}_{\mathcal{A}}, \vec{x} + \vec{0}_{\mathcal{X}}) && \text{aplicando la definición de suma de funciones} \\ &= (\vec{a}, \vec{x}) && \text{puesto que } \mathcal{A} \text{ y } \mathcal{X} \text{ son espacios vectoriales} \\ &&& \text{por tanto } \vec{0} = (\vec{0}_{\mathcal{A}}, \vec{0}_{\mathcal{X}}). \end{aligned}$$

4.

$$\begin{aligned} (\vec{a}, \vec{x}) + (-\vec{a}, -\vec{x}) &= (\vec{a} - \vec{a}, \vec{x} - \vec{x}) && \text{aplicando la definición de suma de funciones} \\ &= (\vec{0}_{\mathcal{A}}, \vec{0}_{\mathcal{X}}) && \text{puesto que } \mathcal{A} \text{ y } \mathcal{X} \text{ son espacios vectoriales} \\ &&& \text{por tanto } (-\vec{a}, -\vec{x}) = -(\vec{a}, \vec{x}). \end{aligned}$$

5.

$$\begin{aligned} \lambda((\vec{a}, \vec{x}) + (\vec{b}, \vec{y})) &= \lambda((\vec{a} + \vec{b}, \vec{x} + \vec{y})) && \text{aplicando la definición de suma de pares} \\ &= (\lambda(\vec{a} + \vec{b}), \lambda(\vec{x} + \vec{y})) && \text{aplicando la definición de escalar por par} \\ &= (\lambda\vec{a} + \lambda\vec{b}, \lambda\vec{x} + \lambda\vec{y}) && \text{puesto que } \mathcal{A} \text{ y } \mathcal{X} \text{ son espacios vectoriales} \\ &= (\lambda\vec{a}, \lambda\vec{x}) + (\lambda\vec{b}, \lambda\vec{y}) && \text{aplicando la definición de suma de pares} \\ &= \lambda(\vec{a}, \vec{x}) + \lambda(\vec{b}, \vec{y}) && \text{aplicando la definición de escalar por par} \end{aligned}$$

6.

$$\begin{aligned}
 (\lambda + \mu)(\vec{a}, \vec{x}) &= ((\lambda + \mu)\vec{a}, (\lambda + \mu)\vec{x}) && \text{aplicando la definición de escalar por par} \\
 &= ((\lambda\vec{a} + \mu\vec{a}), (\lambda\vec{x} + \mu\vec{x})) && \text{puesto que } \mathcal{A} \text{ y } \mathcal{X} \text{ son espacios vectoriales} \\
 &= (\lambda\vec{a}, \lambda\vec{x}) + (\mu\vec{a}, \mu\vec{x}) && \text{aplicando la definición de suma de pares} \\
 &= \lambda(\vec{a}, \vec{x}) + \mu(\vec{a}, \vec{x}) && \text{aplicando la definición de escalar por par}
 \end{aligned}$$

7.

$$\begin{aligned}
 (\lambda\mu)(\vec{a}, \vec{x}) &= ((\lambda\mu)\vec{a}, (\lambda\mu)\vec{x}) && \text{aplicando la definición de escalar por par} \\
 &= (\lambda(\mu\vec{a}), \lambda(\mu\vec{x})) && \text{puesto que } \mathcal{A} \text{ y } \mathcal{X} \text{ son espacios vectoriales} \\
 &= \lambda(\mu\vec{a}, \mu\vec{x}) && \text{aplicando la definición de escalar por par}
 \end{aligned}$$

8.

$$\begin{aligned}
 1(\vec{a}, \vec{x}) &= (1\vec{a}, 1\vec{x}) && \text{aplicando la definición de escalar por par} \\
 &= (\vec{a}, \vec{x}) && \text{puesto que } \mathcal{A} \text{ y } \mathcal{X} \text{ son espacios vectoriales}
 \end{aligned}$$

□

Ejercicio 40. Hay que verificar que se cumplen las ocho propiedades de la Definición 6.1:

Sean las funciones $f: \mathcal{D} \rightarrow \mathcal{V}$, $g: \mathcal{D} \rightarrow \mathcal{V}$, y $h: \mathcal{D} \rightarrow \mathcal{V}$; y sean los escalares λ y μ . Entonces

1.

$$\begin{aligned}
 [f + g](\vec{x}) &= f(\vec{x}) + g(\vec{x}) && \text{aplicando la definición de suma de funciones} \\
 &= g(\vec{x}) + f(\vec{x}) && \text{puesto que } \mathcal{V} \text{ es un espacio vectorial} \\
 &= [g + f](\vec{x}) && \text{aplicando la definición de suma de funciones}
 \end{aligned}$$

2.

$$\begin{aligned}
 [f + (g + h)](\vec{x}) &= f(\vec{x}) + [g + h](\vec{x}) && \text{aplicando la definición de suma de funciones} \\
 &= f(\vec{x}) + (g(\vec{x}) + h(\vec{x})) && \text{aplicando la definición de suma de funciones} \\
 &= (f(\vec{x}) + g(\vec{x})) + h(\vec{x}) && \text{puesto que } \mathcal{V} \text{ es un espacio vectorial} \\
 &= [f + g](\vec{x}) + h(\vec{x}) && \text{aplicando la definición de suma de funciones} \\
 &= [(f + g) + h](\vec{x}) && \text{aplicando la definición de suma de funciones}
 \end{aligned}$$

3. Sea $0: X \rightarrow \mathcal{V}$, entonces
 $\vec{x} \mapsto \vec{0}$

$$\begin{aligned}
 [f + 0](\vec{x}) &= f(\vec{x}) + 0(\vec{x}) && \text{aplicando la definición de suma de funciones} \\
 &= f(\vec{x}) + \vec{0} && \text{pues } 0(\vec{x}) = \vec{0} \\
 &= f(\vec{x})
 \end{aligned}$$

4. Sea $-f: X \rightarrow \mathcal{V}$, entonces
 $\vec{x} \mapsto -f(\vec{x})$

$$\begin{aligned} [f + (-f)](\vec{x}) &= f(\vec{x}) + (-f(\vec{x})) && \text{aplicando la definición de suma de funciones} \\ &= \vec{0} && \text{puesto que en } \mathcal{V} \text{ se verifica que } f(\vec{x}) + (-f(\vec{x})) = \vec{0} \\ &= 0(\vec{x}) && \text{pues } 0(\vec{x}) = \vec{0} \end{aligned}$$

5.

$$\begin{aligned} [\lambda(f + g)](\vec{x}) &= \lambda([f + g](\vec{x})) && \text{aplicando la definición de función por escalar} \\ &= \lambda(f(\vec{x}) + g(\vec{x})) && \text{aplicando la definición de suma de funciones} \\ &= \lambda f(\vec{x}) + \lambda g(\vec{x}) && \text{puesto que } \mathcal{V} \text{ es un espacio vectorial} \\ &= [\lambda f](\vec{x}) + [\lambda g](\vec{x}) && \text{aplicando la definición de función por escalar} \\ &= [\lambda f + \lambda g](\vec{x}) && \text{aplicando la definición de suma de funciones} \end{aligned}$$

6.

$$\begin{aligned} [(\lambda + \mu)f](\vec{x}) &= (\lambda + \mu)f(\vec{x}) && \text{aplicando la definición de función por escalar} \\ &= \lambda f(\vec{x}) + \mu f(\vec{x}) && \text{puesto que } \mathcal{V} \text{ es un espacio vectorial} \\ &= [\lambda f](\vec{x}) + [\mu f](\vec{x}) && \text{aplicando la definición de función por escalar} \\ &= [\lambda f + \mu f](\vec{x}) && \text{aplicando la definición de suma de funciones} \end{aligned}$$

7.

$$\begin{aligned} [(\lambda \cdot \mu)f](\vec{x}) &= (\lambda \cdot \mu)f(\vec{x}) && \text{aplicando la definición de función por escalar} \\ &= \lambda(\mu f(\vec{x})) && \text{puesto que } \mathcal{V} \text{ es un espacio vectorial} \\ &= \lambda([\mu f](\vec{x})) && \text{aplicando la definición de función por escalar} \\ &= [\lambda(\mu f)](\vec{x}) && \text{aplicando la definición de función por escalar} \end{aligned}$$

8.

$$\begin{aligned} [1 \cdot f](\vec{x}) &= 1 \cdot f(\vec{x}) && \text{aplicando la definición de función por escalar} \\ &= f(\vec{x}) && \text{puesto que } \mathcal{V} \text{ es un espacio vectorial} \end{aligned}$$

□

Ejercicio 41(a) Tenemos que comprobar que la función $[g + f]: \mathcal{D} \rightarrow \mathcal{W}$ es lineal.

Dados \vec{x} e \vec{y} de \mathcal{D} se verifica que

$$[g + f](\vec{x} + \vec{y}) = g(\vec{x} + \vec{y}) + f(\vec{x} + \vec{y}) = g(\vec{x}) + g(\vec{y}) + f(\vec{x}) + f(\vec{y}) = g(\vec{x}) + f(\vec{x}) + g(\vec{y}) + f(\vec{y}) = [g + f](\vec{x}) + [g + f](\vec{y}).$$

Dados \vec{x} de \mathcal{D} y $\alpha \in \mathbb{R}$ se verifica que

$$[g + f](\alpha \vec{x}) = g(\alpha \vec{x}) + f(\alpha \vec{x}) = \alpha g(\vec{x}) + \alpha f(\vec{x}) = \alpha(g(\vec{x}) + f(\vec{x})) = \alpha[g + f](\vec{x}).$$

□

Ejercicio 41(b) Tenemos que comprobar que la función $[\alpha \cdot f]: \mathcal{D} \rightarrow \mathcal{W}$ es lineal.

Dados \vec{x} e \vec{y} de \mathcal{D} se verifica que

$$[\alpha \cdot f](\vec{x} + \vec{y}) = \alpha f(\vec{x} + \vec{y}) = \alpha(f(\vec{x}) + f(\vec{y})) = \alpha f(\vec{x}) + \alpha f(\vec{y}) = [\alpha \cdot f](\vec{x}) + [\alpha \cdot f](\vec{y}).$$

Dados \vec{x} de \mathcal{D} y $\gamma \in \mathbb{R}$ se verifica que

$$[\alpha \cdot f](\gamma \vec{x}) = \alpha f(\gamma \vec{x}) = \alpha(\gamma f(\vec{x})) = \gamma(\alpha f(\vec{x})) = \gamma[\alpha \cdot f](\vec{x}).$$

□

Ejercicio 42. Basta comprobar que cualquier combinación de soluciones de $\mathbf{A}\mathbf{x} = \mathbf{0}$ es también solución.

Sean \mathbf{x} y \mathbf{y} dos vectores de $\mathcal{N}(\mathbf{A})$, es decir, $\mathbf{A}\mathbf{x} = \mathbf{0}$ y $\mathbf{A}\mathbf{y} = \mathbf{0}$, y sean b y c dos números reales; entonces

$$\mathbf{A}(b\mathbf{x} + c\mathbf{y}) = \mathbf{A}b\mathbf{x} + \mathbf{A}c\mathbf{y} = b\mathbf{A}\mathbf{x} + c\mathbf{A}\mathbf{y} = b\mathbf{0} + c\mathbf{0} = \mathbf{0} \Rightarrow (b\mathbf{x} + c\mathbf{y}) \in \mathcal{N}(\mathbf{A}).$$

□

Ejercicio 43. Sea $\boldsymbol{\rho} = (\rho_1, \dots, \rho_r)$ la lista de índices de las r columnas de \mathbf{K} con pivote, y sea $\boldsymbol{\nu}$ la lista de índices de las $n - r$ columnas nulas, es decir, $\mathbf{K}_{|\boldsymbol{\nu}} = \mathbf{0}$.

Que en $\mathbf{K}\mathbf{x}$ sean nulos los coeficientes correspondientes a las columnas no nulas significa que el subvector $(_{\boldsymbol{\rho}}\mathbf{x})$ es $\mathbf{0}$. Así, usando la Ecuación 3.7 en la página 40 deducimos que necesariamente

$$\mathbf{K}\mathbf{x} = (\mathbf{K}_{|\boldsymbol{\rho}}) \underbrace{(_{\boldsymbol{\rho}}\mathbf{x})}_{\mathbf{0}} + (\mathbf{K}_{|\boldsymbol{\nu}}) \underbrace{(_{\boldsymbol{\nu}}\mathbf{x})}_{\mathbf{0}} = \mathbf{0}.$$

Recíprocamente, supongamos que $(_{\boldsymbol{\rho}}\mathbf{x}) \neq \mathbf{0}$, es decir, que algún coeficiente es distinto de cero para alguna columna no nula de \mathbf{K} ; y sea j el menor índice de $\boldsymbol{\rho}$ tal que $_{j|\mathbf{x}} \neq 0$. Entonces, para todo $h < j$, el producto $(\mathbf{K}_{|h})(_{h|\mathbf{x}}) = \mathbf{0}$ (pues son nulos $\mathbf{K}_{|h}$ ó $_{h|\mathbf{x}}$). Por tanto, $\mathbf{K}\mathbf{x} = (\mathbf{K}_{|j})x_j + \dots + (\mathbf{K}_{|n})x_n$. Si i es la posición de pivote de $\mathbf{K}_{|j}$ tendremos que $_{i|}((\mathbf{K}_{|j})x_j + \dots + (\mathbf{K}_{|n})x_n) = (_{i|\mathbf{K}_{|j}})x_j \neq 0$; pues todas las componentes a la derecha del pivote $_{i|\mathbf{K}_{|j}}$ son nulas por estar \mathbf{K} pre-escalada. Consecuentemente $\mathbf{K}\mathbf{x} \neq \mathbf{0}$.

□

Ejercicio 44(a)

$$\mathbf{Z}(\mathbf{b} + \mathbf{c}) = \sum_{k=1}^n \left(\mathbf{Z}_{|k}(b_k + c_k) \right) = \sum_{k=1}^n \left((\mathbf{Z}_{|k})b_k + (\mathbf{Z}_{|k})c_k \right) = \sum_{k=1}^n (\mathbf{Z}_{|k})b_k + \sum_{k=1}^n (\mathbf{Z}_{|k})c_k = \mathbf{Z}\mathbf{b} + \mathbf{Z}\mathbf{c}.$$

□

Ejercicio 44(b) $\mathbf{Z}(\lambda \mathbf{b}) = \sum_{k=1}^n (\mathbf{Z}_{|k})\lambda b_k = \lambda \sum_{k=1}^n (\mathbf{Z}_{|k})b_k = \lambda(\mathbf{Z}\mathbf{b}).$

□

Ejercicio 45(a)

$$\mathbf{Z}(\mathbf{B}\mathbf{c}) = \mathbf{Z} \left(\sum_{k=1}^n (\mathbf{B}_{|k})c_k \right) = \sum_{k=1}^n \left(\mathbf{Z}(\mathbf{B}_{|k})c_k \right) = \sum_{k=1}^n \left((\mathbf{Z}\mathbf{B})_{|k} \right) c_k = \left[\mathbf{Z}\mathbf{B}_{|1}; \dots; \mathbf{Z}\mathbf{B}_{|n} \right] \mathbf{c} = (\mathbf{Z}\mathbf{B})\mathbf{c}.$$

□

Ejercicio 45(b) Puesto que $\mathbf{C}_{|j}$ es un vector de \mathbb{R}^q , aplicando las definiciones de producto de sistema por matriz, y matriz por matriz junto con la proposición anterior, tenemos:

$$\left(Z(\mathbf{BC}) \right)_{|j} = Z(\mathbf{BC})_{|j} = Z\left(\mathbf{B}(\mathbf{C}_{|j}) \right) \underset{(\text{Prop. 9.1.3})}{=} \left(Z\mathbf{B} \right)(\mathbf{C}_{|j}) = \left((Z\mathbf{B})\mathbf{C} \right)_{|j}, \quad \forall j = \{1, \dots, n\}.$$

□

Ejercicio 46. Sea la lista de índices menores que i , $\gamma_1 = (h \mid h < i)$ y la lista de índices mayores que i , $\gamma_2 = (h \mid h > i)$, de manera que la lista de índices $(1 : n)$ de 1 a n es la concatenación de la lista γ_1 , con la lista de un solo elemento $(i,)$ y la lista γ_2 ; es decir, $(1 : n) = (\gamma_1 \# (i,) \# \gamma_2)$.

Por una parte, si \vec{z}_i es combinación lineal del resto de vectores de Z (los índice menor que i y los de índice mayor que i):

$$Z_{|i} = (Z_{|\gamma_1})\mathbf{b} + (Z_{|\gamma_2})\mathbf{c};$$

entonces restando $Z_{|i}$ de ambos lados (y recordando la Ecuación 3.7 en la página 40) tenemos que

$$\vec{0} = (Z_{|\gamma_1})\mathbf{b} + \underbrace{(Z_{|(i,)})(-1,)}_{-Z_{|i}} + (Z_{|\gamma_2})\mathbf{c} = Z\mathbf{a},$$

donde \mathbf{a} es la concatenación de los vectores \mathbf{b} , $(-1,)$ y \mathbf{c} , es decir $\mathbf{a} = \mathbf{b} \# (-1,) \# \mathbf{c}$.

Por otra parte, si $\vec{0} = Z\mathbf{a}$, con $\mathbf{a} \neq \mathbf{0}$, entonces existe un i tal que ${}_i\mathbf{a} \neq 0$, y por lo tanto

$$\vec{0} = Z\mathbf{a} = (Z_{|\gamma_1})(\gamma_1|\mathbf{a}) + \underbrace{(Z_{|(i,)})({}_{(i,)}|\mathbf{a})}_{a_i Z_{|i}} + (Z_{|\gamma_2})(\gamma_2|\mathbf{a}).$$

Restando $(Z_{|(i,)})({}_{(i,)}|\mathbf{a})$ de ambos lados tenemos que

$$-a_i Z_{|i} = (Z_{|\gamma_1})(\gamma_1|\mathbf{a}) + (Z_{|\gamma_2})(\gamma_2|\mathbf{a}) \implies Z_{|i} = \frac{-1}{a_i} (Z_{|(\gamma_1 \# \gamma_2)})({}_{(\gamma_1 \# \gamma_2)}|\mathbf{a}).$$

□

Ejercicio 47. Basta expresar la imagen de f_Z con la notación funcional (6.6) y comprobar que coincide con la notación matricial de la Definición 9.2 de espacio engendrado por Z (9.1):

$$\text{imagen}(f_Z) = \left\{ \vec{v} \in \mathcal{V} \mid \text{existe } \mathbf{a} \in \mathbb{R}^n \text{ tal que } \vec{v} = Z\mathbf{a} \right\} = \mathcal{L}(Z)$$

□

Ejercicio 48. Recordemos que $\mathbf{b} = f_{\mathbf{A}}(\mathbf{x})$ es equivalente a decir que $(\mathbf{x}, \mathbf{b}) \in f_{\mathbf{A}}$ (véase *notación funcional* en la Página 88). Así pues, la imagen de $f_{\mathbf{A}}$ es

$$\begin{aligned} \text{imagen}(f_{\mathbf{A}}) &= \{ \mathbf{b} \mid \text{existe } \mathbf{x} \text{ tal que } (\mathbf{x}, \mathbf{b}) \in f_{\mathbf{A}} \} \\ &= \{ \mathbf{b} \mid \text{existe } \mathbf{x} \in \mathbb{R}^n \text{ tal que } \mathbf{b} = f_{\mathbf{A}}(\mathbf{x}) = \mathbf{A}\mathbf{x} \} = \mathcal{C}(\mathbf{A}). \end{aligned}$$

□

Ejercicio 49. Sean $\vec{x}, \vec{y} \in \mathcal{N}(f)$, entonces dados $\alpha, \beta \in \mathbb{R}$

$$f(\alpha\vec{v} + \beta\vec{w}) = f(\alpha\vec{v}) + f(\beta\vec{w}) = \alpha f(\vec{v}) + \beta f(\vec{w}) = \alpha\vec{0} + \beta\vec{0} = \vec{0}.$$

Por tanto $(\alpha\vec{v} + \beta\vec{w}) \in \mathcal{N}(f)$. □

Ejercicio 50. Hay que demostrar que la imagen es cerrada para las combinaciones lineales.

Sea $f: \mathcal{V} \rightarrow \mathcal{W}$ lineal. Sabemos que la imagen de f es no vacía, pues al ser f lineal, su dominio \mathcal{V} es subespacio y $f(\vec{0}) = \vec{0}$.

Sean $\vec{x}, \vec{y} \in \text{imagen}(f)$, entonces existen $\vec{v}, \vec{w} \in \mathcal{V}$ tales que $f(\vec{v}) = \vec{x}$ y $f(\vec{w}) = \vec{y}$. Así, dados $\alpha, \beta \in \mathbb{R}$

$$f(\alpha\vec{v} + \beta\vec{w}) = f(\alpha\vec{v}) + f(\beta\vec{w}) = \alpha f(\vec{v}) + \beta f(\vec{w}) = \alpha\vec{x} + \beta\vec{y} \in \text{imagen}(f).$$

□

Ejercicio 51. Supongamos que $\vec{z} \in \text{imagen}(g \circ f)$. Entonces tiene que existir \vec{x} tal que $(\vec{x}, \vec{y}) \in f$ y $(\vec{y}, \vec{z}) \in g$. Por tanto $\vec{z} \in \text{imagen}(g)$. □

Ejercicio 52. Por una parte sabemos que $\text{imagen}(g \circ f) \subset \text{imagen}(g)$.

Por otra, puesto que $\text{imagen}(g) = \text{imagen}(g \circ f \circ f^{-1})$, aplicando la proposición anterior tenemos que

$$\text{imagen}(g) = \text{imagen}([g \circ f] \circ f^{-1}) \subset \text{imagen}(g \circ f).$$

□

Ejercicio 53. La función $f \circ f_B$ es lineal y va de \mathbb{R}^n a \mathcal{W} . Por tanto existe un sistema Z de vectores de \mathcal{W} tal que $f \circ f_B = f_Z$. Por tanto

$$f(\vec{v}) = f(B(\vec{v}_B)) = f(f_B(\vec{v}_B)) = f \circ f_B(\vec{v}_B) = f_Z(\vec{v}_B) = Z(\vec{v}_B).$$

□

Ejercicio 54.

Librería NAcAL para Python

```
A = Sistema([Vector([1, 0, 1, 0]), Vector([0, -1, 0, -1])])
B = Sistema([Vector([1, 1, 1, 1]), Vector([1, 0, 0, 0])])
C = A.concatena(B)      # concatenación de A y B
Homogenea(C, 1)
```

$$\begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & -1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{\begin{matrix} [(-1)\mathbf{1}+\mathbf{3}] \\ [(-1)\mathbf{1}+\mathbf{4}] \end{matrix}} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & 0 \\ \hline 1 & 0 & -1 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{[(1)\mathbf{2}+\mathbf{3}]} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & -1 & 0 & 0 \\ \hline 1 & 0 & -1 & -1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

De donde deducimos que las dos primeras componentes de la única solución especial nos dicen cómo combinar los vectores del sistema A (es decir, que la diferencia de vectores de dicho sistema es una base de la intersección). Fíjese también que las últimas dos componentes de la única solución especial nos dicen cómo combinar los vectores del sistema B (es decir, que el primer vector de dicho sistema es una base de la intersección). Comprobación:

SubEspacio(A) & SubEspacio(B) # *SubEspacio intersección*

$$\left\{ \mathbf{v} \in \mathbb{R}^4 \mid \exists \mathbf{p} \in \mathbb{R}^1, \mathbf{v} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \mathbf{p} \right\} = \left\{ \mathbf{v} \in \mathbb{R}^4 \mid \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix} \mathbf{v} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right\}$$

Librería NAcAL para Python

□

Ejercicio 55(a) Sean $\vec{x}, \vec{y} \in \mathcal{V}$, entonces

$$\begin{aligned} f(\vec{x}) &= (\vec{x}_a, \vec{x}_b) && \text{con } \vec{x}_a \in \mathcal{A}; \text{ y } \vec{x}_b \in \mathcal{B}, \text{ tales que } \vec{x} = \vec{x}_a + \vec{x}_b \\ f(\vec{y}) &= (\vec{y}_a, \vec{y}_b) && \text{con } \vec{y}_a \in \mathcal{A}; \text{ e } \vec{y}_b \in \mathcal{B}, \text{ tales que } \vec{y} = \vec{y}_a + \vec{y}_b. \end{aligned}$$

Consecuentemente

$$\alpha \vec{x} + \beta \vec{y} = \underbrace{(\alpha \vec{x}_a + \beta \vec{y}_a)}_{\in \mathcal{A}} + \underbrace{(\alpha \vec{x}_b + \beta \vec{y}_b)}_{\in \mathcal{B}}.$$

Y como, por la Proposición 10.1.4, sabemos que dicha descomposición es única, entonces sabemos que

$$\begin{aligned} f(\alpha \vec{x} + \beta \vec{y}) &= (\alpha \vec{x}_a + \beta \vec{y}_a, \alpha \vec{x}_b + \beta \vec{y}_b) && \text{aplicando } f \text{ y dado que la descomposición es única} \\ &= (\alpha \vec{x}_a, \alpha \vec{x}_b) + (\beta \vec{y}_a, \beta \vec{y}_b) && \text{aplicando la suma de productos cartesianos} \\ &= \alpha(\vec{x}_a, \vec{x}_b) + \beta(\vec{y}_a, \vec{y}_b) && \text{aplicando el producto por escalar del producto cartesiano} \\ &= \alpha f(\vec{x}) + \beta f(\vec{y}). \end{aligned}$$

□

Ejercicio 55(b) La demostración es como la anterior, pero considerando únicamente las primeras componentes de los productos cartesianos. Es decir: $f_{\mathcal{A}}(\alpha \vec{x} + \beta \vec{y}) = \alpha \vec{x}_a + \beta \vec{y}_a = \alpha f_{\mathcal{A}}(\vec{x}) + \beta f_{\mathcal{A}}(\vec{y})$.

□

Ejercicio 55(c) Debemos demostrar que $\mathcal{N}(f_{\mathcal{A}}) = \{ \vec{v} \in \mathcal{V} \mid f(\vec{v}) = \vec{0} \} = \mathcal{B}$.

Por una parte, sea $\vec{x} = \vec{x}_a + \vec{x}_b$ con $\vec{x}_a \in \mathcal{A}$ y $\vec{x}_b \in \mathcal{B}$; entonces

$$\text{si } \vec{x} \in \mathcal{N}(f_{\mathcal{A}}) \Rightarrow f(\vec{x}) = f(\vec{x}_a + \vec{x}_b) = \vec{0} \Rightarrow \vec{x}_a = \vec{0} \Rightarrow \vec{x} = \vec{x}_b \in \mathcal{B}.$$

Por otra parte, sea $\vec{x} \in \mathcal{B}$, entonces $f(\vec{x}) = f(\vec{0} + \vec{x}) = \vec{0}$ por tanto $\vec{x} \in \mathcal{N}(f_{\mathcal{A}})$.

□

Ejercicio 56. Sea $\varphi = (\varphi_1, \dots, \varphi_r)$ la lista de posiciones de los r pivotes de \mathbf{K}

$$\varphi = (h \mid h \text{ es la posición de algún pivote de } \mathbf{K}),$$

donde no hay índices repetidos y los índices φ_i están dispuestos de menor a mayor. Entonces la matriz $\varphi | \mathbf{K}$ de orden r por n es de rango completo por filas. Para demostrar que sus filas son linealmente independientes vamos a demostrar que si combinación lineal de las filas es nula, $\mathbf{x}(\varphi | \mathbf{K}) = \mathbf{0}$, entonces $\mathbf{x} = \mathbf{0} \in \mathbb{R}^r$.

Sea j la columna correspondiente al *último* pivote (posición r); como $\mathbf{x} \cdot (\varphi|\mathbf{K})_{|j} = \mathbf{0}_{|j}$ y como la única componente no nula de dicha columna es el pivote, la componente de \mathbf{x} que multiplica al pivote tiene que ser cero: $\mathbf{x}_{|r} = 0$.

Sea j la columna correspondiente al *penúltimo* pivote (posición $r-1$); como $\mathbf{x} \cdot (\varphi|\mathbf{K})_{|j} = \mathbf{0}_{|j}$ y como las componentes de \mathbf{x} con posición mayor que $r-1$ son nulas, la componente que multiplica al pivote tiene que ser cero: $\mathbf{x}_{|r-1} = 0$.

Sea j la columna correspondiente al pivote *ante-penúltimo* (posición $r-2$), entonces repitiendo el argumento se deduce que $\mathbf{x}_{|r-2} = 0$ Y así con todas las componentes de \mathbf{x} , es decir, $\mathbf{x}_{|h} = 0$ para $h = 1 : r$.

Para ver que cada fila sin pivote es combinación lineal de las filas con pivote que la preceden, basta aplicar la eliminación “de arriba a abajo” sobre las filas de \mathbf{K} para anular las componentes por debajo de los pivotes. De esta manera las filas sin pivote quedan anuladas, es decir, son combinación lineal de las filas con pivote⁴.

□

Ejercicio 57. Evidentemente $\mathbf{0}$ pertenece a A^\perp , así que el subconjunto A^\perp no es vacío.

Sean $\mathbf{c}_1, \mathbf{c}_2 \in A^\perp$ y $x, y \in \mathbb{R}$. Entonces $\mathbf{c}_1 \cdot \mathbf{a} = 0$ y $\mathbf{c}_2 \cdot \mathbf{a} = 0$, para cualquier $\mathbf{a} \in A$. Consecuentemente

$$(x\mathbf{c}_1 + y\mathbf{c}_2) \cdot \mathbf{a} = x(\mathbf{c}_1 \cdot \mathbf{a}) + y(\mathbf{c}_2 \cdot \mathbf{a}) = x \cdot 0 + y \cdot 0 = 0.$$

y por tanto $(x\mathbf{c}_1 + y\mathbf{c}_2) \in A^\perp$. Es decir, A^\perp es cerrado para las combinaciones lineales.

□

Ejercicio 58(a) $\mathbf{x} \in A^\perp$ si y solo si $\mathbf{x} \cdot \mathbf{b} = 0$ para cada $\mathbf{b} \in B$; (y como $A \subset B$) en particular también $\mathbf{x} \cdot \mathbf{a} = 0$ para cada $\mathbf{a} \in A$, es decir $\mathbf{x} \in A^\perp$.

□

Ejercicio 58(b) Tenemos que demostrar la inclusión de estos subespacios en ambos sentidos.

Inclusión en un sentido: como $A \subset \mathcal{L}(A)$, por el apartado anterior sabemos que $\mathcal{L}(A)^\perp \subset A^\perp$.

Inclusión en el otro sentido: Sabemos que $\mathbf{x} \in A^\perp$ si y solo si $\mathbf{x} \cdot \mathbf{a} = 0$ para cada $\mathbf{a} \in A$. Sea $\mathbf{z} \in \mathcal{L}(A)$, entonces existen $\alpha_1, \dots, \alpha_k \in \mathbb{R}$ y $\mathbf{a}_1, \dots, \mathbf{a}_k \in A$ tales que $\mathbf{z} = \alpha_1 \mathbf{a}_1 + \dots + \alpha_k \mathbf{a}_k$. Entonces $\mathbf{x} \cdot \mathbf{z} = \mathbf{x} \cdot (\alpha_1 \mathbf{a}_1 + \dots + \alpha_k \mathbf{a}_k) = \alpha_1 \mathbf{x} \cdot \mathbf{a}_1 + \dots + \alpha_k \mathbf{x} \cdot \mathbf{a}_k = \alpha_1 0 + \dots + \alpha_k 0 = 0$; es decir $\mathbf{x} \in \mathcal{L}(A)^\perp$. Por tanto $A^\perp \subset \mathcal{L}(A)^\perp$.

□

Ejercicio 58(c) Sea $\mathbf{a} \in \mathcal{A} \cap \mathcal{A}^\perp$; como $\mathbf{a} \in \mathcal{A}$ es perpendicular a todo vector en \mathcal{A}^\perp , entonces $\mathbf{a} \cdot \mathbf{a} = 0 \Rightarrow \mathbf{a} = \mathbf{0}$.

□

Ejercicio 59. Supongamos que tanto A como B contienen al menos un vector no nulo. Entonces sea \mathbf{A} una matriz cuyas k columnas son vectores de A que forman una base de $\mathcal{L}(A)$ y sea \mathbf{B} otra matriz cuyas p columnas son vectores de B que forman una base de $\mathcal{L}(B)$. Entonces, para todo $\mathbf{a} \in \mathcal{L}(A)$ existe un único $\mathbf{x} \in \mathbb{R}^k$ tal que $\mathbf{a} = \mathbf{A}\mathbf{x}$, y para todo $\mathbf{b} \in \mathcal{L}(B)$ existe un único $\mathbf{y} \in \mathbb{R}^p$ tal que $\mathbf{b} = \mathbf{B}\mathbf{y}$.

⁴aplicar la eliminación “de arriba a abajo” sobre las filas es equivalente a aplicar transformaciones elementales “de izquierda a derecha” sobre el sistema de vectores $[_1|\mathbf{L}, \dots, _m|\mathbf{L}]$.

Veamos que $A \perp B \Rightarrow \mathcal{L}(A) \perp \mathcal{L}(B)$.

$A \perp B \Rightarrow \mathbf{A}^T \mathbf{B} = \mathbf{0} \Rightarrow$ Para todo $\mathbf{a} \in \mathcal{L}(A)$ y $\mathbf{b} \in \mathcal{L}(B)$ tenemos que $\mathbf{a} \cdot \mathbf{b} = \mathbf{x} \mathbf{A}^T \mathbf{B} \mathbf{y} = \mathbf{a} \mathbf{0} \mathbf{y} = 0$.

Veamos que $\mathcal{L}(A) \perp \mathcal{L}(B) \Rightarrow A \perp B$.

$\mathcal{L}(A) \perp \mathcal{L}(B) \Rightarrow$ para todo $\mathbf{a} \in \mathcal{L}(A)$ y $\mathbf{b} \in \mathcal{L}(B)$, $\mathbf{a} \cdot \mathbf{b} = 0$.

En particular, también para todo $\mathbf{a} \in A$ y $\mathbf{b} \in B$.

Ahora veamos qué ocurre cuando alguno de los conjuntos es vacío (o solo contiene el vector cero). Como $A \perp B$ implica que todo vector en A es perpendicular a todo vector en B . Entonces $A \perp B$ es falso solo cuando existe algún vector en \mathbf{a} en A y algún vector \mathbf{b} en B que no son perpendiculares ($\mathbf{a} \cdot \mathbf{b} \neq 0$). Por lo tanto, si se da el caso de que A es vacío (o $\{\mathbf{0}\}$), tanto A como $\mathcal{L}(A) = \{\mathbf{0}\}$ son ortogonales a cualquier otro subconjunto de vectores (pues no existe tal par de vectores \mathbf{a} y \mathbf{b}). Lo mismo ocurre si B es vacío (o $\{\mathbf{0}\}$). □

Ejercicio 60. Basta comprobar que todas las propiedades del producto punto (Página 2.1) también se verifican si se divide el producto punto por n . □

Ejercicio 61. Hay que demostrar que se verifican todos los axiomas:

- **Simetría:** $\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a} \mathbf{D} \mathbf{b} = \mathbf{b} (\mathbf{D}^T) \mathbf{a} = \mathbf{b} \mathbf{D} \mathbf{a} = \langle \mathbf{b}, \mathbf{a} \rangle$, pues \mathbf{D} es simétrica.
- **Linealidad respecto al primer argumento:**
 1. $\langle (\alpha \mathbf{a}), \mathbf{b} \rangle = (\alpha \mathbf{a}) \mathbf{D} \mathbf{b} = \alpha (\mathbf{a} \mathbf{D} \mathbf{b}) = \alpha \langle \mathbf{a}, \mathbf{b} \rangle$.
 2. $\langle (\mathbf{a} + \mathbf{b}), \mathbf{c} \rangle = (\mathbf{a} + \mathbf{b}) \mathbf{D} \mathbf{c} = \mathbf{a} \mathbf{D} \mathbf{c} + \mathbf{b} \mathbf{D} \mathbf{c} = \langle \mathbf{a}, \mathbf{c} \rangle + \langle \mathbf{b}, \mathbf{c} \rangle$.
- **Positivo:** $\langle \mathbf{a}, \mathbf{a} \rangle = \mathbf{a} \mathbf{D} \mathbf{a} = \sum_{i=1}^n a_i d_{ii} a_i = \sum_{i=1}^n d_{ii} a_i^2 \geq 0$, ya que $d_{ii} > 0$ y $a_i^2 \geq 0$.
- **Definido:** $\langle \mathbf{a}, \mathbf{a} \rangle = \sum_{i=1}^n d_{ii} a_i^2 = 0$, entonces $a_i^2 = 0$ ya que $d_{ii} > 0$, por tanto $\mathbf{a} = \mathbf{0}$. □

Ejercicio 62. El vector nulo: $\vec{0}$. □

Ejercicio 63. El vector nulo: $\vec{x} = \vec{0}$. □

Ejercicio 64. Si \vec{x} e \vec{y} son *perpendiculares*, entonces $\langle \vec{x}, \vec{y} \rangle = 0$ y

$$\begin{aligned}
 \|\vec{x} + \vec{y}\|^2 &= \langle (\vec{x} + \vec{y}), (\vec{x} + \vec{y}) \rangle = \langle \vec{x}, (\vec{x} + \vec{y}) \rangle + \langle \vec{y}, (\vec{x} + \vec{y}) \rangle \\
 &= \langle (\vec{x} + \vec{y}), \vec{x} \rangle + \langle (\vec{x} + \vec{y}), \vec{y} \rangle \\
 &= \langle \vec{x}, \vec{x} \rangle + \langle \vec{y}, \vec{x} \rangle + \langle \vec{x}, \vec{y} \rangle + \langle \vec{y}, \vec{y} \rangle \\
 &= \langle \vec{x}, \vec{x} \rangle + 2(\langle \vec{x}, \vec{y} \rangle) + \langle \vec{y}, \vec{y} \rangle \\
 &= \langle \vec{x}, \vec{x} \rangle + 0 + \langle \vec{y}, \vec{y} \rangle = \|\vec{x}\|^2 + \|\vec{y}\|^2.
 \end{aligned}$$
□

Ejercicio 65. Sean \vec{x} y \vec{y} vectores de \mathcal{V} y sea el producto escalar $\langle _, _ \rangle: \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}$. Veamos que siempre se verifica que $|\langle \vec{x}, \vec{y} \rangle| \leq \|\vec{x}\| \|\vec{y}\|$

Caso trivial: \vec{x} y \vec{y} son linealmente dependientes

En tal caso, o bien $\vec{x} = \vec{0}$ o bien existe $\alpha \in \mathbb{R}$ tal que $\alpha \vec{x} = \vec{y}$. Si ocurre lo primero $\langle \vec{x}, \vec{y} \rangle = 0$ y $\|\vec{x}\| \|\vec{y}\| = 0$. Si ocurre lo segundo

$$|\langle \vec{x}, \vec{y} \rangle| = |\langle \vec{x}, \alpha \vec{x} \rangle| = |\alpha| \|\vec{x}\|^2 = \|\vec{x}\| |\alpha| \|\vec{x}\| = \|\vec{x}\| \|\vec{y}\|.$$

Caso no trivial: \vec{x} y \vec{y} son linealmente independientes

En tal caso existe un vector \vec{h} ortogonal a \vec{x} tal que $L(\vec{x}, \vec{h}) = L(\vec{x}, \vec{y})$. Dicho vector lo construimos de la forma $\vec{h} = \vec{y} + \alpha \vec{x}$, con lo que garantizamos que $L(\vec{x}, \vec{h}) = L(\vec{x}, \vec{y})$ e imponemos que

$$0 = \langle \vec{x}, \vec{h} \rangle = \langle \vec{x}, \vec{y} + \alpha \vec{x} \rangle = \langle \vec{x}, \vec{y} \rangle + \alpha \langle \vec{x}, \vec{x} \rangle,$$

con lo que $\alpha = -\frac{\langle \vec{x}, \vec{y} \rangle}{\langle \vec{x}, \vec{x} \rangle}$.

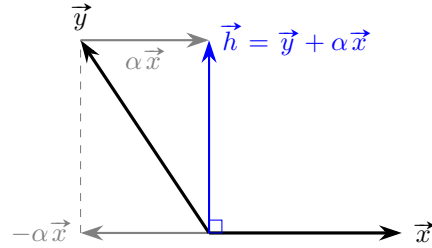


Figura A.1: Representación esquemática donde \vec{h} es combinación lineal de \vec{x} y \vec{y} y perpendicular a \vec{x} .

Por tanto tendremos que

$$\begin{aligned} \|\vec{y}\|^2 &= \langle \vec{y}, \vec{y} \rangle = \langle \vec{h} - \alpha \vec{x}, \vec{h} - \alpha \vec{x} \rangle = \langle \vec{h}, \vec{h} \rangle - 2\alpha \langle \vec{h}, \vec{x} \rangle + \alpha^2 \langle \vec{x}, \vec{x} \rangle \\ &= \|\vec{h}\|^2 + \alpha^2 \|\vec{x}\|^2 > \alpha^2 \|\vec{x}\|^2 = \frac{\langle \vec{x}, \vec{y} \rangle^2}{\langle \vec{x}, \vec{x} \rangle^2} \|\vec{x}\|^2 = \frac{\langle \vec{x}, \vec{y} \rangle^2}{\|\vec{x}\|^2}, \end{aligned}$$

de donde $\langle \vec{x}, \vec{y} \rangle^2 < \|\vec{x}\|^2 \|\vec{y}\|^2$ y por consiguiente $|\langle \vec{x}, \vec{y} \rangle| < \|\vec{x}\| \|\vec{y}\|$.

□

Ejercicio 66.

$$\begin{aligned} \|\vec{x} + \vec{y}\|^2 &= \langle (\vec{x} + \vec{y}), (\vec{x} + \vec{y}) \rangle = \langle \vec{x}, \vec{x} \rangle + 2\langle \vec{x}, \vec{y} \rangle + \langle \vec{y}, \vec{y} \rangle \\ &= \|\vec{x}\|^2 + 2\langle \vec{x}, \vec{y} \rangle + \|\vec{y}\|^2 \\ &\leq \|\vec{x}\|^2 + 2\|\vec{x}\| \|\vec{y}\| + \|\vec{y}\|^2 \quad \text{puesto que } |\langle \vec{x}, \vec{y} \rangle| \leq \|\vec{x}\| \|\vec{y}\| \\ &= (\|\vec{x}\| + \|\vec{y}\|)^2 \end{aligned}$$

Consecuentemente $\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$.

□

Ejercicio 67. Evidentemente $\vec{0}$ pertenece a A^\perp , así que el subconjunto A^\perp no es vacío.

Sean $\vec{c}_1, \vec{c}_2 \in A^\perp$ y $x, y \in \mathbb{R}$. Entonces $\langle \vec{c}_1, \vec{a} \rangle = 0$ y $\langle \vec{c}_2, \vec{a} \rangle = 0$, para cualquier $\vec{a} \in A$. Consecuentemente

$$\langle (x\vec{c}_1 + y\vec{c}_2), \vec{a} \rangle = x\langle \vec{c}_1, \vec{a} \rangle + y\langle \vec{c}_2, \vec{a} \rangle = x \cdot 0 + y \cdot 0 = 0.$$

y por tanto $(x\vec{c}_1 + y\vec{c}_2) \in A^\perp$. Es decir, A^\perp es cerrado para las combinaciones lineales. □

Ejercicio 68(a) $\vec{x} \in B^\perp$ si y solo si $\langle \vec{x}, \vec{b} \rangle = 0$ para cada $\vec{b} \in B$; (y como $A \subset B$) en particular también $\langle \vec{x}, \vec{a} \rangle = 0$ para cada $\vec{a} \in A$, es decir $\vec{x} \in A^\perp$. □

Ejercicio 68(b) Tenemos que demostrar la inclusión de estos subespacios en ambos sentidos.

Inclusión en un sentido: como $A \subset \mathcal{L}(A)$, por el apartado anterior sabemos que $\mathcal{L}(A)^\perp \subset A^\perp$.

Inclusión en el otro sentido: Sabemos que $\vec{x} \in A^\perp$ si y solo si $\langle \vec{x}, \vec{a} \rangle = 0$ para cada $\vec{a} \in A$. Sea $\vec{z} \in \mathcal{L}(A)$, entonces existen $\alpha_1, \dots, \alpha_k \in \mathbb{R}$ y $\vec{a}_1, \dots, \vec{a}_k \in A$ tales que $\vec{z} = \alpha_1 \vec{a}_1 + \dots + \alpha_k \vec{a}_k$. Entonces $\langle \vec{x}, \vec{z} \rangle = \langle \vec{x}, (\alpha_1 \vec{a}_1 + \dots + \alpha_k \vec{a}_k) \rangle = \alpha_1 \langle \vec{x}, \vec{a}_1 \rangle + \dots + \alpha_k \langle \vec{x}, \vec{a}_k \rangle = \alpha_1 0 + \dots + \alpha_k 0 = 0$; es decir $\vec{x} \in \mathcal{L}(A)^\perp$. Por tanto $A^\perp \subset \mathcal{L}(A)^\perp$. □

Ejercicio 68(c) Sea $\vec{a} \in \mathcal{A} \cap \mathcal{A}^\perp$; como $\vec{a} \in \mathcal{A}$ es perpendicular a todo vector en \mathcal{A}^\perp , entonces $\langle \vec{a}, \vec{a} \rangle = 0 \Rightarrow \vec{a} = \vec{0}$. □

Ejercicio 69. $(\mathbf{I} - \mathbf{P})^\top = \mathbf{I}^\top - \mathbf{P}^\top = \mathbf{I} - \mathbf{P}$. □

Ejercicio 70. $(\mathbf{I} - \mathbf{P})^2 = (\mathbf{I} - \mathbf{P})(\mathbf{I} - \mathbf{P}) = \mathbf{I} - \mathbf{P} - \mathbf{P} + \mathbf{P}^2 = \mathbf{I} - \mathbf{P} - \mathbf{P} + \mathbf{P} = \mathbf{I} - \mathbf{P}$. □

Ejercicio 71(a) Por una parte $\mathcal{C}(\mathbf{P}) \subset \mathcal{C}(\mathbf{A})$, pues para todo $\mathbf{z} \in \mathcal{C}(\mathbf{P})$

existe $\mathbf{y} \in \mathbb{R}^m$ tal que $\mathbf{z} = \mathbf{P}\mathbf{y} = \mathbf{A}(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{y} = \mathbf{A}\hat{\mathbf{x}} \in \mathcal{C}(\mathbf{A})$, donde $\hat{\mathbf{x}} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{y}$,

y por otra $\mathcal{C}(\mathbf{A}) \subset \mathcal{C}(\mathbf{P})$, pues para todo $\mathbf{z} \in \mathcal{C}(\mathbf{A})$

existe $\mathbf{x} \in \mathbb{R}^n$: $\mathbf{z} = \mathbf{A}\mathbf{x} = \mathbf{A} \underbrace{(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{A}}_{\mathbf{I}} \mathbf{x} = \mathbf{P}\mathbf{A}\mathbf{x} = \mathbf{P}\mathbf{z} \in \mathcal{C}(\mathbf{P})$, donde $\mathbf{z} = \mathbf{A}\mathbf{x}$. □

Ejercicio 71(b) Por una parte $\mathcal{C}(\mathbf{M}) \subset \mathcal{N}(\mathbf{A}^\top)$, pues para todo $\mathbf{z} \in \mathcal{C}(\mathbf{M})$ existe $\mathbf{y} \in \mathbb{R}^m$ tal que

$$\mathbf{z} = \mathbf{M}\mathbf{y} = (\mathbf{I} - \mathbf{P})\mathbf{y} = \mathbf{y} - \mathbf{P}\mathbf{y} \in \mathcal{C}(\mathbf{A})^\perp = \mathcal{N}(\mathbf{A}^\top)$$

por ser $\mathbf{P}\mathbf{y}$ la proyección ortogonal sobre $\mathcal{C}(\mathbf{A})$. Por otra parte $\mathcal{N}(\mathbf{A}^\top) \subset \mathcal{C}(\mathbf{M})$, pues para todo $\mathbf{y} \in \mathcal{N}(\mathbf{A}^\top)$ sabemos que $\mathbf{y} \perp \mathcal{C}(\mathbf{A})$, y por tanto $\mathbf{P}\mathbf{y} = \mathbf{0}$. Así pues

$$\mathbf{y} = \mathbf{y} + \mathbf{0} = \mathbf{y} + \mathbf{P}\mathbf{y} = (\mathbf{I} + \mathbf{P})\mathbf{y} = \mathbf{M}\mathbf{y} \Rightarrow \mathbf{y} \in \mathcal{C}(\mathbf{M}).$$

□

Ejercicio 72. Puesto que $\mathbf{x}\mathbf{A} = \mathbf{A}^\top\mathbf{x} = \mathbf{A}\mathbf{x}$, tenemos que $\langle \mathbf{x}, \mathbf{A}\mathbf{y} \rangle = \mathbf{x}\mathbf{A}\mathbf{y} = \langle \mathbf{x}\mathbf{A}, \mathbf{y} \rangle = \langle \mathbf{A}\mathbf{x}, \mathbf{y} \rangle$.

□

Ejercicio 73(a) Por ser autoadjunta e idempotente: $\langle f(\vec{u}), f(\vec{w}) \rangle = \langle f(f(\vec{u})), \vec{w} \rangle = \langle f(\vec{u}), \vec{w} \rangle$.

□

Ejercicio 73(b) Para la demostración usaremos el resultado del anterior apartado (a). Es sencillo pero pesado.

Por una parte tenemos que demostrar que $f(\vec{u} + \vec{v}) = f(\vec{u}) + f(\vec{v})$, y lo haremos demostrando que la diferencia es $\vec{0}$, pues su norma es cero, es decir, $\|f(\vec{u} + \vec{v}) - f(\vec{u}) - f(\vec{v})\|^2 = 0$:

$$\begin{aligned} \|f(\vec{u} + \vec{v}) - f(\vec{u}) - f(\vec{v})\|^2 &= \langle f(\vec{u} + \vec{v}) - f(\vec{u}) - f(\vec{v}), f(\vec{u} + \vec{v}) - f(\vec{u}) - f(\vec{v}) \rangle \\ &= \langle f(\vec{u} + \vec{v}), f(\vec{u} + \vec{v}) \rangle - \langle f(\vec{u} + \vec{v}), f(\vec{u}) \rangle - \langle f(\vec{u} + \vec{v}), f(\vec{v}) \rangle \\ &\quad - \langle f(\vec{u}), f(\vec{u} + \vec{v}) \rangle + \langle f(\vec{u}), f(\vec{u}) \rangle + \langle f(\vec{u}), f(\vec{v}) \rangle \\ &\quad - \langle f(\vec{v}), f(\vec{u} + \vec{v}) \rangle + \langle f(\vec{v}), f(\vec{u}) \rangle + \langle f(\vec{v}), f(\vec{v}) \rangle \quad (\text{desarrollando}) \\ &= \langle f(\vec{u} + \vec{v}), \vec{u} + \vec{v} \rangle - \langle f(\vec{u} + \vec{v}), \vec{u} \rangle - \langle f(\vec{u} + \vec{v}), \vec{v} \rangle \\ &\quad - \langle f(\vec{u}), \vec{u} + \vec{v} \rangle + \langle f(\vec{u}), \vec{u} \rangle + \langle f(\vec{u}), \vec{v} \rangle \\ &\quad - \langle f(\vec{v}), \vec{u} + \vec{v} \rangle + \langle f(\vec{v}), \vec{u} \rangle + \langle f(\vec{v}), \vec{v} \rangle \quad (\text{por (a)}) \\ &= \langle f(\vec{u} + \vec{v}), \vec{u} \rangle + \langle f(\vec{u} + \vec{v}), \vec{v} \rangle - \langle f(\vec{u} + \vec{v}), \vec{u} \rangle - \langle f(\vec{u} + \vec{v}), \vec{v} \rangle \\ &\quad - \langle f(\vec{u}), \vec{u} \rangle - \langle f(\vec{u}), \vec{v} \rangle + \langle f(\vec{u}), \vec{u} \rangle + \langle f(\vec{u}), \vec{v} \rangle \\ &\quad - \langle f(\vec{v}), \vec{u} \rangle - \langle f(\vec{v}), \vec{v} \rangle + \langle f(\vec{v}), \vec{u} \rangle + \langle f(\vec{v}), \vec{v} \rangle \quad (\text{desarrollando}) \\ &= 0 \quad (\text{cancelando términos}). \end{aligned}$$

Y por otra, tenemos que demostrar que $f(\alpha\vec{u}) = \alpha f(\vec{u})$; y seguiremos una estrategia similar, es decir, demostraremos que $\|f(\alpha\vec{u}) - \alpha f(\vec{u})\|^2 = 0$.

$$\begin{aligned} \|f(\alpha\vec{u}) - \alpha f(\vec{u})\|^2 &= \langle f(\alpha\vec{u}) - \alpha f(\vec{u}), f(\alpha\vec{u}) - \alpha f(\vec{u}) \rangle \\ &= \langle f(\alpha\vec{u}), f(\alpha\vec{u}) \rangle - \langle f(\alpha\vec{u}), \alpha f(\vec{u}) \rangle - \langle \alpha f(\vec{u}), f(\alpha\vec{u}) \rangle + \langle \alpha f(\vec{u}), \alpha f(\vec{u}) \rangle \\ &= \langle f(\alpha\vec{u}), \alpha\vec{u} \rangle - \langle f(\alpha\vec{u}), \alpha f(\vec{u}) \rangle - \langle \alpha f(\vec{u}), f(\alpha\vec{u}) \rangle + \langle \alpha f(\vec{u}), \alpha f(\vec{u}) \rangle \quad (\text{por (a)}) \\ &= \alpha \langle f(\alpha\vec{u}), \vec{u} \rangle - \alpha \langle f(\alpha\vec{u}), f(\vec{u}) \rangle - \alpha \langle f(\vec{u}), f(\alpha\vec{u}) \rangle + \alpha^2 \langle f(\vec{u}), f(\vec{u}) \rangle \\ &= \alpha \langle f(\alpha\vec{u}), \vec{u} \rangle - \alpha \langle f(\alpha\vec{u}), \vec{u} \rangle - \alpha \langle f(\vec{u}), \alpha\vec{u} \rangle + \alpha^2 \langle f(\vec{u}), \vec{u} \rangle \quad (\text{por (a)}) \\ &= \alpha \langle f(\alpha\vec{u}), \vec{u} \rangle - \alpha \langle f(\alpha\vec{u}), \vec{u} \rangle - \alpha^2 \langle f(\vec{u}), \vec{u} \rangle + \alpha^2 \langle f(\vec{u}), \vec{u} \rangle \\ &= 0 \quad (\text{cancelando términos}). \end{aligned}$$

□

Ejercicio 74. Sea \vec{p}_y la proyección de \vec{y} sobre \mathcal{H} y tomemos un vector \vec{v} cualquiera de \mathcal{H} . Veamos que \vec{y} está más lejos de \vec{v} que de su proyección ortogonal \vec{p}_y .

Como \vec{v} y \vec{p}_y están en \mathcal{H} , su diferencia $(\vec{p}_y - \vec{x})$ está en \mathcal{H} (pues \mathcal{H} es subespacio) y es ortogonal a $(\vec{y} - \vec{p}_y)$ (por ser \vec{p}_y la proyección ortogonal sobre \mathcal{H}). Y como la suma de ambos vectores perpendiculares es $(\vec{y} - \vec{p}_y) + (\vec{p}_y - \vec{x}) = (\vec{y} - \vec{x})$, por el Teorema de Pitágoras concluimos que

$$\|\vec{y} - \vec{x}\|^2 = \|\vec{y} - \vec{p}_y\|^2 + \|\vec{p}_y - \vec{x}\|^2 \geq \|\vec{y} - \vec{p}_y\|^2.$$

Por tanto, $\boxed{\|\vec{y} - \vec{x}\| \geq \|\vec{y} - \vec{p}_y\|}$ para todo $\vec{x} \in \mathcal{H}$.

□

Ejercicio 75(a) Puesto que como $\mathbf{P}\mathbf{y} = \mathbf{y}\mathbf{P}$ (por ser \mathbf{P} simétrica):

$$(\mathbf{y} - \mathbf{P}\mathbf{y})\mathbf{A} = (\mathbf{y} - \mathbf{y}\mathbf{P})\mathbf{A} = \mathbf{y}\mathbf{A} - \mathbf{y}\mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{A} = \mathbf{y}\mathbf{A} - \mathbf{y}\mathbf{A} = \mathbf{0}.$$

□

Ejercicio 75(b) Si $\mathbf{y} = \mathbf{A}\mathbf{x}$ para algún $\mathbf{x} \in \mathbb{R}^n$:

$$f_{\mathbf{P}}(\mathbf{y}) = \mathbf{P}\mathbf{y} = \mathbf{P}\mathbf{A}\mathbf{x} = \mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{x} = \mathbf{y};$$

□

Ejercicio 75(c) Si $\mathbf{A}^\top\mathbf{y} = \mathbf{0}$, entonces

$$f_{\mathbf{P}}(\mathbf{y}) = \mathbf{P}\mathbf{y} = \mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{y} = \mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{0} = \mathbf{0}.$$

□

Ejercicio 75(d) Como $\mathcal{C}(\mathbf{A})$ es el complemento ortogonal de $\mathcal{N}(\mathbf{A}^\top)$, demostrar que para todo $\mathbf{y} \in \mathbb{R}^m$ la diferencia $(\mathbf{y} - f_{\mathbf{M}}(\mathbf{y}))$ es perpendicular a $\mathcal{N}(\mathbf{A}^\top)$ es lo mismo que demostrar que dicha diferencia pertenece a $\mathcal{C}(\mathbf{A})$; así

$$\mathbf{y} - f_{\mathbf{M}}(\mathbf{y}) = \mathbf{y} - (\mathbf{y} - f_{\mathbf{P}}(\mathbf{y})) = f_{\mathbf{P}}(\mathbf{y}) \in \mathcal{C}(\mathbf{A}).$$

□

Ejercicio 75(e) Por una parte, $f_{\mathbf{M}}$ actúa como la función identidad con todo $\mathbf{y} \in \mathcal{N}(\mathbf{A}^\top)$, pues en tal caso $f_{\mathbf{P}}(\mathbf{y}) = \mathbf{0}$, y por tanto:

$$f_{\mathbf{M}}(\mathbf{y}) = \mathbf{y} - f_{\mathbf{P}}(\mathbf{y}) = \mathbf{y} - \mathbf{0} = \mathbf{y}.$$

Por otra parte, $f_{\mathbf{M}}$ actúa como la función nula con todo $\mathbf{y} \in \mathcal{C}(\mathbf{A})$, pues en tal caso $f_{\mathbf{P}}(\mathbf{y}) = \mathbf{y}$, y por tanto:

$$f_{\mathbf{M}}(\mathbf{y}) = \mathbf{y} - f_{\mathbf{P}}(\mathbf{y}) = \mathbf{y} - \mathbf{y} = \mathbf{0}.$$

□

Ejercicio 76(a) Debemos demostrar que cada uno de los conjuntos está contenido en el otro.

$\boxed{\mathcal{N}(\mathbf{A}) \subset \mathcal{N}(\mathbf{A}^\top\mathbf{A})}$: Si $\mathbf{x} \in \mathcal{N}(\mathbf{A})$ entonces $\mathbf{A}\mathbf{x} = \mathbf{0}$. Multiplicando ambos lados de la igualdad por \mathbf{A}^\top tenemos $\mathbf{A}^\top\mathbf{A}\mathbf{x} = \mathbf{A}^\top\mathbf{0} = \mathbf{0}$; y por tanto $\mathbf{x} \in \mathcal{N}(\mathbf{A}^\top\mathbf{A})$.

$\boxed{\mathcal{N}(\mathbf{A}^\top \mathbf{A}) \subset \mathcal{N}(\mathbf{A})}$: Si $\mathbf{x} \in \mathcal{N}(\mathbf{A}^\top \mathbf{A})$ entonces $\mathbf{A}^\top \mathbf{A} \mathbf{x} = \mathbf{0}$. Multiplicando por \mathbf{x} tenemos que $\mathbf{x} \mathbf{A}^\top \mathbf{A} \mathbf{x} = \mathbf{x} \cdot \mathbf{0} = 0$. Por tanto $\mathbf{A} \mathbf{x}$ tiene norma cero, ya que $\mathbf{x} \mathbf{A}^\top \mathbf{A} \mathbf{x} = (\mathbf{A} \mathbf{x}) \cdot (\mathbf{A} \mathbf{x}) = \|\mathbf{A} \mathbf{x}\|^2$. Como el único vector de norma cero es $\mathbf{0}$, tenemos que $\mathbf{A} \mathbf{x} = \mathbf{0}$; es decir, $\mathbf{x} \in \mathcal{N}(\mathbf{A})$.

□

Ejercicio 76(b) Como $\mathbf{A}^\top \mathbf{A}$ y \mathbf{A} tienen n columnas e idéntico espacio nulo $\mathcal{N}(\mathbf{A}^\top \mathbf{A}) = \mathcal{N}(\mathbf{A})$, ambas matrices tienen el mismo rango: $k = n - \dim \mathcal{N}(\mathbf{A})$.

□

Ejercicio 76(c) Sea \mathbf{A} de rango k . El sistema $(\mathbf{A}^\top \mathbf{A}) \mathbf{x} = \mathbf{A}^\top \mathbf{b}$ tiene solución si el rango de la matriz de coeficientes $\mathbf{A}^\top \mathbf{A}$ es igual que el rango de la matriz ampliada $[\mathbf{A}^\top \mathbf{A} \mid \mathbf{A}^\top \mathbf{b}]$; es decir, si $\mathbf{A}^\top \mathbf{b}$ es combinación lineal de las columnas de $\mathbf{A}^\top \mathbf{A}$. Veámoslo:

Como al añadir una columna a una matriz, el rango (el número de pivotes) no puede disminuir, sabemos que $\text{rg}([\mathbf{A}^\top \mathbf{A} \mid \mathbf{A}^\top \mathbf{b}]) \geq \text{rg}(\mathbf{A}^\top \mathbf{A}) = \text{rg}(\mathbf{A}) = k$ (por el Corolario 12.3.4).

Por otra parte, $[\mathbf{A}^\top \mathbf{A} \mid \mathbf{A}^\top \mathbf{b}] = \mathbf{A}^\top [\mathbf{A} \mid \mathbf{b}]$ es el producto de \mathbf{A}^\top por la matriz ampliada $[\mathbf{A} \mid \mathbf{b}]$, y por la Proposición 5.2.6 en la página 73 sabemos que $\text{rg}(\mathbf{A}^\top [\mathbf{A} \mid \mathbf{b}]) \leq \text{rg}(\mathbf{A}^\top) = \text{rg}(\mathbf{A}) = k$.

Así, como por una parte $\text{rg}([\mathbf{A}^\top \mathbf{A} \mid \mathbf{A}^\top \mathbf{b}]) \geq k$ y por la otra $\text{rg}([\mathbf{A}^\top \mathbf{A} \mid \mathbf{A}^\top \mathbf{b}]) \leq k$, se deduce que $\text{rg}([\mathbf{A}^\top \mathbf{A} \mid \mathbf{A}^\top \mathbf{b}]) = k$; que es el rango de la matriz de coeficientes $\mathbf{A}^\top \mathbf{A}$ del sistema $\mathbf{A}^\top \mathbf{A} \mathbf{x} = \mathbf{A}^\top \mathbf{b}$, y por tanto el sistema siempre tiene solución, es decir, para todo $\mathbf{b} \in \mathbb{R}^m$ y toda $\mathbf{A} \in \mathbb{R}^{m \times n}$ se verifica que $\mathbf{A}^\top \mathbf{b} \in \mathcal{C}(\mathbf{A}^\top \mathbf{A})$.

□

Ejercicio 77. Dado que \mathbf{Q} es de rango r , existen matrices $\mathbf{A} = \mathbf{I}_{\tau_1 \dots \tau_i}$ y $\mathbf{B} = \mathbf{I}_{\tau_{i+1} \dots \tau_j}$ de orden n tales que

$$\mathbf{A} \mathbf{Q} \mathbf{B} = \begin{bmatrix} \mathbf{I}_{r \times r} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \text{y por tanto} \quad \mathbf{Q} = \mathbf{A}^{-1} \begin{bmatrix} \mathbf{I}_{r \times r} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{B}^{-1} = \mathbf{A}^{-1} \begin{bmatrix} \mathbf{I}_{r \times r} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{I}_{r \times r} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{B}^{-1} = (\mathbf{A}^{-1})_{|(1:r)} ((1:r) \mathbf{B}^{-1});$$

donde las dos últimas matrices son de rango r y tienen órdenes $n \times r$ y $r \times n$ respectivamente. Consecuentemente, existen \mathbf{C} y \mathbf{D} invertibles, y de orden r , tales que

$$\mathbf{C}(\mathbf{A}^{-1})_{|(1:r)} = \mathbf{I}_{r \times r} \quad \text{y} \quad ((1:r) \mathbf{B}^{-1}) \mathbf{D} = \mathbf{I}_{r \times r}.$$

Como $\mathbf{Q} \mathbf{Q} = \mathbf{Q}$, si denotamos $(\mathbf{A}^{-1})_{|(1:r)}$ con \mathbf{A}_1 y $((1:r) \mathbf{B}^{-1})$ con \mathbf{B}_1 , tenemos que $\mathbf{A}_1 \mathbf{B}_1 = \mathbf{A}_1 \mathbf{B}_1 \mathbf{A}_1 \mathbf{B}_1$. Y multiplicado por \mathbf{C} y \mathbf{D} a izquierda y derecha respectivamente, también tenemos que $\mathbf{I}_{r \times r} = \mathbf{B}_1 \mathbf{A}_1$, pues como

$$\mathbf{C} \mathbf{A}_1 = \mathbf{I}_{r \times r} = \mathbf{B}_1 \mathbf{D},$$

$$\mathbf{C} \mathbf{A}_1 \mathbf{B}_1 \mathbf{D} = \mathbf{C} \mathbf{A}_1 \mathbf{B}_1 \mathbf{A}_1 \mathbf{B}_1 \mathbf{D} \implies \mathbf{I}_{r \times r} = \mathbf{B}_1 \mathbf{A}_1;$$

Así que, usando la Ecuación 3.8 en la página 40: $\text{tr}(\mathbf{Q}) = \text{tr}(\mathbf{A}_1 \mathbf{B}_1) = \text{tr}(\mathbf{B}_1 \mathbf{A}_1) = \text{tr}(\mathbf{I}_{r \times r}) = r = \text{rg}(\mathbf{Q})$.

□

Ejercicio 78. El vector cero $\mathbf{0}$ de \mathbb{R}^n es múltiplo de cualquier otro vector \mathbf{x} de \mathbb{R}^n , pues $\mathbf{0} = 0 \cdot \mathbf{x}$; y por tanto, por la propiedad P-3 el determinante de \mathbf{A} es cero:

$$\det[\mathbf{A}_1; \dots; \mathbf{0}; \dots; \mathbf{A}_n] = \det[\mathbf{A}_1; \dots; 0\mathbf{x}; \dots; \mathbf{A}_n] = 0 \cdot \det[\mathbf{A}_1; \dots; \mathbf{x}; \dots; \mathbf{A}_n] = 0$$

□

Ejercicio 79(a) Como $\mathbf{A}_{\tau_1 \dots \tau_k} = \mathbf{A}(\mathbf{I}_{\tau_1 \dots \tau_k}) = \mathbf{A}(\mathbf{I}_{\tau_1}) \cdots (\mathbf{I}_{\tau_k})$, aplicando repetidamente (13.2) tenemos

$$\begin{aligned} \det(\mathbf{A}_{\tau_1 \dots \tau_k}) &= |\mathbf{A}(\mathbf{I}_{\tau_1}) \cdots (\mathbf{I}_{\tau_k})| \\ &= |\mathbf{A}(\mathbf{I}_{\tau_1}) \cdots (\mathbf{I}_{\tau_{(k-1)}})| \cdot |\mathbf{I}_{\tau_k}| \\ &= |\mathbf{A}(\mathbf{I}_{\tau_1}) \cdots (\mathbf{I}_{\tau_{(k-2)}})| \cdot |\mathbf{I}_{\tau_{(k-1)}}| \cdot |\mathbf{I}_{\tau_k}| \\ &\quad \vdots \\ &= |\mathbf{A}| \cdot |\mathbf{I}_{\tau_1}| \cdots |\mathbf{I}_{\tau_k}|. \end{aligned}$$

□

Ejercicio 79(b)

$$|\mathbf{B}| = \det(\mathbf{I}_{\tau_1 \dots \tau_k}) = |\mathbf{I}| \cdot |\mathbf{I}_{\tau_1}| \cdots |\mathbf{I}_{\tau_k}|.$$

y como los determinantes de las matrices elementales son distintos de cero, necesariamente $|\mathbf{B}| \neq 0$.

□

Ejercicio 79(c) Si \mathbf{B} es de rango completo entonces $\mathbf{B} = \mathbf{I}_{\tau_1 \dots \tau_k}$. Así

$$\det(\mathbf{AB}) = \det(\mathbf{A} \mathbf{I}_{\tau_1 \dots \tau_k}) = \det(\mathbf{A}_{\tau_1 \dots \tau_k}) = |\mathbf{A}| \cdot (|\mathbf{I}_{\tau_1}| \cdots |\mathbf{I}_{\tau_k}|) = |\mathbf{A}| \cdot |\mathbf{B}|.$$

□

Ejercicio 80. Un intercambio de columnas es una sucesión de transformaciones elementales *Tipo I* y una única transformación elemental de *Tipo II* que multiplica por -1 una columna (EJERCICIO 31 en la página 49).

□

Ejercicio 81(a) Puesto que ambas son matrices elementales del mismo tipo, $\det(\mathbf{I}_{\tau}) = \det(\tau \mathbf{I})$.

□

Ejercicio 81(b) Puesto que $\mathbf{B} = \mathbf{I}_{\tau_1 \dots \tau_k} = (\mathbf{I}_{\tau_1}) \cdots (\mathbf{I}_{\tau_k})$, su determinante es el producto de los determinantes $\det(\mathbf{I}_{\tau_i})$

$$|\mathbf{B}| = \det(\mathbf{I}_{\tau_1}) \cdots \det(\mathbf{I}_{\tau_k}) = \prod_{i=1}^k \det(\mathbf{I}_{\tau_i}).$$

Pero también sabemos que $\mathbf{B}^T = {}_{\tau_k \dots \tau_1} \mathbf{I} = ({}_{\tau_k} \mathbf{I}) \cdots ({}_{\tau_1} \mathbf{I})$, por lo que su determinante es

$$|\mathbf{B}^T| = \prod_{i=1}^k \det({}_{\tau_i} \mathbf{I}) = \prod_{i=1}^k \det(\mathbf{I}_{\tau_i}) = |\mathbf{B}|.$$

□

Ejercicio 82(a) Como la matriz de orden n es de rango completo, los n elementos de la diagonal principal son pivotes (i.e., distintos de cero).

$$\mathbf{L} = \begin{bmatrix} *_1 & & & \\ \vdots & *_2 & & \\ & \vdots & \ddots & \\ \vdots & \vdots & \vdots & *_{n-1} \\ & & & \vdots & *_n \end{bmatrix} \quad \text{donde } *_j \text{ son números distintos de cero.}$$

Como sabemos que $|\mathbf{L}| = \begin{vmatrix} \mathbf{L} & \\ & 1 \end{vmatrix}$, vamos a realizar una secuencia de transformaciones elementales sobre las n primeras columnas esta última matriz, pero aplicaremos transformaciones elementales sobre la última *fila* para mantener constante el valor del determinante:

Primero dividimos cada una de las n primeras columnas por su correspondiente pivote $*_j$ ($j = 1 : n$) para normalizar dichos pivotes al valor 1 (así que multiplicamos la última *fila* por cada uno de los n primeros pivotes para compensar dichas transformaciones *Tipo II* y mantener constante el valor de el determinante). De esta manera logramos una matriz en la que todos los pivotes (salvo en de la última columna) son 1. Después aplicamos la eliminación de derecha a izquierda para anular todo lo que queda a la izquierda de los pivotes (como las transformaciones empleadas en este segundo paso son de *tipo I*, ahora basta con multiplicar la última fila por 1).

$$\left[\begin{array}{cccc|c} *_1 & & & & \\ \vdots & *_2 & & & \\ & \vdots & \ddots & & \\ \vdots & \vdots & \vdots & & \\ & & & *_{n-1} & \\ & & & \vdots & *_{n-1} \end{array} \right] \xrightarrow{\begin{array}{l} \tau \\ [(\frac{1}{*_1})\mathbf{1}] \\ [(\frac{1}{*_2})\mathbf{n}] \\ \vdots \\ [(\frac{1}{*_n})(\mathbf{n}+1)] \end{array}} \left[\begin{array}{cccc|c} 1 & & & & \\ & 1 & & & \\ & \vdots & \ddots & & \\ & \vdots & \vdots & & \\ & \vdots & \vdots & & 1 \\ & & & *_{n-1} & *_{n-1} \end{array} \right] \xrightarrow{\begin{array}{l} \tau_1 \dots \tau_q \\ \text{(de Tipo I)} \\ [(1)(\mathbf{n}+1)] \end{array}} \left[\begin{array}{cccc|c} 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ & & & & *_{n-1} \end{array} \right]$$

De esta manera hemos obtenido una matriz como la identidad salvo por su última columna, que está multiplicada por el producto de los pivotes de \mathbf{L} . Así pues, el determinante de esta última matriz es igual al producto de los pivotes \mathbf{L} , y como en todo momento hemos mantenido el valor del determinante, concluimos que el determinante de una matriz triangular inferior de rango completo es igual al producto de sus pivotes; es decir, al producto de los elementos de la diagonal.

$$\det(\mathbf{L}) = \text{producto de los elementos de la diagonal}$$

□

Ejercicio 82(b) Una matriz de orden n y triangular solo puede ser de rango completo si los n elementos de la diagonal son distintos de cero. Así, si hay algún cero en su diagonal, la matriz es singular. Consecuentemente su determinante es cero, es decir, es igual al producto de los elementos de la diagonal (donde uno de ellos es cero).

□

Ejercicio 82(c)

$$\det(\mathbf{U}) = \det(\mathbf{U}^T) = \text{producto de los elementos de la diagonal}$$

por ser \mathbf{U}^T triangular inferior.

□

Ejercicio 83(a) Puesto que $\begin{bmatrix} \mathbf{A} & \\ & \mathbf{B} \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \\ & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \\ & \mathbf{B} \end{bmatrix}$ entonces $\begin{vmatrix} \mathbf{A} & \\ & \mathbf{B} \end{vmatrix} = \begin{vmatrix} \mathbf{A} & \\ & \mathbf{I} \end{vmatrix} \cdot \begin{vmatrix} \mathbf{I} & \\ & \mathbf{B} \end{vmatrix} = |\mathbf{A}| \cdot |\mathbf{B}|.$

□

Ejercicio 83(b) Puesto que

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{I} \\ \mathbf{C} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{B} \\ & \mathbf{I} \end{bmatrix}$$

y puesto que mediante una sucesión τ_1, \dots, τ_k de transformaciones elementales *Tipo I* es posible la transformación $\begin{bmatrix} \mathbf{A} & \mathbf{I} \\ \mathbf{C} & \mathbf{I} \end{bmatrix} \xrightarrow{\tau_1, \dots, \tau_k} \begin{bmatrix} \mathbf{A} & \mathbf{I} \\ \mathbf{C} & \mathbf{I} \end{bmatrix}$ tenemos que $\begin{vmatrix} \mathbf{A} & \mathbf{I} \\ \mathbf{C} & \mathbf{I} \end{vmatrix} = \begin{vmatrix} \mathbf{A} & \mathbf{I} \\ \mathbf{C} & \mathbf{I} \end{vmatrix}$; y por tanto

$$\begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{I} \end{vmatrix} = \begin{vmatrix} \mathbf{A} & \mathbf{I} \\ \mathbf{C} & \mathbf{I} \end{vmatrix} \begin{vmatrix} \mathbf{I} & \mathbf{B} \\ & \mathbf{I} \end{vmatrix} = \begin{vmatrix} \mathbf{A} & \mathbf{I} \\ \mathbf{C} & \mathbf{I} \end{vmatrix} \begin{vmatrix} \mathbf{I} & \mathbf{B} \\ & \mathbf{I} \end{vmatrix} = |\mathbf{A}| \cdot |\mathbf{B}|.$$

□

Ejercicio 84. Desarrollando por la segunda columna tenemos

$$\det \mathbf{A} = -0 \begin{vmatrix} 5 & 2 & 4 \\ 3 & 1 & 2 \\ 5 & 2 & 1 \end{vmatrix} + \begin{vmatrix} 2 & 3 & 2 \\ 3 & 1 & 2 \\ 5 & 2 & 1 \end{vmatrix} - 0 \begin{vmatrix} 2 & 3 & 2 \\ 5 & 2 & 4 \\ 5 & 2 & 1 \end{vmatrix} + 3 \begin{vmatrix} 2 & 3 & 2 \\ 5 & 2 & 4 \\ 3 & 1 & 2 \end{vmatrix} = 0 + 17 - 0 + 3 \times 4 = 29$$

□

Ejercicio 85.

$$\begin{aligned} \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} &= -b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + e \begin{vmatrix} a & c \\ g & i \end{vmatrix} - h \begin{vmatrix} a & c \\ d & f \end{vmatrix} \\ &= -b(-fg + id) + e(-cg + ia) - h(-cd + fa) \\ &= bfg - bdi - ecg + eia + hcd - hfa && \text{desarrollando términos} \\ &= a(ei - hf) - d(bi - hc) + g(bf - ec) && \text{reordenando} \\ &= aei + bfg + dhc - gec - dbi - hfa && \text{reordenando.} \end{aligned}$$

(esta fórmula se denomina regla de Sarrus)

□

Ejercicio 86. Sea λ un autovalor de \mathbf{A} y sean \mathbf{x} y \mathbf{y} dos autovectores correspondientes al autovalor λ , es decir, tales que $\mathbf{Ax} = \lambda\mathbf{x}$ y $\mathbf{Ay} = \lambda\mathbf{y}$; entonces

$$\mathbf{A}(a\mathbf{x} + b\mathbf{y}) = a\mathbf{Ax} + b\mathbf{Ay} = \lambda a\mathbf{x} + \lambda b\mathbf{y} = \lambda(a\mathbf{x} + b\mathbf{y}).$$

□

Ejercicio 87. Extendemos el sistema $[\mathbf{q}]$ hasta formar una base ortogonal de \mathbb{R}^n

$$[\mathbf{q}] \longrightarrow [\mathbf{q}; \mathbf{z}_1; \dots; \mathbf{z}_{(n-1)}];$$

y luego colocamos \mathbf{q} en la última posición y normalizamos todos los vectores; obteniendo la siguiente base ortonormal de \mathbb{R}^n :

$$\left[\frac{1}{\|\mathbf{z}_1\|}(\mathbf{z}_1); \dots; \frac{1}{\|\mathbf{z}_{(n-1)}\|}(\mathbf{z}_{(n-1)}); \mathbf{q} \right].$$

□

Ejercicio 88. ${}_i(\mathbf{Q}^\top)\mathbf{Q}_{|j} = (\mathbf{Q}_{|i}) \cdot (\mathbf{Q}_{|j}) = \begin{cases} 1 & \text{cuando } i = j \\ 0 & \text{cuando } i \neq j \end{cases}$, es decir $(\mathbf{Q}^\top)\mathbf{Q} = \underset{n \times n}{\mathbf{I}}$.

□

Ejercicio 89(a) $(\mathbf{QR})^\top = \mathbf{R}^\top(\mathbf{Q}^\top) = \mathbf{R}^{-1}(\mathbf{Q}^{-1}) = (\mathbf{QR})^{-1}$.

□

Ejercicio 89(b) $(\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q})^\top = \mathbf{Q}^\top\mathbf{A}(\mathbf{Q}^{-1})^\top = \mathbf{Q}^\top\mathbf{A}(\mathbf{Q}^\top)^\top = \mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}$.

□

Ejercicio 90. Si $\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q} = \mathbf{D}$ es diagonal, entonces $\mathbf{A} = \mathbf{Q}\mathbf{D}\mathbf{Q}^\top$. Por tanto, $\mathbf{A}^\top = (\mathbf{Q}\mathbf{D}\mathbf{Q}^\top)^\top = \mathbf{Q}\mathbf{D}\mathbf{Q}^\top = \mathbf{A}$.

□

Ejercicio 91(a) Si \mathbf{A} es definida positiva, entonces para cualquier $\mathbf{x} \in \mathbb{R}^n$ no nulo se verifica que $\mathbf{x}\mathbf{A}\mathbf{x} > 0$. En particular, si \mathbf{x} es la columna j -ésima de la matriz identidad, tenemos

$${}_j|\mathbf{A}\mathbf{I}|_j = {}_j|\mathbf{A}|_j > 0.$$

□

Ejercicio 91(b) $\mathbf{x}(\mathbf{A} + \mathbf{B})\mathbf{x} = (\mathbf{x}\mathbf{A} + \mathbf{x}\mathbf{B})\mathbf{x} = \mathbf{x}\mathbf{A}\mathbf{x} + \mathbf{x}\mathbf{B}\mathbf{x} > 0$; (suma de dos números positivos).

□

Ejercicio 92. \mathbf{A} es invertible y, como es simétrica, es ortogonalmente diagonalizable $\mathbf{A} = \mathbf{Q}\mathbf{D}(\mathbf{Q}^{-1}) = \mathbf{Q}\mathbf{D}(\mathbf{Q}^\top)$; así

$$\mathbf{A}^{-1} = (\mathbf{Q}\mathbf{D}(\mathbf{Q}^{-1}))^{-1} = (\mathbf{Q}^{-1})^{-1}(\mathbf{D}^{-1})(\mathbf{Q}^{-1}) = \mathbf{Q}(\mathbf{D}^{-1})(\mathbf{Q}^\top);$$

y como los autovalores de \mathbf{A} son positivos, sus inversos (los autovalores de \mathbf{A}^{-1} que se encuentran en la diagonal de \mathbf{D}^{-1}) también lo son.

□

Ejercicio 93(a) Sea $\mathbf{z} \in \mathbb{R}^k$ tal que $\mathbf{z}(\mathbf{X}^\top)\mathbf{A}\mathbf{X}\mathbf{z} = (\mathbf{X}\mathbf{z})\mathbf{A}(\mathbf{X}\mathbf{z}) \leq 0$; entonces necesariamente $\mathbf{X}\mathbf{z} = \mathbf{0}$ (pues \mathbf{A} es definida positiva), y como \mathbf{X} es de rango completo por columnas, necesariamente $\mathbf{z} = \mathbf{0}$.

□

Ejercicio 93(b) Sea $\mathbf{X} = \mathbf{I}_{|\alpha|}$, donde los elementos del vector α son números enteros tales que $1 \leq \alpha_1 < \alpha_2 < \dots < \alpha_k \leq n$; entonces, por la proposición anterior, $(\mathbf{X}^\top)\mathbf{A}\mathbf{X} = {}_{\alpha}|\mathbf{A}|_{\alpha}$ es definida positiva.

□

Ejercicio 93(c) Comencemos con el resultado propuesto como pista. Consideremos la submatriz principal

$${}_{(i,j)}|\mathbf{A}|_{(i,j)} = \begin{bmatrix} a_{ii} & a_{ij} \\ a_{ij} & a_{jj} \end{bmatrix},$$

que es simétrica, pues \mathbf{A} también lo es. Además, como \mathbf{A} es definida positiva, esta submatriz también: así que su determinante es positivo; es decir: $a_{ii}a_{jj} - a_{ij}^2 > 0$. Por tanto $\boxed{a_{ii}a_{jj} > a_{ij}^2}$.

Ahora supongamos erróneamente que a_{ij} (con $i \neq j$) es la componente más grande en valor absoluto, en particular $|a_{ij}| > a_{ii}$ y $|a_{ij}| > a_{jj}$; consecuentemente $a_{ij}^2 > a_{ii}a_{jj}$. Pero esto es imposible, pues contradice el resultado del recuadro de arriba.

□

Ejercicio 94. La demostración es análoga a la del Apartado (a) del EJERCICIO 91 en la página 200.

□

Ejercicio 95. Sea $\mathbf{K} = \mathbf{A}_{\tau_1 \dots \tau_k}$ la matriz pre-escalada de \mathbf{A} obtenida por eliminación de izquierda a derecha; entonces $\mathbf{U} = (\mathbf{I}_{\tau_1 \dots \tau_k})^{-1}$ es triangular superior unitaria. Como todos los menores principales superiores de \mathbf{A} son de rango completo, por la proposición anterior sabemos que $\mathbf{A} = \mathbf{KU}$ es la factorización LU y que todas las componentes de la diagonal de \mathbf{K} son pivotes.

Sea \mathbf{D} la matriz diagonal con los pivotes de \mathbf{K} , es decir, tal que ${}_i\mathbf{D}_{|i} = {}_i\mathbf{K}_{|i}$. Puesto que la diagonal de \mathbf{K} no tiene ceros, la matriz \mathbf{D} es invertible y por tanto la matriz $\mathbf{L} = \mathbf{K}(\mathbf{D}^{-1})$ es triangular inferior unitaria. Así, $\mathbf{A} = \mathbf{KU} = \mathbf{K}(\mathbf{D}^{-1})\mathbf{DU} = \mathbf{LDU}$.

La unicidad de esta factorización proviene de la unicidad de la factorización LU del teorema anterior.

□

Ejercicio 96. Sea $\mathbf{K} = \mathbf{A}_{\tau_1 \dots \tau_k}$ la matriz pre-escalada de \mathbf{A} obtenida por eliminación de izquierda a derecha; entonces $\mathbf{U} = (\mathbf{I}_{\tau_1 \dots \tau_k})^{-1}$ es triangular superior unitaria. Sea \mathbf{D} diagonal tal que ${}_i\mathbf{D}_{|i} = {}_i\mathbf{K}_{|i}$, entonces $\mathbf{L} = \mathbf{K}(\mathbf{D}^{-1})$ es triangular inferior unitaria. Como todos los menores principales superiores de \mathbf{A} son de rango completo, por la proposición anterior sabemos que $\mathbf{A} = \mathbf{LDU}$ es la factorización LDU.

Consideremos la matriz

$$\underbrace{(\mathbf{U}^\top)^{-1}\mathbf{A}(\mathbf{U}^{-1})}_{\text{simétrica}} = \underbrace{(\mathbf{U}^\top)^{-1}\mathbf{K}}_{\text{triangular inferior}} = (\mathbf{U}^\top)^{-1}\mathbf{LD};$$

que por ser simultáneamente simétrica y triangular, es diagonal. Como \mathbf{D} es diagonal, $(\mathbf{U}^\top)^{-1}\mathbf{L}$ lo es, pero además es triangular inferior unitaria. Por tanto, $(\mathbf{U}^\top)^{-1}\mathbf{L} = \mathbf{I}$, lo que implica que $\mathbf{L} = \mathbf{U}^\top$.

□

Ejercicio 97.

Linealidad respecto al primer argumento:

1. $\langle \alpha x, y \rangle_{\mathbf{A}} = \alpha x \mathbf{A} y = \alpha \langle x, y \rangle_{\mathbf{A}}$.
2. $\langle (x + y), z \rangle_{\mathbf{A}} = (x + y) \mathbf{A} z = x \mathbf{A} z + y \mathbf{A} z = \langle x, z \rangle_{\mathbf{A}} + \langle y, z \rangle_{\mathbf{A}}$.

Por ser \mathbf{A} simétrica:

Simetría: $\langle x, y \rangle_{\mathbf{A}} = x \mathbf{A} y = x (\mathbf{A}^\top) y = y \mathbf{A} x = \langle y, x \rangle_{\mathbf{A}}$.

Por ser \mathbf{A} definida positiva:

Positivo: $\langle x, x \rangle_{\mathbf{A}} = x \mathbf{A} x \geq 0$

Definido: $\langle x, x \rangle_{\mathbf{A}} = x \mathbf{A} x = 0 \Leftrightarrow x = 0$.

□

Ejercicio 98. Sea \mathbf{P} la matriz proyección correspondiente a la proyección ortogonal sobre $\mathcal{L}([1])$ y, consecuentemente, sea $(\mathbf{I} - \mathbf{P})$ la matriz proyección correspondiente a la proyección ortogonal sobre $\mathcal{L}([1])^\perp$; entonces, puesto que $\mathbf{P}\mathbf{1} = \mathbf{1}$ y que $(\mathbf{I} - \mathbf{P})\mathbf{1} = \mathbf{0}$, tenemos que

- La media de $(\mathbf{y} + \mathbf{1}a)$ es el valor por el que hay que multiplicar $\mathbf{1}$ para lograr *el vector constante más próximo* a $(\mathbf{y} + \mathbf{1}a)$; es decir,

$$\mathbf{P}(\mathbf{y} + \mathbf{1}a) = \mathbf{P}\mathbf{y} + \mathbf{P}\mathbf{1}a = \mathbf{1}\mu_{\mathbf{y}} + \mathbf{1}a = (\mu_{\mathbf{y}} + a)\mathbf{1} \implies \mu_{(\mathbf{y} + \mathbf{1}a)} = \mu_{\mathbf{y}} + a.$$

- La desviación típica de $(\mathbf{y} + \mathbf{1}a)$ es la longitud de la proyección ortogonal

$$(\mathbf{I} - \mathbf{P})(\mathbf{y} + \mathbf{1}a) = (\mathbf{I} - \mathbf{P})\mathbf{y} + (\mathbf{I} - \mathbf{P})\mathbf{1}a = (\mathbf{I} - \mathbf{P})\mathbf{y} = \mathbf{y} - \mu_{\mathbf{y}},$$

que es la componente de \mathbf{y} ortogonal a $\mathbf{1}$, cuya longitud hemos denominado $\sigma_{\mathbf{y}}$.

□

Ejercicio 99(a) Sea \mathbf{y} un vector de \mathbb{R}^m , entonces: $\mu_{\mathbf{y}} = 0 \iff \frac{1}{m}(\mathbf{y} \cdot \mathbf{1}) = 0 \iff \mathbf{y} \perp \mathbf{1}$.

□

Ejercicio 99(b) Sea \mathbf{x} tal que $\mu_{\mathbf{x}} = 0$; y sea \mathbf{P} la matriz proyección correspondiente a la proyección ortogonal sobre $\mathcal{L}([1])$ y, consecuentemente, sea $(\mathbf{I} - \mathbf{P})$ la matriz proyección correspondiente a la proyección ortogonal sobre $\mathcal{L}([1])^\perp$; entonces, puesto que \mathbf{x} es perpendicular a $\mathbf{1}$ tenemos que $\mathbf{P}\mathbf{x} = \mathbf{0}$ y que $(\mathbf{I} - \mathbf{P})\mathbf{x} = \mathbf{x}$; así

- La media de $(\mathbf{y} + \mathbf{x})$ es el valor por el que hay que multiplicar $\mathbf{1}$ para lograr *el vector constante más próximo* a $(\mathbf{y} + \mathbf{x})$, es decir,

$$\mathbf{P}(\mathbf{y} + \mathbf{x}) = \mathbf{P}\mathbf{y} + \mathbf{P}\mathbf{x} = \mathbf{1}\mu_{\mathbf{y}} + \mathbf{0} = \mathbf{1}\mu_{\mathbf{y}} \implies \mu_{(\mathbf{y} + \mathbf{x})} = \mu_{\mathbf{y}}.$$

- La desviación típica de $(\mathbf{y} + \mathbf{x})$ es la longitud de la proyección ortogonal

$$(\mathbf{I} - \mathbf{P})(\mathbf{y} + \mathbf{x}) = (\mathbf{I} - \mathbf{P})\mathbf{y} + (\mathbf{I} - \mathbf{P})\mathbf{x} = (\mathbf{y} - \mu_{\mathbf{y}}) + \mathbf{x},$$

que es un vector distinto del vector $(\mathbf{y} - \mu_{\mathbf{y}})$ siempre que $\mathbf{x} \neq \mathbf{0}$. Consecuentemente la longitud del vector $(\mathbf{y} - \mu_{\mathbf{y}} + \mathbf{x})$ es *en general* distinta de $\sigma_{\mathbf{y}}$.

(Fíjese que si $\mathbf{x} = -2(\mathbf{y} - \mu_{\mathbf{y}})$ entonces $(\mathbf{y} - \mu_{\mathbf{y}} + \mathbf{x}) = -(\mathbf{y} - \mu_{\mathbf{y}})$ y por tanto, en este caso particular la desviación típica tampoco habría cambiado: $\sigma_{\mathbf{y}} = \sigma_{\mathbf{y} + \mathbf{x}}$).

□

Ejercicio 100.

$$z \cdot (\mathbf{x} - a\mathbf{1}) = (z \cdot \mathbf{x}) - a \underbrace{\mathbf{y} \cdot \mathbf{1}}_0 = (z \cdot \mathbf{x}).$$

□

Ejercicio 101.

$$\sigma_{\mathbf{x}\mathbf{y}} = \frac{1}{m}\mathbf{x} \cdot (\mathbf{y} - \mu_{\mathbf{y}}) = \frac{1}{m}\mathbf{x} \cdot \mathbf{y} - \frac{1}{m}\mathbf{x} \cdot \mu_{\mathbf{y}} = \frac{1}{m}\mathbf{x} \cdot \mathbf{y} - \frac{1}{m}\mathbf{x} \cdot \mathbf{1}\mu_{\mathbf{y}} = \frac{1}{m}\mathbf{x} \cdot \mathbf{y} - \mu_{\mathbf{x}}\mu_{\mathbf{y}}.$$

□

Ejercicio 102. En este caso las ecuaciones normales $\frac{1}{m}(\mathbf{X}^\top)\mathbf{X}\hat{\boldsymbol{\beta}} = \frac{1}{m}(\mathbf{X}^\top)\mathbf{y}$ se reducen a una única ecuación

$$\frac{1}{m}[\mathbf{1}]^\top[\mathbf{1}]\hat{\boldsymbol{\beta}} = \frac{1}{m}[\mathbf{1}]^\top\mathbf{y},$$

donde $\hat{\boldsymbol{\beta}}$ es un vector de \mathbb{R}^1 (vector con una sola componente). Puesto que

$$\frac{1}{m}[\mathbf{1}]^\top[\mathbf{1}] = \left[\frac{m}{m}\right] = \mathbf{1}_{1 \times 1} \quad \text{y} \quad \frac{1}{m}[\mathbf{1}]^\top\mathbf{y} = \left(\frac{1}{m}(\mathbf{1} \cdot \mathbf{y}),\right) = (\mu_{\mathbf{y}});$$

sustituyendo tenemos que: $\hat{\boldsymbol{\beta}} = (\mu_{\mathbf{y}})$; y consecuentemente $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = [\mathbf{1}](\mu_{\mathbf{y}}) = \mathbf{1}\mu_{\mathbf{y}} = \mu_{\mathbf{y}}$. □

Ejercicio 103. Puesto que la matriz de regresores es $\mathbf{X} = [\mathbf{1} \quad \mathbf{x}] = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$, las ecuaciones normales son

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix},$$

es decir

$$\begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} 5 \\ 11 \end{pmatrix};$$

que resolvemos por eliminación

$$\left[\begin{array}{cc|c} 3 & 6 & -5 \\ 6 & 14 & -11 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array} \right] \xrightarrow{\begin{matrix} [(-2)\mathbf{1}+\mathbf{2}] \\ [(\frac{1}{3})\mathbf{3}] \\ [(\frac{5}{3})\mathbf{1}+\mathbf{3}] \end{matrix}} \left[\begin{array}{cc|c} 3 & 0 & 0 \\ 6 & 2 & -3 \\ 1 & -2 & 5 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{array} \right] \xrightarrow{\begin{matrix} [(\frac{2}{3})\mathbf{3}] \\ [(\frac{2}{3})\mathbf{2}+\mathbf{3}] \end{matrix}} \left[\begin{array}{cc|c} 3 & 0 & 0 \\ 6 & 2 & 0 \\ 1 & -2 & 4 \\ 0 & 1 & 3 \\ 0 & 0 & 6 \end{array} \right] \xrightarrow{[(\frac{1}{6})\mathbf{3}]} \left[\begin{array}{cc|c} 3 & 0 & 0 \\ 6 & 2 & 0 \\ 1 & -2 & \frac{2}{3} \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 1 \end{array} \right]$$

Es decir, la ordenada en el origen es $\hat{a} = \frac{2}{3}$ y la pendiente $\hat{b} = \frac{1}{2}$.

Por otra parte $\hat{\mathbf{y}} = \mathbf{X} \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} \frac{7}{6} \\ \frac{5}{3} \\ \frac{13}{6} \end{pmatrix}$. Por tanto $(\mathbf{y} - \hat{\mathbf{y}}) = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} - \begin{pmatrix} \frac{7}{6} \\ \frac{5}{3} \\ \frac{13}{6} \end{pmatrix} = \begin{pmatrix} -\frac{1}{6} \\ \frac{1}{3} \\ -\frac{1}{6} \end{pmatrix}$. □

Bibliografía

- Arvesú Carballo, J., Marcellán Español, F., and Sánchez Ruiz, J. (2005). *Problemas resueltos de Álgebra Lineal*. Thomson Learning, Madrid. España. ISBN 84-9732-284-3.
- Cullen, C. G. (1972). *Matrices and Linear Transformations*. Dover publications, Inc., New York, USA., second ed.
- Godement, R. (1968). *Algebra*. Hermann. ISBN 9780901665003.
- Larson, R., Edwards, B. H., and Falvo, D. C. (2004). *Álgebra lineal*. Ediciones Pirámide, Madrid. España, fifth ed. ISBN 84-368-1878-4. Título de la obra original: Elementary Linear Algebra. Houghton Mifflin Company.
- Lay, D. C. (2007). *Álgebra Lineal y sus aplicaciones*. Pearson Educación, Inc., third ed.
- Poole, D. (2004). *Álgebra lineal. Una introducción moderna*. Thomson Learning, Mexico D.F. ISBN 970-686-272-2.
- Strang, G. (2003). *Introduction to Linear Algebra*. Wellesley-Cambridge Press, Wellesley, Massachusetts. USA, third ed. ISBN 0-9614088-9-8.
- Strang, G. (2007). *Álgebra Lineal y sus Aplicaciones*. Thomsom Learning, Inc, Santa Fe, México, D. F., fourth ed. ISBN 970686609-4.

C

combinación lineal de los vectores de Z ($Z\mathbf{a}$).
 complemento ortogonal del subespacio \mathcal{A} (\mathcal{A}^\perp).
 concatenación de los sistemas Y y Z ($Y \# Z$).
 conjunto de matrices de orden $m \times n$ $\mathbb{R}^{m \times n}$.
 conjunto de numeros complejos (\mathbb{C}).
 conjunto de numeros reales (\mathbb{R}).
 conjunto de sistemas de n números reales (vectores) \mathbb{R}^n .
 conjunto ortogonal al conjunto A (A^\perp).

E

espacio columna de \mathbf{A} ($\mathcal{C}(\mathbf{A})$).
 espacio nulo de \mathbf{A} ($\mathcal{N}(\mathbf{A})$).
 espacios vectoriales ($\mathcal{C}, \mathcal{N}, \mathcal{X}, \mathcal{V}, \mathcal{W}$).

M

matriz ($\mathbb{R}^{m \times n}$) ($\mathbf{A}, \mathbf{B}, \mathbf{C}$).
 elemental (\mathbf{I}_τ).
 Tipo I ($\mathbf{I}_{\tau_{[(\lambda)i+j]}}$).
 Tipo II ($\mathbf{I}_{\tau_{[(\alpha)i]}}$).
 escalonada reducida (por columnas) (\mathbf{R}).
 identidad (\mathbf{I}).
 intercambio ($\mathbf{I}_{\tau_{[p \rightleftharpoons q]}}$).
 inversa (de \mathbf{A}) (\mathbf{A}^{-1}).
 invertible, de rango completo, o no singular ($\mathbf{I}_{\tau_1 \dots \tau_k}$).
 determinante de \mathbf{A} ($|\mathbf{A}|$).
 nula ($\mathbf{0}$).
 permutación ($\mathbf{I}_{\tau_{[\hat{\sigma}]}}$).
 pre-escalonada (\mathbf{K}).
 rango de \mathbf{A} ($\text{rg}(\mathbf{A})$).
 transpuesta (\mathbf{A}^\top ; transpuesta de \mathbf{A}).
 triangular inferior (\mathbf{L}).
 triangular superior (\mathbf{U}).

O

operador concatenación de listas o sistemas ($\#$).
 operador selector de componentes ($|$).

por la derecha selecciona el elemento i -ésimo de un sistema (\mathbf{I}) .
 por la izquierda selecciona el elemento i -ésimo de un vector o la fila i -ésima de una matriz (\mathbf{I}) .
 operador transposición (\mathbf{T}) .

P

producto punto (producto escalar usual en \mathbb{R}^n) $(\mathbf{a} \cdot \mathbf{b})$.

S

Sistema de ecuaciones lineales $(\mathbf{Ax} = \mathbf{b})$.

sistemas de vectores de un subespacio genérico $\mathcal{V}(\mathbf{A}, \mathbf{B}, \dots, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$.

subespacio engendrado por un sistema $\mathcal{Z}(\mathcal{L}(\mathcal{Z}))$.

suma directa de los subespacios \mathcal{A} y \mathcal{B} $(\mathcal{A} \oplus \mathcal{B})$.

T

transformación elemental $(\boldsymbol{\tau})$.

de las columnas de \mathbf{A} $(\mathbf{A}_{\boldsymbol{\tau}})$.

elemental Tipo I; suma λ veces el elemento i -ésimo al elemento j -ésimo $\left(\begin{smallmatrix} \boldsymbol{\tau} \\ [(\lambda)i+j] \end{smallmatrix}\right)$.

elemental Tipo II; multiplica por α el elemento i -ésimo $\left(\begin{smallmatrix} \boldsymbol{\tau} \\ [(\alpha)i] \end{smallmatrix}\right)$.

intercambio entre los elementos i -ésimo y j -ésimo $\left(\begin{smallmatrix} \boldsymbol{\tau} \\ [i \rightleftharpoons j] \end{smallmatrix}\right)$.

inversa de $\boldsymbol{\tau}$ $(\boldsymbol{\tau}^{-1})$.

permutación (reordenamiento) de elementos $\left(\begin{smallmatrix} \boldsymbol{\tau} \\ [\mathfrak{S}] \end{smallmatrix}\right)$.

V

vectores de \mathbb{R}^n $(\mathbf{a}, \mathbf{b}, \mathbf{c})$.

columna j -ésima de \mathbf{A} $(\mathbf{A}_{|j})$.

combinación lineal de las columnas de \mathbf{A} (\mathbf{Ab}) .

combinación lineal de las filas de \mathbf{A} (\mathbf{aB}) .

fila i -ésima de \mathbf{A} $({}_i|\mathbf{A})$.

nulo $(\mathbf{0})$.

vectores de un espacio vectorial genérico $\mathcal{V}(\vec{x}, \vec{y}, \vec{z})$.

conjuntos

$\{X, Y, X, \dots\}$ conjunto.
 \mathbb{C} conjunto de numeros complejos.
 \mathbb{R} conjunto de numeros reales.
 $\mathbb{R}^{m \times n}$ conjunto de matrices de orden $m \times n$.
 \mathbb{R}^n conjunto de sistemas de n números reales (vectores).

espacios y subespacios vectoriales

$\mathcal{C}(\mathbf{A})$ espacio columna de \mathbf{A} .
 \mathcal{A}^\perp complemento ortogonal del subespacio \mathcal{A} .
 A^\perp conjunto ortogonal al conjunto A .
 $\mathcal{C}, \mathcal{N}, \mathcal{X}, \mathcal{V}, \mathcal{W}$ espacios vectoriales.
 $\mathcal{N}(\mathbf{A})$ espacio nulo de \mathbf{A} .
 $\mathcal{L}(\mathbf{Z})$ subespacio engendrado por un sistema \mathbf{Z} .
 $\mathcal{A} \oplus \mathcal{B}$ suma directa de los subespacios \mathcal{A} y \mathcal{B} .

matrices

$[a \ b \ c \ \dots]$ con $a, b, c \in \mathbb{R}^m$; matriz de $\mathbb{R}^{m \times n}$.
 $\mathbf{A}, \mathbf{B}, \mathbf{C}$ matriz ($\mathbb{R}^{m \times n}$).
 $|\mathbf{A}|$ determinante de \mathbf{A} .
 \mathbf{R} matriz escalonada reducida (por columnas).
 \mathbf{A}^{-1} inversa (de \mathbf{A}).
 \mathbf{I}_τ matriz elemental.
 $\mathbf{I}_{\tau_{[(\lambda)\mathbf{i}+\mathbf{j}]}}$ matriz elemental Tipo I.
 $\mathbf{I}_{\tau_{[(\alpha)\mathbf{i}]}}$ matriz elemental Tipo II.
 \mathbf{I} matriz identidad.
 $\mathbf{I}_{\tau_{[p \rightleftharpoons q]}}$ matriz intercambio.
 $\mathbf{I}_{\tau_1 \dots \tau_k}$ matriz invertible, de rango completo, o no singular.
 $\mathbf{0}$ matriz nula.
 $\mathbf{I}_{\tau_{[\mathfrak{S}]}}$ matriz permutación.
 \mathbf{A}^\top transpuesta de \mathbf{A} .
 \mathbf{K} matriz pre-escalonada.
 $\text{rg}(\mathbf{A})$ rango de \mathbf{A} .
 \mathbf{L} triangular inferior.
 \mathbf{U} triangular superior.

operaciones

$a \cdot b$ producto punto (producto escalar usual en \mathbb{R}^n).

operadores

$\#$ operador concatenación de listas o sistemas.

\top operador transposición.

$|$ operador selector de componentes.

$i|$ por la izquierda.

$|i$ por la derecha.

Sistema de ecuaciones lineales

$\mathbf{A}x = b$ Sistema de ecuaciones lineales.

sistema de vectores de un espacio vectorial generico

$Y \# Z$ concatenación de los sistemas Y y Z .

$A, B, \dots X, Y, Z$ sistemas de vectores de un subespacio genérico \mathcal{V} .

sistemas genéricos

$[X; Y; Z; \dots;]$ sistema.

transformaciones elementales (o sucesiones de transformaciones elementales)

τ^{-1} inversa de τ .

τ transformación elemental.

$\tau_{[i \rightleftharpoons j]}$ intercambio entre los elementos i -ésimo y j -ésimo.

$\tau_{[\sigma]}$ permutación (reordenamiento) de elementos.

\mathbf{A}_τ de las columnas de \mathbf{A} .

$\tau_{[(\alpha)i]}$ elemental Tipo II; multiplica por α el elemento i -ésimo.

$\tau_{[(\lambda)i+j]}$ elemental Tipo I; suma λ veces el elemento i -ésimo al elemento j -ésimo.

vectores

$\vec{x}, \vec{y}, \vec{z}$, vectores de un espacio vectorial genérico \mathcal{V} .

a, b, c , vectores de \mathbb{R}^n .

$\mathbf{A}_{|j}$ vector columna j -ésima de \mathbf{A} .

$\mathbf{A}b$ vector combinación lineal de las columnas de \mathbf{A} .

$a\mathbf{B}$ vector combinación lineal de las filas de \mathbf{A} .

$i|\mathbf{A}$ vector fila i -ésima de \mathbf{A} .

(a, b, c, \dots) vector de \mathbb{R}^n ; con $a, b, c, \dots \in \mathbb{R}$.

$\mathbf{0}$ vector nulo.

Za vector combinación lineal de los vectores de Z .

- ajuste MCO, *véase* regresión lineal
- aplicación lineal, *véase* función lineal
- autoespacio, 179, 197
- autovalor, 179, 181, 200, 205
 - espectro, 179
 - multiplicidad algebraica, 181, 192
 - multiplicidad geométrica, 181, 192
 - producto, 191
 - suma, 191
- autovector, 179
- Cauchy-Schwarz*, *véase* desigualdad, 146
- combinación lineal, 26, 29, 40, 80, 107
- concatenación de sistemas, 37, 114
- conjunto, 3
 - de vectores linealmente dependiente, 118
 - también, *véase* sistema de vectores
 - generador de un subespacio, 118
 - también, *véase* sistema de vectores
 - vacío, 90
- conjunto de sucesiones casi nulas, 149
 - conjunto de polinomios de una variable, 150
- conjunto ortogonal de vectores, 138, 156
- coordenadas, 120
- correlación, 215
- covarianza, 215
- desigualdad
 - Cauchy-Schwarz*, 136, 146
 - triangular, 138, 146
- desviación típica, 213
- determinante, 162, 191
 - cofactor, 172
 - cálculo de la inversa, 174
 - cálculo por eliminación Gaussiana, 168
 - de matrices diagonales por bloques, 168
 - desarrollo por cofactores, 170
 - expansión de Laplace, 170, 172
 - menor, 172, 206
 - producto de pivotes, 206
 - regla de Cramer, 173
 - regla de Sarrus, 170
- diagonalización
 - matriz simétrica, 195
 - ortogonal, 195
 - por bloques triangulares, 185
 - por bloques triangulares, 192
 - por congruencia, 201, 204
 - por semejanza, *véase* por bloques triangulares, 204
- ecuaciones cartesianas, 93, 140
- ecuaciones paramétricas, 97, 140
- ecuación lineal, 91
- eliminación, 51, 94
 - de izquierda a derecha, 51, 114
 - pre-escalonar, 51
 - Gauss-Jordan, 53
 - escalonar y reducir, 53
 - gaussiana, 52, 53
 - escalonar, 53
- eliminación por filas, 54
- escalares, 3
 - números complejos, 195
 - números reales, 3
- espacio afín, 142
 - hiperplano en \mathbb{R}^n , 142
 - plano en \mathbb{R}^n , 142
 - punto en \mathbb{R}^n , 142
 - recta en \mathbb{R}^n , 142
- espacio euclideo, 147
 - \mathbb{R}^n , 147
- espacio pre-Hilbert, 147, 155
- espacio vectorial, 78
 - subespacio, *véase* subespacios
- espectro, 179, 185
- factorización de una matriz, 206
 - factorización KU, 206
- factorización de una matriz cuadrada
 - factorización LDU, 207
 - factorización LU, 206
- factorización de una matriz cuadrada y simétrica
 - factorización de Cholesky, 207

- factorización de LDU, 207
- forma cuadrática, 199, 201
 - completar el cuadrado, 205
 - tipos, 201
 - definida positiva, 199
- función, 87
 - autoadjunta, 154
 - composición de funciones, 89
 - idempotente, 154
 - invertible, 89
 - notación, 88
 - producto por un escalar, 84
 - suma de funciones, 84
- función determinante, *véase* determinante
- función lineal, 80
 - composición, 81
 - imagen, 119
 - invertible, 81
 - núcleo o espacio nulo, 119
- Gram-Schmidt, *véase* método de Gram-Schmidt
- inercia, *véase* ley de inercia
- inversa
 - de una función, 89
 - de una matriz, 61
 - transpuesta, 62
 - triangular, 69
 - de una transformación elemental, 62
 - espejo*, 59
- ley de inercia, 204
- linealidad, *véase* operador lineal
- matriz, 11
 - adjunta, *véase* determinante
 - columna de una matriz, 13
 - columna libre de una matriz, 97
 - columna pivote de una matriz, 97
 - componente (o elemento) de una, 12
 - congruente, 201, 205
 - cuadrada, 18
 - de regresores, 215
 - definida positiva, 199, 205
 - producto escalar, 208
 - diagonal, 18, 205
 - diagonalizable ortogonalmente, 196, 197
 - diagonalizable por semejanza, 192
 - elemental, 43
 - elemental de Tipo I, 44
 - elemental de Tipo II, 45
 - escalonada, 52, 54
 - escalonada por filas, 57
 - escalonada reducida, 53, 54
 - extendida, 167
 - fila de una matriz, 13
 - fila pivote de una matriz, 127
 - idempotente, 153, 154
 - identidad, 18
 - igualdad, 14
 - intercambio, 49
 - inversa, 61
 - invertible, 61, 63, 65, 73
 - invertible por la derecha, 61
 - invertible por la izquierda, 61
 - matriz columna*, 11
 - matriz columna*, 15, 17
 - matriz fila*, 17
 - múltiplo de, 19
 - notación, 12
 - nula, 17
 - opuesta, 17
 - orden de una matriz, 12, 18
 - ortogonal, 195
 - permutación, 50
 - pre-escalonada, 51, 54
 - producto de matrices, 31, 33, 35
 - producto por un escalar, 19
 - proyección, 153
 - rango, 71
 - rango completo, 73, 117
 - rango completo por columnas, 73
 - rango completo por filas, 73
 - rectangular, 18
 - selección de un elemento (o componente), 14
 - selección de una columna, 13, 14
 - selección de una fila, 13, 14
 - selección de una submatriz, 22
 - semejante, 183, 191
 - simétrica, 18, 155, 197, 201
 - singular, 61
 - suma, 19
 - transpuesta, 16, 34
 - traza, 40, 184, 191
 - propiedades, 40
 - triangular
 - por bloques, 168, 187
 - triangular inferior, 35
 - triangular superior, 35
 - triangular unitaria, 35
- media aritmética, 212
- método de eliminación, *véase* eliminación
- método de Gram-Schmidt, 193, 198
- nomenclatura de G. Strang, 97

- números complejos, 195
 - parte imaginaria, 195
 - parte real, 195
- operador lineal, 6, 19–21, 25, 28, 30, 33, 48, 108
- permutación, 50
- Pitágoras* (Teorema), 136, 146, 214
- pivote, 51, 206
 - columna pivote, 97
 - posición de pivote, 51
- plano de regresión, 218
- polinomio, 150, 199
 - grado, 199
 - raíces, 201
- polinomio característico, 180
 - matriz triangular por bloques, 187
 - raíces, 181
- pre-escalonar, 51
- producto cartesiano, 83
- producto escalar, 25, 144, 208, 211
- producto punto, 25
- proyección ortogonal, 151, 155, 215
 - en \mathbb{R}^n , 151
 - en un espacio pre-Hilbert \mathcal{V} , 155
 - vector constante más próximo, 212
- rango, *véase* matriz
- recta de regresión, 217
- regresión lineal, 215
- Rouché-Frobenius* (Teorema), 103
- selector (*operador*), 22
- selector (*operador*), 4, 13, 14, 22, 31, 107, 108, 114
- sistema, 3, 11, 90, 107
 - vacío, 90, 112, 118
- sistema de ecuaciones lineales, 91
 - compatible determinado, 102
 - compatible indeterminado, 102
 - equivalentes, 141, 152
 - homogéneos, 92
 - algoritmo de resolución, 94, 97
 - solución trivial, 93
 - incompatible, 102
 - matriz de coeficientes, 92
 - matriz de coeficientes ampliada, 99
 - no homogéneos, 99
 - regla de Cramer, 173
 - resolución de, 102
 - resolver un sistema, 92
 - soluciones especiales, 95, 97
 - solución de un sistema, 92
 - variables libres (o exógenas), 97
 - variables pivote (o endógenas), 97
 - vector de incógnitas, 92
 - vector del lado derecho, 92
- sistema de ecuaciones normales, 152
- sistema de vectores de \mathcal{V}
 - base ortogonal, 194
 - generador de un subespacio, 109, 110
 - sistemas equivalentes, 110, 193
 - linealmente dependiente, 112
 - linealmente independiente, 112, 113, 120, 193
 - ortogonal, 193
 - ortonormal, 194
 - selección de varias componentes, 108
 - sistemas acoplados, 114
 - sistemas equivalentes, 111, 193
- subespacios, 82
 - base, 112
 - complementos ortogonales, 139, 147, 149
 - dimensión, 113, 122
 - infinita, 149
 - espacio columna de \mathbf{A} , 104
 - espacio fila de \mathbf{A} , 124
 - espacio nulo de \mathbf{A} , 93
 - espacio nulo de un sistema de vectores \mathbf{Z} , 112
 - espacio nulo de una función lineal, 119
 - espacio nulo por la izquierda de \mathbf{A} , 125
 - generado por el conjunto vacío, 118
 - generado por un conjunto, $\mathcal{L}(Z)$, 118
 - generado por un sistema, $\mathcal{L}(Z)$, 109
 - generado por un sistema, $\mathcal{L}(Z)$, 110
 - iguales, 110, 113
 - imagen de una función lineal, 119
 - intersección de, 83
 - núcleo de una función lineal, 119
 - suma de subespacios, 121
 - suma directa de subespacios, 123
 - suplementarios, 123, 139, 149, 156
- submatriz, 22
 - principal, 23, 200
 - principal superior, 23
- Teorema Espectral, 197, 204
- Teorema Fundamental del Álgebra, 181
- transformación elemental, 43
 - de las columnas de una matriz, 44, 46
 - de las filas de una matriz, 54
 - de los vectores de un sistema \mathbf{Z} , 111
 - intercambio, 49
 - inversa, 62
 - permutación, 37, 50
 - secuencia de transformaciones, 48
 - Tipo I, 43

- Tipo II, 45
- transpuesta, 54
- traza, *véase* matriz
- varianza, 214
- vectores de \mathcal{V} , 78
 - combinación lineal, 80, 107
 - coseno del ángulo formado por, 146
 - distancia entre vectores, 146
 - longitud o norma, 145
 - múltiplo de, 78
 - ortogonales (o perpendiculares), 145
 - producto por escalares de, 78
 - suma de, 78
 - unitario, 146
- vectores de \mathbb{R}^n , 3
 - alineados, 138
 - combinación lineal, 26, 29, 40
 - coseno del ángulo formado por, 137
 - de medias, 212
 - distancia entre vectores, 138
 - igualdad, 4
 - longitud o norma, 135, 211
 - múltiplo de, 6
 - notación, 4
 - nulo, 5
 - opuesto, 5
 - ortogonales (o perpendiculares), 135
 - producto escalar usual en \mathbb{R}^n , *véase* producto
 - punto
 - producto por un escalar de, 6
 - selección de una componente, 4, 14, 15
 - selección de varias componentes, 40
 - suma, 6
 - unitario, 136, 194, 212

Índice de figuras

6.1. Representación esquemática de que $\mathcal{V} \times \mathcal{W} = f \oplus (\{\vec{0}\} \times \mathcal{W})$	86
10.1. Esquema de descomposición de los cuatro subespacios fundamentales de una matriz (parte 1)	125
11.1. Longitud de un vector en \mathbb{R}^2 y en \mathbb{R}^3	133
11.2. Relación entre el ángulo θ y su coseno (el círculo exterior es de radio uno)	134
11.3. Triángulo formado por $\mathbf{0}$ y dos vectores \mathbf{a} y \mathbf{b} no nulos.	134
11.4. Representación esquemática donde \mathbf{h} es combinación lineal de \mathbf{a} y \mathbf{b} y perpendicular a \mathbf{a}	137
11.5. Representación esquemática del triángulo formado por $\vec{0}$ y dos vectores \vec{x} y \vec{y} no nulos de \mathcal{V}	146
11.6. Representación esquemática desde dos ángulos distintos. El subespacio \mathcal{W} es de dimensión n , el “suelo” \mathcal{H} es un subespacio de dimensión $n+1$ que contiene a \mathcal{W} , y el espacio vectorial ambiente \mathcal{V} (posiblemente de dimensión infinita) está representado por el espacio tridimensional (a la derecha se muestra la visión desde la vertical sobre \mathcal{H}).	148
11.7. Representación esquemática. La recta \mathcal{W} es un subespacio de dimensión n , el plano (el suelo) \mathcal{H} es un subespacio de dimensión $n+1$ que contiene a \mathcal{W} , a $\mathcal{L}(\vec{z})$ (que es el único subespacio de la figura que con seguridad es de dimensión 1), y el espacio vectorial ambiente \mathcal{V} (posiblemente de dimensión infinita) está representado por el espacio tridimensional (si \mathcal{V} es de dimension infinita, entonces \mathcal{H}^\perp también).	148
12.1. Representación esquemática de la proyección ortogonal de \mathbf{y} sobre \mathcal{V}	151
13.1. $\text{Area}[\mathbf{l}_1; \mathbf{l}_2] = 1$ (izquierda) y $\text{Volumen}[\mathbf{l}_1; \mathbf{l}_2; \mathbf{l}_3] = 1$ (derecha).	161
13.2. $\text{Area}[\mathbf{a}; \mathbf{b}] = \text{Area}[\mathbf{a}; (\alpha\mathbf{a}+\mathbf{b})]$ (izquierda) y $\text{Vol}[\mathbf{a}; \mathbf{b}; \mathbf{c}] = \text{Vol}[\mathbf{a}; (\alpha\mathbf{a}+\mathbf{b}); \mathbf{c}]$ (derecha).	161
13.3. $ \alpha \cdot \text{Area}[\mathbf{a}; \mathbf{b}] = \text{Area}[\alpha\mathbf{a}; \mathbf{b}]$ (izquierda) y $ \alpha \cdot \text{Vol}[\mathbf{a}; \mathbf{b}; \mathbf{c}] = \text{Vol}[\alpha\mathbf{a}; \mathbf{b}; \mathbf{c}]$ (derecha).	162
19.1. Representaciones esquemáticas de la proyección de \mathbf{y} sobre el subespacio de vectores constantes (izquierda); y de la proyección de los vectores \mathbf{v} y \mathbf{x} sobre el subespacio de vectores constantes (derecha).	212
19.2. Representaciones esquemáticas de la proyección de \mathbf{y} sobre el subespacio ortogonal a los vectores constantes (izquierda); y de la proyección de los vectores \mathbf{v} y \mathbf{x} sobre el subespacio ortogonal a los vectores constantes (derecha).	213
19.3. Sumar un vector constante no nulo, $a\mathbf{1}$, cambia la media pero no la desviación típica.	214
19.4. Sumar un vector de media nula (perpendicular a $\mathbf{1}$) no cambia la media pero puede cambiar la desviación típica. En la figura $\sigma_{\mathbf{y}} \neq \sigma_{\mathbf{y}+\mathbf{x}}$, pero sin embargo $\sigma_{\mathbf{y}} = \sigma_{\mathbf{y}+\mathbf{z}}$	214
19.5. Visión desde dos perspectivas distintas de la parte derecha de la Figura 19.2 con la representación del ángulo θ cuyo coseno es la correlación entre los vectores \mathbf{v} y \mathbf{x}	215
19.6. Representación esquemática de la proyección sobre el espacio columna de $\mathbf{X} = [\mathbf{1}; \mathbf{x}]$; con $\mathbf{x} \notin \mathcal{L}([\mathbf{1}])$. El “suelo” de la figura representa el subespacio $\mathcal{C}(\mathbf{X}) \subset \mathbb{R}^m$. La figura de la derecha es una visión desde “arriba” (desde esta perspectiva $\hat{\mathbf{y}}$ está detrás de \mathbf{y}).	217
19.7. Diagrama de dispersión de los vectores $\mathbf{x} = (1, 2, 3)$ e $\mathbf{y} = (1, 2, 2)$	217

19.8. Diagrama de dispersión de $\mathbf{x} = (1, 2, 3,)$ y la combinación lineal $\mathbf{z} = \mathbf{1} + \frac{1}{3}\mathbf{x}$. Los puntos se sitúan sobre la recta cuya ordenada en el origen es 1 y cuya pendiente es $\frac{1}{3}$	218
19.9. Ajuste de la recta de regresión $\hat{\mathbf{y}} = \hat{a}\mathbf{1} + \hat{b}\mathbf{x}$ donde $\mathbf{x} = (1, 2, 3,)$ e $\mathbf{y} = (1, 2, 2,)$	218
19.10 Representación esquemática de la proyección de $\mathbf{y} \in \mathbb{R}^m$ sobre el subespacio $\mathcal{C}(\mathbf{X})$ que contienen el subespacio de vectores constantes $\mathcal{L}([\mathbf{1};])$. El “suelo” de la figura representa el subespacio $\mathcal{C}(\mathbf{X})$	219
19.11 Ajuste del plano de regresión $\hat{\mathbf{y}} = \hat{a}\mathbf{1} + \hat{b}\mathbf{x} + \hat{c}\mathbf{z}$ que minimiza los errores de ajuste. Arriba: las barras verticales arrancan de los puntos (x_i, z_i) . La altura de los cuadrados rojos corresponde a los correspondientes valores y_i y la altura de los círculos azules (en el extremo superior de las barras) a los valores ajustados \hat{y}_i (dichos círculos están contenidos en el plano de regresión). Las diferencias entre ambas alturas, $y_i - \hat{y}_i$, son los errores de ajuste (los componentes de $\hat{\mathbf{e}} = \mathbf{y} - \hat{\mathbf{y}}$). Abajo: Representación de dichos errores de ajuste mediante barras continuas si son positivos, y discontinuas si son negativos.	220
A.1. Representación esquemática donde \vec{h} es combinación lineal de \vec{x} y \vec{y} y perpendicular a \vec{x}	254

Índice de cuadros

1.1. Precedencia de operadores (detalles aquí : 6.16 <i>Orden de evaluación</i> y 6.17 <i>Prioridad de operador</i>)	7
---	---